WAPD-TM-364
AEC RESEARCH AND
DEVELOPMENT REPORT

NUMERICAL METHODS AND TECHNIQUES USED IN THE TWO-DIMENSIONAL NEUTRON-DIFFUSION PROGRAM PDQ-5

February 1963

CONTRACT AT-11-1-GEN-14

BETTIS ATOMIC POWER LABORATORY
PITTSBURGH, PENNSYLVANIA

Operated for the
U.S. ATOMIC ENERGY COMMISSION
by WESTINGHOUSE
ELECTRIC CORPORATION

DISCLAIMER

WAPD-TM-364

UC-32: Mathematics and Computers
TID-4500 (18th Ed.)

NUMERICAL METHODS AND TECHNIQUES USED IN THE TWO-DIMENSIONAL NEUTRON-DIFFUSION PROGRAM PDQ-5

L. A. Hageman

CONTRACT AT-11-1-GEN-14

February 1963

Printed in USA. Price $1.75. Available from the Office of Technical Services, Department of Commerce, Washington 25, D. C.

BETTIS ATOMIC POWER LABORATORY PITTSBURGH, PENNSYLVANIA

Operated for the U. S. Atomic Energy Commission by
Westinghouse Electric Corporation

STANDARD EXTERNAL DISTRIBUTION

IC-32: Mathematics and Computers, TID-4500, 18th Edition	617

SPECIAL EXTERNAL DISTRIBUTION

Director, Development Division, PNRO	3
Argonne National Laboratory, W. F. Miller	2
Brookhaven National Laboratory, J. Chernick	2
Case Institute of Technology, R. S. Varga	1
David Taylor Model Basin, H. Polachek	2
Harvard University, G. Birkhoff	1
Knolls Atomic Power Laboratory, R. Ehrlich	10
Los Alamos Scientific Laboratory, B. Carlson	2
New York University, R. Richtmyer	2
Oak Ridge National Laboratory, A. S. Householder	1
Union Carbide Nuclear Company, K-25, V. E. Anderson	1
University of California Radiation Laboratory, Livermore, S. Fernbach	2
TOTAL	646

ACKNOWLEDGMENT

The person primarily responsible for the development of the mathematical theory and techniques employed by the PDQ program is Professor R. S. Varga of Case Institute of Technology.

TABLE OF CONTENTS

		Page
I.	INTRODUCTION	1
II.	NOTATIONAL CONVENTIONS AND DEFINITIONS	4
III.	PASSAGE TO THE DISCRETE PROBLEM	6
	A. Derivation of the Difference Equations	6
	B. Statement of the Matrix Problem	15
IV.	OUTER ITERATIONS	22
	A. The Use of Chebyshev Polynomials	25
V.	INNER ITERATIONS	33
	A. Stationary Iterative Techniques	34
	B. The Cyclically Reduced Three-Line Successive Overrelaxation Method	41
VI.	ESTIMATION OF THE PARAMETERS $\bar{\sigma}$ AND ω	54
	A. The Initial Estimate for $\bar{\sigma}$	55
	B. Updating the Estimates for $\bar{\sigma}$	58
APPENDIX A:	The Square Root Method for Symmetric Positive Definite Matrices	64
APPENDIX B:	The Inner-Outer Iteration Problem	67
	A. Error Propagation in the Outer Iterations	68
	B. A New Eigenvalue Problem	75
REFERENCES		81

In this report a description is given of the mathematical theory and methods used in the PDQ-5 program to obtain the numerical solution to the few-group, time-independent, neutron diffusion equations in two dimensions. The finite difference approximations of the continuous problem for both x-y and r-z geometries and the properties of the resulting matrix problem are given. The mathematical features and the practical applications of the numerical methods used to solve the matrix problem are discussed in detail.

NUMERICAL METHODS AND TECHNIQUES USED IN THE TWO-DIMENSIONAL NEUTRON-DIFFUSION PROGRAM PDQ-5

I. INTRODUCTION

This report describes the numerical solution of the few-group, time-independent, neutron-diffusion equations in two dimensions utilized in the PDQ-5 program [4][1]. The neutron-diffusion group equations solved by PDQ-5 are of the form

$$(1.1)\quad \left\{ -\operatorname{div}\left[D_g(r)\ \operatorname{grad}\ \varphi_g(\underline{r})\right] + \sigma_g(\underline{r})\varphi_g(\underline{r}) - \sigma^s_{g-1}(\underline{r})\varphi_{g-1}(\underline{r}) = \frac{\chi_g}{\lambda} \sum_{g=1}^{G} \nu\sigma^f_g(\underline{r})\varphi_g(\underline{r}) \right\}_{g=1}^{g=G},$$

where

$\underline{r}$ = the spatial vector whose set of components denotes the x-y and r-z coordinates.

---

[1] Numbers in brackets refer to the list of references given at the end of the report.

g = the lethargy group index.

G = the number of lethargy groups.

$\varphi_g(\underline{r})$ = the neutron flux.

$D_g(\underline{r})$ = the diffusion coefficient and $D_g(\underline{r}) > 0$.

$\sigma_g^a(\underline{r})$ = the absorption macroscopic cross section.

$\sigma_g^s(\underline{r})$ = the scattering macroscopic cross section from group g to group $g+1$. $\sigma_0^s(\underline{r}) = \sigma_G^s(\underline{r}) = 0$ and $\sigma_g^s(\underline{r}) \geq 0$.

B_z^2 = the geometric buckling.

$\sigma_g(\underline{r}) = \sigma_g^a(\underline{r}) + \sigma_g^s(\underline{r}) + D_g(\underline{r})B_z^2$. $\sigma_g(\underline{r}) \geq 0$.

χ_g = the integral of the fission spectrum over the lethargy range represented by group g. $\chi_g \geq 0$ and $\chi_1 > 0$.

$\nu\sigma_g^f(\underline{r})$ = the fission macroscopic cross section times the average number of neutrons released per fission. $\nu\sigma_g^f(\underline{r}) \geq 0$.

λ = the eigenvalue.

We further assume that the following conditions are satisfied:

(a) The domain of interest is a rectangular region R in the x-y or r-z plane.

(b) The region R may be divided into a finite number of subregions R_i such that D_g, σ_g^a, σ_g^s, and σ_g^f are constant within each subregion R_i.

(c) $\varphi_g(\underline{r})$ is continuous in R and $D_g(\underline{r}) \frac{\partial \varphi_g(\underline{r})}{\partial n}$ is continuous across interfaces between subregions[1].

---

[1] $\frac{\partial \varphi}{\partial n}$ refers to the normal derivative.

(d) When $\underline{r} \varepsilon \Gamma$, the external boundary of R, then either $\varphi_g(\underline{r}) = 0$ or $\frac{\partial \varphi_g(\underline{r})}{\partial n} = 0$. There is no mixing of boundary conditions on any one external side of R.

(e) For $g = 1, 2, \ldots, G$, $\sigma_g(\underline{r}) > 0$ for some subregion R_i or else $\varphi_g(\underline{r}) = 0$ on some boundary segment of R. Also, $\sigma_g^f(\underline{r}) > 0$ for some subregion R_i and some g.

With the homogeneous boundary condition (d), the problem stated above then defines an eigenvalue problem and we seek to determine solutions of (1.1) corresponding to the largest (in modulus) eigenvalue λ of (1.1)[1]. For complicated reactor designs, we can only hope to find approximate solutions to this problem by the use of numerical methods. The PDQ-5 program was written[2] for the Philco-2000 digital computer to solve this problem numerically.

The purpose of this report is to give a complete mathematical analysis and discussion of the numerical methods used in the PDQ-5 program. This report is one of three reports being written which will describe the PDQ-5 program. The other reports will cover (1) program description [4] and (2) auxiliary subroutines [15].

---

[1] For a proof of the existence and uniqueness of this largest eigenvalue and its corresponding eigenfunction, see Ref. 10.

[2] The PDQ-5 program is written in the FORTRAN language.

II. NOTATIONAL CONVENTIONS AND DEFINITIONS

The notation $A \equiv (a_{i,j})$ means that $a_{i,j}$ is the (i,j) entry of the matrix A. Similarly, the notation $A \equiv (A_{i,j})$ means that $A_{i,j}$ is the (i,j) submatrix in the partitioned form of the matrix A. If $A \equiv (a_{i,j})$ is an n by m matrix, then we say A is a nonnegative matrix if $a_{i,j} \geq 0$ for all $1 \leq i \leq n$ and $1 \leq j \leq m$. Similarly, we say A is a positive matrix if $a_{i,j} > 0$ for all $1 \leq i \leq n$ and $1 \leq j \leq m$. The transpose of the matrix A will be denoted by A^T and the conjugate transpose by A^*.

A vector in n-dimensional space over the complex field will be indicated by a symbol with a bar under it, such as $\underline{x}$. In this paper, vectors are considered as n by 1 matrices. Thus, $\underline{x}^T$, $\underline{x}^*$ and the nonnegative and positive properties for vectors are defined as above for matrices. The inner product of the vector $\underline{x}$ with the vector $\underline{y}$ is defined as the vector product $\underline{x}^*\underline{y}$. We shall denote the inner product of $\underline{x}$ with $\underline{y}$ as $[\underline{x}, \underline{y}]$. A matrix A is said to be positive definite if the inner product $[\underline{x}, A\underline{x}]$ is positive for every nonzero vector $\underline{x}$.

We now define several matrix properties which will be of interest to us.

Definition 2.1. The n by n matrix $A \equiv (a_{i,j})$ is said to be irreducible $\Longleftrightarrow$ for any ordered pair of integers i and j, $1 \leq i, j \leq n$, there exists a sequence of nonzero entries of A of the form

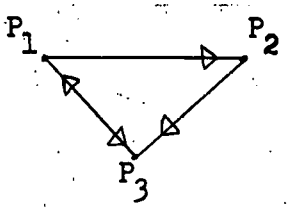
$$a_{i,i_1}, a_{i_1,i_2}, \cdots a_{i_{m-2},i_{m-1}}, a_{i_{m-1},j} .$$

Definition 2.2. The real n by n matrix $A \equiv (a_{i,j})$ with $a_{i,j} \leq 0$ for all $i \neq j$ is called a Stieltjes matrix if A is symmetric and positive definite.

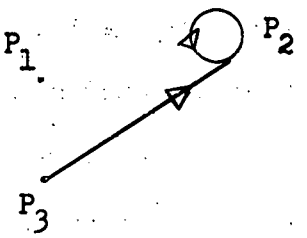
The concept of the directed graph [21, p. 19] of a matrix will be used in Chapter V. Let $A \equiv (a_{i,j})$ be an n by n matrix and consider any n distinct points $P_1, P_2, \ldots, P_n$ in the plane. The points P_k are called nodes. Now for every nonzero entry $a_{i,j}$, we connect the node P_i to the node P_j with an arrow pointed towards P_j. In this way, a finite <u>directed graph</u> can be associated with every n by n matrix. The directed graphs of the matrices

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \qquad \text{and} \qquad B = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 2 & 0 \end{pmatrix}$$

are given in Fig. 2.1.



Directed Graph of A



Directed Graph of B

FIGURE 2.1

Finally, by $f(x) = a(x) + o(x)$, we shall mean that

$$\lim_{x \to 0} \frac{f(x) - a(x)}{x} = 0.$$

III. PASSAGE TO THE DISCRETE PROBLEM

Having stated the continuous problem in Chapter I, we now proceed to form and state the discrete problem. To do this we first impose a nonuniform mesh of horizontal and vertical lines on our rectangular region R such that all internal interfaces and external boundaries lie exactly on mesh lines. The intersections of the horizontal and vertical mesh lines then define the set, H, of mesh points on R and we seek the solution $\varphi_g(\underline{r})$ only at the mesh points of R. If the point $\tilde{\underline{r}}$ is in the set H and is the point at the intersection of the i-th column and the j-th row of the mesh lines, then we let $\varphi_g(i,j)$ denote an approximation for $\varphi_g(\tilde{\underline{r}})$. If $\varphi_g(\tilde{\underline{r}})$ is not known, then we call the mesh point (i,j) a nodal point. In Fig. 3.1, there are (M + 2)(N) mesh points but only MN nodal points. The $\varphi_g(i,j)$ at nodal points are the unknowns for the discrete problem. If now, for every nodal point, we replace the differential equation (1.1) by a certain finite-difference expression[1], then the discrete problem will be completely defined. A finite-difference equation which approximates the differential equation at a nodal point is by no means unique. We shall derive the finite-difference equations used in the PDQ-5 program for the r-z geometry and just state the finite-difference equations used for x-y geometry. For a derivation of the difference equations in x-y geometry, see Varga [17].

A. Derivation of the Difference Equations

In applying the diffusion equations to reactor calculations, we are actually interested in a three dimensional model. The two-dimensional model

---

[1] A finite-difference expression is simply a linear expression in the $\varphi_g(i,j)$.

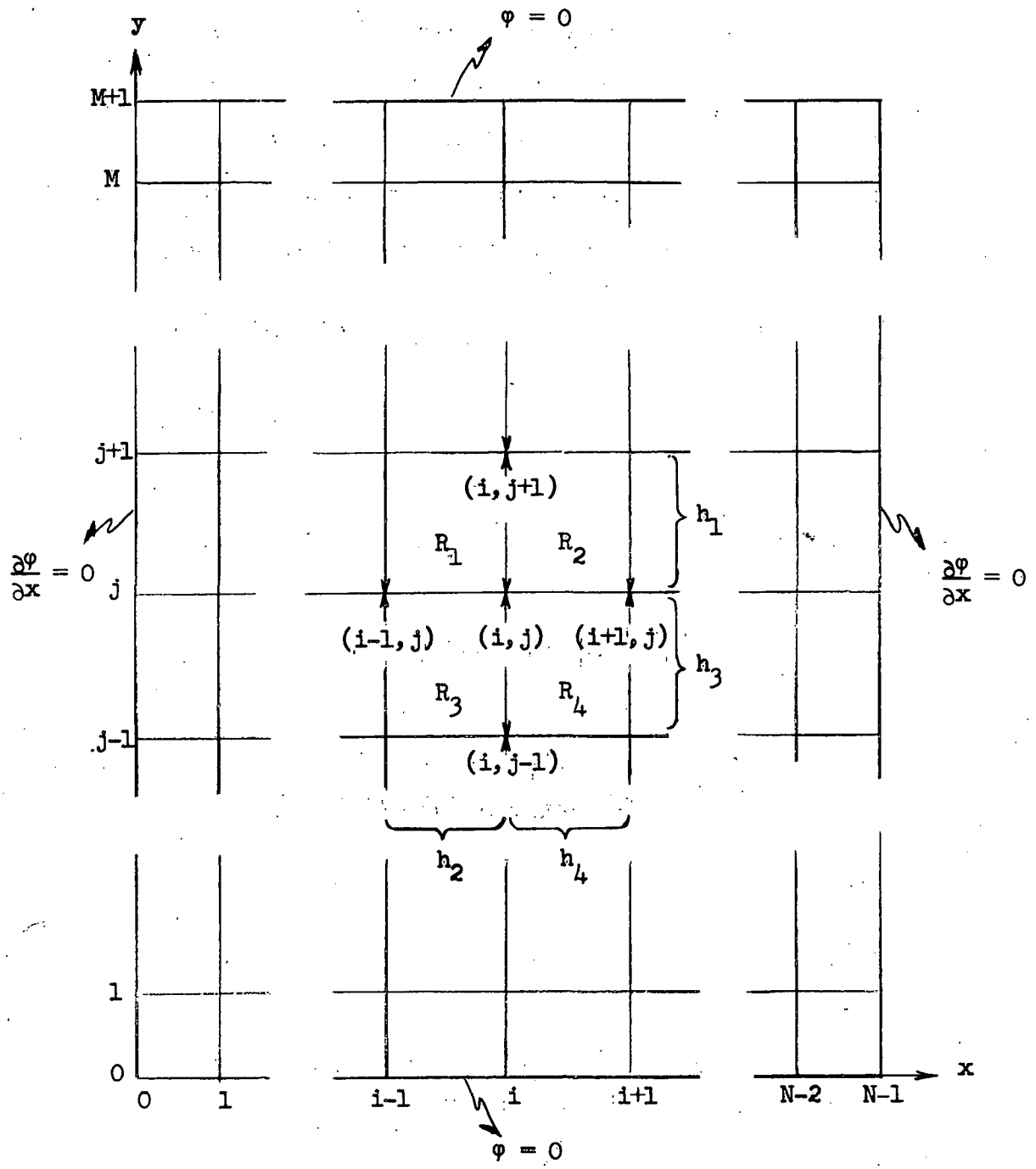

y
φ = 0
M+1
M
j+1
(i, j+1)
h1
R1
R2
∂φ/∂x = 0
j
∂φ/∂x = 0
(i-1, j)
(i, j)
(i+1, j)
h3
R3
R4
j-1
(i, j-1)
h2
h4
1
0
0
1
i-1
i
i+1
N-2
N-1
x
φ = 0

FIGURE 3.1

is obtained by assuming that the flux φ_g is a function of only two of the three space variables. In order to derive the difference equations for r-z geometry, we shall consider the three-dimensional r-z-θ model but make use of the fact that φ_g is not a function of the variable θ.

