NUMERICAL METHODS FOR SOLVING MULTI-DIMENSIONAL MULTIGROUP DIFFUSION EQUATIONS

By

RICHARD S. VARGA

Reprinted from
Proceedings of Symposia in Applied Mathematics, vol. XI
Nuclear Reactor Theory, pp. 164-189
1961

NUMERICAL METHODS FOR SOLVING MULTI-DIMENSIONAL MULTIGROUP DIFFUSION EQUATIONS

BY

RICHARD S. VARGA

1. **Introduction.** It is well known [48] that numerical solutions of multigroup diffusion equations have played, and will continue to play, an important part in the design of nuclear reactors. With the availability of digital computing machines with ever-increasing speed, the scope of nuclear design problems which can be numerically attacked with the diffusion approximation has also been increasing, bringing forth interesting new questions for numerical analysts.

We shall concentrate on surveying the available numerical methods for solving the multi-dimensional multigroup diffusion equations. Since Dr. Ehrlich has already sketched the numerical methods available for treating the case of one space variable, we shall therefore concentrate on the cases of several space variables, although in general our theoretical discussions will be phrased independently of the number of space variables. In all cases we shall attempt to discuss both the rigorous mathematical features and the practical applications of these various numerical methods to both the time independent and time dependent diffusion equations.

2. **Statements of the problems.** Let the domain of the nuclear reactor R be a finite connected region in n-dimensional Euclidean space, $n \leqq 3$, with exterior boundary Γ. We assume that there are a finite number of subregions $R_1, R_2, \cdots, R_l$, and internal boundaries[1] γ, between the regions R_i, which together constitute R. This is illustrated in Figure 1. If the number of lethargy groups is m, then [17, p. 291] the time dependent multigroup diffusion equations are:

$$(2.1)\qquad \left\{\frac{1}{v_i}\left(\frac{\partial \varphi_i(x,t)}{\partial t}\right) = \operatorname{div}(D_i(x)\cdot \operatorname{grad} \varphi_i(x,t)) - \sigma_i(x)\varphi_i(x,t) + \sum_{j\neq i} \sigma_{i,j}^{(r)}(x)\varphi_j(x,t)\right\}_{i=1}^{m}, \qquad x \in R_k,\ 1 \leqq k \leqq l,$$

where

$$(2.2)\qquad \sigma_i(x) \equiv \sigma_i^{(a)}(x) + \sum_j \sigma_{j,i}^{(r)}(x).$$

The quantity $\varphi_i(x,t)$ is the neutron flux in the ith lethargy group, and v_i is the average velocity of the neutrons in this group; $D_i(x)$ is the diffusion coefficient, $\sigma_i(x)$ is the total cross-section,[2] and $\sigma_{i,j}^{(r)}(x)$ is the cross-section which removes neutrons from the jth lethargy group to the ith lethargy group, and thus takes into account both up- and down-scattering, as well as fission. We assume that these latter quantities are time dependent, so that changes due to depletion, poisoning, etc., are ignored. By virtue of their physical definitions, we have, for $1 \leqq i,j \leqq m$, that these quantities are continuous in each R_k, $1 \leqq k \leqq l$, with at most finite discontinuities on γ, and satisfy, for $1 \leqq i,j \leqq m$,

$$(2.3)\qquad \begin{cases} 1. & D_i(x) \geqq \delta > 0 \text{ for all } x \in \bar{R}, \\ 2. & \sigma_i^{(a)}(x) \geqq 0 \text{ for all } x \in \bar{R}, \text{ and} \\ 3. & \sigma_{i,j}^{(r)}(x) \geqq 0 \quad \text{while } \sigma_{i,i-1}^{(r)}(x) \geqq \delta_2 > 0 \text{ for all } x \in R. \end{cases}$$

We further assume, for $1 \leqq i \leqq m$, that

$$(2.4)\qquad \begin{cases} 1. & \varphi_i(x,t) \text{ is continuous in } x,\ x \in \bar{R}, \text{ for all } t \geqq 0, \text{ and} \\ 2. & D_i(x)\dfrac{\partial \varphi_i(x,t)}{\partial n} \text{ is continuous across any internal boundary,} \end{cases}$$

$$x \in \gamma, \text{ for all } t \geqq 0.$$

On Γ, the external boundary of R, we assume for simplicity the extrapolated (homogeneous) boundary condition[3] [17, p. 103]

$$(2.5)\qquad \varphi_i(x,t) + a_i(x)\frac{\partial \varphi_i(x,t)}{\partial n} = 0, \qquad x \in \Gamma,$$

where $a_i(x)$ is continuous and non-negative on Γ. To complete the statement of the time dependent multigroup diffusion equations, we are given as initial conditions the functions $\varphi_i(x,0)$, $1 \leqq i \leqq m$.

[1] More precisely, $\gamma = \cup_{i=1}^{l}\{\bar{R}_i - R_i\} - \Gamma$, where $\bar{R}_i$ denotes the closure of R_i.

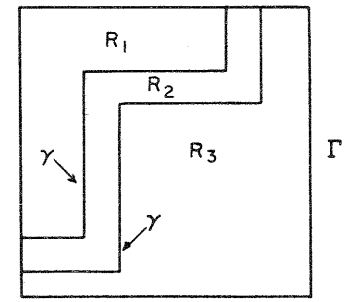

FIGURE 1.

[2] If $n < 3$, then $\sigma_i(x)$ includes also the contributions of buckling terms, arising from assumptions of separability. See [17, Chapter 7].

[3] The normal here refers to the outward pointing normal.

For the time independent multigroup diffusion equations, we have

$$\left\{-\operatorname{div}\left(D_i(x)\cdot\operatorname{grad}\varphi_i(x)\right)+\sigma_i(x)\varphi_i(x)=\sum_{j<i}\sigma_{i,j}^{(r)}(x)\varphi_j(x)+\frac{\chi_i\psi(x)}{\lambda}\right\}_{i=1}^{m}, \tag{2.6}$$

where the fission source of neutrons, $\psi(x)$, is defined by

$$\psi(x)\equiv\sum_i[\nu\sigma_f(x)]_i\varphi_i(x). \tag{2.7}$$

The probabilities χ_i are non-negative scalars with $\sum_i\chi_i=1$. The quantity $[\nu\sigma_f(x)]_i$ is related to the macroscopic fission cross-section, and is thus continuous in each R_k, $1\leqq k\leqq l$, with at most finite discontinuities on γ, and is non-negative. The homogeneous boundary conditions of (2.5), and the continuity conditions of (2.4), are also assumed to apply, in the time independent problem, to the fluxes $\varphi_i(x)$, as well as to the fission source $\psi(x)$.

For the time dependent problem, we seek solutions of (2.1) for given initial conditions, and we are especially interested in the behavior of $\varphi_i(x,t)$ for large positive values of t. For the time independent problem, which is an eigenvalue problem, we seek to determine solutions of (2.6) corresponding to the largest (in modulus) eigenvalue λ of (2.6).

The theoretical results for discrete approximations to (2.1) and (2.6) are strongly dependent on the Perron-Frobenius theory [**28**; **14**; **50**; **7**] of non-negative matrices, and suitable extensions thereof.[4] It is interesting, but not surprising, that the practical numerical methods which have been used to solve the discrete approximations to (2.1) and (2.6) are also greatly influenced by this same theory of non-negative matrices. In later sections, we shall give specific numerical applications which result from this theory.

3. The discrete space matrix problem for the time independent multigroup diffusion equations. We now pass from the continuous space problem of §2 to the discrete space diffusion equations, which are actually used in numerical computations. To derive the set of coupled matrix equations approximating (2.6), we assume for simplicity that we are using cartesian coordinates in the plane, $n=2$, and that the subregions R_i are finite sums of rectangles,[5] as shown in Figure 1. This enables us to impose a mesh[6] Λ of horizontal and vertical line segments on the plane in such a way that all internal and external boundaries segments of γ and Γ coincide with segments of Λ. With the mesh Λ, the unknowns, numbering N, for the ith lethargy group in the discrete case are then defined to be the approximate values of φ_i at the inter-

[4] Since non-negative square matrices leave the positive hyperoctant invariant, the application of the Perron-Frobenius theory of non-negative matrices in the discrete case is closely related to the abstractions in the papers in this same volume by G. Birkhoff, and G. Habetler and M. A. Martino.

[5] If cylindrical coordinates are used in the plane, for example, a corresponding treatment of the derivation of difference equations can be given.

[6] The mesh Λ need not be uniform in either coordinate direction.

section of the horizontal and vertical line segments of Λ. By replacing the differential equation of (2.6) with difference equations in the unknowns of the discrete space, we will have the discrete time independent diffusion equations.

