

A Method of Normalized Block Iteration*

ELIZABETH H. CUTHILL

David Taylor Model Basin, U. S. Navy Department, Carderock, Md.

AND

RICHARD S. VARGA

Westinghouse Electric Corporation, Pittsburgh, Pa.

1. Introduction

The Young-Frankel successive point overrelaxation scheme [14, 4] has been shown [14] to be applicable in solving partial difference equations of elliptic type arising from discrete approximations to general partial differential equations of elliptic type. More recently, Arms, Gates, and Zondek [1] generalized the successive point overrelaxation scheme of Young-Frankel to what is called the *successive block overrelaxation scheme*, and they stated that, with certain additional assumptions, a theoretical advantage in the rates of convergence is always obtained in using successive block overrelaxation rather than successive point overrelaxation. In particular, for the numerical solution of the Dirichlet problem of uniform mesh size h in a rectangle, the successive block overrelaxation scheme [1] is asymptotically faster by a factor of $2^{\frac{1}{2}}$ than the successive point overrelaxation scheme, as $h \rightarrow 0$. Despite that advantage, the successive block overrelaxation scheme has not been widely used, mainly because the usual computing machine application of block overrelaxation requires more arithmetic operations than point overrelaxation, and this increase in the number of arithmetic operations would appear to cancel any gains in the rates of convergence.

We shall show for a large class of matrix problems that, when the equations are suitably normalized, the successive block overrelaxation scheme can be applied in the *same* number of arithmetic operations per iteration as that required by the successive point overrelaxation scheme. Therefore, in this computing machine application of block relaxation, the full advantage in the rates of convergence of block vs. point relaxation will be obtained. We shall also show that the same normalization of our equations in general gives rise to an essential reduction in the number of entries of the coefficient matrix which is used in the computing machine application of the iterative block relaxation scheme.

2. Basic Assumptions

We seek the solution vector \mathbf{x} of the matrix problem

$$A\mathbf{x} = \mathbf{k}, \quad (2.1)$$

* Received September, 1958.

where the coefficient matrix $A = \|a_{i,j}\|$ is a given real $n \times n$ matrix. We shall assume that A is partitioned into

$$A = \begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,N} \\ \vdots & \vdots & & \vdots \\ A_{N,1} & A_{N,2} & \cdots & A_{N,N} \end{pmatrix}, \quad (2.2)$$

where the diagonal blocks $A_{i,i}$ are $n_i \times n_i$ submatrices of A , $1 \leq i \leq N$, and $\sum_{i=1}^N n_i = n$. We further assume that

(a) A is symmetric (2.3)

(b) $a_{i,i} \leq 0$ for $i \neq j$, $1 \leq i, j \leq n$.

(c) A is *irreducible* [5], i.e., there exists no permutation matrix P such that

$$PAP^{-1} = \begin{pmatrix} Q & R \\ 0 & S \end{pmatrix},$$

where Q and S are square submatrices.

(d) $\sum_{j=1}^n a_{i,j} \geq 0$ for all $1 \leq i \leq n$, with strict inequality for some i .

(e) Each $A_{i,i}$ is *tridiagonal*, i.e., if $A_{i,i} = \|a_{k,l}^{(i)}\|$ then $a_{k,l}^{(i)} = 0$ for $|k - l| > 1$, $1 \leq k, l \leq n_i$.

(f) A has *Property A** [1, p. 221], i.e., there exist two disjoint subsets S and T of W , the set of the first N integers, such that $S \cup T = W$, and if $A_{i,j}$ does not have all zero entries, then either $i = j$, or $i \in S$ and $j \in T$, or $i \in T$ and $j \in S$.

We may assume, without loss of generality, that A is (consistently) ordered [1, p. 221]. We remark that from (a), (b), (c), and (d) of (2.3), it follows [11] that the matrices A and $A_{i,i}$, $1 \leq i \leq N$, are all symmetric and positive definite.

We shall show in section 5 that the above assumptions are fulfilled for a large class of matrix problems, especially those occurring in the numerical solution of self-adjoint partial differential equations.