Let us consider an arbitrary interior nodal point (i,j) in the r-z plane (see Fig. 3.2). For each of the mesh volumes, V_ℓ, surrounding the point (i,j) (see Fig. 3.3), the diffusion equation may be written as[1]

$$(3.1)\quad \left\{ - D_{g,\ell} \operatorname{div} (\operatorname{grad} \varphi_g(r,z)) + \sigma_{g,\ell}\, \varphi_g(r,z) - \sigma^{o}_{g-1,\ell}\, \varphi_{g-1}(r,z) = \frac{\chi_g}{\lambda} \sum_{g=1}^{G} \nu\sigma^{f}_{g,\ell}\, \varphi_g(r,z) \right\}_{\ell=1}^{\ell=4}$$

We now integrate Eq. (3.1) over each of the mesh volumes, V_ℓ.

$$(3.2)\quad \left\{ - D_{g,\ell} \int_{V_\ell} \operatorname{div} (\operatorname{grad} \varphi_g(r,z))dV + \sigma_{g,\ell} \int_{V_\ell} \varphi_g(r,z)dV - \sigma^{s}_{g-1,\ell} \int_{V_\ell} \varphi_{g-1}(r,z)dV = \frac{\chi_g}{\lambda} \sum_{g=1}^{G} \left[\nu\sigma^{f}_{g,\ell} \int_{V_\ell} \varphi_g(r,z)dV \right] \right\}_{\ell=1}^{\ell=4}$$

By the divergence theorem, the first term of (3.2) can be reduced to a surface integral of[2] $\overline{\frac{\partial \varphi_g}{\partial n}}$ over the six surfaces which enclose V_ℓ. Since φ_g is not a

---

[1] The physical parameters $D_g(\underline{r})$, $\sigma_g(\underline{r})$, etc. are constant in each mesh volume V_ℓ and we denote these constants by $D_{g,\ell}$, $\sigma_{g,\ell}$, etc.

[2] Here $\overline{\frac{\partial \varphi_g}{\partial n}}$ represents the derivative of φ_g in the direction of the outward normal to the surface. Hence, $\overline{\frac{\partial \varphi_g}{\partial n}} = \pm \frac{\partial \varphi_g}{\partial n}$.

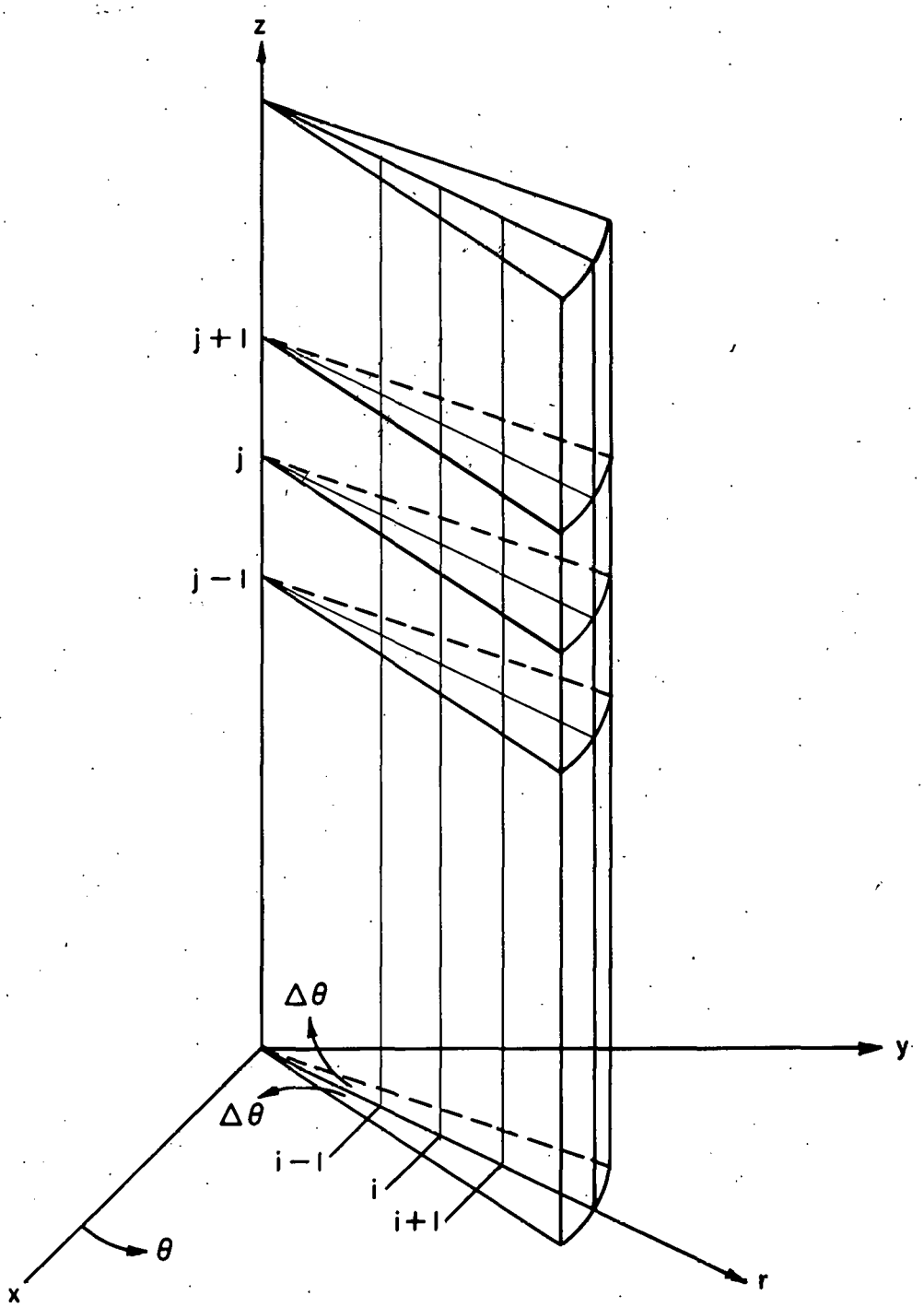


FIGURE 3.2

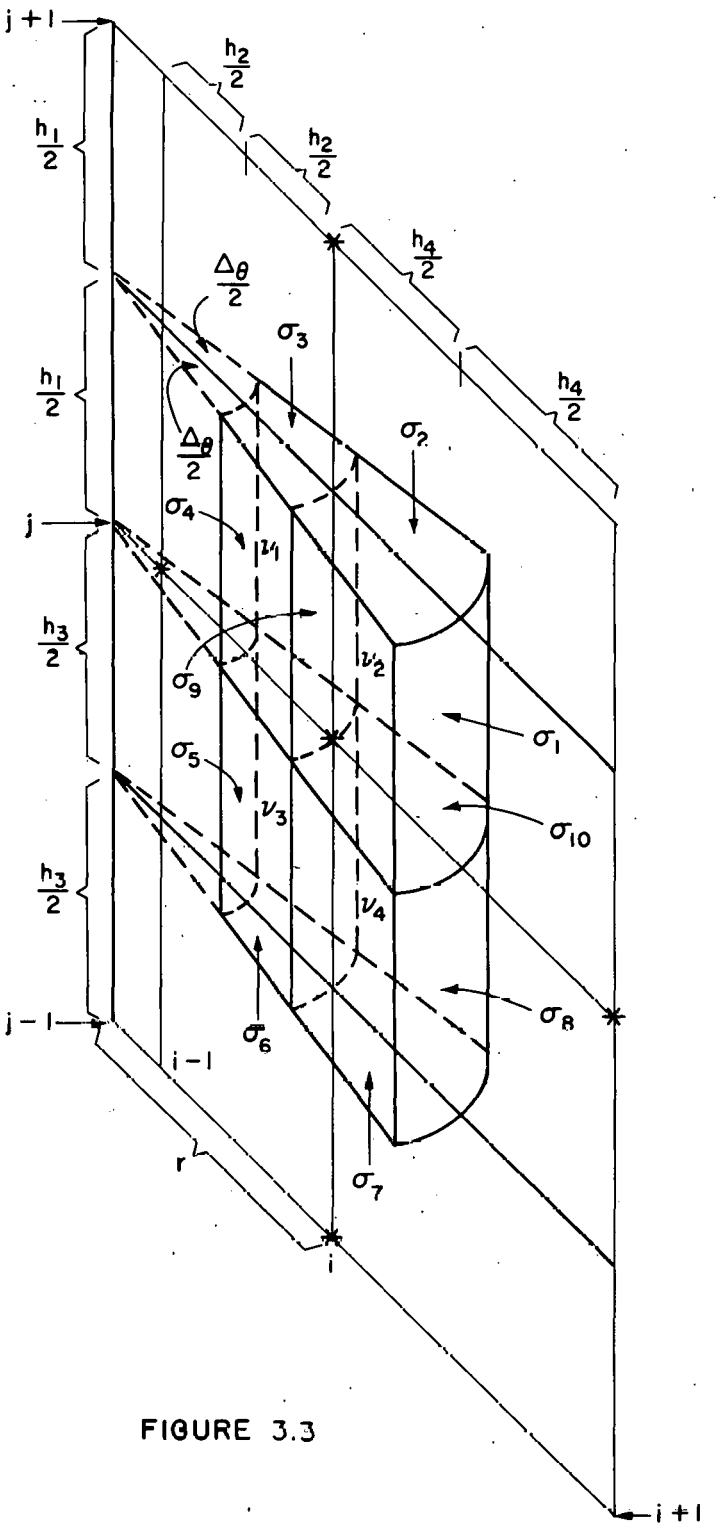

j + 1
h2/2
h1/2
h2/2
h4/2
Δθ/2
σ3
h1/2
h4/2
σ2
Δθ/2
σ4
j
ν1
h3/2
σ9
ν2
σ1
σ5
ν3
σ10
h3/2
ν4
σ8
j − 1
σ6
i − 1
r
σ7
i
i + 1

FIGURE 3.3

function of θ, $\overline{\frac{\partial \varphi_g}{\partial n}}$ is zero over the two vertical plane surfaces which enclose V_ℓ. Hence, say for volume element 2, Eq. (3.2) may be written as

$$(3.3) \qquad - D_{g,2} \left\{ \int_{\sigma_1} \overline{\frac{\partial \varphi_g}{\partial n}}\, d\sigma + \int_{\sigma_2} \overline{\frac{\partial \varphi_g}{\partial n}}\, d\sigma + \int_{\sigma_9} \overline{\frac{\partial \varphi_g}{\partial n}}\, d\sigma + \int_{\sigma_{10}} \overline{\frac{\partial \varphi_g}{\partial n}}\, d\sigma \right\}$$

$$+ \sigma_{g,2} \int_{V_2} \varphi_g(r,z)dV - \sigma^s_{g-1,2} \int_{V_2} \varphi_{g-1}(r,z)dV$$

$$= \frac{\chi_g}{\lambda} \sum_{g=1}^{G} \left[\nu\sigma^f_{g,2} \int_{V_2} \varphi_g(r,z)dV \right] .$$

Equations similar to (3.3) may be written for V_ℓ, $\ell = 1, 3, 4$. Since the neutron current, $- D \frac{\partial \varphi_g}{\partial n}$, is assumed to be continuous across interfaces, the surface integrals over the common surfaces cancel when the four expressions of (3.2) are added. Hence summing (3.2) over the four volume elements, we obtain

$$(3.4) \qquad - D_{g,2} \sum_{k=1}^{2} \int_{\sigma_k} \overline{\frac{\partial \varphi_g}{\partial n}}\, d\sigma - D_{g,1} \sum_{k=3}^{4} \int_{\sigma_k} \overline{\frac{\partial \varphi_g}{\partial n}}\, d\sigma$$

$$- D_{g,3} \sum_{k=5}^{6} \int_{\sigma_k} \overline{\frac{\partial \varphi_g}{\partial n}}\, d\sigma - D_{g,4} \sum_{k=7}^{8} \int_{\sigma_k} \overline{\frac{\partial \varphi_g}{\partial n}}\, d\sigma$$

$$+ \sum_{\ell=1}^{4} \left[\sigma_{g,\ell} \int_{V_\ell} \varphi_g(r,z)dV \right] - \sum_{\ell=1}^{4} \left[\sigma^s_{g-1,\ell} \int_{V_\ell} \varphi_{g-1}(r,z)dV \right]$$

$$= \frac{\chi_g}{\lambda} \sum_{g=1}^{G} \left[\sum_{\ell=1}^{4} \nu\sigma^f_{g,\ell} \int_{V_\ell} \varphi_g(r,z)dV \right] .$$

We now make numerical approximations to the integrals in (3.4) in order to obtain our finite-difference equation at mesh point (i,j). Integrals of the normal derivative such as $\int_{\sigma_1} \frac{\overline{\partial \varphi}_g}{\partial n} d\sigma$ are approximated by

$$(3.5) \qquad \int_{\sigma_1} \frac{\overline{\partial \varphi}_g}{\partial n} d\sigma \simeq \frac{\left[\varphi_g(i+1,j) - \varphi_g(i,j)\right]}{h_4} \int_{\sigma_1} d\sigma$$

$$\simeq \frac{\left[\varphi_g(i+1,j) - \varphi_g(i,j)\right]}{h_4} \left[\Delta\theta \frac{h_1}{2} (r + \frac{h_4}{2})\right]$$

and integrals such as $\int_{V_2} \varphi_g(r,z)dV$ are approximated by

$$(3.6) \qquad \int_{V_2} \varphi_g(r,z)dV \simeq \varphi_g(i,j) \int_{V_2} dV \simeq \varphi_g(i,j) \left[\frac{\Delta\theta\, h_1 h_4 (r + \frac{h_4}{4})}{4}\right] .$$

Using the above approximations, the general finite-difference equation at nodal point (i,j) may be written as

$$(3.7) \qquad \alpha_{g,1}\, \varphi_g(i,j+1) + \alpha_{g,2}\, \varphi_g(i-1,j) + \alpha_{g,3}\, \varphi_g(i,j-1)$$

$$+ \alpha_{g,4}\, \varphi_g(i+1,j) + \alpha_{g,0}\, \varphi_g(i,j) - \alpha_{g-1,5}\, \varphi_{g-1}(i,j)$$

$$= \frac{\chi_g}{\lambda} \sum_{g=1}^{G} \alpha_{g,6}\, \varphi_g(i,j) \quad ,$$

where

$$\alpha_{g,1} = - \frac{D_{g,1}\, h_2(r - \frac{h_2}{4}) + D_{g,2}\, h_4(r + \frac{h_4}{4})}{2h_1}$$

$$\alpha_{g,3} = - \frac{D_{g,3}\, h_2(r - \frac{h_2}{4}) + D_{g,4}\, h_4(r + \frac{h_4}{4})}{2h_3}$$

$$\alpha_{g,2} = - (\frac{r}{2h_2} - \frac{1}{4}) \left[D_{g,1}\, h_1 + D_{g,3}\, h_3 \right]$$

$$\alpha_{g,4} = - (\frac{r}{2h_4} + \frac{1}{4}) \left[D_{g,2}\, h_1 + D_{g,4}\, h_3 \right]$$

$$\alpha_{g,0} = - \left[\alpha_{g,1} + \alpha_{g,2} + \alpha_{g,3} + \alpha_{g,4} \right] + \frac{1}{4} \left\{ \sigma_{g,1}\, h_1 h_2 (r - \frac{h_2}{4}) \right.$$

$$\left. + \sigma_{g,2}\, h_1 h_4 (r + \frac{h_4}{4}) + \sigma_{g,3}\, h_2 h_3 (r - \frac{h_2}{4}) + \sigma_{g,4}\, h_3 h_4 (r + \frac{h_4}{4}) \right\}$$

$$\alpha_{g-1,5} = \frac{1}{4} \left\{ \sigma^s_{g-1,1}\, h_1 h_2 (r - \frac{h_2}{4}) + \sigma^s_{g-1,2}\, h_1 h_4 (r + \frac{h_4}{4}) \right.$$

$$\left. + \sigma^s_{g-1,3}\, h_2 h_3 (r - \frac{h_2}{4}) + \sigma^s_{g-1,4}\, h_3 h_4 (r + \frac{h_4}{4}) \right\}$$

$$\alpha_{g,6} = \frac{1}{4} \left\{ \nu\sigma^f_{g,1}\, h_1 h_2 (r - \frac{h_2}{4}) + \nu\sigma^f_{g,2}\, h_1 h_4 (r + \frac{h_4}{4}) \right.$$

$$\left. + \nu\sigma^f_{g,3}\, h_2 h_3 (r - \frac{h_2}{4}) + \nu\sigma^f_{g,4}\, h_3 h_4 (r + \frac{h_4}{4}) \right\} .$$

If the nodal point (i,j) lies on a segment of the boundary where $\frac{\partial \varphi}{\partial n} = 0$, then the constants $D_{g,\ell}$, $\sigma_{g,\ell}$, $\sigma^s_{g-1,\ell}$, and $\sigma^f_{g,\ell}$ for those regions which are outside of R are set to zero. No finite-difference equations are needed for mesh points on a segment of the boundary where $\varphi_g = 0$.

The finite-difference equation (see Fig. 3.1) used in the PDQ-5 program for x-y geometry at a nodal point (i,j) is[1]

$$(3.8)\qquad \alpha_{g,1}\,\varphi_g(i,j+1) + \alpha_{g,2}\,\varphi_g(i-1,j) + \alpha_{g,3}\,\varphi_g(i,j-1)$$

$$+ \alpha_{g,4}\,\varphi_g(i+1,j) + \alpha_{g,0}\,\varphi_g(i,j) - \alpha_{g-1,5}\,\varphi_{g-1}(i,j)$$

$$= \frac{\chi_g}{\lambda} \sum_{g=1}^{G} \alpha_{g,6}\,\varphi_g(i,j) \quad ,$$

where

$$\alpha_{g,1} = -\frac{D_{g,1}\,h_2 + D_{g,2}\,h_4}{2h_1}$$

$$\alpha_{g,2} = -\frac{D_{g,1}\,h_1 + D_{g,3}\,h_3}{2h_2}$$

$$\alpha_{g,3} = -\frac{D_{g,3}\,h_2 + D_{g,4}\,h_4}{2h_3}$$

$$\alpha_{g,4} = -\frac{D_{g,4}\,h_3 + D_{g,2}\,h_1}{2h_4}$$

$$\alpha_{g,0} = -\left[\alpha_{g,1} + \alpha_{g,2} + \alpha_{g,3} + \alpha_{g,4}\right] + \frac{1}{4}\left\{\sigma_{g,1}\,h_1h_2 + \sigma_{g,3}\,h_2h_3 + \sigma_{g,4}\,h_3h_4 + \sigma_{g,2}\,h_4h_1\right\}$$

---

[1] For a derivation of these equations, see Varga [17].

$$\alpha_{g-1,5} = \frac{1}{4}\left\{\sigma^s_{g-1,1}\ h_1h_2 + \sigma^s_{g-1,3}\ h_2h_3 + \sigma^s_{g-1,4}\ h_3h_4 + \sigma^s_{g-1,2}\ h_4h_1\right\}$$

$$\alpha_{g,6} = \frac{1}{4}\left\{\nu\sigma^f_{g,1}\ h_1h_2 + \nu\sigma^f_{g,3}\ h_2h_3 + \nu\sigma^f_{g,4}\ h_3h_4 + \nu\sigma^f_{g,2}\ h_4h_1\right\}$$

The difference equations at nodal boundary points for the case of x-y geometry are modified in the same manner as that described above for r-z geometry.