To explicitly derive difference equations approximating (2.6), the derivation which is based on integrating (2.6) seems to be more widely used, and gives rise to certain matrix properties which are very useful. Integrating (2.6) over the mesh region $S(x_0)$ indicated in Figure 2, we have, using Green's theorem,

$$-\oint D(x)\frac{\partial\varphi_i(x)}{\partial n}\,dl+\int_{S(x_0)}\sigma_i(x)\varphi_i(x)\,dA=\sum_{j<i}\int_{S(x_0)}\sigma_{i,j}^{(r)}(x)\varphi_j(x)dA+\frac{\chi_i}{\lambda}\int_{S(x_0)}\psi(x)dA, \tag{3.1}$$

where the line integral is taken over the boundary of $S(x_0)$. Approximating the left-hand side of (3.1) by a five-point formula, which linearly relates $\varphi_i(x_0)$ with $\varphi_i(x_1),\cdots,\varphi_i(x_4)$, we obtain [**40**] the following discrete approximation to (2.6):

$$\left\{A_i\boldsymbol{\varphi}_i=\sum_{j<i}B_{i,j}\boldsymbol{\varphi}_j+\frac{1}{\lambda}\mathrm{X}_i\boldsymbol{\psi}\right\}_{i=1}^{m} \tag{3.2}$$

where

$$\boldsymbol{\psi}=\sum_{i=1}^{m}V_i\boldsymbol{\varphi}_i. \tag{3.3}$$

As shown in [**40**], making use of the boundary conditions of (2.5), and the conditions of (2.3)–(2.4), the quantities A_i, $B_{i,j}$, X_i, and V_i are all $N\times N$ matrices, with the following properties:

(3.4)
- (a) $B_{i,j}$, X_i, and V_i are all non-negative diagonal matrices, with $B_{1,j}\equiv 0$ for $1\leqq j\leqq m$.
- (b) If $A_i\equiv(a_{k,l}^{(i)})$, then $a_{k,k}^{(i)}>0$, and $a_{k,l}^{(i)}\leqq 0$, and $a_{k,l}^{(i)}=a_{l,k}^{(i)}$ for all $1\leqq k,l\leqq N$, $1\leqq i\leqq m$.
- (c) $a_{k,r}^{(i)}>\sum_{l\neq k}|a_{k,l}^{(i)}|$ for all $1\leqq k\leqq N$, $1\leqq i\leqq m$.

But even more can be proved about the matrices A_i. First, we make the following

DEFINITION 1. An $n\times n$ matrix $M=(m_{i,j})$ is irreducible[7] if, for any i and j, $1\leqq i,j\leqq n$, there exists a finite sequence of integers

$$k(0)=i,k(1),\cdots,k(r)=j,$$

such that $m_{k(h-1),\,k(h)}\neq 0$ for $1\leqq h\leqq r$.

[7] This is also called *indecomposable*, as in [**7**; **31**]. An equivalent definition of irreducibility is given by Geiringer in [**16**].

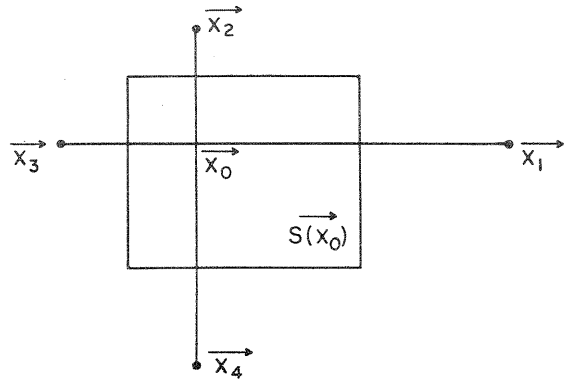

FIGURE 2.

In other words, irreducibility implies that every integer i is connnected to every integer j, $1 \leqq i,j \leqq n$, through non-zero coefficients of the matrix M. Thus, remembering that the domain of our nuclear reactor R is connected, then with a sufficiently fine mesh Λ, it follows that, since every mesh point is coupled to its four adjacent mesh points, as in Figure 2, every mesh point of Λ is ultimately connected to every other mesh point of Λ through non-zero coefficients.

THEOREM 1. *The $N \times N$ matrices $A_i = (a_{k,l}^{(i)})$ are irreducible, symmetric and positive definite. Moreover, each matrix A_i^{-1} is a positive matrix, i.e., every entry of A_i^{-1} is a positive real number.*

The positive definite nature of A_i follows [**38**] directly from c of (3.4). The conclusion about A_i^{-1} is an extension [**15**; **7**; **2**; **26**] of an old result due to Stieltjes [**37**], and is related to the positivity of the Green's function for the differential operator defined by the left-hand side of (2.6).

Although in the next section we shall make use of still more properties of the matrices A_i, we can with the results above demonstrate that the largest (in modulus) eigenvalue of the discrete eigenvalue problem of (3.2) is unique. Since $B_{1,j} \equiv 0$, then from (3.2),

$$\phi_1 = A_1^{-1}\mathrm{X}_1(\psi/\lambda)\,; \phi_2 = A_2^{-1}\{B_{2,1}\,A_1^{-1}\mathrm{X}_1 + \mathrm{X}_2\}(\psi/\lambda),$$

and in general

$$\phi_i = L_i(\psi/\lambda), \qquad 1 \leqq i \leqq m. \tag{3.5}$$

If we define

$$T \equiv \sum_{i=1}^{m} V_i L_i, \tag{3.6}$$

then, with the use of Equations (3.3) and (3.5), the eigenvalue problem of (3.2)–(3.3) takes the form

$$T\psi = \lambda\psi. \tag{3.7}$$

It is clear from the derivation of the diagonal matrices X_i that the diagonal entries of X_i are precisely χ_i times the mesh areas $S(x_j)$ about each mesh point x_j. Since, without loss of generality, we may assume $\chi_1 > 0$, it follows that X_1 is a positive diagonal matrix. Similarly, from (2.3) we also have that the matrices $B_{i,i-1}$, $1 < i \leqq m$, are positive diagonal matrices. Thus, since each A_i^{-1} is a positive matrix from Theorem 1, we can inductively show that each L_i, $1 \leqq i \leqq m$, is also a positive matrix. By definition, each matrix V_i is a non-negative diagonal matrix. We now assume that at least one of the sub-regions R_k contains some fissionable material. By the same permutation of rows and columns of the matrices V_i, we can write

$$V_i = \begin{pmatrix} d_1(i) & & & \\ & d_2(i) & & \\ & & \ddots & \\ & & & d_N(i) \end{pmatrix}, \tag{3.8}$$

and there exists a non-negative integer r such that $d_l(i) = 0$ for $1 \leqq i \leqq r$, for *all* $1 \leqq l \leqq m$, and for each i, $r + 1 \leqq i \leqq N$, there is at least one $l(i)$ such that $d_i(l(i)) > 0$. From the definition of the matrix T in (3.6) we have

$$T = \left(\begin{array}{c|c} 0 & 0 \\ \hline T_1 & T_2 \end{array}\right), \tag{3.9}$$

where T_1 is an $(N - r) \times r$ matrix, and T_2 is a square $(N - r) \times (N - r)$ matrix. It thus follows that T_2 is a positive matrix, using the fact that each L_i is a positive matrix. Now, the eigenvalues of T are the $(N - r)$ eigenvalues of T_2 and an r-fold zero eigenvalue.[8] By a theorem of Perron [**28**], T_2 possesses an eigenvalue λ^* which is positive, simple, and greater in modulus than all other eigenvalues of T_2. Moreover, to λ^* can be associated a unique[9] eigenvector Ψ, of T_2 with positive components. This is the basis of our

THEOREM 2. *The largest (in modulus) eigenvalue λ^* of T is positive, simple, and its corresponding unique eigenvector Ψ can be chosen to have non-negative components. Furthermore, for any arbitrary positive vector ψ_0, the iteration procedure*

$$T\psi_n = S_{n+1}; \quad \lambda_{n+1} = \frac{(S_{n+1}, \psi_n)}{(\psi_n, \psi_n)}; \quad \psi_{n+1} \equiv S_{n+1}/\lambda_{n+1}, \qquad n \geqq 0 \tag{3.10}$$

is convergent, and

$$\lim_{n\to\infty} \psi_n = c\Psi, \quad \lim_{n\to\infty} \lambda_n = \lambda^* \tag{3.11}$$

where c is some positive scalar.

[8] It is interesting to point out that the number r of zero eigenvalues is at least as large as the number of mesh points in non-fissionable regions.

[9] Up to scalar factors.

The first assertion of the above theorem is a direct consequence of Perron's theorem. The iterative method of (3.10) is commonly referred to as the *power method* of iteration, and the convergence of this iterative procedure is guaranteed by the fact that λ^* exceeds in modulus all other eigenvalues of T.

COROLLARY. *If*

$$\beta_{n+1} = \max_i \left(\frac{S_{n+1,i}}{\psi_{n,i}}\right), \quad \textit{and} \quad \alpha_{n+1} = \min_i \left(\frac{S_{n+1,i}}{\psi_{n,i}}\right),$$

where $S_{n+1,i}$ denotes the ith component of $\boldsymbol{S}_{n+1}$, and the subscript i varies over the set of indices for which $\psi_{n,i} \neq 0$, then

$$\alpha_{n+1} \leqq \alpha_{n+2} \leqq \beta_{n+2} \leqq \beta_{n+1}, \qquad n \geqq 0, \tag{3.12}$$

and

$$\alpha_{n+1} \leqq \lambda^* \leqq \beta_{n+1}, \qquad n \geqq 0. \tag{3.13}$$

The inequalities $\alpha_{n+1} \leqq \lambda^* \leqq \beta_{n+1}$ follows from a result due to Collatz [4]. The inclusion property (3.12) of the bounds α_n and β_n for λ^* follows [40] from the fact that T_2 is a positive matrix.