3. Factorization

The following well-known result [6, pp. 20–22] gives a representation for the matrices $A_{i,i}$:

LEMMA 1. *If $C = \|c_{i,j}\|$ is a real $n \times n$ symmetric and positive definite matrix, then there exists a unique positive diagonal matrix D and a unique real upper-triangular¹ matrix T with unit diagonal entries such that*

$$C = DT'TD \quad (3.1)$$

where T' denotes the transpose of T .

¹ An $n \times n$ matrix $T = \|t_{i,j}\|$ is *upper triangular* if $t_{i,j} = 0$ for $i > j$, $1 \leq i, j \leq n$.

The matrices $A_{i,i}$ in addition to being symmetric and positive definite are, by hypothesis (e) of (2.3), also tridiagonal. For this type of matrix the following corollary can be proved inductively:

COROLLARY 1. *Let*

$$C = \begin{bmatrix} b_1 & c_1 & & & \\ c_1 & b_2 & c_2 & & \\ & c_2 & b_3 & \ddots & \\ & & \ddots & \ddots & c_{n-1} \\ & & & c_{n-1} & b_n \end{bmatrix} \quad (3.2)$$

be a real symmetric and positive definite tridiagonal matrix. Then C has the unique factorization $C = DT'TD$, where

$$D \equiv \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{bmatrix}, \quad T \equiv \begin{bmatrix} 1 & e_1 & & & \\ & 1 & e_2 & & \\ & & \ddots & \ddots & \\ & & & 1 & e_{n-1} \\ & & & & 1 \end{bmatrix}, \quad (3.3)$$

and

$$d_1 = b_1^{\frac{1}{2}}; \quad d_j = \left\{ b_j - \left(\frac{c_{j-1}}{d_{j-1}} \right)^2 \right\}^{\frac{1}{2}}, \quad 2 \leq j \leq n, \quad (3.4)$$

and

$$e_j = \frac{c_j}{d_j d_{j+1}}, \quad 1 \leq j \leq n-1. \quad (3.4')$$

Assuming that C is a tridiagonal symmetric and positive definite matrix, we seek to solve the matrix problem

$$C\mathbf{u} = \mathbf{k}. \quad (3.5)$$

Corollary 1 implies that we can write: $DT'TD\mathbf{u} = \mathbf{k}$, from which we obtain

$$T'T(D\mathbf{u}) = D^{-1}\mathbf{k}. \quad (3.6)$$

Letting $\mathbf{y} = D\mathbf{u}$, and $\mathbf{g} = D^{-1}\mathbf{k}$, our matrix problem is reduced to

$$T'T\mathbf{y} = \mathbf{g}. \quad (3.7)$$

This can be solved directly for \mathbf{y} in terms of the auxiliary vector \mathbf{h} , where

$$h_1 = g_1, \quad h_{j+1} = g_{j+1} - e_j h_j, \quad 1 \leq j \leq n-1, \quad (3.8)$$

and

$$y_n = h_n, \quad y_j = h_j - e_j y_{j+1}, \quad 1 \leq j \leq n-1. \quad (3.9)$$

From (3.8) and (3.9), it is clear that at most two multiplications and two additions are needed per component in finding directly the solution \mathbf{y} of (3.7).

4. Normalized Block Relaxation Applied to the Matrix A

Let A satisfy the conditions of (2.3), and let the column vectors \mathbf{x} and \mathbf{k} of (2.1) be partitioned in a form consistent with (2.2) so that (2.1) can be written as

$$\begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,N} \\ A_{2,1} & & & \vdots \\ \vdots & & & \\ A_{N,1} & \cdots & & A_{N,N} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{pmatrix} = \begin{pmatrix} K_1 \\ K_2 \\ \vdots \\ K_N \end{pmatrix}. \quad (4.1)$$

Here, X_i and K_i are column vectors with n_i components.