B. Statement of the Matrix Problem

In this section we wish to form the matrix problem which results from the above discrete approximation of (1.1). Unless otherwise stated, the discussions and results given for the rest of this paper are independent of whether the geometry is x-y or r-z. Also, as is the case for Fig. (3.1), we shall assume that there are MN nodal points which are determined by M horizontal and N vertical mesh lines.

For each group, using the above approximations, we have a linear equation for each of the nodal points. In order to represent this system of equations in matrix form, we need to order the equations and unknowns in some specific way. We shall do this by assigning an integral index number from 1 to MN to each nodal point and then order the equations and unknowns for group g such that the k-th equation corresponds to the finite-difference equation at the nodal point of index k and the k-th unknown corresponds to the $\varphi_g(i,j)$ at the nodal point of index k. For example, if the nodal points are indexed consecutively by rows, as shown by Fig. 3.4, we may express the discrete approximation to Eq. (1.1) in the matrix form

$$\left\{ A_g \underline{\varphi}_g = R_{g-1} \underline{\varphi}_{g-1} = \frac{X_g}{\lambda} \sum_{k=1}^{G} F_k \underline{\varphi}_k \right\}_{g=1}^{G} , \tag{3.9}$$

where A_g, R_{g-1}, X_g, and F_g are MN by MN matrices and the $\underline{\varphi}_g$ are MN-th order column vectors, the k-th term of which is the $\varphi_g(i,j)$ at the nodal point of index k. The matrix $A_g \equiv (a^g_{k,\ell})$ is given in Fig. (3.5). The $a^g_{k,k}$, $a^g_{k,k+1}$, $a^g_{k,k-1}$, $a^g_{k,k+N}$ and $a^g_{k,k-N}$ are, respectively, the $\alpha_{g,0}$, $\alpha_{g,4}$, $\alpha_{g,2}$, $\alpha_{g,1}$, and $\alpha_{g,3}$ given in the finite difference expression for the nodal point of index k. R_{g-1}, X_g, and F_g are nonnegative diagonal matrices with R_0 being the null matrix. The diagonal elements of the R_{g-1} matrix are the $\alpha_{g-1,5}$'s and the diagonal elements of F_g are the $\alpha_{g,6}$'s. X_g is a scalar matrix with χ_g on the diagonal.

Independent of the indexing of the nodal points, the discrete approximation to Eq. (1.1) may always be put in the matrix form of (3.9) and one can show [20] that

Theorem 3.1 The MN by MN matrices R_{g-1}, X_g, and F_g are all nonnegative diagonal matrices, with R_0 being the null matrix. The MN by MN matrices A_g are irreducible Stieltjes matrices. Moreover, the inverse of each A_g has all positive entries, i.e., $A_g^{-1} > 0$.

We now rewrite Eqs. (3.9) in the following matrix form:

$$E \underline{\Phi} = \frac{XF}{\lambda} \underline{\Phi} , \tag{3.10}$$

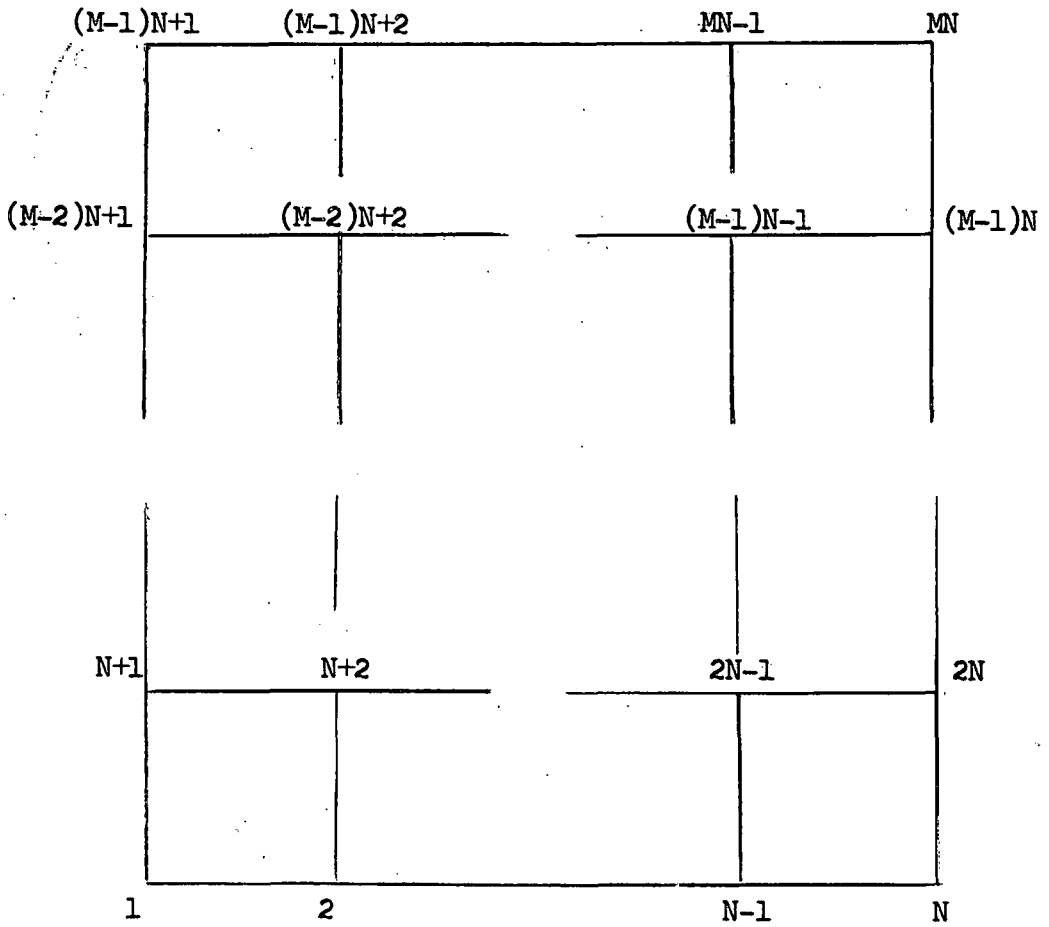


FIGURE 3.4

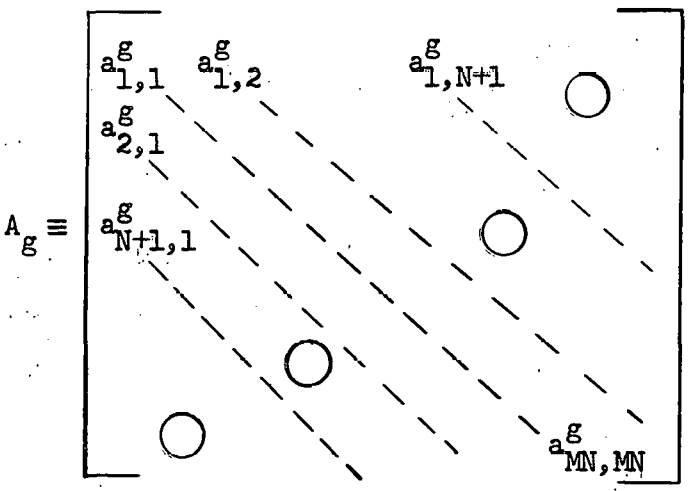


FIGURE 3.5

where

$$\Phi \equiv \begin{pmatrix} \underline{\varphi}_1 \\ \underline{\varphi}_2 \\ \vdots \\ \underline{\varphi}_g \end{pmatrix}, \qquad E \equiv \begin{pmatrix} A_1 & & & 0 \\ -R_1 & A_2 & & \\ & \ddots & \ddots & \\ 0 & & -R_{G-1} & A_G \end{pmatrix},$$

(3.11)

$$X = \begin{pmatrix} X_1 & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & X_G \end{pmatrix}, \text{ and } F = \begin{pmatrix} F_1 & F_2 & \cdots & F_G \\ F_1 & F_2 & \cdots & F_G \\ \vdots & \vdots & & \vdots \\ F_1 & F_2 & \cdots & F_G \end{pmatrix}$$

Since each A_g is nonsingular, the matrix E is nonsingular so that Eq. (3.10) may be written as

$$\lambda \Phi = E^{-1} XF \Phi \tag{3.12}$$

The discrete problem is then to determine the largest (in modulus) eigenvalue of (3.12) and its corresponding eigenvector. Birkhoff and Varga [3] have shown that

Theorem 3.2 The eigenvalue problem (3.12) possesses a simple, positive, largest (in modulus) fundamental eigenvalue, λ_1. Moreover, if R_{g-1} and X_g are not both

identically the null matrix in any group g[1], then the eigenvector $\underline{\Phi}_1$ corresponding to this fundamental eigenvalue may be chosen to have all positive components. Further, any positive eigenvector of $E^{-1}XF$ is a scalar multiple of $\underline{\Phi}_1$.

Thus, the discrete problem is well defined. We remind the reader that the fundamental solution of (3.12) is only an approximation to the desired solution of the continuous problem. The question as to whether the fundamental solution of (3.12) approaches the solution of (1.1) as the mesh is suitable refined has not been completely answered and will not be discussed here.

We note that the matrix $E^{-1}XF$ is of order GMN and hence has GMN eigenvalues. But the rank of XF, and hence also that of $E^{-1}XF$, is at most MN so that the matrix $E^{-1}XF$ has at most MN nonzero eigenvalues. Since we are not interested in the zero eigenvalues, we now shall obtain an equivalent eigenvalue problem of order MN for which these (G-1)MN zero eigenvalues have been eliminated.

To do this, we first introduce the fission source vector $\underline{\Psi}$, which is a vector of order MN and is defined by

$$\underline{\Psi} \equiv \frac{1}{\lambda} \sum_{g=1}^{G} F_g \underline{\varphi}_g \quad . \tag{3.13}$$

If the MN by MN matrices L_g are now defined recursively by

$$\begin{cases} L_g \equiv A_g^{-1}(X_g + R_{g-1}\, L_{g-1}) \quad , \; g = 1, 2, \ldots, G, \\ \text{where } L_0 \text{ is the null matrix} \end{cases} \tag{3.14}$$

---

[1] Henceforth, we shall assume that $X_g + R_{g-1}$ is not the null matrix.

then from Eqs. (3.10) and (3.11) we have

$$\underline{\varphi}_g = L_g \underline{\Psi} \qquad (3.15)$$

From the definition of $\underline{\Psi}$, we then obtain

$$\lambda \underline{\Psi} = T \underline{\Psi} \qquad (3.16)$$

where T is a MN by MN matrix and

$$T = \sum_{g=1}^{G} F_g L_g \quad .$$

Therefore, the fission source vector, $\underline{\Psi}$, must satisfy (3.16). Thus, if $\underline{\Phi}$ is an eigenvector of $E^{-1}XF$ corresponding to the nonzero eigenvalue λ, then $\underline{\Psi}$, obtained from $\underline{\Phi}$ using (3.13), is an eigenvector of T with the corresponding eigenvalue λ. Conversely, if $\underline{\Psi}$ is an eigenvector of T corresponding to the nonzero eigenvalue λ and if $\underline{\Phi}$ is obtained from $\underline{\Psi}$ using (3.15), then $\underline{\Phi}$ is an eigenvector of $E^{-1}XF$ with corresponding eigenvalue λ.

Using a similarity transformation, we now shall show that the nonzero eigenvalue spectrum of T is identical to that of $E^{-1}XF$. The reader may readily convince himself that the matrix $E^{-1}XF$ may be written as

$$E^{-1}XF = \begin{pmatrix} L_1F_1 & L_1F_2 & \cdots\cdots & L_1F_g \\ L_2F_1 & L_2F_2 & \cdots\cdots & L_2F_g \\ \vdots & \vdots & & \vdots \\ L_GF_1 & L_GF_2 & \cdots\cdots & L_GF_G \end{pmatrix} \qquad (3.17)$$

where the L_g are defined by (3.14). Since $\chi_1 > 0$, the matrix L_1 is nonsingular so that the matrix

$$P \equiv \begin{pmatrix} L_1^{-1} & 0 & \cdots & 0 \\ -L_2L_1^{-1} & I & 0 & \cdots & 0 \\ -L_3L_1^{-1} & 0 & I & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ -L_GL_1^{-1} & 0 & & & I \end{pmatrix}$$

exists and is nonsingular. It is easily verified that

$$P(E^{-1}XF)P^{-1} = \begin{pmatrix} T & F_2 & F_3 & \cdots & F_G \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

Hence, the nonzero eigenvalues of $E^{-1}XF$ are the same as those of $P(E^{-1}XF)P^{-1}$ which in turn are identical to the nonzero eigenvalues of T.

The L_g are positive matrices so that from Thm. 3.2 and Eq. (3.13) it follows that the eigenvector $\underline{\Psi}_1$ corresponding to the fundamental eigenvalue λ_1 may be chosen to have nonnegative components. Moreover, since the positive eigenvector of $E^{-1}XF$ is unique[1], any nonnegative eigenvector of T is either a scalar multiple of $\underline{\Psi}_1$ or else has a corresponding eigenvalue of zero.

Thus, the lower order eigenvalue problem $\lambda\underline{\Psi} = T\underline{\Psi}$ possesses essentially the same properties as does (3.12). In the next chapter we shall describe the iterative technique used to obtain the largest (in modulus) eigenvalue of T and its corresponding eigenvector.

---

[1] Up to a scalar factor.

IV. OUTER ITERATIONS

For the rest of this paper we shall assume that the eigenvalues of T are given by $\{\lambda_i\}_{i=1}^{MN}$, where $\lambda_1 > |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_{MN}|$, and that $\underline{\Psi}_i$ is the eigenvector associated with λ_i; i.e., $\lambda_i \underline{\Psi}_i = T\underline{\Psi}_i$.

The PDQ-5 program uses the "power iterative method" to obtain approximations[1] to the fundamental eigenvector and eigenvalue of the eigenvalue problem (3.16). Given an arbitrary nonnegative guess $\underline{\Psi}(0)$, the power method generates successive estimates for the fundamental eigenvector $\underline{\Psi}_1$ and eigenvalue λ_1 by the process

$$(4.1) \qquad \begin{cases} \underline{S}(\ell) = T\,\underline{\Psi}(\ell-1) \\ \lambda(\ell) = \dfrac{\left[\underline{S}(\ell),\ \underline{S}(\ell)\right]}{\left[\underline{S}(\ell),\ \underline{\Psi}(\ell-1)\right]} \\ \underline{\Psi}(\ell) = \dfrac{\underline{S}(\ell)}{\lambda(\ell)} \end{cases} , \quad \ell = 1,\ 2,\ \ldots$$

In this chapter we shall assume that we have ways to obtain the matrix-vector product $T\underline{\Psi}(\ell-1)$ needed to carry out the power method (4.1). We shall describe how this product $T\underline{\Psi}$ is actually obtained in Chapter V.

In the solution of the multi-group neutron diffusion problem the iterations (4.1) are generally called <u>outer iterations</u> and ℓ is called the outer iteration index. Since the largest (in modulus) eigenvalue of T is simple and real, the power method (4.1) is guaranteed to converge; i.e., for an arbitrary nonnegative guess vector $\underline{\Psi}(0)$,

---

[1] Theoretically, convergent iterative methods give the exact solution only after an infinite number of iterations.

$$\lim_{\ell \to \infty} \lambda(\ell) = \lambda_1 \quad \text{and} \quad \lim_{\ell \to \infty} \underline{\Psi}(\ell) = c\underline{\Psi}_1 \quad ,$$

where c is some positive constant.

The power method essentially involves repeated multiplication by the matrix T. For if we assume that the eigenvalue estimates $\lambda(\ell)$ of λ_1 are sufficiently accurate, then the process (4.1) gives $\underline{\Psi}(1) = \frac{T}{\lambda_1}\underline{\Psi}(0)$, $\underline{\Psi}(2) = \frac{T}{\lambda_1}\underline{\Psi}(1) = (\frac{T}{\lambda_1})^2\underline{\Psi}(0)$ and in general

$$\underline{\Psi}(\ell) = (\frac{T}{\lambda_1})^{\ell}\underline{\Psi}(0) \quad . \tag{4.2}$$

In order to see how quickly $\underline{\Psi}(\ell)$ approaches $\underline{\Psi}_1$ in (4.2), let us assume that $\underline{\Psi}(0)$ may be expanded in terms of the eigenvectors of T so that $\underline{\Psi}(0)$ may be written as $\underline{\Psi}(0) = \sum_{i=1}^{MN} c_i\underline{\Psi}_i$ for suitable scalars c_i. Hence, we may write

$$\underline{\Psi}(\ell) = c_1\underline{\Psi}_1 + \sum_{i=2}^{MN} c_i \left(\frac{\lambda_i}{\lambda_1}\right)^{\ell} \underline{\Psi}_i \quad .$$

Since $\left|\frac{\lambda_i}{\lambda_1}\right| < 1$ for $i > 1$, we see that $\underline{\Psi}(\ell)$ approaches $c_1\underline{\Psi}_1$ as ℓ tends toward infinity and that the convergence rate of $\underline{\Psi}(\ell)$ to $\underline{\Psi}_1$ depends on how well separated the fundamental eigenvalues λ_1 is from the other eigenvalues of T. In other words, the convergence rate of $\underline{\Psi}(\ell)$ depends on the <u>dominance ratio</u>

$$\bar{\sigma} = \max_{i \neq 1} \frac{|\lambda_i|}{\lambda_1} \quad . \tag{4.3}$$

Normally, the smaller this ratio is, the faster the convergence. The most slowly decaying component of the error with respect to the initial guess $\underline{\Psi}(0)$ decays as $(\bar{\sigma})^{\ell}$.