We remark that Theorem 2 assures the existence of a largest (in modulus) eigenvalue λ^* of the time independent problem (3.2). Moreover, machine computations based on (3.10) are necessarily convergent, and Equation (3.13) of the Corollary gives for each iteration of (3.10) non-trivial upper and lower bounds on λ^*, which is of considerable practical use, since λ^* corresponds physically [17, p. 86] to the effective multiplication factor K_{eff}. The bounds of (3.13) are also of numerical importance in determining when the iterative procedure of (3.10) has been carried far enough.[10]

Theoretically, we can perform the iterative process of (3.10) by solving successively the system of equations:

$$\left\{A_i \boldsymbol{\Phi}_i^{(n+1)} = \sum_{j<i} B_{i,j} \boldsymbol{\Phi}_j^{(n+1)} + \frac{\mathrm{X}_i \boldsymbol{\psi}^{(n)}}{\lambda_n}\right\}_{i=1}^{m}. \tag{3.14}$$

By defining

$$\boldsymbol{\psi}^{(n+1)} = \sum_i V_i \boldsymbol{\Phi}_i^{(n+1)}; \quad \lambda^{n+1} \equiv \lambda_n \frac{(\boldsymbol{\psi}^{(n+1)}, \boldsymbol{\psi}^{(n)})}{(\boldsymbol{\psi}^{(n)}, \boldsymbol{\psi}^{(n)})}, \tag{3.15}$$

it follows that $\boldsymbol{\psi}^{(n+1)} = T(\boldsymbol{\psi}^{(n)}/\lambda_n)$. Thus, we can carry out the iterative process of (3.10) without the explicit determination of the matrix T if we can solve matrix equations of the form

$$A_k \boldsymbol{x} = \boldsymbol{S}, \tag{3.16}$$

[10] As mentioned by Dr. Ehrlich in his paper, $(\beta_n - \alpha_n)/\lambda_n < \epsilon$ is a useful criterion for terminating the iterative procedure of (3.10).

where A_k is a matrix with properties as given in Theorem 1, and $\boldsymbol{S}$ is a given column vector. The numerical solution of such systems of linear equations will be discussed in the next section.

From a numerical point of view, it is fortunate that the matrix T of (3.6) need not be explicitly calculated in order to carry out the iterative procedure of (3.10), since the explicit computing machine storage of the coefficients of T for typical cases presents a serious problem. This storage difficulty stems from the fact that, by Theorem 1, the matrices A_i have positive inverses, and thus the matrix T has of the order of N^2 non-zero entries, where the matrices T and A_i all are $N \times N$ matrices.[11]

4. Iterative numerical methods for solving systems of linear equations. We now concentrate on the numerical problems of approximating the solution of

$$A\boldsymbol{x} = \boldsymbol{k}, \tag{4.1}$$

where the $N \times N$ matrix A satisfies the conditions of Theorem 1, and $\boldsymbol{k}$ is a given column vector with N components. If D is the positive diagonal matrix, composed of the diagonal entries of A, then we define the square matrix $M = (m_{i,j})$ as

$$M \equiv I - D^{-1} A; \tag{4.2}$$

we shall call M the (point) *Jacobi matrix* derived from the matrix A. By definition, M has zero diagonal entries, and, from Theorem 1, all the entries in M are non-negative real numbers, which we indicate by $M \geqq 0$, and, from (4.2) and Theorem 1, M is irreducible.

DEFINITION 2. The *spectral radius* $\bar{\mu}[C]$ of an arbitrary square matrix C is the maximum of the absolute values of its eigenvalues. If $\bar{\mu}[C] < 1$, then C is *convergent*, and

$$R[C] = -\ln \bar{\mu}[C] \tag{4.3}$$

is the *rate of convergence* [52] of the matrix C.

From Theorem 1, it is easily verified that the eigenvalues λ_k of M are real, with $-1 < \lambda_k < +1$, and thus M is convergent. We now decompose M into

$$M = L + U, \tag{4.4}$$

[11] At the present time, for two-dimensional calculations based on (3.14), several codes exist for the numerical solution of the time independent diffusion equations. One such code, PDQ-4 of the Westinghouse Atomic Power Laboratory, written for the Philco-2000 allows as many as $N = 20{,}000$ mesh points, and at most four energy groups. Typical computing times, for $\beta_n - \alpha_n/\lambda_n < .0025$, are of the order of one-half hour.

where L and U are respectively strictly[12] lower and upper triangular matrices. With $D^{-1}k \equiv g$, we write (4.1) as

$$x = (L + U)x + g = Mx + g. \tag{4.5}$$

Given an initial approximation $x^{(0)}$ to the unique solution x^* of (4.1) the iterative method defined by

$$x^{(n+1)} = (L + U)x^{(n)} + g, \qquad n \geqq 0, \tag{4.6}$$

is called the *Jacobi* or *total-step method*. If $\epsilon^{(n)} \equiv x^{(n)} - x^*$ denotes the error for any iterate, then it follows that

$$\epsilon^{(n)} = M\epsilon^{(n-1)} = \cdots = M^n\epsilon^{(0)}. \tag{4.7}$$

Since $\epsilon^{(n)}$ tends to the zero vector $\mathbf{0}$ if and only if M is convergent [**25**], it follows that the iterative procedure of (4.6) is necessarily convergent. However, other iterative methods based on (4.5) may be more rapidly convergent. Consider the iterative method defined by

$$x^{(n+1)} = Lx^{(n+1)} + Ux^{(n)} + g, \qquad n \geqq 0, \tag{4.8}$$

which is called the *Gauss-Seidel* or *single-step method*. Since L is strictly lower triangular, then $(I - L)$ is non-singular, and we can write (4.8) in the equivalent form:

$$x^{(n+1)} = (I - L)^{-1}Ux^{(n)} + (I - L)^{-1}g, \qquad n \geqq 0. \tag{4.8'}$$

Using the fact that the matrix M has non-negative entries and is convergent, it follows from the work of Stein and Rosenberg [**34**] that the matrix $(I - L)^{-1}U$ is also convergent, and moreover

$$R[(I - L)^{-1}U] > R[M]. \tag{4.9}$$

The proof of this last statement uses only the non-negative irreducible and convergent nature of the matrix M. In order to sharpen this last result, as well as introduce the basis for the successive overrelaxation iterative method of Young and Frankel [**52**; **12**], we make the following definition.

DEFINITION 3. A square matrix C is *cyclic of index* 2 [**31**] if and only if there exists a permutation matrix P such that

$$PCP^{-1} = \begin{pmatrix} 0 & C_1 \\ C_2 & 0 \end{pmatrix}, \tag{4.10}$$

where the null diagonal submatrices are square.

Cyclic matrices have been extensively discussed in the literature [**14**; **31**; **49**], and it can be shown [**40**; **52**] that the Jacobi matrices M_i derived from the matrices A_i of Theorem 1, are cyclic of index 2. In the familiar terminology of Young [**52**], the matrices A_i are said to satisfy *property* (A). These

[12] The matrix $L_1 = (l_{i,j})$, for example, is a strictly lower triangular matrix if $l_{i,j} = 0$ for all $j \geqq i$.

concepts are equivalent in the sense that if the matrix A has non-zero diagonal entries, then the matrix A satisfies property (A) of Young if and only if its (point) Jacobi matrix M of (4.2) is cyclic of index 2 [**41**].

We now turn to the *successive overrelaxation iterative method* of Young and Frankel [**52**; **12**], which is defined, in terms of (4.5), by

$$x^{(n+1)} = x^{(n)} + \omega\{Lx^{(n+1)} + Ux^{(n)} + g - x^{(n)}\}, \qquad n \geqq 0, \tag{4.11}$$

which we can write equivalently as

$$x^{(n+1)} = (I - \omega L)^{-1}\{\omega U + (1 - \omega)I\}x^{(n)} + \omega(I - \omega L)^{-1}g. \tag{4.12}$$

The quantity ω in (4.11) is called the *relaxation factor*. We observe that, for $\omega = 1$, this iterative method reduces to the Gauss-Seidel iterative method of (4.8)–(4.8'). For reasons of brevity, we shall say that a matrix C, which is cyclic of index 2, is *consistently ordered* [**52**] if it is the form of (4.10). With the concept of a consistent ordering, Young [**52**] established the following general relationship between the eigenvalues λ of the successive overrelaxation matrix

$$\mathscr{L}_\omega \equiv (I - \omega L)^{-1}\{\omega U + (1 - \omega)\cdot I\}, \tag{4.13}$$

and the eigenvalues μ of the Jacobi matrix M of (4.2):

$$(\lambda + \omega - 1)^2 = \lambda\omega^2\mu^2. \tag{4.14}$$

With

$$\omega_b \equiv \frac{2}{1 + \sqrt{1 - \bar{\mu}^2[M]}}, \tag{4.15}$$

Young [**52**] established the following fundamental result.