Using the results of lemma 1, we write $A_{i,i} = D_i T'_i T_i D_i$, $1 \leq i \leq N$, where D_i is a positive diagonal matrix, and T_i is an upper triangular matrix of the form (3.3). Letting

$$D_i X_i = Y_i, \quad D_i^{-1} K_i = M_i, \quad 1 \leq i \leq N, \quad (4.2)$$

this matrix problem reduces to

$$\begin{pmatrix} \tilde{A}_{1,1} & \tilde{A}_{1,2} & \cdots & \tilde{A}_{1,N} \\ \tilde{A}_{2,1} & & & \vdots \\ \vdots & & & \\ \tilde{A}_{N,1} & \cdots & & \tilde{A}_{N,N} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{pmatrix} = \begin{pmatrix} M_1 \\ M_2 \\ \vdots \\ M_N \end{pmatrix}, \quad (4.3)$$

or equivalently

$$\tilde{A} \mathbf{y} = \mathbf{M}, \quad (4.3')$$

where

$$\tilde{A}_{i,j} = D_i^{-1} A_{i,j} D_j^{-1} \quad 1 \leq i, j \leq N. \quad (4.4)$$

In particular,

$$\tilde{A}_{i,i} = T'_i T_i, \quad 1 \leq i \leq N. \quad (4.5)$$

From (2.3), it is clear that \tilde{A} satisfies (2.3), except possibly for (d). While (d) of (2.3) was used in establishing that the matrices A and $A_{i,i}$, $1 \leq i \leq N$, are positive definite, we obtain the same result for the matrices A and $A_{i,i}$, $1 \leq i \leq N$, directly from (4.4) and (4.5).

If superscripts denote iteration indices, then we define the *normalized block relaxation scheme* applied to (4.1) as

$$Y_i^{(l+1)} = \omega \left[(T_i' T_i)^{-1} \left\{ \sum_{j=1}^{i-1} (-\tilde{A}_{i,j}) Y_j^{(l+1)} + \sum_{j=i+1}^N (-\tilde{A}_{i,j}) Y_j^{(l)} + M_i \right\} - Y_i^{(l)} \right] + Y_i^{(l)}, \quad (4.6)$$

where ω is the overrelaxation factor. We can write (4.6) in the form:

$$Y_i^{(l+1)} = \omega [Y_i^{(l+1)*} - Y_i^{(l)}] + Y_i^{(l)}, \quad (4.7)$$

where

$$(T_i' T_i) Y_i^{(l+1)*} \equiv - \left(\sum_{j=1}^{i-1} \tilde{A}_{i,j} Y_j^{(l+1)} + \sum_{j=i+1}^N \tilde{A}_{i,j} Y_j^{(l)} \right) + M_i. \quad (4.7')$$

Having evaluated the right-hand side of (4.7'), (4.7') represents a matrix problem of the form (3.7), where \mathbf{g} is known. As previously mentioned, the solution $Y_i^{(l+1)*}$ of (4.7') can be found directly, with at most two multiplications and two additions per component. We remark that having found the solution \mathbf{y} of (4.3'), the solution \mathbf{x} of (2.1) can be found from (4.2). Note that passing from \mathbf{y} to \mathbf{x} requires but one multiplication per component. Hence, if the number of iterations of (4.6) is large, the work in passing from \mathbf{y} to \mathbf{x} will be quite negligible in comparison.

5. Self-adjoint Partial Differential Equations

We consider the numerical solution of two-dimensional elliptic partial difference equations, arising from discrete approximations to the self-adjoint partial differential equation²

$$-\operatorname{div}\{D(\mathbf{u}) \operatorname{grad} \phi(\mathbf{u})\} + \sigma(\mathbf{u}) \phi(\mathbf{u}) = S(\mathbf{u}), \quad \mathbf{u} \in R, \quad (5.1)$$

where R is a finite connected region in two dimensions, subject to the boundary conditions

$$\alpha(\mathbf{u}) \phi(\mathbf{u}) + \beta(\mathbf{u}) \frac{\partial \phi(\mathbf{u})}{\partial n} = g(\mathbf{u}), \quad \mathbf{u} \in \Gamma, \quad (5.2)$$

where Γ is the exterior boundary of R . Here, the normal derivative refers to the outward normal.

We assume that

- (a) $\alpha(\mathbf{u})$ and $\beta(\mathbf{u})$ are piecewise continuous, and $\alpha > 0$, $\beta \geq 0$ on Γ . (5.3)
- (b) $D(\mathbf{u}) > 0$ in R .
- (c) $\sigma(\mathbf{u}) \geq 0$ in R .