The next theorem [20] provides the basis for the convergence criterion used in the PDQ-5 program to terminate the iterative procedure (4.1).

Theorem 4.1 If $\underline{\Psi}(0)$ is a nonnegative vector, then the iterative procedure (4.1) is convergent; i.e., $\lim_{\ell \to \infty} \lambda(\ell) = \lambda_1$ and $\lim_{\ell \to \infty} \underline{\Psi}(\ell) = c\underline{\Psi}_1$, where c is a positive scalar. Moreover, if the k-th components of $\underline{\Psi}(\ell)$ and $\underline{S}(\ell)$ are denoted by $\Psi_k(\ell)$ and $S_k(\ell)$ and if for $\Psi_k(\ell-1) \neq 0$

$$\bar{\lambda}(\ell) \equiv \max_k \frac{S_k(\ell)}{\Psi_k(\ell-1)} \quad , \quad \underline{\lambda}(\ell) \equiv \min_k \frac{S_k(\ell)}{\Psi_k(\ell-1)} \quad , \tag{4.4}$$

then

$$\begin{cases} \bar{\lambda}(\ell) \geq \lambda_1 \geq \underline{\lambda}(\ell) \; ; \; \bar{\lambda}(\ell) \geq \lambda(\ell) \geq \underline{\lambda}(\ell) \; ; \text{ and} \\ \lim_{\ell \to \infty} \bar{\lambda}(\ell) = \lim_{\ell \to \infty} \underline{\lambda}(\ell) = \lambda_1 \end{cases} \tag{4.5}$$

The estimates $\lambda(\ell)$ are obtained by considering the components of the eigenvector estimates in the aggregate while $\bar{\lambda}(\ell)$ and $\underline{\lambda}(\ell)$ are obtained by considering these components individually. Hence, in addition to giving upper and lower bounds for the desired eigenvalue λ_1, $\bar{\lambda}(\ell)$ and $\underline{\lambda}(\ell)$ also give a good indication as to how well the eigenvector estimate $\underline{\Psi}(\ell)$ is converged. Thus, a practical criterion for terminating the iterative procedure (4.1) is that the inequality

$$\frac{\bar{\lambda}(\ell) - \underline{\lambda}(\ell)}{2\lambda(\ell)} \leq \varepsilon \tag{4.6}$$

be satisfied. The positive quantity ε is an input parameter.

In the next section we shall describe the use of Chebyshev polynomials in accelerating the rate of convergence of the basic power method.

A. The Use of Chebyshev Polynomials

Experience has shown that the eigenvalue estimates $\lambda(\ell)$ tend to converge faster[1] than the vector estimates $\underline{\Psi}(\ell)$ in the straight power method (4.1). Therefore, in an attempt to accelerate the convergence of the $\underline{\Psi}(\ell)$, we take linear combinations of the $\underline{\Psi}(\ell)$; i.e., for the ℓ-th iterate we take

$$\tilde{\underline{\Psi}}(\ell) = \sum_{p=0}^{\ell} a_{\ell p}\, \underline{\Psi}(p) \tag{4.7}$$

as the ℓ-th estimate for $\underline{\Psi}_1$. It is hoped that by a suitable choice of the constants $a_{\ell p}$, the vector $\tilde{\underline{\Psi}}(\ell)$ is a much better approximation to $\underline{\Psi}_1$ than $\underline{\Psi}(\ell)$ is.

In order to determine the "optimum"[2] $a_{\ell p}$, we shall assume that the eigenvalues λ_i of T are all real and nonnegative and that the corresponding eigenvectors form a basis for the associated vector space of T. Since we have assumed that the eigenvectors of T are complete, we may write the initial guess as

$$\underline{\Psi}(0) = \sum_{i=1}^{MN} c_i \underline{\Psi}_i \,. \tag{4.8}$$

---

[1] If the matrix T in (4.1) were symmetric, then the $\lambda(\ell)$ would converge at twice the rate of the $\underline{\Psi}(\ell)$. See Bilodeau and Hageman [2].

[2] In what sense the coefficients $a_{\ell p}$ are optimized will become clear later.

If we now assume that the eigenvalue estimates $\lambda(\ell)$ are sufficiently close to λ_1, then from Eqs. (4.1) and (4.8) we may write Eq. (4.7) as

$$\tilde{\underline{\Psi}}(\ell) \cong \sum_{i=1}^{MN} c_i \sum_{p=0}^{\ell} a_{\ell p} \left|\frac{\lambda_i}{\lambda_1}\right|^p \underline{\Psi}_i$$

or equivalently

$$\tilde{\underline{\Psi}}(\ell) \cong P_\ell\left|\frac{T}{\lambda_1}\right| \underline{\Psi}(0) = c_1 P_\ell(1)\underline{\Psi}_1 + \sum_{i=2}^{MN} c_i P_\ell\left|\frac{\lambda_i}{\lambda_1}\right| \underline{\Psi}_i \quad , \tag{4.9}$$

where $P_\ell(x) \equiv \sum_{p=0}^{\ell} a_{\ell p} x^p$.

The sum $\sum_{i=2}^{MN} c_i P_\ell\left|\frac{\lambda_i}{\lambda_1}\right|\underline{\Psi}_i$ is our error and hence we would like to choose $P_\ell(x)$ such that $P_\ell(1) = 1$ and $\sum_{i=2}^{MN} c_i P_\ell\left|\frac{\lambda_i}{\lambda_1}\right|\underline{\Psi}_i$ is minimized (in modulus). Since the scalars c_i are arbitrary[1] and the eigenvalues λ_i of T are not known, a true minimization is not feasible. However, a practical "optimum" minimization of $\sum_{i=2}^{MN} c_i P_\ell\left|\frac{\lambda_i}{\lambda_1}\right|\underline{\Psi}_i$ is obtained if we could determine $P_\ell(x)$, under the restriction that $P_\ell(1) = 1$, such that the maximum of $|P_\ell(x)|$ is minimized over the range $0 \le x \le \bar{\sigma}$, where $\bar{\sigma}$ is defined by Eq. (4.3). The solution to this practical minimization problem is well known [7] and can be given explicitly in terms of Chebyshev polynomials

$$P_\ell(x) = \frac{C_\ell\left|\frac{2x}{\bar{\sigma}} - 1\right|}{C_\ell\left|\frac{2}{\bar{\sigma}} - 1\right|} \tag{4.10}$$

---

[1] The c_i's are arbitrary since the initial guess $\underline{\Psi}(0)$ is arbitrary. We only assume that $c_i \neq 0$.

where the $C_\ell(y)$ are Chebyshev polynomials of degree ℓ and for $\ell \geq 0$

$$C_\ell(y) = \begin{cases} \cos(\ell \cos^{-1} y) & |y| \leq 1 \\ \cosh(\ell \cosh^{-1} y) & y \geq 1 \end{cases}$$

From the well-known three-term recurrence relation which the Chebyshev polynomials satisfy

$$(4.11) \qquad \begin{cases} C_{\ell+1}(y) = 2yC_\ell(y) - C_{\ell-1}(y) & , \ \ell \geq 1 , \\ C_0(y) = 1 \ ; \ C_1(y) = y \end{cases}$$

it is possible to obtain the three-term relation

$$(4.12) \qquad \begin{cases} P_{\ell+1}(x) = 2 \left(\dfrac{\cosh [\ell\gamma]}{\cosh [(\ell+1)\gamma]} \right) \left(\dfrac{2x}{\bar{\sigma}} - 1 \right) P_\ell(x) \\ \qquad\qquad - \left(\dfrac{\cosh [(\ell-1)\gamma]}{\cosh [(\ell+1)\gamma]} \right) P_{\ell-1}(x), \quad \ell \geq 1 , \end{cases}$$

where $P_0(x) = 1$, $P_1(x) = \dfrac{\left(\frac{2x}{\bar{\sigma}} - 1 \right)}{\left(\frac{2}{\bar{\sigma}} - 1 \right)}$, and $\gamma = \cosh^{-1} \left| \frac{2}{\bar{\sigma}} - 1 \right|$,

for the $P_\ell(x)$. Thus, using the recurrence relation (4.12), the $\tilde{\Psi}(\ell)$ vectors of Eq. (4.9) may be obtained using the iterative procedure[1]

---

[1]The identity $1 - \alpha_\ell + \beta_\ell \equiv - \dfrac{\bar{\sigma}}{2} \alpha_\ell$ is also used in obtaining the three-term relation for $\tilde{\Psi}(\ell)$ given in (4.13).

$$(4.13)\quad \begin{cases} \tilde{\underline{S}}(\ell) = T\tilde{\underline{\Psi}}(\ell-1) \\ \lambda(\ell) = \dfrac{[\tilde{\underline{S}}(\ell),\ \tilde{\underline{S}}(\ell)]}{[\tilde{\underline{S}}(\ell),\ \tilde{\underline{\Psi}}(\ell-1)]} \\ \tilde{\underline{\Psi}}(\ell) = \tilde{\underline{\Psi}}(\ell-1) + \alpha_\ell\left[\dfrac{\tilde{\underline{S}}(\ell)}{\lambda(\ell)} - \tilde{\underline{\Psi}}(\ell-1)\right] + \beta_\ell\left[(\tilde{\underline{\Psi}}(\ell-1) + \tilde{\underline{\Psi}}(\ell-2)\right] \end{cases}$$

for $\ell \geq 1$ and where

$$\alpha_1 = \frac{2}{2 - \bar{\sigma}} \ ; \quad \beta_1 = 0 \ ;$$

$$\alpha_\ell = \frac{4}{\bar{\sigma}}\left[\frac{\cosh\,[(\ell-1)\gamma]}{\cosh\,[\ell\gamma]}\right], \quad \beta_\ell = \frac{\cosh\,[(\ell-2)\gamma]}{\cosh\,[\ell\gamma]} \quad \text{for } \ell \geq 2 \ .$$

When the iterative procedure (4.13) is used, $\bar{\lambda}(\ell)$ and $\underline{\lambda}(\ell)$ may be computed as before (using, of course, $\tilde{\underline{S}}(\ell)$ and $\tilde{\underline{\Psi}}(\ell-1)$ instead of $\underline{S}(\ell)$ and $\underline{\Psi}(\ell-1)$) and $\lim_{\ell\to\infty} \bar{\lambda}(\ell) = \lim_{\ell\to\infty} \underline{\lambda}(\ell) = \lambda_1$. Moreover, if $\tilde{\underline{\Psi}}(\ell-1)$ is a nonnegative vector, then the inequalities in (4.5) are also valid.

Figure 4.1 illustrates the effect of the polynomial $P_4\left(\frac{T}{\lambda_1}\right)$ with $\bar{\sigma} = .9$ operating on an arbitrary vector $\underline{\Psi}(0)$. For if $\underline{\Psi}(0)$ is expanded in terms of the eigenvectors of T as in Eq. (4.8), then from Fig. 4.1 we see that $P_4\left(\frac{T}{\lambda_1}\right)$ operating on $\underline{\Psi}(0)$ has the effect of multiplying the coefficients of all eigenvectors except $\underline{\Psi}_1$ by a factor of .145 or less while the coefficient of $\underline{\Psi}_1$ remains unchanged. Also shown in Fig. 4.1 is the effect on the coefficients produced by the operator $\left(\frac{T}{\lambda_1}\right)^4$, which corresponds to performing four power method iterations (4.1).

Using the Chebyshev polynomial method of iteration (4.13), we see from Eq. (4.9) that the most slowly decaying component of the error in the initial

GRAPH OF $P_4\left(\frac{\lambda}{\lambda_1}\right)$ WITH $\bar{\sigma} = 0.9$

AND GRAPH OF $\left(\frac{\lambda}{\lambda_1}\right)^4$

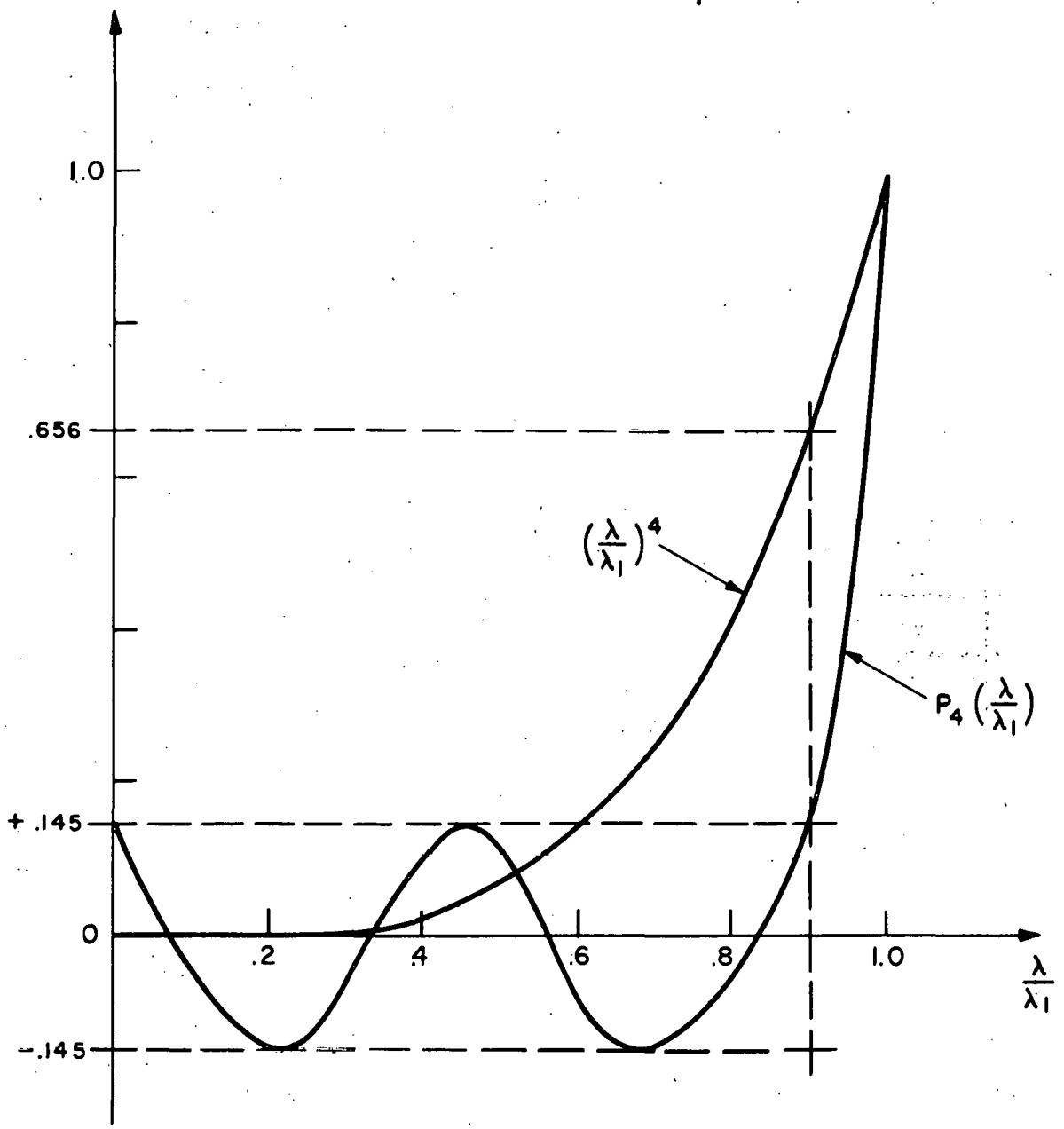


FIGURE 4.1

guess expansion decays as

$$\max_{i \neq 1} \left| P_\ell\left(\frac{\lambda_i}{\lambda_1}\right) \right| \leq \max_{0 \leq x \leq \bar{\sigma}} \left| P_\ell(x) \right| = \frac{1}{C_\ell\left(\frac{2}{\bar{\sigma}} - 1\right)} \quad . \tag{4.14}$$

For the straight power method (4.1), as we have seen previously, the most slowly decaying component decays as $(\bar{\sigma})^\ell$. Table 4.1 gives some indication of the gain in the decay of the most slowly decaying component of the error one obtains by using the Chebyshev polynomial method of iteration.