THEOREM 3. *Let A be a symmetric and positive definite matrix, and let its corresponding Jacobi matrix M be cyclic of index 2 and consistently ordered. Then*

$$R(\mathscr{L}_{\omega_b}) > R(\mathscr{L}_\omega) \text{ for any } \omega \neq \omega_b. \tag{4.16}$$

Moreover, as $\bar{\mu}[M] \to 1-$, then

$$R(\mathscr{L}_{\omega_b}) \sim 2[R(\mathscr{L}_{\omega=1})]^{1/2} = 2\sqrt{2}[R(M)]^{1/2}. \tag{4.17}$$

We remark that, by using the fact that the Jacobi matrix M is cyclic of index 2 and consistently ordered, the inequality of (4.9) can be sharpened to

$$R[\mathscr{L}_1] = R[(I - L)^{-1}U] = 2R[M], \tag{4.9'}$$

by employing (4.14). Moreover, the asymptotic result of (4.17) shows that the successive overrelaxation iterative method is considerably faster in rate of convergence than either the (point) Jacobi or Gauss-Seidel iteration methods, when the latter methods are slowly convergent. To illustrate this,

we consider the numerical solution of the Dirichlet problem for the unit square over a uniform mesh of side $h = 1/p$. Then, as shown in [52], $\bar{\mu}[M] = \cos \pi h$, and $R(M) \sim \pi^2 h^2/2$, as $h \to 0$. But from (4.17), $R(\mathscr{L}_{\omega_b}) \sim 2\pi h$, as $h \to 0$, and thus the successive overrelaxation method (with optimum relaxation factor) has a rate of convergence for this problem which is an order of magnitude greater, as $h \to 0$, than the rates of convergence of either the corresponding (point) Jacobi or Gauss-Seidel iterative method. Since the application of the successive overrelaxation method requires only a small increase in arithmetic operations over, say, the Gauss-Seidel iterative method, the overall gain in computing time can be great.

The relaxation factor ω_b, by virtue of its definition, satisfies $1 \leqq \omega_b < 2$, and while it has been first demonstrated by Young [52] that overestimating ω_b is not as detrimental as underestimating ω_b, it has nevertheless been a difficult problem, from a numerical point of view, to estimate ω_b [33; 55]. Clearly, estimating ω_b is equivalent, from (4.15), to estimating $\bar{\mu}[M]$. Now, we can make use of the fact that the matrix M is non-negative and irreducible, and we express its spectral radius $\bar{\mu}[M]$ as a minimax, in the following way. If $\boldsymbol{x}$ is a vector with positive components x_i, and $M\boldsymbol{x} = \boldsymbol{y}$, then [49]

$$\max_{x \in P} \{\min_i (y_i/x_i)\} = \bar{\mu}[M] = \min_{x \in P} \{\max_i (y_i/x_i)\}, \tag{4.18}$$

where P is the set of all vectors $\boldsymbol{u}$ with positive components. It is clear how this expression can be used numerically to find non-trivial upper and lower bounds for $\bar{\mu}[M]$, which in turn from (4.15) gives non-trivial bounds for ω_b [40; 6]. Similiar remarks hold for the Gauss-Seidel matrix

$$\mathscr{L}_1 = (I - L)^{-1} U,$$

which is a non-negative matrix since

$$(I - L)^{-1} = I + L + L^2 + \cdots + L^{n-1}.$$

In fact, it can be shown that Theorem 2 and its Corollary are valid if the matrix T is replaced by the Gauss-Seidel matrix $\mathscr{L}_1$. This gives a convergent method for obtaining upper and lower bounds for $\bar{\mu}[\mathscr{L}_1]$, which is equal to $\bar{\mu}^2[M]$ by (4.9′).

To improve the rate of convergence of the successive overrelaxation method, we consider now the combination of partitioning and factoring of certain matrices. Here, we attempt to directly invert smaller simultaneous systems of linear equations, combined with an iterative method. In theory, several results indicate, for matrices having the properties as given in (3.4) and Theorem 1, that the direct inversion of larger submatrices, coupled to a successive overrelaxation iterative method, does indeed improve the rate of convergence [1; 10; 19; 20; 43]. At this point, we consider solving (4.1), under the assumption that the matrix $A \equiv (a_{i,j})$ is symmetric, positive definite, and *tridiagonal*, i.e., $a_{i,j} = 0$ for $|i - j| > 1$. For the case of a single space variable, the matrices A_i of (3.2), derived in the manner of §3,

couple each mesh point only to its adjacent neighbors, so that these assumptions are fulfilled. Since A is by hypothesis symmetric and positive definite, we can write [23]

$$A = T'T, \tag{4.19}$$

where $T \equiv (t_{i,j})$ is a uniquely determined upper bidiagonal matrix, i.e., $t_{i,j} = 0$ if $i > j$, or $i + 1 < j$, and the diagonal entries $t_{i,i}$ are positive real numbers. The entries of the matrix T are easily found by means of a recurrence relation, and we solve (4.1) by solving first

$$T'\boldsymbol{y} = \boldsymbol{k} \tag{4.20}$$

directly for $\boldsymbol{y}$, followed by

$$T\boldsymbol{x} = \boldsymbol{y}, \tag{4.20′}$$

which is solved directly for $\boldsymbol{x}$. As Dr. Ehrlich has pointed out, in one-dimensional problems, the solution of the matrix equation (4.1) can be efficiently carried out directly.[13] In the case of two and three space variables, an iterative solution of the matrix problem (4.1) is generally used in actual computing machine codes. The iterations of this latter type are called *inner iterations*, in contrast to the iterations of (3.9), which are called *outer iterations*.

Returning to the case of two space variables, we can think of a two-dimensional mesh Λ as a coupled system of horizontal (or vertical) mesh lines. This partitions the matrix A and the vectors $\boldsymbol{x}$ and $\boldsymbol{k}$ of (4.1) in a natural way,

$$A = \begin{bmatrix} B_1 & C_1 & & & \\ C_1' & B_2 & C_2 & & \\ & \ddots & \ddots & \ddots & \\ & & & C_{r-1} & \\ & & C_{r-1}' & B_r & \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ \cdot \\ X_r \end{bmatrix}, \quad \vec{K} = \begin{bmatrix} K_1 \\ K_2 \\ \cdot \\ \cdot \\ \cdot \\ K_r \end{bmatrix}, \tag{4.21}$$

where the diagonal submatrices B_i are square, and the entries of B_i correspond to the coupling of mesh points on the same horizontal mesh line. Thus, the matrices B_i are symmetric and positive definite tridiagonal matrices, and the iterative method

[13] The familiar method of Gauss elimination for solving (4.1) is similar to the particular method of factoring the tridiagonal matrix A above, since in Gauss elimination the matrix A is factored into the form $A = DU$, where D is a lower bidiagonal matrix, and U is an upper bidiagonal matrix with unit diagonal entries. Both direct methods are numerically efficient.

$$B_j\overset{*}{X}_j^{(n+1)} = -C'_{j-1}X_{j-1}^{(n+1)} - C_jX_{j+1}^{(n)} + K_j,\ 1 \leqq j \leqq r, \qquad n \geqq 0, \tag{4.22}$$

$$X_j^{(n+1)} = X_j^{(n)} + \omega\{\overset{*}{X}_j^{(n+1)} - X_j^{(n)}\}, \qquad 1 \leqq j \leqq r, \tag{4.22'}$$

which we call the *successive line overrelaxation method* (SLOR), can be carried out [**1**; **6**; **13**; **20**]. The theory for the iterative method SLOR is a generalization [**1**] of Young's work [**52**]. To indicate the possible gains in rate of convergence of the iterative method SLOR over that of the (point) successive overrelaxation method (SOR), we again consider the numerical solution of the Dirichlet problem for the unit square with uniform mesh spacings $h = 1/p$. It is known [**1**; **13**; **20**] that if $\mathscr{L}_{\omega_b}^{(2)}$ denotes the iteration matrix corresponding to the iterative method SLOR, then $R(\mathscr{L}_{\omega_b}^{(2)}) \sim 2\sqrt{2}\pi h$, as $h \to 0$, and a gain of a factor $\sqrt{2}$ in the asymptotic rate of convergence is achieved. While other iterative methods exist [**43**] which are based on the direct inversion of larger submatrix equations, the iterative method SLOR shows generally how these are applied, and what type improvements in rate of convergence can be obtained.

In terms of actual arithmetic operations, the direct inversion of large submatrices combined with an iterative method tends to increase the amount of arithmetic operations per mesh point. This is, of course, to be balanced by an increase in the rate of convergence. In the case of the iterative method SLOR, either in two or three space dimensions, it fortunately can be shown [**6**] by suitable normalization of equations, that no additional arithmetic operations are required for the successive line overrelaxation method (SLOR) over what is required by the successive point overrelaxation method, while an improvement in rate of convergence is always obtained.

The direct inversion of tridiagonal matrix equations leads us to newer iterative methods, called the implicit alternating direction (IAD) methods. In the case of two space variables, we can decompose the matrix A, derived from a five-point approximation in the plane, into

$$A = H + V + \Sigma. \tag{4.23}$$

Here, the matrices H and V are symmetric and positive definite matrices, which are each, after suitable permutation of indices, tridiagonal matrices. The matrix Σ is a non-negative diagonal matrix. Recalling that tridiagonal matrix equations are efficiently solved by the Gauss elimination method, we consider now the Peaceman-Rachford iterative method [**27**], a particular variant of the IAD methods, which is defined by

$$(H + \Sigma/2 + \rho_nI)\,x^{(n+1/2)} = k + (\rho_nI - \Sigma/2 - V)x^{(n)}, \tag{4.24}$$

$$(V + \Sigma/2 + \rho_nI)x^{(n+1)} = k + (\rho_nI - \Sigma/2 - H)x^{(n+1/2)}. \tag{4.24'}$$

Combining the above equations, we have

$$x^{(n+1)} = T_{\rho_n}x^{(n)} + h_{\rho_n}, \tag{4.25}$$

where

$$T_\rho \equiv (V + \Sigma/2 + \rho I)^{-1}(\rho I - \Sigma/2 - H)\cdot(H + \Sigma/2 + \rho I)^{-1}(\rho I - \Sigma/2 - V), \tag{4.25'}$$

and

$$h_\rho = (V + \Sigma/2 + I)^{-1}\{I + (\rho I - \Sigma/2 - H)(\rho I + \Sigma/2 + H)^{-1}\}k. \tag{4.25''}$$

We shall call T the Peaceman-Rachford matrix.