² Problems of this type occur in the multigroup neutron diffusion approximation to the neutron transport equation of reactor physics. See for example [9] and [12]

- (d) Both $\phi(\mathbf{u})$ and $D(\mathbf{u}) \text{ grad } \phi(\mathbf{u})$ are continuous in $R \cup \Gamma$.
 (e) $S(\mathbf{u})$ and $g(\mathbf{u})$ are piecewise continuous in R and Γ , respectively.

We merely state that a matrix problem of the form (2.1) is obtained from the discrete approximation to the above problem by imposing a rectangular³ mesh Λ on $R \cup \Gamma$, and approximating the partial differential equation (5.1) by a five-point formula in two dimensions. Based on (5.3), the matrix A can be derived [12], moreover, in such a way that hypotheses (a), (b), and (d) are satisfied, and that the matrix A satisfies Young's (point) Property A. When a sufficiently fine mesh is chosen, the connectivity of R implies the irreducibility of A , so that (c) of (2.3) is satisfied. If the mesh points of Λ are numbered in the usual manner (cf. [12]) and n_i is the number of interior mesh points in the i th row (line) of Λ , then partitioning, as in (2.2), leads to square submatrices on the diagonal of A which are tridiagonal, and (e) of (2.3) is satisfied. With this partitioning in the two-dimensional case, the associated matrix \tilde{A} of section 4 is of the special tridiagonal block form:

$$\tilde{A} = \begin{pmatrix} \tilde{A}_{1,1} & \tilde{A}_{1,2} & & & \\ \tilde{A}_{2,1} & \tilde{A}_{2,2} & \tilde{A}_{2,3} & & 0 \\ & \diagdown & \diagdown & \diagdown & \\ 0 & & & \tilde{A}_{N-1,N} & \\ & & \diagup & \diagup & \diagup \\ & & & \tilde{A}_{N,N-1} & \tilde{A}_{N,N} \end{pmatrix} \quad (5.4)$$

The matrix \tilde{A} clearly satisfies Property A^* , (f) of (2.3), as well as (point) Property A. Therefore, \tilde{A} satisfies all the conditions of (2.3), so that the results of sections 3-4 are applicable.

Since the approximation to (5.1) is by means of a five-point formula, the block matrices $\tilde{A}_{i,i+1}$ are such that there is at most one nonzero entry per row of $\tilde{A}_{i,i+1}$. Using the symmetry of \tilde{A} , and the fact that $\tilde{A}_{i,i} = T_i' T_i$, where T_i is of the form given in (3.3), it is clear that *only two* coefficients are needed per mesh point to completely specify the matrix \tilde{A} for which the normalized block relaxation scheme is defined. On the other hand, even with symmetry, one would in general⁴ need three coefficients per mesh point to specify the matrix A for the usual application of the block relaxation scheme. This means that iteration of the block (line) relaxation scheme based on the matrix \tilde{A} of (5.4) can in general be

³ The mesh spacings in each coordinate direction need not be constant.

⁴ In reactor problems and various heat conduction problems, nonconstant mesh spacings and nonhomogeneous composition are the rule rather than the exception. It is for these problems that the above remarks concerning coefficient reductions are of interest. In the solution of the Dirichlet problem on a uniform mesh, the above reduction is certainly not gained, since all the coefficients of A are either unity or one-fourth in magnitude.

applied to numerical problems with more mesh points for the same internal storage of a given computing machine.

The normalized block (line) relaxation scheme as applied to the matrix \tilde{A} requires, from (4.6), five multiplications and six additions per mesh point for each iteration. It is *precisely* the same for the point relaxation scheme. Since the matrix \tilde{A} satisfies Property A, Property A*, and the hypotheses of (2.3), we can conclude that the rate of convergence of the block relaxation scheme is at least as fast, and in general faster, than the rate of convergence of the point relaxation scheme applied to \tilde{A} [1, p. 228]. Hence, for the matrix \tilde{A} , an increase in the rate of convergence is in general obtained in passing from point to block relaxation, without obtaining a corresponding increase in the number of arithmetic operations.