	1 ITERATION		5 ITERATIONS		10 ITERATIONS	
$\bar{\sigma}$	$(\bar{\sigma})^1$	$\left[C_1\left(\frac{2}{\bar{\sigma}} - 1\right)\right]^{-1}$	$(\bar{\sigma})^5$	$\left[C_5\left(\frac{2}{\bar{\sigma}} - 1\right)\right]^{-1}$	$(\bar{\sigma})^{10}$	$\left[C_{10}\left(\frac{2}{\bar{\sigma}} - 1\right)\right]^{-1}$
.5	.5	.33333	.03125	.00030	.00098	---
.8	.8	.66667	.32768	.01626	.10770	.00013
.9	.9	.81818	.59049	.07556	.34868	.00286
.98	.98	.96080	.90392	.45533	.81707	.11565

TABLE 4.1

The efficiency of the Chebyshev polynomial method of iteration depends on a prior knowledge of $\bar{\sigma}$ and λ_1. Generally, of course, these constants are not known. However, practical numerical means do exist for estimating these constants. For example, before starting the Chebyshev method of iteration (4.13), four or five straight power iterations (4.1) may be performed in order to obtain an initial estimate for λ_1. Further, if a low degree Chebyshev polynomial is repeatedly applied instead of trying to apply one high degree

polynomial, then the estimates of λ_1 may be continually updated. For example, after performing the initial ℓ_1 power iterations, instead of applying one high degree Chebyshev polynomial

$$\tilde{\underline{\Psi}}(\ell) = P_{\ell-\ell_1}\left(\frac{T}{\lambda(\ell_1)}\right)\left(\frac{T}{\lambda(0)}\right)^{\ell_1}\underline{\Psi}(0)$$

we do[1] instead

$$\tilde{\underline{\Psi}}(\ell) = P_{\ell_n}\left(\frac{T}{\lambda(\ell-\ell_n)}\right) \cdots\cdots P_{\ell_3}\left(\frac{T}{\lambda(\ell_1+\ell_2)}\right) P_{\ell_2}\left(\frac{T}{\lambda(\ell_1)}\right)\left(\frac{T}{\lambda(0)}\right)^{\ell_1}\underline{\Psi}(0) ,$$

where $\ell_1 + \ell_2 + \cdots \ell_n = \ell$. The numerical methods used to obtain initial and updated estimates for the dominance ratio, $\bar{\sigma}$, will be described in Chapter VI.

The assumptions that the eigenvectors of T form a complete set and that the eigenvalues are real and nonnegative[2] have not been shown to be valid except for special cases. However, the method as given above has been used quite successfully. This seems to indicate that the above assumptions are very nearly satisfied.

We remark that the results of this chapter apply equally well to the eigenvalue problem $\lambda\underline{\Phi} = E^{-1}XF\,\underline{\Phi}$. In other words, the results of this chapter are valid if $\underline{\Psi}$ were replaced by $\underline{\Phi}$ and the matrix T replaced by $E^{-1}XF$. From a

---

[1]There are other reasons (see Chapter VI and Appendix B) for using repeated applications of low degree Chebyshev polynomials instead of one high degree polynomial.

[2]The fact that the eigenvalues are nonnegative is not crucial in the application of the Chebyshev polynomial method of iteration. To take into account negative eigenvalues, one needs only to change the argument of the Chebyshev polynomials. See Bilodeau and Hageman [2]. Also, the restriction that the eigenvalues be real may be weakened somewhat if only low degree Chebyshev polynomials are used. As the degree of the Chebyshev polynomial is increased, then the restriction on the eigenvalues becomes more nearly that they be real. See Varga [18].

practical point of view, the eigenvalue problem $\lambda\underline{\Psi} = T\underline{\Psi}$ has the advantage in that $\underline{\Psi}$ is a lower order vector than $\underline{\Phi}$.

V. INNER ITERATIONS

In this chapter we shall describe how the matrix-vector product $T\underline{\Psi}(\ell-1)$ may be obtained (or at least approximated).

From Eqs. (3.15 and 3.16), $T\underline{\Psi}(\ell-1)$ may be written as

$$T\underline{\Psi}(\ell-1) = \sum_{g=1}^{G} F_g L_g \underline{\Psi}(\ell-1) = \sum_{g=1}^{G} F_g \underline{\varphi}_g(\ell) \tag{5.1}$$

where

$$\underline{\varphi}_g(\ell) \equiv L_g \underline{\Psi}(\ell-1) \tag{5.2}$$

Hence, the matrix-vector product $T\underline{\Psi}(\ell-1)$ is readily obtained once the $\underline{\varphi}_g(\ell)$ are known. But from the definition of L_g (see Eq. (3.14)), the $\underline{\varphi}_g(\ell)$ can be determined by solving successively the system of group equations

$$\left\{ A_g \underline{\varphi}_g(\ell) = X_g \underline{\Psi}(\ell-1) + R_{g-1} \underline{\varphi}_{g-1}(\ell) \right\}_{g=1}^{G} , \tag{5.3}$$

where R_0 is the null matrix. Thus, the product $T\underline{\Psi}(\ell-1)$ can be determined if we can solve matrix equations of the form

$$A_g \underline{\varphi}_g(\ell) = \underline{k}_g(\ell) , \tag{5.4}$$

where A_g is the MN by MN matrix given in Theorem 3.1 and $\underline{k}_g(\ell)$ is a known nonzero column vector which is defined by $\underline{k}_g(\ell) \equiv X_g \underline{\Psi}(\ell-1) + R_{g-1} \underline{\varphi}_{g-1}(\ell)$.

For one-dimensional problems, direct inversions of the A_g are possible[1] so that the group fluxes $\underline{\varphi}_g(\ell)$ may be obtained without any complications. For two- and three-dimensional problems, however, direct inversions are not feasible

---

[1] See, for example, Marlowe and Suggs [14].

and the $\underline{\varphi}_g(\ell)$ must be approximated by some iterative process. The iterations used to obtain an approximation for the $\underline{\varphi}_g(\ell)$ are called inner iterations.

Before describing in detail the iterative techniques used by the PDQ-5 program to obtain approximations for the $\underline{\varphi}_g(\ell)$, we shall give a short discussion on the mechanics of stationary iterative techniques and shall define some particular iterative processes.

A. Stationary Iterative Techniques

We are interested in obtaining by iterative means a good approximation to the solution of the matrix problem

$$A\underline{x} = \underline{g} \tag{5.5}$$

where A is a MN by MN nonsingular matrix, $\underline{g}$ is a known column vector with MN components and $\underline{x}$ is the MN-th order column vector of unknowns.

In general, a stationary first degree linear iterative procedure for obtaining successive approximations to A^{-1} may be described as

$$\underline{x}^{(m)} = P\underline{x}^{(m-1)} + H\underline{g} \ , \ m = 1, 2, \ldots, \tag{5.6}$$

where P and H are MN by MN matrices, m is the iteration index[1], and $\underline{x}^{(0)}$ is an arbitrary guess vector. If $\underline{x}^{(m-1)}$ is equal to the unique solution $A^{-1}\underline{g}$, then for $\underline{x}^{(m)}$ to also equal the unique solution we further demand that $P + HA = I$, where I is the MN by MN identity matrix.

We now ask under what conditions does the infinite sequence of vectors $\{\underline{x}^{(m)}\}$ generated by Eq. (5.6) converge to the unique solution $\underline{x} = A^{-1}\underline{g}$. Let

---

[1] In this paper, the superscript (m) will always denote the inner iteration cycle number. As used previously, (ℓ) will always denote the outer iteration cycle number.

$\underline{E}^{(m)} = \underline{x}^{(m)} - \underline{x}$ be the error vector at the end of the m-th iteration. Since $P + HA = I$, we have from (5.6) that

$$\underline{E}^{(m)} = P\underline{E}^{(m-1)} = P^m\underline{E}^{(0)} \tag{5.7}$$

For an arbitrary initial error $\underline{E}^{(0)}$, it is clear from (5.7) that the error vector becomes arbitrarily small as $m \to \infty \iff P^m$ converges to the null matrix as $m \to \infty$. Thus, we make the following definition:

Definition 5.1 The iteration procedure defined by Eq. (5.6) is said to be convergent $\iff$ the sequence $\left\{P^m\right\}_{m=1}^{\infty}$ converges to the null matrix. Otherwise the iteration procedure is said to be divergent.

If $\sigma_1, \ldots, \sigma_{MN}$ are the eigenvalues of the MN by MN matrix P, then

$$\rho(P) \equiv \max_i |\sigma_i|$$

is called the spectral radius of the matrix P and [21].

Theorem 5.1 A necessary and sufficient condition for the iterative process (5.6) to be convergent is that $\rho(P) < 1$.

The spectral radius, $\rho(P)$, can also be used as a measure of the effectiveness of the iterative process. In general, the smaller the ρ the faster the process will converge and

$$R_\infty(P) \equiv -\ell n\ \rho(P)$$

is called the asymptotic rate of convergence of the matrix P. The significance of this definition is that the number of iterations required to reduce the initial

error vector by a certain fraction is approximately (in a certain asymptotic sense) inversely proportional to the asymptotic rate of convergence. Thus, if one iterative process has an asymptotic rate of convergence twice as big as another, it will require roughly half as many iterations for the same degree of convergence[1].

In order to define some basic stationary iterative processes we let the MN by MN matrix A be partitioned in the form

$$A = \begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,Q} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,Q} \\ \vdots & & & \vdots \\ A_{Q,1} & \cdots & \cdots & A_{Q,Q} \end{pmatrix}, \tag{5.8}$$

where the diagonal blocks $A_{i,i}$ are nonsingular "easily solvable" submatrices[2] of A, and let the matrices D, E, and F be defined by

---

[1] For more detailed discussions of convergence rates, the reader is referred to Varga [21, p. 61] and Keller [13].

[2] By saying that a nonsingular matrix D is "easily solvable" we mean that if $D\underline{y} = \underline{k}$, then $\underline{y}$ may be easily obtained by a direct method. We do not mean that D^{-1} is easily available. For example, a tri-diagonal matrix is "easily solvable".

$$D \equiv \begin{pmatrix} A_{1,1} & & \bigcirc \\ & A_{2,2} & \\ & \ddots & \\ \bigcirc & & A_{Q,Q} \end{pmatrix} ,$$

(5.9)

$$E \equiv \begin{pmatrix} \bigcirc & & \bigcirc \\ -A_{2,1} & \ddots & \\ \vdots & & \\ -A_{Q,1} & -A_{Q,2} \cdots & \bigcirc \end{pmatrix}, \quad F \equiv \begin{pmatrix} \bigcirc & -A_{1,2} \cdots -A_{1,2} \\ & \ddots & \vdots \\ \bigcirc & & \bigcirc \end{pmatrix} .$$

We may write $A\underline{x} = \underline{g}$ as $D\underline{x} = (E + F)\underline{x} + \underline{g}$ and this suggests the iteration method

$$D\underline{x}^{(m+1)} = (E + F)\underline{x}^{(m)} + \underline{g} \quad . \tag{5.10}$$

We shall refer to the method defined by Eq. (5.10) as the Jacobi iterative method and the matrix

$$J = D^{-1}(E + F) \tag{5.11}$$

as the Jacobi iteration matrix associated with the matrix A. This scheme is also known as iteration by simultaneous displacements or total steps by Geiringer [9].

If $x_i^{(m+1)}$ is the i-th component of the vector $\underline{x}^{(m+1)}$, then the Jacobi method is to solve for each component $x_i^{(m+1)}$ using only the previous iterate

$\underline{x}^{(m)}$. It seems plausible that if one always uses the latest value of all other components when solving for a component $x_i^{(m+1)}$, that the rate of convergence might be increased. With this in mind, the Gauss-Seidel iteration method is defined as

$$D\underline{x}^{(m+1)} = E\underline{x}^{(m+1)} + F\underline{x}^{(m)} + \underline{g} \quad . \tag{5.12}$$

The matrix

$$\mathcal{L}_1 = (D - E)^{-1}F \tag{5.13}$$

is called the Gauss-Seidel iteration matrix associated with the matrix A. The Gauss-Seidel method is also known as iteration by successive displacements or single steps by Geiringer [9] and the Liebman method by Frankel [8].

The third basic iteration method with which we are concerned is formed by modifying the Gauss-Seidel method as follows:

$$D\underline{x}^{(m+1)} = \omega\left\{E\underline{x}^{(m+1)} + F\underline{x}^{(m)} + \underline{g}\right\} + (1 - \omega)D\underline{x}^{(m)} \quad . \tag{5.14}$$

We shall follow Young [24] and call this method the successive overrelaxation method and shall refer to the matrix

$$\mathcal{L}_\omega = (D - \omega E)^{-1}\left\{\omega F + (1 - \omega)D\right\} \tag{5.15}$$

as the successive overrelaxation iteration matrix associated with the matrix A. The parameter ω is called the relaxation factor. When $\omega = 1$, the scheme defined by Eq. (5.14) reduces to the Gauss-Seidel method. This iterative method is also called the accelerated Liebman method by Frankel [8] and the extrapolated Gauss-Seidel method by Kahan, [12]. It should be noted that in using the successive overrelaxation method we are faced with two questions:

first, for which values of ω (if any) does the method defined by Eq. (5.14) converge and second, if the method does converge for some values of ω, which value of ω gives the highest rate of convergence? In order to answer the above questions for our specific problem, we introduce the following definitions:

Definition 5.2 If the MN by MN matrix A is partitioned in the form of (5.8), then the matrix A is 2-cyclic (relative to the partitioning of (5.8)) $\Longleftrightarrow$ there exist two disjoint nonempty subsets S and T such that $S \cup T = \{1, 2, \ldots, Q\}$ and if $A_{i,j} \not\equiv 0$ and $i \neq j$, then $i \in S$ and $j \in T$ or $j \in S$ and $i \in T$[1].

Definition 5.3 If the MN by MN matrix A is partitioned in the form of (5.8) and is 2-cyclic, then A is consistently ordered $\Longleftrightarrow$ there exists a vector $\underline{\gamma} \equiv (\gamma_1, \gamma_2, \ldots, \gamma_Q)$ with integral components such that if $A_{i,j} \not\equiv 0$ and $i \neq j$, then

$$\left\{ \begin{array}{ll} \gamma_j - \gamma_i = 1 & \text{for } j > i \\ \gamma_j - \gamma_i = -1 & \text{for } i > j \end{array} \right\}$$

For any n x n matrix A, $n \geq 2$, there exists a partitioning such that A is 2-cyclic. For example, the matrix

$$A = \begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix}$$

---

[1] The 2-cyclic property defined here is called property A^{Π} by Arms, Gates and Zondek [1] and reduces to Young's [24] property (A) when the diagonal submatrices are all one by one matrices.

is 2-cyclic. Also, not all 2-cyclic matrices are consistently ordered but for any 2-cyclic matrix there exists an ordering which is consistent. The concept of directed graphs [21, pp. 48 and 121] is very useful in determining if a matrix with a particular partitioning is 2-cyclic and is consistently ordered.

We now state the main result for consistently ordered, 2-cyclic matrices.

Theorem 5.2[1] If A, partitioned in the form (5.8), is a consistently ordered, 2-cyclic matrix, then $\rho^2(J) = \rho(\mathcal{L}_1)$. Moreover, if the eigenvalues of the Gauss-Seidel iteration matrix $\mathcal{L}_1$ are nonnegative and less than unity and if

$$\omega_b = \frac{2}{1 + \sqrt{1 - \rho(\mathcal{L}_1)}}, \tag{5.16}$$

then

$$\begin{cases} \rho(\mathcal{L}_\omega) > \rho(\mathcal{L}_{\omega_b}) & \text{if } \omega \neq \omega_b \quad \text{and} \\ \rho(\mathcal{L}_{\omega_b}) = \omega_b - 1 \end{cases} \tag{5.17}$$

If A is partitioned in the form (5.8), then point iterative methods correspond to the case when the $A_{i,i}$, for $i = 1, 2, \ldots, MN$, are one by one matrices. Block iteration methods correspond to the case where no special restrictions are placed on the order of the submatrices $A_{i,i}$. In point methods only one single unknown is modified at each step of the iterative procedure while for block methods a group of unknowns may be modified simultaneously.

[1] Young [24] did the basic analysis for the application of the successive overrelaxation method to the point 2-cyclic matrix. Arms, Gates and Zondek [1] generalized Young's results to block methods and Varga [21] later extended the theory to the more general p-cyclic matrices.

In the next section, we shall describe the "cyclically reduced three-line successive overrelaxation method" [11] which is the inner iteration method used in the PDQ-5 program.

B. The Cyclically Reduced Three-Line Successive Overrelaxation Method

We now concentrate on a particular iterative method used to obtain approximations to the solution of the matrix equation

$$A_g \underline{\varphi}_g(\ell) = \underline{k}_g(\ell) \quad , \tag{5.18}$$

where A_g is the MN by MN matrix of Theorem 3.1 and $\underline{k}_g(\ell)$ is a known column vector given by $\underline{k}_g(\ell) \equiv X_g \underline{\psi}(\ell-1) + R_{g-1} \underline{\varphi}_{g-1}(\ell)$. In what follows we shall drop the group subscript g and the outer iteration index (ℓ) in Eq. (5.18).

In order to completely specify the matrix equation (5.18), we need to give an ordering for the equations and unknowns. As in Chapter III, we shall do this by indexing the MN nodal points. We first split the nodal points into what we shall call <u>square nodal points</u> and <u>circle nodal points</u>. We do this by first making the nodal point 1 of Fig. 3.4 a square point and then proceeding by making circle points of the four (or fewer if near or on a boundary) nearest neighbors of the square points and making square points of the four nearest neighbors of the circle points. We now index the nodal points by indexing first all the square nodal points consecutively by rows and then all the circle nodal points consecutively by rows. For example,

see Fig. 5.1[1]. The matrix $A \equiv (a_{i,j})$ is given in Fig. 5.2. The nonzero

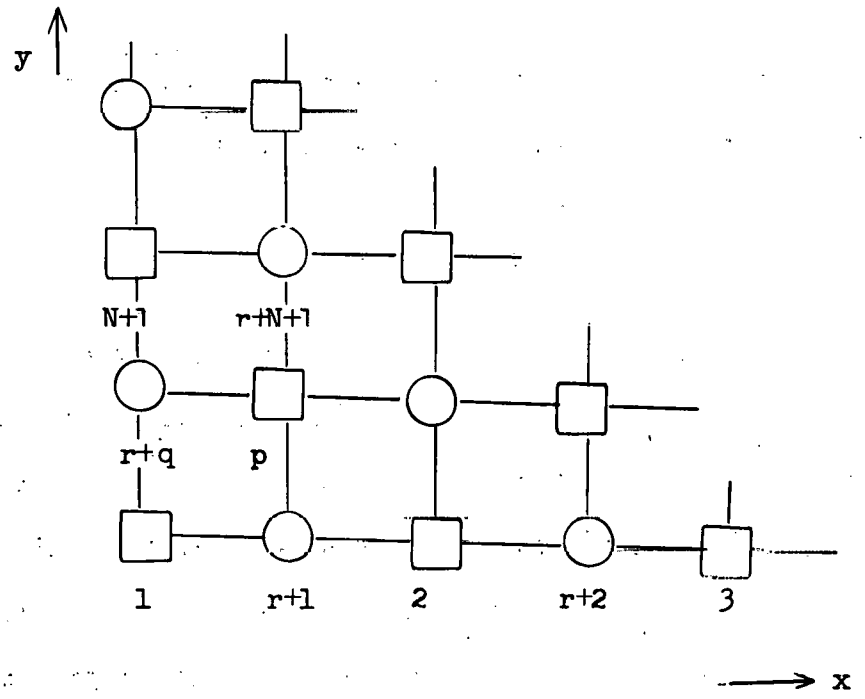


FIGURE 5.1

---

[1]In Fig. 5.1, if MN is even, then $r = \frac{MN}{2}$ and if MN is odd, then $r = \frac{MN+1}{2}$. Similarly, if N is even, then $p = q = \frac{N}{2} + 1$ and if N is odd, then $p = \frac{N}{2} + \frac{3}{2}$ and $q = \frac{N}{2} + \frac{1}{2}$.