If the matrices H, V, and Σ all commute with one another, i.e.,

$$HV = VH,\ H\Sigma = \Sigma H,\ V\Sigma = \Sigma H, \tag{4.26}$$

then there exists a common basis of eigenvectors $\{\alpha_j\}_{j=1}^N$ for which

$$H\alpha_j = \sigma_j\alpha_j,\ V\alpha_j = \tau_j\alpha_j,\ \Sigma\alpha_j = \nu_j\alpha_j, \qquad 1 \leqq j \leqq N, \tag{4.27}$$

and we can prove [**27**; **46**] the following result.

THEOREM 4. *If* (4.26) *is valid, then the Euclidean norm of the error vector* ϵ_m *after* m *applications of the Peaceman-Rachford process* (4.24)–(4.24') *satisfies*

$$\frac{\|\epsilon_m\|}{\|\epsilon_0\|} \leqq \sup_k \prod_{r=1}^{m} \left| \frac{\sigma_k - \rho_r}{\sigma_k + \rho_r} \cdot \frac{\tau_k - \rho_r}{\tau_k + \rho_r} \right|. \tag{4.28}$$

Unfortunately, the equalities in (4.26) are the exception, rather than the rule [**3**], in discrete approximations to typical reactor problems. Nevertheless, some positive results can be proved [**3**; **47**] which we include.

THEOREM 5. *If* H *and* V *are symmetric and positive definite matrices, and* Σ *is a non-negative definite matrix, then the Peaceman-Rachford matrix* T_ρ *is convergent for any fixed* $\rho > 0$.

For purposes for comparison, we again consider the numerical solution of the Dirichlet problem for the unit square with uniform mesh spacings $h = 1/p$. For this problem, it can be shown that (4.26) is valid, and that the average rate of convergence $R[T_{\{\rho_i\}}]$ of the Peaceman-Rachford iterative method, with suitably selected parameters ρ_i, is [**9**; **56**]

$$R[T_{\{\rho_i\}}] = 0\left(\frac{1}{|\ln h|}\right), \text{ as } h \to 0,$$

which is an order of magnitude faster in rate of convergence than the iterative methods SLOR and SOR, (with optimum ω's), for this problem. Numerical experience with the Peaceman-Rachford iterative method suggests that it compares favorably in rate of convergence with the successive overrelaxation iterative method for very slowly convergent problems, even when (4.26) does not hold. For the special case in which $\Sigma \equiv 0$ and mesh spacings are uniform, a partial theoretical corroboration of this can be found in [**44**].

Other variants of the IAD methods exist [**8**; **47**], but the results that we

have described for the particular variant, the Peaceman-Rachford iterative method, generally hold for the other variants [3]. Another variant, the Douglas-Rachford iterative method [8], generalizes to the case of three space variables, whereas the Peaceman-Rachford method does not. Further research in this area is definitely in order, and it is hoped that a fuller theoretical understanding of this approach to the iterative solution of the matrix equation (4.1) will soon be reached.

We briefly mention the use of orthogonal polynomials [21; 36] to accelerate the basic iterative method of (4.6). If the vectors $x^{(n)}$ are defined from (4.6), we consider the sequence of vectors

$$\zeta^{(n)} \equiv a_{n,0}x^{(n)} + a_{n,1}x^{(n-1)} + \cdots + a_{n,n}x^{(0)}, \qquad n \geqq 0. \tag{4.29}$$

If $x^{(0)}$ is the unique solution of (4.1), then $x^{(n)} = x^{(0)}$ for all n; if the same is to be true for the vectors $\zeta^{(n)}$, then

$$\sum_{j=0}^{n} a_{n,j} = 1 \text{ for all } n \geqq 0. \tag{4.30}$$

With (4.30), the error vector $\overset{*}{\epsilon}{}^{(n)}$ for the vector iterates $\zeta^{(n)}$ are, from (4.7),

$$\overset{*}{\epsilon}{}^{(n)} = \sum_{j=0}^{n} a_{n,j}\,\epsilon^{(j)} = \left(\sum_{j=0}^{n} a_{n,j}M^j\right)\epsilon^{(0)}. \tag{4.31}$$

Thus, if $p_n(x) \equiv \sum_{j=0}^{n} a_{n,j}x^j$, then (4.30) asserts that $p_n(1) = 1$. Now the eigenvectors β_k of M, where $M\beta_k = \lambda_k\beta_k$, span the vector space $V_N(R)$ over the real numbers, and the eigenvalues λ_k of M are real and satisfy

$$-1 < \lambda_k < +1.$$

If $\epsilon^{(0)} = \sum_k c_k\beta_k$, then $\overset{*}{\epsilon}{}^{(n)} = \sum_k c_k p_n(\lambda_k)\beta_k$, and we thus seek to minimize

$$\max_k |p_n(\lambda_k)|.$$

Since all the eigenvalues λ_k of M are in general unknown, we seek, then, to minimize instead

$$\max_{-\bar{\mu}[M] \leqq x \leqq +\bar{\mu}[M]} \{|p_n(x)|\}, \tag{4.32}$$

under the restriction that $p_n(1) = 1$. As is well known [11], the polynomials which minimize (4.32) are given explicitly by

$$\tilde{p}_n(x) = \frac{C_n(x/\bar{\mu}[M])}{C_n(1/\bar{\mu}[M])}, \qquad n \geqq 0 \tag{4.33}$$

where $C_n(x) = \cos[n\cos^{-1}x]$ for $|x| \leqq 1$ is the Chebyshev polynomial of degree n, and the semi-iterative method of (4.29) corresponding to the polynomials of (4.33) is called the Chebyshev semi-iterative method with respect to the Jacobi method [42].

It is natural to ask how this Chebyshev semi-iterative method compares with the successive overrelaxation method. By defining an average rate of convergence [53; 42] for the semi-iterative method (4.29), we prove[14] [42]

THEOREM 6.[15] *The successive overrelaxation method with optimum relaxation factor converges at least twice as fast as the Chebyshev semi-iterative method with respect to the Jacobi method, and therefore at least twice as fast as any semi-iterative method with respect to the Jacobi method. Furthermore, as the number of iterations tends to infinity, the successive overrelaxation method becomes exactly twice as fast as the Chebyshev semi-iterative method.*

Other results concerning the use of semi-iterative methods are known [18; 42; 53], but will not, for reasons of brevity, be covered here.

As a final remark, it is clear that some criterion must be used to terminate the iterations of these various iterative methods. In general, the vector k of (4.1) is related physically to a source of neutrons, and thus has non-negative components. Remembering that A^{-1} has all its entries positive, it follows that the unique solution of (4.1) must have positive components, and this gives rise to a simple check as to whether enough inner iterations have been performed.

5. **Acceleration of the outer iterations.** Returning to the numerical problem of approximating the solution of the time independent multigroup diffusion equations, we saw by the results of §3 that we in turn seek, for the corresponding matrix problem, the largest (in modulus) eigenvalue λ^* of the matrix T of §3 and its associated eigenvector Ψ, all of whose components are non-negative. Thus, in solving

$$T\Psi = \lambda^*\Psi. \tag{5.1}$$

We proved that the iterative method

$$T\left(\frac{\psi^{(n)}}{\lambda_{n+1}}\right) \equiv \psi^{(n+1)}; \qquad \lambda_{n+2} \equiv \lambda_{n+1}\frac{(\psi^{(n+1)},\psi^{(n)})}{(\psi^{(n)},\psi^{(n)})}, \qquad n \geqq 0, \tag{5.2}$$

was necessarily convergent for any initial positive vector $\psi^{(0)}$. Thus, the computational power method gives us our first iterative method for solving (5.1). We now suppose that the eigenvalue estimates λ_n of λ^* are sufficiently accurate, so that we can consider the iterative process

$$\left(\frac{T}{\lambda^*}\right)\psi^{(n)} = \psi^{(n+1)}, \qquad n \geqq m_0. \tag{5.3}$$

---

[14] In [54], this same result was proved in a different way only for the case in which the diagonal entries of the matrix A of (4.1) were all equal, and was conjectured to be true in general. A result of Forsythe and Straus [Proc. Amer. Math. Soc. vol. 6 (1955) pp. 340–345] supplied the proof of this conjecture.

[15] *Added in proof:* A new iterative method, called the *modified Chebyshev semi-iterative method*, eliminated this factor of two in the cyclic case, and is more rapidly convergent than the successive over-relaxation iterative method. See [18a].