For the three-dimensional problem based on a seven-point difference formula, an analogous argument can be made. In this case, only three coefficients are needed per mesh point to specify the matrix \tilde{A} , and point and block relaxation applied to \tilde{A} both require seven multiplications and eight additions per mesh point per iteration.

6. Estimation of the Optimum Relaxation Factor ω

Under the assumptions of (2.3), we shall show how upper and lower bounds can be found for the optimum overrelaxation factor, ω_0 , associated with the successive block overrelaxation scheme, much as has been done for the successive point overrelaxation scheme [12, p. 58–61].

We first consider the square matrix

$$\tilde{B} \equiv \begin{pmatrix} 0 & -\tilde{A}_{1,1}^{-1} \tilde{A}_{1,2} & \cdots & -\tilde{A}_{1,1}^{-1} \tilde{A}_{1,N} \\ -\tilde{A}_{2,2}^{-1} \tilde{A}_{2,1} & 0 & & -\tilde{A}_{2,2}^{-1} \tilde{A}_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ -\tilde{A}_{N,N}^{-1} \tilde{A}_{N,1} & -\tilde{A}_{N,N}^{-1} \tilde{A}_{N,2} & \cdots & 0 \end{pmatrix}, \quad (6.1)$$

obtained from the matrix \tilde{A} of (4.3), which exists since the matrices $\tilde{A}_{i,i}$ are, by (3.5), nonsingular. Letting $\mu[B]$ denote the *spectral radius*⁵ of \tilde{B} , i.e., $\mu[B] = \max_k |\lambda_k|$ where λ_k is an eigenvalue of \tilde{B} , we have

LEMMA 2. *If A satisfies (2.3), then the matrix \tilde{B} is a non-negative matrix, i.e., every entry of \tilde{B} is a non-negative real number. Moreover, $\mu[\tilde{B}] < 1$.*

PROOF. From the assumptions of (2.3), it follows that the entries of $\tilde{A}_{i,j}$, $i \neq j$ are nonpositive real numbers. To prove that \tilde{B} is a non-negative matrix, it suffices to prove that each $\tilde{A}_{i,i}^{-1}$ is a non-negative matrix. Since $\tilde{A}_{i,i}$ is symmetric and positive definite with nonpositive off-diagonal entries, then $\tilde{A}_{i,i}^{-1}$ has non-negative entries⁶ by an early result of Stieltjes [10].⁷

⁵ This is also called the *spectral norm* of a matrix by Young [14, p. 94], although it is not a norm in the usual sense. For \tilde{B} a non-negative matrix, as is the case by lemma 2, this quantity is called the *Jacobi constant for the matrix \tilde{A}* by Ostrowski [7, p. 182].

⁶ If $A_{i,i}$ is in addition irreducible, it can be shown that $A_{i,i}^{-1}$ has every entry positive

⁷ See also [7, p. 188] and [3].

Since \tilde{A} and $\tilde{A}_{i,i}$, $1 \leq i \leq N$, are all symmetric and positive definite, and \tilde{A} satisfies (f) of (2.3), it follows [1, pp. 224-225] that $\bar{\mu}[\tilde{B}] < 1$, which completes the proof.

The Perron-Frobenius theory of non-negative matrices can now be applied to \tilde{B} in order to estimate the optimum value of ω , ω_b , in (4.6). It is known that this optimum overrelaxation factor ω_b , producing the fastest convergence in (4.6), is given explicitly [1] by the formula

$$\omega_b = \frac{2}{1 + \sqrt{1 - \bar{\mu}^2[\tilde{B}]}}. \quad (6.2)$$

THEOREM. Let A satisfy (2.3), and let α be any vector with positive components. If $\tilde{B}\alpha \equiv \beta$, and if $\mu_1(\alpha) \equiv \min_j(\beta_j/\alpha_j)$, $\mu_2(\alpha) \equiv \max_j(\beta_j/\alpha_j)$, then

$$\mu_1(\alpha) \leq \bar{\mu}[\tilde{B}] \leq \mu_2(\alpha). \quad (6.3)$$

Moreover, if $\mu_2(\alpha) \leq 1$, then

$$\frac{2}{1 + \sqrt{1 - \mu_1^2(\alpha)}} \leq \omega_b \leq \frac{2}{1 + \sqrt{1 - \mu_2^2(\alpha)}}. \quad (6.4)$$