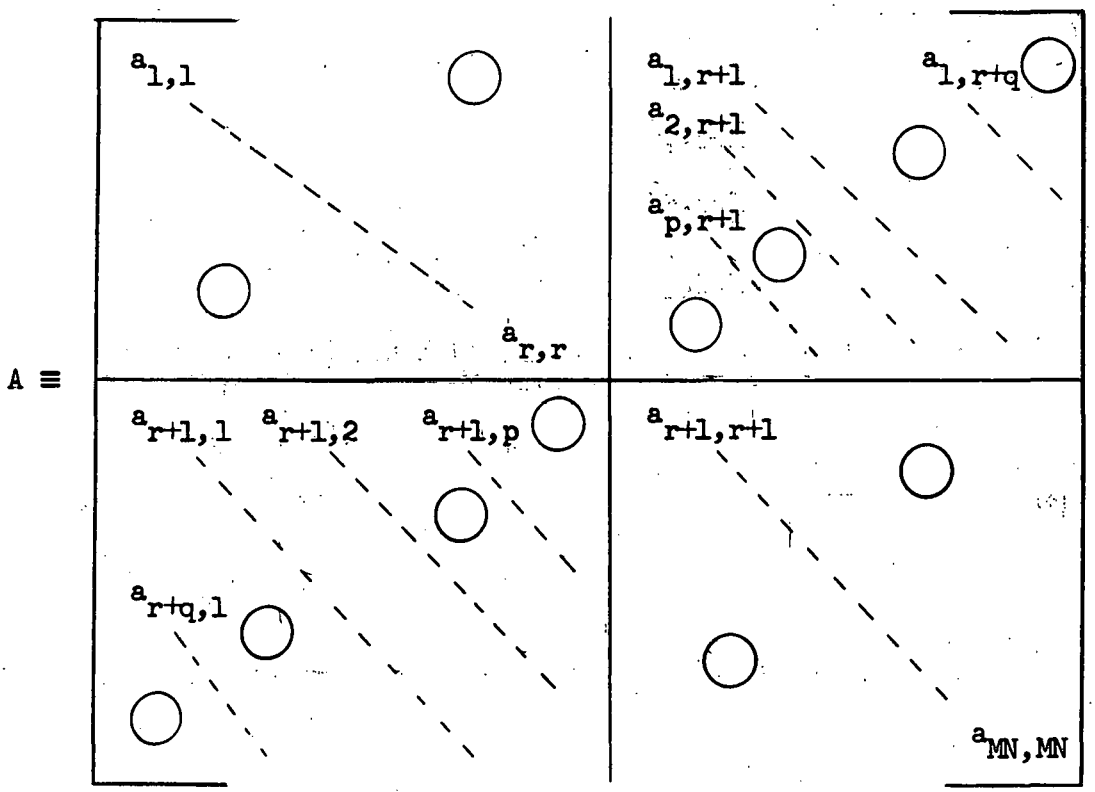


FIGURE 5.2

elements of A are the $\alpha_{g,0}$, $\alpha_{g,1}$, $\alpha_{g,2}$, $\alpha_{g,3}$, and $\alpha_{g,4}$ given in the finite difference expression (3.7) or (3.8).

If we now let

(5.19)

$$D \equiv \begin{pmatrix} \sqrt{a_{1,1}} & & \\ & \ddots & \\ & & \sqrt{a_{MN,MN}} \end{pmatrix}$$

then Eq. (5.18) may be expressed as

$$(I - B)\underline{\alpha} = \underline{g} \quad , \tag{5.20}$$

where

$$\underline{\alpha} = D\underline{\varphi}, \quad (I - B) = D^{-1}AD^{-1}, \quad \underline{g} = D^{-1}\underline{k} \tag{5.21}$$

Since A is symmetric (see Thm. 3.1), the matrix B may be written as

$$B \equiv \begin{pmatrix} 0 & B_1 \\ B_1^T & 0 \end{pmatrix} \tag{5.22}$$

We note that the matrix $B_1 \equiv (b_{i,j})$ is a r by (MN-r) matrix and is explicitly given by

$$b_{i,j} = \frac{-a_{i,j+r}}{\sqrt{a_{i,i}}\,\sqrt{a_{j+r,j+r}}} \qquad \begin{matrix} i = 1, 2, \ldots, r \\ j = 1, 2, \ldots, MN - r \end{matrix}$$

Since $a_{i,j} > 0$ if $i = j$ and $a_{i,j} \leq 0$ if $i \neq j$, it follows that B_1 is a nonnegative matrix.

If

$$\underline{\alpha} \equiv \begin{pmatrix} \underline{x} \\ \underline{\alpha}_1 \end{pmatrix} \quad \text{and} \quad \underline{g} \equiv \begin{pmatrix} \underline{g}_1 \\ \underline{g}_2 \end{pmatrix} \tag{5.23}$$

are partitionings for $\underline{\alpha}$ and $\underline{g}$ which are consistent with the partitioning of B in (5.23), then Eq. (5.20) can be written as

$$(5.24) \qquad \begin{pmatrix} I & -B_1 \\ -B_1^T & I \end{pmatrix} \begin{pmatrix} \underline{x} \\ \underline{a}_1 \end{pmatrix} = \begin{pmatrix} \underline{g}_1 \\ \underline{g}_2 \end{pmatrix}$$

and multiplying both sides of this equation by $(I + B)$ gives

$$(I - B^2)\underline{a} = (I + B)\underline{g}$$

or equivalently

$$(5.25) \qquad \begin{pmatrix} I - B_1 B_1^T & 0 \\ 0 & I - B_1^T B_1 \end{pmatrix} \begin{pmatrix} \underline{x} \\ \underline{a}_1 \end{pmatrix} = \begin{pmatrix} I & B_1 \\ B_1^T & I \end{pmatrix} \begin{pmatrix} \underline{g}_1 \\ \underline{g}_2 \end{pmatrix} .$$

Eq. (5.25) represents two uncoupled systems[1] of equations so that the solution to the original matrix equation (5.20) may be obtained by solving a lower order matrix problem. Thus, we may obtain the solution to (5.20) by first solving by iterative means the system of r equations (square points)

$$(5.26) \qquad (I - B_1 B_1^T)\underline{x} = \underline{g}_1 + B_1 \underline{g}_2 \equiv \underline{e}$$

and then obtaining the reamining (MN-r) unknowns (circle points) explicitly from

$$(5.27) \qquad \underline{a}_1 = B_1^T \underline{x} + \underline{g}_2 \quad .$$

The number of unknowns which we must determine by iterative means has been reduced from MN to roughly $\frac{MN}{2}$.

---

[1] In essence, we have uncoupled the square points from the circle points and vice-versa.

We now consider the directed graphs of $\hat{B}_1$ and $\hat{B}_1^T$, where

$$\hat{B}_1 \equiv \begin{pmatrix} 0 & B_1 \\ 0 & 0 \end{pmatrix} ,$$

in order to show how the elements of the unknown vector $\underline{x}$ are coupled to each other through the matrix $(B_1 B_1^T)$.

If $B_1 \equiv (b_{i,j})$, $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, MN-r$, then the directed graph of $\hat{B}_1$ is obtained as follows: if $b_{i,j} \neq 0$, then we connect the square nodal point of index i to the circle nodal point of index $j + r$ by means of an arrow with the value of $b_{i,j}$ being indicated in some manner. The directed graph of $\hat{B}_1^T$ is obtained, if $b_{i,j} \neq 0$, by connecting the circle nodal point of index $j + r$ to the square nodal point of index i. The directed graphs of $\hat{B}_1$ and $\hat{B}_1^T$ are given in Fig. 5.3 for an arbitrary section of the mesh net. The directed graph of the product $\hat{B}_1 \hat{B}_1^T$ may be obtained very easily from the directed graphs of $\hat{B}_1$ and $\hat{B}_1^T$. The nodal point of index u is connected to the nodal point of index v in the directed graph of $\hat{B}_1 \hat{B}_1^T$ if the nodal point u is connected to the nodal point w in the graph of $\hat{B}_1$ and the nodal point w is connected to the nodal point v in the directed graph of $\hat{B}_1^T$. The directed graph of $\hat{B}_1 \hat{B}_1^T$ also gives us the directed graph of $B_1 B_1^T$ since

$$\hat{B}_1 \hat{B}_1^T = \begin{pmatrix} 0 & 0 \\ 0 & B_1 B_1^T \end{pmatrix}$$

The directed graph of $B_1 B_1^T$ is given in Fig. 5.4 for the same section of the mesh net as in Fig. 5.3. Therefore, for any square nodal point of index C there corresponds an equation, in terms of the elements of the discrete function $\underline{x}$ of the type[1]

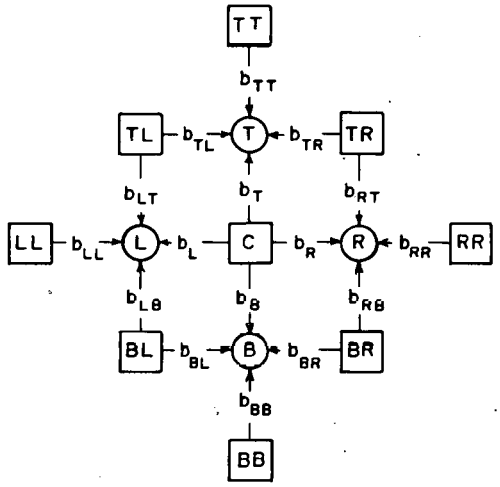
$$(1 - p_C)x_C - p_{RR}x_{RR} - p_{TT}x_{TT} - p_{LL}x_{LL} - p_{BB}x_{BB}$$
$$- p_{TR}x_{TR} - p_{TL}x_{TL} - p_{BL}x_{BL} - p_{BR}x_{BR} = e_C \quad . \tag{5.27}$$

Note that we now have a 9-point formula for the r square nodal points instead of the 5-point formula for the MN nodal points of the original system. Since there are r square nodal points, equation (5.27) defines a system of r equations in r unknowns. We may index the square nodal points and represent the system (5.27) as a matrix equation just as was done for the original system of equations (3.7) or (3.8). The matrix equation (5.26) corresponds to the indexing of the square nodal points consecutively by rows.

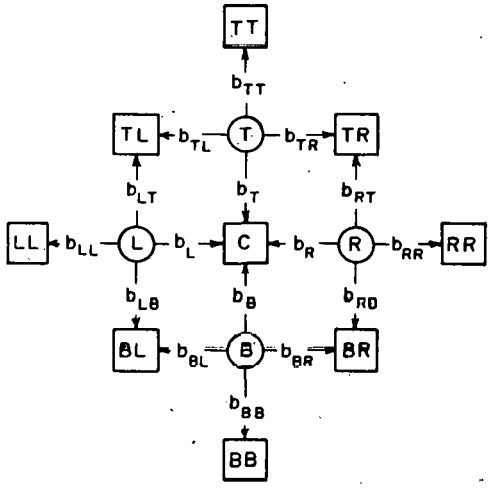
To give the indexing of the square nodal points for the 3-line block method, we proceed in two steps. First, we index consecutively blocks of successive three horizontal mesh lines (see Fig. 5.5) and express the system of equations (5.27) in the block matrix form

---

[1] In Eq. (5.27), the inhomogeneous term e_C is explicitly defined as $e_C = g_C + b_R g_R + b_T g_T + b_L g_L + b_B g_B$. The modifications required when the square nodal point C is near or on a boundary are obvious. For example, if C = 1 in Fig. 5.1, then $p_C = b_R^2 + b_T^2$ and $p_{LL} = p_{BB} = p_{BL} = p_{BR} = 0$.

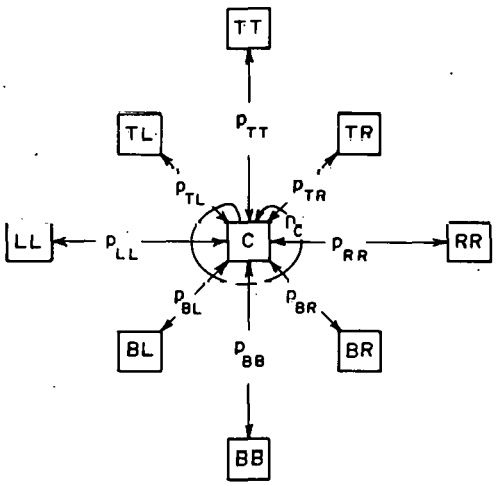


DIRECTED GRAPH OF $\hat{B}_1$



DIRECTED GRAPH OF $\hat{B}_1^T$

FIGURE 5.3



DIRECTED GRAPH OF $B_1 B_1^T$

$$p_C = b_R^2 + b_T^2 + b_L^2 + b_B^2$$

$$p_{TT} = b_T b_{TT}$$

$$p_{LL} = b_L b_{LL}$$

$$p_{BB} = b_B b_{BB}$$

$$p_{RR} = b_R b_{RR}$$

$$p_{TR} = b_R b_{RT} + b_T b_{TR}$$

$$p_{TL} = b_L b_{LT} + b_T b_{TL}$$

$$p_{BL} = b_L b_{LB} + b_B b_{BL}$$

$$p_{RR} = b_B b_{BR} + b_R b_{RB}$$

FIGURE 5.4

$$(5.28)\quad \tilde{A}\underline{x} \equiv \begin{pmatrix} \tilde{A}_{1,1} & \tilde{A}_{1,2} & & & \\ \tilde{A}_{2,1} & \tilde{A}_{2,2} & \tilde{A}_{2,3} & & \\ & \ddots & \ddots & \ddots & \\ & & \tilde{A}_{s,s-1} & \tilde{A}_{s,s} \end{pmatrix} \begin{pmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \vdots \\ \underline{x}_s \end{pmatrix} = \begin{pmatrix} \underline{e}_1 \\ \underline{e}_2 \\ \vdots \\ \underline{e}_s \end{pmatrix}.$$

We here assume that M is divisible by three so that $s = \frac{M}{3}$ is an integer. The diagonal submatrices $\tilde{A}_{i,i}$ correspond to the coupling of square nodal points in

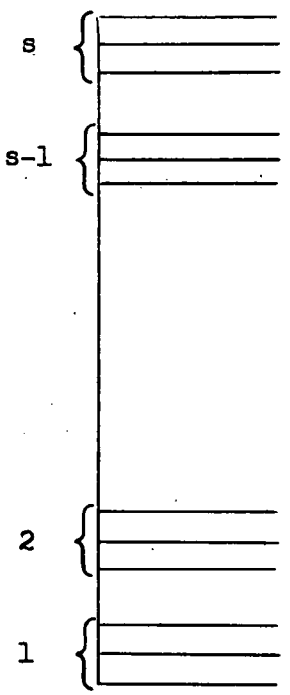


FIGURE 5.5

the block of successive 3-lines with index i. The matrix equation (5.28) will be completely specified if we now index the square nodal points within each successive triplet of horizontal mesh lines as given in Fig. 5.6.

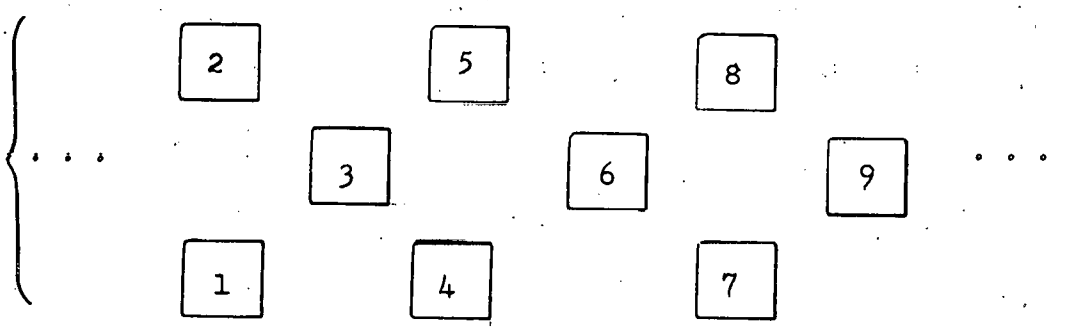


FIGURE 5.6

The matrix $\tilde{A}$ of (5.28) possesses essentially the same properties [see Ref. 11] as does the matrix A_g of Theorem 3.1; i.e., $\tilde{A}$ is an irreducible Stieltjes matrix and $\tilde{A}^{-1} > 0$. Moreover, $\tilde{A}$ is a consistently ordered 2-cyclic matrix. The eigenvalues of the Gauss-Seidel matrix $\mathcal{L}_1$ associated with (5.28) are nonnegative and less than unity [11] so that the results of Theorem 5.2 are valid when the successive overrelaxation method is applied to (5.28). The three line successive overrelaxation method may be carried out for $m \geq 1$ by

$$(5.29)\quad \left\{ \begin{array}{l} \tilde{A}_{i,i}\, \hat{\underline{x}}_i^{(m)} = -\tilde{A}_{i,i-1}\, \underline{x}_{i-1}^{(m)} - \tilde{A}_{i,i+1}\, \underline{x}_{i+1}^{(m-1)} + \underline{e}_i \\ \underline{x}_i^{(m)} = \omega_b \left[\hat{\underline{x}}_i^{(m)} - \underline{x}_i^{(m-1)} \right] + \underline{x}_i^{(m-1)} \end{array} \right\}, \quad 1 \leq i \leq s,$$

where $\omega_b = \dfrac{2}{1 + \sqrt{1 - \rho(\mathcal{L}_1)}}$.