Thus, it follows that

$$\left(\frac{T}{\lambda^*}\right)^r \psi^{(n)} = \psi^{(n+r)}, \qquad n \geqq m_0, r \geqq 0. \tag{5.4}$$

The matrix $(1/\lambda^*)\,T$ has its largest eigenvalue (in modulus) unity. Paralleling the discussion in §4, if we consider all polynomials $p_r(x)$ for which $p_r(+1) = +1$, then the matrix of (5.4) corresponds exactly to $p_r(x) = x^r$, with the variable x replaced by the matrix $(1/\lambda^*)T$. If[16] the eigenvalues λ_j of T are all real and non-negative, and satisfy $0 \leqq \lambda_n \leqq \lambda_{n-1} \leqq \cdots \leqq \lambda_2 < \lambda_1 \equiv \lambda^*$, and if the corresponding eigenvectors ψ_j of T span the vector space $V_N(R)$, then we can express $\psi^{(m_0)}$ as

$$\psi^{(m_0)} = \sum_{j=1}^{N} c_j \psi_j \tag{5.5}$$

for suitable scalars c_j. Thus, from (5.4),

$$\psi^{(m_0+n)} = p_n\left(\frac{T}{\lambda^*}\right)\psi^{(m_0)} = c_1\psi_1 + \sum_{j>1} c_j p_n\left(\frac{\lambda_j}{\lambda^*}\right)\psi_j. \tag{5.6}$$

Since the sum on the right hand side of the equation above is our error, then, as in §4, we minimize the quantity

$$\max_{0 \leqq x \leqq \bar{\sigma}} |p_n(x)|, \quad \bar{\sigma} \equiv \frac{\lambda_2}{\lambda_1} < 1, \tag{5.7}$$

under the restrictions that $p_n(1) = 1$. Again, the solution to this problem is given [**11**] explicitly in terms of Chebyshev polynomials

$$\tilde{p}_n(x) = \frac{C_n\left(\dfrac{2x}{\lambda_2} - 1\right)}{C_n\left(\dfrac{2}{\bar{\sigma}} - 1\right)}, \qquad n \geqq 0. \tag{5.8}$$

From the well known recurrence relation for Chebyshev polynomials

$$C_{n+1}(x) = 2xC_n(x) - C_{n-1}(x), \qquad n \geqq 1, \tag{5.9}$$

where $C_0(x) = 1$, $C_1(x) = x$, the application of Chebyshev polynomials to accelerate the convergence of (5.3) is also defined by means of a recurrence relation. If $T(\zeta^{(m)}/\lambda_m) \equiv \psi^{(m+1)}$, then [**36**; **46**] with $\lambda_m \doteq \lambda^*$,

$$\zeta^{(m+1)} \equiv 2\alpha_{m+1}\left(\frac{2}{\bar{\sigma}}\psi^{(m+1)} - \zeta^{(m)}\right) - \beta_{m+1}\zeta^{(m-1)}, \qquad m \geqq 1, \tag{5.10}$$

where

$$\alpha_{m+1} = \frac{\cosh[m\rho]}{\cosh[(m+1)\rho]}, \; \beta_{m+1} = \frac{\cosh[(m-1)\rho]}{\cosh[(m+1)\rho]}, \qquad m \geqq 1, \tag{5.11}$$

[16] This assumption is apparently made in all time independent diffusion codes which are accelerated by means of Chebyshev polynomials. (See, for example, [**46**].) It has not been proved to be true for general heterogeneous reactor models in n dimensions except for the trivial cases of only one lethargy group and homogeneous bare problems.

and $\rho \equiv \cosh^{-1}(2/\bar{\sigma} - 1)$. To start the process, $\psi^{(0)} \equiv \zeta^{(0)}$, and

$$\zeta^{(1)} = \zeta^{(0)} + \frac{2}{2 - \bar{\sigma}}(\psi^{(1)} - \zeta^{(0)}). \tag{5.10'}$$

It should be pointed out that the use of Chebyshev polynomials for acceleration of the outer iterations does require the storage of an additional vector. Remembering that already $m + 1$ vectors $\varphi_1, \varphi_2, \cdots, \varphi_m$, and ψ are required in the power method, coupled with the necessary storage of the non-zero coefficients of the matrices A_k, B_k, and V_k, the additional storage required for the use of Chebyshev polynomials seems not to be a serious disadvantage. It is well to point out that numerical experience indicates the use of appropriate Chebyshev polynomials does greatly reduce the number of outer iterations necessary for a fixed accuracy for slowing convergent reactor problems.

As in the case of successive overrelaxation, the efficiency of the application of Chebyshev polynomials in accelerating the outer iterations depends upon the accurate estimation of the particular constant, $\bar{\sigma}$, the *dominance ratio* for the matrix T. A practical numerical method for estimating $\bar{\sigma}$ is given in [**45**].

Other methods for accelerating the convergence of the outer iterations exist, and are interesting in their own right. With the definition of the matrices in §3, we now write our discrete time independent eigenvalue problem in the matrix form

$$E\Phi = \frac{1}{\lambda} F\Phi, \tag{5.12}$$

where

$$E \equiv \begin{bmatrix} A_1 & 0 & \cdots 0 \\ -B_{2,1} & A_2 & \cdots 0 \\ \vdots & & \vdots \\ -B_{m,1} & -B_{m,2}\cdots & A_m \end{bmatrix}, \quad F \equiv \begin{bmatrix} X_1V_1 & X_1V_2\cdots X_1V_m \\ X_2V_1 & X_2V_2\cdots X_2V_m \\ \vdots & \vdots \\ 0 & 0 \quad \cdots 0 \end{bmatrix}, \tag{5.13}$$

and

$$\Phi = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_m \end{pmatrix}. \tag{5.14}$$

In this form, E is a non-singular matrix, and E^{-1} and F have non-negative entries. The outer iterations of (5.2) can now be rephrased as

$$E\Phi^{(m+1)} = \frac{1}{\lambda_m} F\Phi^{(m)}. \tag{5.15}$$

If $\tilde{\lambda}$ is an approximation to λ^*, but such that $E^{-1}F - \tilde{\lambda}I$ is non-singular, then, in theory we can consider the iterative method

$$\left(E - \frac{1}{\tilde{\lambda}} F\right)\Phi^{(m+1)} = \left(\frac{1}{\lambda_m} - \frac{1}{\tilde{\lambda}}\right)\Phi^{(m)}, \tag{5.16}$$

which is, in essence, Wielandt's inverse power method [**50**]. This has been recently tried for one-dimensional reactor problems with some success.[17] In principle, this iterative method can also be used to find eigenvalues and their corresponding eigenvectors, of $E^{-1}F$, other than λ^*. The question of the growth of rounding errors is very important here, and instabilities may arise.

As a third method for accelerating the convergence of the outer iterations, we consider an iterative method which stems from the work of Sheldon [**32**]. From (5.12), we have

$$\left(E - \frac{1}{\lambda} F\right)\Phi = 0. \tag{5.17}$$

If we assume that the square diagonal submatrices $A_i - (1/\lambda)X_iV_i \equiv C_i$ are non-singular, then we can write (5.17) as

$$\Phi = M(\lambda)\Phi, \tag{5.18}$$

where

$$M(\lambda) \equiv \begin{bmatrix} 0 & \frac{1}{\lambda}C_1^{-1}X_1V_1 \cdots \frac{1}{\lambda}C_1^{-1}X_1V_m \\ C_2^{-1}B_{2,1} + \frac{1}{\lambda}X_2V_1 & 0 \quad \cdots \frac{1}{\lambda}C_1^{-1}X_2V_m \\ \vdots & \vdots \\ C_m^{-1}B_{m,1} + \frac{1}{\lambda}X_mV_1 & \cdots \quad 0 \end{bmatrix}. \tag{5.19}$$

Since the diagonal entries of $M(\lambda)$ are zero, we can decompose $M(\lambda)$ into

$$M(\lambda) = L(\lambda) + U(\lambda), \tag{5.20}$$

where $L(\lambda)$ and $U(\lambda)$ are respectively strictly lower and upper triangular matrices. It follows from (5.18) that $M(\lambda^*)$ has an eigenvalue of unity, and it is the eigenvector Φ corresponding to this eigenvalue of unity which we seek. If we assume that the partitioned matrix $M(\lambda)$ of (5.19) is cyclic of index 2 and consistently ordered, then we can apply the successive over-relaxation scheme to (5.18):

$$\Phi^{(n+1)} = \Phi^{(n)} + \omega\{L(\lambda)\Phi^{(n+1)} + U(\lambda)\Phi^{(n)} - \Phi^{(n)}\}, \qquad n \geqq 0, \tag{5.21}$$

[17] Personal communication from E. L. Wachspress and G. Habetler of the Knolls Atomic Power Laboratory.

and the relationship between the eigenvalues η of this iterative procedure and the eigenvalues η of $M(\lambda)$ is, as in (4.14),

$$(\nu + \omega - 1)^2 = \nu\omega^2\eta^2. \tag{5.22}$$

For $\lambda = \lambda^*$, there is one eigenvalue $\eta = 1$, and from (5.22), the corresponding eigenvalue ν is also unity. Corresponding to our previous discussion of the successive overrelaxation method, the optimum value for ω is given again by (4.15), except that now we interpret $\bar{\mu}[M(\lambda)]$ as the maximum of the moduli of the eigenvalues η of $M(\lambda)$, *other* than $|\eta| = 1$. By virtue of Theorem 6, both the Chebyshev method and the last method described, have the same order of magnitude improvement upon the rate of convergence of the power method for the outer iterations.