PROOF. If \tilde{B} is a non-negative and irreducible matrix, the inequalities of (6.3) follow from the fact [13] that $\bar{\mu}[\tilde{B}]$ can be expressed as a minimax:

$$\max_{\alpha \in R} \left\{ \min_j \left(\frac{\beta_j}{\alpha_j} \right) \right\} = \bar{\mu}[B] = \min_{\alpha \in R} \left\{ \max_j \left(\frac{\beta_j}{\alpha_j} \right) \right\}, \quad (6.5)$$

where R is the set of all vectors \mathbf{u} with positive components. In the general case where B is only a non-negative matrix, the inequalities of (6.3) follow from a lemma of Debreu and Herstein [2, p. 601]. From the formula

$$\omega(\mu) \equiv \frac{2}{1 + \sqrt{1 - \mu^2}},$$

we see that $\omega(\mu)$ is an increasing function of μ for $0 \leq \mu \leq 1$, from which the inequalities of (6.4) follow. This completes the proof.

It can be shown that if the vector α of the theorem is specifically chosen to be the positive vector whose components are the positive diagonal entries of the matrices D_i , $1 \leq i \leq N$, described in section 4, then $\mu_2(\alpha) \leq 1$. Moreover, it can be shown, based on (2.3c), that \tilde{B} has no row of all zero entries. Thus, it follows [12, p. 59] that the repeated application of the above theorem to $\alpha_k \equiv \tilde{B}^k \alpha_0$, $k = 1, 2, \dots$, gives sequences of nondecreasing lower bounds and nonincreasing upper bounds for $\bar{\mu}[B]$.

BIBLIOGRAPHY

1. R. J. ARMS, L. D. GATES, AND B. ZONDEK, A method of block iteration, *J. Soc. Indust. Appl. Math.* 4 (1956), 220-229.
2. G. DEBREU AND I. N. HERSTEIN, Nonnegative square matrices, *Econometrica* 21 (1953), 598-607.

3. E. EGÉRVÁRY, On a lemma of Stieltjes on matrices, *Acta Sci. Math.* 15 (1954), 99-103
4. S. P. FRANKEL, Convergence rates of iterative treatments of partial differential equations, *Math. Tables Other Aids Comp.* 4 (1950), 65-75
5. G. FROBENIUS, Über Matrizen aus nicht negativen Elementen, *Sitzungsberichte der Akademie der Wissenschaften zu Berlin* (1912), pp. 456-477.
6. F. D. MURNAGHAN, *The Theory of Group Representations*, Johns Hopkins Press, Baltimore (1938)
7. A. OSTROWSKI, Determinanten mit überwiegender Hauptdiagonale und die absolute Konvergenz von linearen Iterationsprozessen, *Comment. Math. Helv.* 30 (1956), 175-210.
8. O. PERRON, Zur Theorie der Matrizen, *Math. Ann.* 64 (1907), 259-263
9. R. H. STARK, Rates of convergence in numerical solution of the diffusion equation, *J. Assoc. Comp. Mach.* 3 (1956), 29-40.
10. T. J. STIELTJES, Sur les racines de l'équation $X_n = 0$, *Acta Math.* 9 (1887), 385-400.
11. O. TAUSKY, A recurring theorem on determinants, *Amer. Math. Month.* 56 (1949), 672-675.
12. R. S. VARGA, Numerical solution of the two-group diffusion equation in x - y geometry, *IRE Trans. of the Professional Group on Nuclear Science*, NS-4 (1957), pp. 52-62.
13. H. WIELANDT, Unzerlegbare, nicht negative Matrizen, *Math. Zeit.* 52 (1950), 642-648.
14. DAVID YOUNG, Iterative methods for solving partial difference equations of elliptic type, *Trans. Amer. Math. Soc.* 76 (1954), 92-111.