Each $\tilde{A}_{i,i}$ is a principal subminor of a Stieltjes matrix and hence is also a Stieltjes matrix. Thus, the matrix equation $\tilde{A}_{i,i}\, \hat{\underline{x}}_i^{(m)} = \underline{f}_i$, which must be solved in order to carry out (5.29), may be solved directly for $\hat{\underline{x}}_i^{(m)}$ using the square root method[1]. The number of arithmetic operations required for this direct inversion may be reduced [5] by normalizing the matrix equation (5.28) in the following way. We let $\tilde{A}_{i,i} = S_{i,i}^T\, S_{i,i}$ be the factorization, as given in Appendix A, of $\tilde{A}_{i,i}$ and let $R_{i,i}$ be a diagonal matrix whose nonzero elements are the diagonal entries of $S_{i,i}$. If

$$\underline{y}_i^{(m)} = R_{i,i}\, \underline{x}_i^{(m)} \tag{5.30}$$

then (5.28) may be written in the normalized form

$$\begin{pmatrix} N_{1,1} & N_{1,2} & & & \\ N_{2,1} & N_{2,2} & N_{2,3} & & \\ & \ddots & \ddots & \ddots & \\ & & N_{s,s-1} & N_{s,s} \end{pmatrix} \begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \vdots \\ \underline{y}_s \end{pmatrix} = \begin{pmatrix} \underline{d}_1 \\ \underline{d}_2 \\ \vdots \\ \underline{d}_s \end{pmatrix}, \tag{5.31}$$

where $N_{i,j} = R_{i,i}^{-1}\, \tilde{A}_{i,j}\, R_{j,j}^{-1}$ and $\underline{d}_i = R_{i,i}^{-1}\, \underline{e}_i$.

The successive overrelaxation method associated with (5.31) may be carried out for $m \geq 1$ by

---

[1] The details of the square root method are given in Appendix A.

$$(5.32)\quad \left\{ \begin{array}{l} N_{i,i}\,\hat{Y}_i^{(m)} = -N_{i,i-1}\,Y_{i-1}^{(m)} - N_{i,i+1}\,Y_{i+1}^{(m-1)} + \underline{d}_i \\ Y_i^{(m)} = \omega_b\left[\hat{Y}_i^{(m)} - Y_i^{(m-1)}\right] + Y_i^{(m-1)} \end{array} \right\}, \quad 1 \le i \le s .$$

The factorization of $N_{i,i}$ is now $N_{i,i} = \hat{S}_{i,i}^T \hat{S}_{i,i}$, where $\hat{S}_{i,i} = S_{i,i} R_{i,i}^{-1}$, so that $\hat{S}_{i,i}$ has unit diagonal entries. Hence, in the direct inversion of $N_{i,i}$, the division operation in equations (A.5) and (A.6) of Appendix A is eliminated.

The numerical method used to obtain an approximation to ω_b for use in (5.32) is given in Chapter VI.

If $y_{i,j}^{(m)}$ is the j-th component of the vector $\underline{y}_i^{(m)}$, then the sum residual $R^{(m)}$ is defined as

$$(5.33)\quad R^{(m)} = \sum_i \sum_j \left| y_{i,j}^{(m)} - y_{i,j}^{(m-1)} \right| .$$

For an eigenvalue problem, the inner iterations (5.32) in PDQ-5 are terminated when

$$(5.34)\quad R^{(m)} \le \delta R^{(1)} ,$$

where δ is some positive constant.

The main mathematical difference between the various PDQ programs is the iteration method used for the inner iterations. The following table indicates the methods used by the various programs. Also given [11] in this table is the asymptotic rate of convergence of these iterative schemes in the numerical solution of the Dirichlet problem on the unit square with the small uniform mesh spacing $h = \frac{1}{N}$.

PROGRAM	ITERATION METHOD	ASYMPTOTIC RATE OF CONVERGENCE
PDQ-2	POINT SUCCESSIVE OVERRELAXATION	$\simeq 2h$
PDQ-3	1-LINE SUCCESSIVE OVERRELAXATION	$\simeq 2\sqrt{2}\,h$
PDQ-4	2-LINE SUCCESSIVE OVERRELAXATION	$\simeq 4h$
PDQ-5	CYCLICALLY REDUCED 3-LINE SUCCESSIVE OVERRELAXATION	$\geq 5.47h$

TABLE 5.1

In Chapter IV, it was assumed that the matrix vector product $T\underline{\Psi}(\ell-1)$ needed to carry out the outer iterations could be obtained exactly. But this product is only approximated when inner iterations are needed. Hence, the results given in Chapter IV need not be strictly valid when inner iterations are used in the solution of the group equations. The reader is referred to Appendix B for a discussion on the effects of the inner iterations on outer iterations.

VI. ESTIMATION OF THE PARAMETERS $\bar{\sigma}$ AND ω

The use of the successive overrelaxation method (5.32) requires a good estimate for the spectral radius of the Gauss-Seidel iteration matrix. A method given by Varga [21, p. 283] is used to estimate $\rho(\mathcal{L}_1)$ in the PDQ-5 program. This method is based on the following theorem:

Theorem 6.1 Let A be an irreducible consistently ordered 2-cyclic Stieltjes matrix and let $\mathcal{L}_1$ be the associated Gauss-Seidel iteration matrix. If $\underline{Y}^{(0)} > 0$ and if

$$\underline{Y}^{(m)} \equiv \mathcal{L}_1 \underline{Y}^{(m-1)}; \qquad \lambda^{(m)} = \frac{\left[\underline{Y}^{(m)}, \underline{Y}^{(m)}\right]}{\left[\underline{Y}^{(m)}, \underline{Y}^{(m-1)}\right]} \tag{6.1}$$

then $\lim_{m \to \infty} \lambda^{(m)} = \rho(\mathcal{L}_1)$. Moreover, if the i-th component of $\underline{Y}^{(m)}$ is denoted by $y_i^{(m)}$ and if for $y_i^{(m-1)} \neq 0$

$$\bar{\lambda}^{(m)} \equiv \max_i \frac{y_i^{(m)}}{y_i^{(m-1)}} \; ; \quad \underline{\lambda}^{(m)} \equiv \min_i \frac{y_i^{(m)}}{y_i^{(m-1)}} \tag{6.2}$$

then

$$\begin{cases} \bar{\lambda}^{(m)} \geq \rho(\mathcal{L}_1) \geq \underline{\lambda}^{(m)} \; ; \; \bar{\lambda}^{(m)} \geq \lambda^{(m)} \geq \underline{\lambda}^{(m)} \; ; \text{ and} \\ \lim_{m \to \infty} \bar{\lambda}^{(m)} = \lim_{m \to \infty} \underline{\lambda}^{(m)} = \rho(\mathcal{L}_1) \; . \end{cases} \tag{6.3}$$

The iteration procedure (6.1) is used to estimate $\rho(\mathcal{L}_1)$ in the PDQ-5 program. The upper and lower bounds, $\bar{\lambda}^{(m)}$ and $\underline{\lambda}^{(m)}$ defined by (6.2), are used to terminate the iteration procedure in the following way. If

$$\bar{\omega}^{(m)} \equiv \frac{2}{1+\sqrt{1-\bar{\lambda}^{(m)}}} \quad ; \quad \underline{\omega}^{(m)} \equiv \frac{2}{1+\sqrt{1-\underline{\lambda}^{(m)}}} \quad ;$$

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

then the iterations are continued until $m = 15$ or

$$|\bar{\omega}^{(m)} - \underline{\omega}^{(m)}| \leq \frac{(2-\omega^{(m)})}{5} \quad . \tag{6.4}$$

The convergence criterion (6.4) implies that

$$|\omega_b - \omega^{(m)}| \leq \frac{2-\omega^{(m)}}{5} \quad ,$$

where ω_b is the optimum overrelaxation factor defined by Eq. (5.16).

We note that the matrix-vector product $\mathcal{L}_1 \underline{Y}^{(m-1)}$ in (6.1) may be obtained using Eq. (5.32) by setting $\underline{d}_i$ equal to the null vector and by setting ω_b equal to unity. Hence, the iterative procedure used to estimate ω_b takes essentially the same coding as that required for the inner iterations.

A. The Initial Estimate for $\bar{\sigma}$

The efficient use of Chebyshev polynomials in accelerating the convergence of the outer iterations requires an accurate estimate of the dominance ratio $\bar{\sigma}$ and the largest eigenvalue λ_1. As already mentioned in Chapter IV, an initial estimate for λ_1 may be obtained by performing four or five straight power iterations (4.1) before starting the Chebyshev polynomial method of

iteration (4.13). By observing the convergence rate[1] of these power iterations one may obtain an initial estimate for $\bar{\sigma}$ at the same time.

There are, of course, many ways of observing the convergence rate of these power iterations. For example, if the error ratio $ER_{k,n}$ is defined as

$$ER_{k,n} \equiv \frac{\bar{\lambda}(n) - \underline{\lambda}(n)}{\bar{\lambda}(k) - \underline{\lambda}(k)} \tag{6.5}$$

for $k < n$, then for the power method of iteration we have

$$\lim_{\ell \to \infty} (ER_{\ell,\ell+1}) = \bar{\sigma} \tag{6.6}$$

Hence, we may obtain estimates for $\bar{\sigma}$ from $ER_{\ell,\ell+1}$.

To see why (6.6) is true, we first write the vector $\underline{\Psi}(\ell)$ as

$$\underline{\Psi}(\ell) = \frac{T^{\ell}\underline{\Psi}(0)}{\lambda(\ell)\cdot\lambda(\ell-1)\cdots\lambda(1)} \tag{6.7}$$

The above expression for $\underline{\Psi}(\ell)$ follows directly from (4.1). If we now assume[2] that the eigenvectors $\underline{\Psi}_i$ of the matrix T are complete, then we may express the initial guess as $\underline{\Psi}(0) = \sum_{h=1}^{MN} c_h \underline{\Psi}_h$ and Eq. (6.7) as

$$\underline{\Psi}(\ell) = \frac{1}{\lambda(\ell)\cdot\lambda(\ell-1)\cdots\lambda(1)} \sum_{h=1}^{MN} c_h \lambda_h^{\ell} \underline{\Psi}_h \tag{6.8}$$

---

[1] We assume that the convergence rate is determined by the decay of the most slowly decaying component of the error in $\underline{\Psi}(\ell)$. For the power method, as we have seen in Chapter IV, the most slowly decaying component of the error decays as powers of $\bar{\sigma}$.

[2] We make this assumption merely for the sake of simplicity. The limit in (6.6) is valid even if the eigenvectors of T are not complete.

If $\underline{e}_i$ is a vector of order MN whose i-th component is unity and all other components zero and if $\underline{e}_i^T\underline{\psi}_1 \neq 0$, then the i-th component $\psi_i(\ell)$ of the vector $\underline{\psi}(\ell)$ may be written as[1]

$$\psi_i(\ell) = \frac{\lambda_1^{\ell}(c_1 \underline{e}_i^T\underline{\psi}_1)}{\lambda(\ell)\cdot\lambda(\ell-1)\cdots\lambda(1)} \left\{ 1 + a_{2,i}\bar{\sigma}^{\ell} + \sum_{h=3}^{MN} a_{h,i} r_h^{\ell} \right\}, \tag{6.9}$$

where

$$r_h = \frac{\lambda_h}{\lambda_1} < \bar{\sigma} \quad \text{and} \quad a_{h,i} = \frac{c_h \underline{e}_i^T\underline{\psi}_h}{c_1 \underline{e}_i^T\underline{\psi}_1}$$

Hence, if[2] $\psi_i(\ell) \neq 0$, then

$$\frac{S_i(\ell+1)}{\psi_i(\ell)} = \lambda_1 \left\{ \frac{1 + a_{2,i}\bar{\sigma}^{\ell+1} + \sum_{h=3}^{MN} a_{h,i} r_h^{\ell+1}}{1 + a_{2,i}\bar{\sigma}^{\ell} + \sum_{h=3}^{MN} a_{h,i} r_h^{\ell}} \right\} \tag{6.10}$$

As ℓ becomes large we may write Eq. (6.10) as

$$\frac{S_i(\ell+1)}{\psi_i(\ell)} = \lambda_1 \left\{ 1 + a_{2,i}(\bar{\sigma} - 1)\bar{\sigma}^{\ell} + o(\bar{\sigma}^{\ell}) \right\} \tag{6.11}$$

Hence, if $\bar{a}_2 \equiv \max_i a_{2,i}$ and $\underline{a}_2 \equiv \min_i a_{2,i}$, then as ℓ becomes large we have

$$\begin{cases} \bar{\lambda}(\ell+1) = \lambda_1 \left\{ 1 + \bar{a}_2(\bar{\sigma} - 1)\bar{\sigma}^{\ell} + o(\bar{\sigma}^{\ell}) \right\} \\ \underline{\lambda}(\ell+1) = \lambda_1 \left\{ 1 + \underline{a}_2(\bar{\sigma} - 1)\bar{\sigma}^{\ell} + o(\bar{\sigma}^{\ell}) \right\} \end{cases} \tag{6.12}$$

---

[1] Again for the sake of simplicity, we assume that $|\lambda_2| > |\lambda_3|$ in writing Eq. (6.9).

[2] If $\underline{\psi}(0)$ is obtained using Eq. (3.13) from a positive guess for the group fluxes, then $\psi_i(\ell)$ and $\underline{e}_i^T\underline{\psi}_1$ are either both zero or both nonzero.

Therefore, using (6.12) we obtain

$$\lim_{\ell \to \infty} (ER_{\ell,\ell+1}) = \bar{\sigma} .$$

Another way of obtaining initial estimates for $\bar{\sigma}$ is by means of the sum residual (defined by Eq. (5.33)). If $R_g^{(m)}$ is the sum residual for group g at the end of inner iteration m during the ℓ-th outer iteration, then the estimate for $\bar{\sigma}$ at the end of the ℓ-th power iteration is

$$\frac{1}{G} \sum_{g=1}^{G} \frac{R_g^{(1)}(\ell)}{R_g^{(1)}(\ell-1)} \tag{6.13}$$

For a discussion on the use of the sum residuals as a means to estimate the initial $\bar{\sigma}$, the reader is referred to Varga [18].

The PDQ-5 program uses the sum residual method for obtaining the initial estimate for $\bar{\sigma}$. Normally, this initial estimate for $\bar{\sigma}$ is too low. Thus, one would like to be able to obtain updated estimates for $\bar{\sigma}$.

B. Updating the Estimates for $\bar{\sigma}$

If instead of applying a single Chebyshev polynomial of high degree, one would repeatedly apply a low degree Chebyshev polynomial then new estimates for $\bar{\sigma}$ may be obtained by comparing the convergence rate actually being obtained with the theoretical convergence rate one would obtain if the $\bar{\sigma}$ being used were correct.

We use the error ratio ER, defined by (6.5), to measure the convergence rate actually being obtained. In order to see what the error ratio measures, let us apply a Chebyshev polynomial of degree t to the vector estimate $\underline{\psi}(\ell^*)$.