6. The discrete space matrix problem for the time dependent multigroup diffusion equations. To derive the discrete space matrix analog of the time dependent multigroup diffusion equation of (2.1), we proceed, in the same manner as in §3, by integrating (2.1) over appropriate rectangular subregions of R. Using the same approximations of §3, we need only consider the treatment of the new term

$$\frac{1}{v_i}\int_{S(x_0)} \frac{\partial\phi_i(x,t)}{\partial t}\, dA = \frac{1}{v_i}\frac{d}{dt}\int_{S(x_0)} \phi_i(x,t)dA, \tag{6.1}$$

where $S(x_0)$ is the mesh region as indicated in Figure 2. If we make the approximation

$$\frac{1}{v_i}\frac{d}{dt}\int_{S(x_0)} \phi_i(x,t)dA \doteq \frac{1}{v_i}\frac{d}{dt}\left[\phi_i(x_0,t)\cdot\int_{S(x_0)} dA\right], \tag{6.2}$$

where $\int_{S(x_0)} dA$ is the mesh area about the point x_0, then the discrete space matrix problem corresponding to (2.1) takes the form

$$\left\{D_i\frac{d\Phi_i(t)}{dt} = -A_i\Phi_i(t) + \sum_{j\neq i} B_{i,j}\Phi_j(t)\right\}_{i=1}^{m} \tag{6.3}$$

where the vectors $\Phi_1(0),\cdots,\Phi_m(0)$ are prescribed as initial vector conditions. The quantities D_i, A_i, and $B_{i,j}$ are again $N \times N$ matrices, and the matrices A_i and $B_{i,j}$ have properties described in (3.4) and Theorem 1. The matrices D_i are diagonal matrices, whose entries are the ratios of mesh areas and lethargy group velocities, and thus the matrices D_i are positive diagonal matrices. Since the matrices D_i are non-singular, we can write (6.3) in the equivalent form

$$\left\{\frac{d\Phi_i}{dt} = -D_i^{-1}A_i\Phi_i + \sum_{j\neq i} D_i^{-1}B_{i,j}\Phi_j\right\}_{i=1}^{m}. \tag{6.4}$$

Using the vector Φ as defined in (5.15), we write (6.4) now in the compact form

$$\frac{d}{dt}\Phi(t) = Q\Phi(t), \tag{6.5}$$

where $\Phi(0)$ is a prescribed initial vector condition, and Q is a $(mN) \times (mN)$ matrix.

As stated in §2, we are primarily interested in the asymptotic behavior for large values of t of the solution of (6.5), subject to the initial vector condition. In considering this topic, we make the

DEFINITION 4. An $n \times n$ matrix $M = (m_{i,j})$ is *essentially positive* if and only if M is irreducible, and $m_{i,j} \geqq 0$ for all $i \neq j$.

It can be shown [2] that an equivalent definition of an essentially positive matrix M is that the matrix $e^{tM} \equiv \sum_{k=0}^{\infty} t^k M^k/k!$ has every entry positive for all $t > 0$. Based on generalizations [2] of the Perron-Frobenius theory of non-negative matrices, we have that any essentially positive matrix M has a unique (up to scalar factors) strictly positive eigenvector Ψ, with real simple eigenvalue $\mu_1 = \gamma$, and $\mu_1 > \mathrm{Re}\{\mu_j\}$ for any other eigenvalue μ_j of Q.

From (3.4) and Theorem 1, it follows that the matrix Q of (6.5′) is essentially positive, and if Ψ is the positive eigenvector of Q, corresponding to the eigenvalue γ, then we have [2].

THEOREM 7. *For $t \to \infty$, the solution of the matrix differential equation* (6.5) *is*

$$\Phi(t) = Ke^{\gamma t}\Psi + o(e^{\mu t}), \tag{6.6}$$

where μ satisfies $\gamma > \mu > \sup_{j>1} \mathrm{Re}\{\mu_j\}$. If F is the positive eigenvector of Q', then

$$K \equiv (F,\Phi(0))/(F,\Psi).$$

As in [2], we make the

DEFINITION 5. For Q essentially positive, the process (6.5) is *subcritical*, *critical*, or *supercritical*, depending on whether $\gamma \leqq 0$, $\gamma = 0$, or $\gamma > 0$.

From this definition, we have as a corollary to Theorem 4,

COROLLARY. *For Q essentially positive and $\Phi(0) > 0$, then*

$$\lim_{t \to +\infty} \Phi(t) = \begin{cases} \begin{pmatrix} +\infty \\ \vdots \\ +\infty \end{pmatrix} \text{ if (6.5) is a supercritical process.} \\ K\Psi \text{ if (6.5) is a critical process.} \\ 0 \text{ if (6.5) is a subcritical process.} \end{cases}$$

For completeness, we consider source problems of the form

$$\frac{d}{dt}\Phi(t) = Q\Phi(t) + S, \tag{6.7}$$

where Q is essentially positive, $\Phi(0)$ is the prescribed initial vector condition,

and S is a time independent source vector. As given in [2], we can deduce the asymptotic behavior of the solution of (6.6) from the results of Theorem 4 and the corollary, and we have

THEOREM 8. *If Q is essentially positive, and $S > 0$, then $\lim_{t\to\infty} \Phi(t)$ is a vector with finite Euclidean norm if and only if Q is subcritical. In this case,*

$$\lim_{t \to \infty} \Phi(t) = -Q^{-1}S. \tag{6.8}$$

We remark that since the matrices $B_{i,j}$ of (6.3) are not necessarily null diagonal matrices for $i < j$, we have treated both the cases of thermal up-scattering, as well as fission.

7. Iterative methods for time dependent multigroup diffusion equations. The numerical solution of time dependent multigroup diffusion equations has not yet received as much attention as the number of numerical techniques, which are presently available for such problems, would suggest. We now consider the time dependent multigroup diffusion equations with an external source

$$\left\{D_i \frac{d\Phi_i(t)}{dt} = -A_i\Phi_i(t) + \sum_{j \neq i} B_{i,j}\Phi_j(t) + S_i\right\}_{i=1}^{m}, \tag{7.1}$$

where the matrices D_i, A_i, and $B_{i,j}$ have time independent entries,[18] the vectors S_i are time independent, and $\Phi_1(0), \cdots, \Phi_m(0)$ are given initial vector conditions. Since the matrices D_i are positive diagonal matrices, we can write (7.1) in the form

$$\left\{\frac{d\Phi_i(t)}{dt} = -D_i^{-1}A_i\Phi_i(t) + \sum_{j \neq i} D_i^{-1}B_{i,j}\Phi_j(t) + D_i^{-1}S_i\right\}_{i=1}^{m} \tag{7.2}$$

and using the definitions of $\Phi(t)$ and Q in (5.15), we can now write (7.2) in the form

$$\frac{d\Phi(t)}{dt} = -Q\Phi(t) + s, \tag{7.3}$$

which is analogous to (6.5), with the exception of the non-homogeneous time independent source vector s. From the properties of the matrices A_i, $B_{i,j}$ and D_i, it follows that the $(mN) \times (mN)$ matrix Q is essentially positive. We assume now that Q is *subcritical*, which implies that all its eigenvalues μ_k satisfy $\mathrm{Re}\mu_k < 0$. Thus, from the subcriticality of Q, the matrix $(I - (t/2)Q)$ is non-singular for all $t \geqq 0$. We consider the Crank-Nicolson approximation [5] to (7.3), which takes the form

$$\frac{\Phi(t + \Delta t) - \Phi(t)}{\Delta t} = -\frac{Q}{2}\{\Phi(t + \Delta t) + \Phi(t)\} + s. \tag{7.4}$$

[18] In an actual burnup calculation, the entries of the matrices D_i, A_i, and $B_{i,j}$ are functions of time.

This is equivalent to

$$\left(I - \frac{\Delta t}{2}Q\right)\Phi(t + \Delta t) = \left(I + \frac{\Delta t}{2}Q\right)\Phi(t) + \Delta t\, s. \tag{7.5}$$

Since the matrix $(I - (\Delta t/2)Q)$ is non-singular for $\Delta t \geqq 0$, this implicit stepping-ahead process from the given initial vector condition $\Phi(0)$ can be in principle carried out.

For the important case in which the matrices $B_{i,j} \equiv 0$ for $j > i$, it follows that the matrix $(I - (\Delta t/2)Q)$ is a block triangular matrix, and in the case of one space variable where the diagonal blocks are tridiagonal, $(I - (\Delta t/2)Q)$ can be directly inverted. For two or more space variables, the process of solving (7.5) for large numbers of mesh points would involve again inner iterations.

8. **Conclusions.** In surveying the numerical methods for solving multi-dimensional multigroup diffusion equations, we have attempted to cover the numerical methods which have been used in existing codes for high-speed digital computers. But many topics of considerable interest, though not treated here, may become more and more valuable as numerical methods in time to come. As an example, except in the case of one space variable, the direct inversions of the discrete time independent matrix eigenvalue problem (3.13), involving no inner iterations, have not been discussed. While such direct inversions have been applied [**24**] to the numerical solution of higher dimensional multigroup diffusion equations, their general use apparently seems practical only for still larger computing machines. Also, the use of higher order point formulas approximating (2.1) and (2.6), as well as questions of truncation errors, have not been discussed, and these topics will undoubtedly play a significant role in the design of future machine codes for solving the multigroup diffusion approximations. It is in these areas, for example, where numerical analysts can find challenging new problems.

Bibliography

1. R. J. Arms, L. D. Gates and B. Zondek, *A method of block iteration*, J. Soc. Indust. Appl. Math. vol. 4 (1956) pp. 220–229.

2. Garrett Birkhoff and Richard S. Varga, *Reactor criticality and nonnegative matrices*, J. Soc. Indust. Appl, Math. vol. 6 (1958) pp. 354–377.

3. ———, *Implicit alternating direction methods*, Trans. Amer. Math. Soc. vol. 92 (1959) pp. 13–24.

3a. Garrett Birkhoff, *Some mathematical problems of nuclear reactor theory*, "Frontiers of Numerical Mathematics," Madison, The University of Wisconsin Press, 1960, pp. 23–42.

4. L. Collatz, *Einschliezungssatz für die charakteristischen Zahlen von Matrizen*, Math. Z. vol. 48 (1942) pp. 221–226.

5. J. Crank and P. Nicolson, *A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type*, Proc. Cambridge Philos. Soc. vol. 43 (1947) pp. 50–67.

6. Elizabeth H. Cuthill and Richard S. Varga, *A method of normalized block iteration*, J. Assoc. Comput. Mach. vol. 6 (1959) pp. 236–244.

7. Gerard Debreu and I. N. Herstein, *Nonnegative square matrices*, Econometrica vol. 21 (1953) pp. 596–607.

8. Jim Douglas, Jr. and H. H. Rachford, Jr., *On the numerical solution of heat conduction problems in two and three space variables*, Trans. Amer. Math. Soc. vol. 82 (1956) pp. 421–439.

9. Jim Douglas, Jr., *A note on the alternating direction implicit method for the numerical solution of heat flow problems*, Proc. Amer. Math. Soc. vol. 8 (1957) pp. 409–412.

10. Miroslav Fiedler and Vlastimil Pták, *Über die Konvergenz des verallgemeinerten Seidelschen Verfahren zur Lösung von Systemen linearer Gleichungen*, Math. Nachr. vol. 15 (1956) pp. 31–38.

11. Donald A. Flanders and George Shortley, *Numerical determination of fundamental modes*, J. Appl. Phys. vol. 21 (1950) pp. 1326–1332.

12. Stanley P. Frankel, *Convergence rates of iterative treatments of partial differential equations*, Math. Tables Aids Comput. vol. 4 (1950) pp. 65–75.

13. B. Friedman, *The iterative solution of elliptic partial difference equations*, AEC Research and Development Report NYO-7698, 1957.

14. G. Frobenius, *Über Matrizen aus nicht negative Elementen*, Sitzungsberichte der Akademie der Wissenschaften zu Berlin (1912) pp. 456–477.

15. F. R. Gantmacher, *Applications of the theory of matrices*, New York, Interscience Publishers, Inc., 1959, p. 83.

16. Hilda Geiringer, *On the solution of systems of linear equations by certain iterative methods*, Reissner Anniversary Volume, J. W. Edwards, Ann Arbor, Michigan, 1944, pp. 365–393.

17. S. Glasstone and M. C. Edlund, *The elements of nuclear reactor theory*, New York, Van Nostrand Company, Inc., 1952.

18. G. H. Golub, *The use of Chebyshev matrix polynomials in the iterative solution of linear equations compared to the method of successive overrelaxation*, Doctoral Thesis, University of Illinois, 1959.

18a. G. H. Golub and R. S. Varga, *Chebyshev semi-iterative methods, successive overrelaxation iterative methods, and second order Richardson iterative methods*, Report No. 1028, Case Institute of Technology, February, 1960.

19. A. S. Householder, *The approximate solution of matrix problems*, J. Assoc. Comput. Mach. vol. 5 (1958) pp. 205–243.

20. Herbert B. Keller, *On some iterative methods for solving elliptic difference equations*, Quart. Appl. Math. vol. 16 (1958) pp. 209–226.

21. Cornelius Lanczos, *Solution of systems of linear equations by minimized iterations*, J. Res. Nat. Bur. Standards vol. 49 (1952) pp. 33–53.

22. R. von Mises and H. Pollaczek-Geiringer, *Praktische Verfahren der Gleichungsauflösung*, Z. Angew. Math. Mech. vol. 9 (1929) pp. 58–77.

23. Francis D. Murnaghan, *The theory of group representations*, Baltimore, Johns Hopkins Press, 1938.

24. J. A. Nohel and W. P. Timlake, *Higher order differences in the numerical solution of two-dimensional neutron diffusion equations*, Proceedings of the Second International Conference on the Peaceful Uses of Atomic Energy, Geneva, 1958, vol. 16, pp. 595–600.

25. Rufus Oldenburger, *Infinite powers of matrices and characteristic roots*, Duke Math. J. vol. 6 (1940) pp. 357–361.

26. A. Ostrowski, *Über die Determinanent mit überwiegender Hauptdiagonale*, Comment. Math. Helv. vol. 10 (1937) pp. 69–96.

27. D. W. Peaceman and H. H. Rachford, Jr., *The numerical solution of parabolic and elliptic differential equations*, J. Soc. Indust. Appl. Math. vol. 3 (1955) pp. 28–41.

28. O. Perron, *Zur Theorie der Matrices*, Math. Ann. vol. 64 (1907) pp. 259–263.

29. Edgar Reich, *On the convergence of the classical iterative method of solving linear simultaneous equations*, Ann. Math. Statist. vol. 20 (1949) pp. 448–451.

30. L. F. Richardson, *The approximate arithmetical solution by finite differences of physical problems involving differential equations, with an application to the stresses in a masonry dam*, Philos. Trans. Roy. Soc. London Ser. A vol. 210 (1910) pp. 307–357.

31. V. Romanovsky, *Recherches sur les chaînes de Markoff*, Acta Math. vol. 66 (1936) pp. 147–251.

32. J. W. Sheldon, *Numerical solution of BORE flux equations*, Mimeographed notes from the Computer Usage Company.

33. R. H. Stark, *Rates of convergence in numerical solution of the diffusion equation*, J. Assoc. Comput. Mach. vol. 3 (1956) pp. 29–40.

34. P. Stein and R. L. Rosenberg, *On the solution of linear simultaneous equations by iteration*, J. London Math. Soc. vol. 23 (1948) pp. 111–118.

35. E. Stiefel, *On solving Fredholm integral equations*, J. Soc. Indust. Appl. Math. vol. 4 (1956) pp. 63–85.

36. ———, *Kernel polynomials in linear algebra and their numerical applications*, Nat. Bur. Standards Applied Math. Series 49, U.S. Government Printing Office, Washington, D.C., 1958, pp. 1–22.

37. T. J. Stieltjes, *Sur les racines de l'équation* $X_n = 0$, Acta Math. vol. 9 (1887) pp. 385–400.

38. Olga Taussky, *A recurring theorem on determinants*, Amer. Math. Monthly vol. 56 (1949) pp. 52–62.

39. R. S. Varga and M. A. Martino, *The theory for the numerical solution of time-dependent and time-independent multigroup diffusion equations*, Proceedings of the Second International Conference on the Peaceful Uses of Atomic Energy, Geneva, 1958, vol. 16, pp. 570–577.

40. Richard 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.

41. ———, *P-cyclic matrices: a generalization of the Young-Frankel successive over-relaxation scheme*, Pacific J. Math. vol. 9 (1959) pp. 617–628.

42. ———, *A comparison of the successive overrelaxation method and semi-iterative methods using Chebyshev polynomials*, J. Soc. Indust. Appl. Math. vol. 5 (1957) pp. 39–46.

43. ———, *Factorization and normalized iterative methods*, Boundary Problems in Differential Equations, Madison, The University of Wisconsin Press, 1960, pp. 121–142.

44. ———, *Overrelaxation applied to implicit alternating direction methods*, to appear in the Proceedings of the First International Conference on Information Processing, Paris, June 13, 1959.

45. ———, *On estimating rates of convergence in multigroup diffusion problems*, Report WAPD-TM-41, Bettis Atomic Power Division of the Westinghouse Electric Corp., 1957.

46. E. L. Wachspress, *CURE: A generalized two-space-dimension multigroup coding for the* IBM-704, Report KAPL-1724, Knolls Atomic Power Laboratory of the General Electric Company, 1957.

47. E. L. Wachspress and G. J. Habetler, *An alternating-direction-implicit iterative technique*, J. Soc. Indust. Appl. Math. vol. 8 (1960) pp. 403–424.

48. Alvin M. Weinberg and Eugene P. Wigner, *The physical theory of neutron chain reactors*, Chicago, The University of Chicago Press, 1958.

49. Helmut Wielandt, *Unzerbegbare, nicht negative Matrizen*, Math. Z. vol. 52 (1950) pp. 642–648.

50. H. Wielandt, *Bestimmung höherer Eigenwerte durch gebrochene Iteration*, Ber. Aerodynamischen Versuchsanstalt Göttingen, 44, J. 37 (1944).

51. Y. K. Wong, *Some properties of the proper values of a matrix*, Proc. Amer. Math. Soc. vol. 6 (1955) pp. 891–899.

52. David Young, *Iterative methods for solving partial difference equations of elliptic type*, Trans. Amer. Math. Soc. vol. 76 (1954) pp. 92–111.

53. ———, *On Richardson's method for solving linear systems with positive definite matrices*, J. Math. Phys. vol. 32 (1953) pp. 243–255.

54. ———, *On the solution of linear systems by iteration*, Proceedings of the Sixth Symposium in Applied Mathematics, New York, McGraw-Hill, 1956, pp. 283–298.

55. ———, *ORDVAC solutions of the Dirichlet problem*, J. Assoc. Comput. Mach. vol. 2 (1955) pp. 137–161.

56. David Young and Louis Ehrlich, *Some numerical studies of iterative methods for solving elliptic difference equations*, Boundary Problems in Differential Equations, Madison, The University of Wisconsin Press, 1960, pp. 143–162.

WESTINGHOUSE ELECTRIC CORPORATION,
PITTSBURGH, PENNSYLVANIA