We assume that λ_1 is known and shall denote the present estimate for $\bar{\sigma}$ by σ_0. From Eqs. (4.9) and (4.10), we may write

$$\tilde{\underline{\Psi}}\,(\ell^* + t) = P_{t,\sigma_0}(\frac{T}{\lambda_1})\,\underline{\Psi}(\ell^*) \qquad , \tag{6.14}$$

where

$$P_{t,\sigma_0}(x) = \frac{C_t(\frac{2x}{\sigma_0} - 1)}{C_t(\frac{2}{\sigma_0} - 1)} \quad .$$

If $\underline{\Psi}(\ell^*)$ is expanded in terms of the eigenvectors of T, $\underline{\Psi}(\ell^*) = \sum_{h=1}^{MN} c_h \underline{\psi}_h$, then $\tilde{\underline{\Psi}}\,(\ell^* + t)$ may be written as

$$\tilde{\underline{\Psi}}\,(\ell^* + t) = \sum_{h=1}^{MN} c_h P_{t,\sigma_0}(\frac{\lambda_h}{\lambda_1})\,\underline{\psi}_h \quad .$$

As done previously, we now write the i-th component of the vector $\tilde{\underline{\Psi}}(\ell^* + t)$ as

$$\tilde{\psi}_i(\ell^* + t) = c_1 \underline{e}_i^T \underline{\psi}_1 \left\{ 1 + a_{2,i} P_{t,\sigma_0}(\bar{\sigma}) + \sum_{h=3}^{MN} a_{h,i} P_{t,\sigma_0}(r_h) \right\} , \tag{6.15}$$

where

$$r_h = \frac{\lambda_h}{\lambda_1} < \bar{\sigma} \quad \text{and} \quad a_{h,i} = \frac{c_h \underline{e}_i^T \underline{\psi}_h}{c_1 \underline{e}_i^T \underline{\psi}_1} \quad .$$

If we assume that $\sum_{h=3} a_{h,i} P_{t,\sigma_0}(r_h)$ is small relative to $a_{2,i} P_{t,\sigma_0}(\bar{\sigma})$[1], then (6.15) may be written as

$$\tilde{\psi}_i(\ell^* + t) \simeq c_1 \underline{e}_i^T \underline{\psi}_1 \left\{ 1 + a_{2,i} P_{t,\sigma_0}(\bar{\sigma}) \right\}$$

---

[1] This assumption is generally valid if $\bar{\sigma}$ is pretty well separated from the r_h's and if σ_0 is an underestimate for $\bar{\sigma}$.

and

$$(6.16)\qquad \begin{cases} \bar{\lambda}(\ell^* + t + 1) \simeq \lambda_1 \left\{ 1 + \bar{a}_2(\bar{\sigma} - 1)P_{t,\sigma_0}(\bar{\sigma}) \right\} \\ \underline{\lambda}(\ell^* + t + 1) \simeq \lambda_1 \left\{ 1 + \underline{a}_2(\bar{\sigma} - 1)P_{t,\sigma_0}(\bar{\sigma}) \right\} \end{cases},$$

where

$$\bar{a}_2 = \max_i a_{2,i} \quad \text{and} \quad \underline{a}_2 = \min_i a_{2,i} .$$

Therefore, we have for $0 \leq k < t$

$$(6.17)\qquad ER_{\ell^*+k+1,\ell^*+t+1} \simeq \frac{P_{t,\sigma_0}(\bar{\sigma})}{P_{k,\sigma_0}(\bar{\sigma})}$$

and in particular for $k = 0$

$$(6.18)\qquad ER_{\ell^*+1,\ell^*+t+1} \simeq P_{t,\sigma_0}(\bar{\sigma})$$

Therefore, ER_{ℓ^*+1,ℓ^*+t+1} gives a measure of $P_{t,\sigma_0}(\bar{\sigma})$. For notational purposes, we shall denote ER_{ℓ^*+1,ℓ^*+t+1} just by $(ER)_t$. We now shall examine $(ER)_t$ to see if σ_0 is a good approximation for $\bar{\sigma}$.

Case 1:
$$(ER)_t \leq \frac{1}{C_t(\frac{2}{\sigma_0} - 1)} .$$

From Fig. 6.1, this case implies that $\bar{\sigma} \leq \sigma_0$ so that we are getting the convergence rate expected by using σ_0. Hence, we continue to use σ_0 in the generation of the next Chebyshev polynomial.

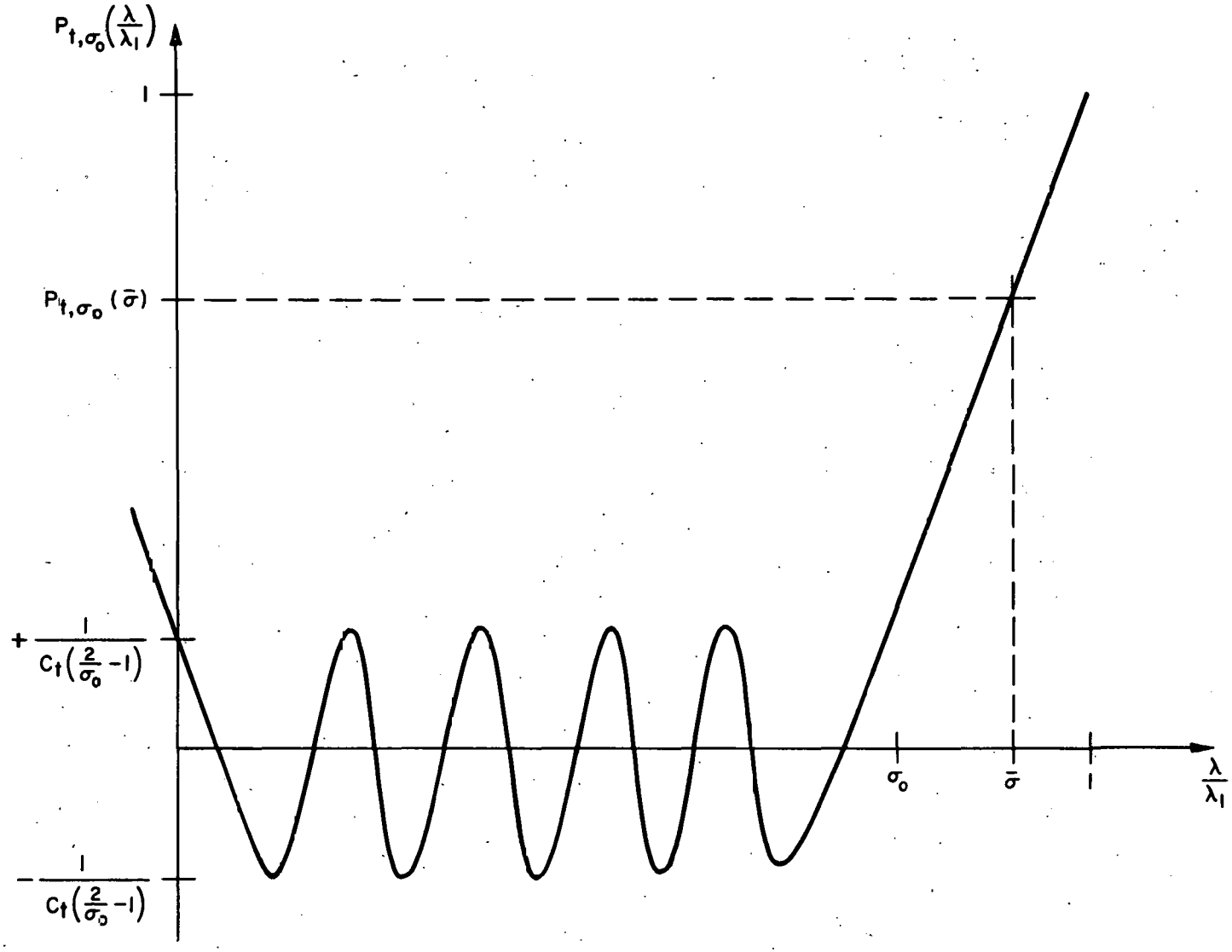

$P_{t,\sigma_0}\left(\frac{\lambda}{\lambda_1}\right)$
1
$P_{t,\sigma_0}(\bar{\sigma})$
$+\frac{1}{C_t\left(\frac{2}{\sigma_0}-1\right)}$
$-\frac{1}{C_t\left(\frac{2}{\sigma_0}-1\right)}$
σ_0
$\bar{\sigma}$
1
$\frac{\lambda}{\lambda_1}$

FIGURE 6.1

Case 2: $$1 > (ER)_t > \frac{1}{C_t(\frac{2}{\sigma_0} - 1)}$$

Again from Fig. 6.1, this case implies that $\bar{\sigma} > \sigma_0$ and we are not obtaining the expected convergence rate. Thus, a new estimate for $\bar{\sigma}$ should be used in the generation of the next Chebyshev polynomial. To obtain this new estimate for $\bar{\sigma}$ we use the result given by (6.18). Using the definition of $P_{t,\sigma_0}(\bar{\sigma})$, we may write (6.18) as

$$(ER)_t \simeq \frac{C_t(\frac{2\bar{\sigma}}{\sigma_0} - 1)}{C_t(\frac{2}{\sigma_0} - 1)}$$

and since $C_t(y) = \cosh(t \cosh^{-1} y)$ for $y \geq 1$ we have

$$\bar{\sigma} \simeq \sigma_0 \left\{ \frac{\cosh\left[\frac{\cosh^{-1}\left\{(ER)_t C_t(\frac{2}{\sigma_0} - 1)\right\}}{t}\right] + 1}{2} \right\} \tag{6.19}$$

The right side of expression (6.19) is then used as the new estimate for $\bar{\sigma}$. If σ_{00} is this new estimate obtained from (6.19), one may easily show that

$$\sigma_0 < \sigma_{00} < 1 \quad .$$

Case 3: $$(ER)_t > 1$$

If $(ER)_t > 1$, then there has been no error reduction and something is obviously wrong.

Another technique for updating the estimates for $\bar{\sigma}$ would be to perform power iterations between the generation of the low degree Chebyshev polynomials. However, in addition to slowing down the convergence rate of the overall problem, it has been found experimentally that the estimates for $\bar{\sigma}$ obtained by this method are generally not as good as those obtained by (6.19).

APPENDIX A

The Square-Root Method for Symmetric Positive Definite Matrices

In this appendix, we shall describe a systematic method to solve directly the matrix equation

$$A\underline{X} = \underline{F} \quad , \tag{A.1}$$

where A is an N x N symmetric positive definite matrix, $\underline{X}$ is the unknown vector and $\underline{F}$ is the source vector.

We have [6, 23].

Theorem A.1 Let $A \equiv (a_{i,j})$ be an N x N symmetric positive definite matrix. There exists a real upper triangular matrix $S \equiv (s_{i,j})$ with positive diagonal entries such that

$$A = S^T S \quad . \tag{A.2}$$

The elements $s_{i,j}$ are given by

$$\begin{cases} s_{1,1} = \sqrt{a_{1,1}} \ , \quad s_{1,j} = \dfrac{a_{1,j}}{s_{1,1}} \\ \\ s_{i,i} = \sqrt{a_{i,i} - \sum_{\ell=1}^{i-1} s_{\ell,i}^2} \ , \ i > 1 \ ; \\ \\ s_{i,j} = \dfrac{a_{i,j} - \sum_{\ell=1}^{i-1} s_{\ell,i}\, s_{\ell,j}}{s_{i,i}} \ , \ j > i \ , \\ \\ s_{i,j} = 0 \ , \ i > j \ . \end{cases} \tag{A.3}$$

Therefore, the solution of matrix equation (A.1) can be reduced to the solution of two triangular systems

$$\text{(A.4)} \qquad \begin{cases} S^T \underline{Y} = \underline{F} \\ S\underline{X} = \underline{Y} \end{cases} .$$

Both systems in (A.4) may be solved directly by

$$\text{(A.5)} \qquad y_1 = \frac{f_1}{s_{1,1}} \; ; \; y_i = \frac{f_i - \sum_{\ell=1}^{i-1} s_{\ell,i}\, y_\ell}{s_{i,i}} \; , \; i > 1 \; .$$

$$\text{(A.6)} \qquad x_N = \frac{y_N}{s_{N,N}} \; ; \; x_i = \frac{y_i - \sum_{\ell=i+1}^{N} s_{i,\ell}\, x_\ell}{s_{i,i}} \; , \; i < N \; .$$

We note that the matrix S is independent of the source vector $\underline{F}$. Hence, if the matrix equation $A\underline{X} = \underline{F}$ must be solved for many different vectors $\underline{F}$, the matrix S need only be computed once.

We now seek to determine the number of multiplications and divisions required by the square-root method to obtain the solution to (A.1) when A is a particular sparse matrix.

Definition A.1 The matrix $A \equiv (a_{i,j})$ is an n-diagonal matrix[1] if n is the smallest nonnegative integer such that $a_{i,j} = 0$ if $|j - i| > n$.

From Eq. (A.3), it follows that if A is an n-diagonal matrix, then S is also an n-diagonal matrix. Hence, from equations (A.5) and (A.6) we have

---

[1] An n-diagonal matrix is sometimes called a band or striped matrix.

Theorem A.2 Let A be a general irreducible n-diagonal Stieltjes matrix. If the S matrix is at hand, then the solution of $A\underline{X} = \underline{F}$ by the square-root method requires an average of 2 n multiplies and 2 divisions per unknown.

In order to carry out the cyclically reduced three-line successive overrelaxation method described in Chapter V, the matrix equation

$$\tilde{A}_{i,i}\,\underset{\sim}{\hat{x}}_i^{(m)} = \underline{f}_i \tag{A.7}$$

is solved directly using the square root method described above. If the square nodal points within each successive triplet of horizontal mesh lines are ordered as in Fig. 5.6, then it is easily seen that the $\tilde{A}_{i,i}$ are 3-diagonal matrices. Hence, from Thm. A.2, the solution of (A.7) by the square root method requires 6 multiplies and 2 divisions per unknown. As described in Chapter V, the division operation may be eliminated by the normalization of Eq. (A.7).

APPENDIX B

The Inner-Outer Iteration Problem

The basic problem in multigroup diffusion problems is to determine the fundamental eigenvalue and corresponding eigenvector for the eigenvalue system

$$\lambda \underline{\Phi} = E^{-1} X F \underline{\Phi} \quad . \tag{B.1}$$

We saw in Chapter III that the problem (B.1) may be restated in terms of the fission source[1] as

$$\lambda \underline{\Psi} = T \underline{\Psi} \quad . \tag{B.2}$$

As described in Chapter IV, the power method (4.1) and the Chebyshev polynomial method (4.13) may be used to obtain by iterative means approximations to the fundamental eigenvalue and eigenvector of (B.2). All the results of Chapter IV are based on the assumption that the matrix-vector product $T\underline{\Psi}$ may be obtained exactly. But in Chapter V we saw that this product is only approximated when inner iterations are required in the solution of the group equations. Hence, the results of Chapter IV are not strictly valid. Two related questions which then confront us are (1) What effect do the inner iterations have on the outer iterations in the solution of the eigen-problem? (2) What accuracy is required in the inner iterations in order to minimize the total computer time required to obtain a satisfactory numerical solution to the eigenvalue problem?

The inter-relations between inner and outer iterations seem to be very complex. We will not even attempt to give direct answers to the above questions.

---

[1] The definitions of the vectors and matrices given in Eqs. (B.1) and (B.2) are as given in Chapter III.

Instead we seek only to present several approaches which hopefully will shed a little light on the answers to these questions. The first approach is simply an error analysis designed to determine how the errors created in PDQ-type problems by not obtaining $T\underline{\Psi}$ exactly are propagated in the outer iterations. The second approach is to obtain a new eigenvalue problem which takes into account the fact that inner iterations are being performed.

A. Error Propagation in the Outer Iterations

In this section we seek to determine how the errors created by not obtaining $T\underline{\Psi}$ exactly are propagated in the outer iterations. For this purpose we need only consider the two-group problem which may be described as

$$\text{(B.3)} \qquad \begin{cases} A_1\underline{\varphi}_1 = \dfrac{X_1}{\lambda}\underline{\Psi} \\ A_2\underline{\varphi}_2 = R_1\underline{\varphi}_1 + \dfrac{X_2}{\lambda}\underline{\Psi} \end{cases} ,$$

where $$\underline{\Psi} = F_1\underline{\varphi}_1 + F_2\underline{\varphi}_2 \quad .$$

Thus, the eigenvalue problem (B.3) can be written in terms of the fission source as

$$\text{(B.4)} \qquad \lambda\underline{\Psi} = T\underline{\Psi} \quad ,$$

where $T = F_1A_1^{-1}X_1 + F_2A_2^{-1}\left[X_2 + R_1A_1^{-1}X_1\right]$.

For notational purposes let

$\underline{\varphi}_1(\ell)$, $\underline{\varphi}_2(\ell)$, $\underline{\Psi}(\ell)$, $\underline{S}(\ell)$ be the values of the group fluxes, fission source, and $T\underline{\Psi}(\ell-1)$ which one would have at the end of the ℓ-th outer iteration if the $T\underline{\Psi}$ could be obtained exactly. (The vector $\underline{S}(\ell)$ is defined in Eq. (4.1).)

$\hat{\underline{\varphi}}_1(\ell)$, $\hat{\underline{\varphi}}_2(\ell)$, $\hat{\underline{\Psi}}(\ell)$, $\hat{\underline{S}}(\ell)$ be the approximate values of the group fluxes, fission source, and $\hat{T}\underline{\Psi}(\ell-1)$ which one would have at the end of the ℓ-th outer iteration when inner iterations are needed to approximate the matrix-vector product $T\underline{\Psi}$.

$\underline{\varepsilon}_1(\ell)$, $\underline{\varepsilon}_2(\ell)$ be the inner iteration error vectors which are defined by $\underline{\varepsilon}_1(\ell) \equiv A_1^{-1}X_1\hat{\underline{\Psi}}(\ell-1) - \hat{\underline{\varphi}}_1(\ell)$
$\underline{\varepsilon}_2(\ell) \equiv A_2^{-1}[R_1\hat{\underline{\varphi}}_1(\ell) + X_2\hat{\underline{\Psi}}(\ell-1)] - \hat{\underline{\varphi}}_2(\ell)$

$\underline{E}(\ell)$ be the outer iteration error vector defined by

$$\underline{E}(\ell) \equiv \frac{\hat{\underline{S}}(\ell) - \hat{T}\hat{\underline{\Psi}}(\ell-1)}{\lambda(\ell)} .$$

Using the above definitions we have

$$\hat{\underline{\varphi}}_1(\ell) = A_1^{-1}X_1\hat{\underline{\Psi}}(\ell-1) - \underline{\varepsilon}_1(\ell)$$

$$\hat{\underline{\varphi}}_2(\ell) = A_2^{-1}[R_1\hat{\underline{\varphi}}_1(\ell) + X_2\hat{\underline{\Psi}}(\ell-1)] - \underline{\varepsilon}_2(\ell)$$

$$\hat{\underline{S}}(\ell) = \hat{T}\hat{\underline{\Psi}}(\ell-1) - F_1\underline{\varepsilon}_1(\ell) - F_2[\underline{\varepsilon}_2(\ell) + A_2^{-1}R_1\underline{\varepsilon}_1(\ell)]$$

and

$$\underline{E}(\ell) = -\left\{ \frac{F_1\underline{\varepsilon}_1(\ell) + F_2\left[\underline{\varepsilon}_2(\ell) + A_2^{-1}R_1\underline{\varepsilon}_1(\ell)\right]}{\lambda(\ell)} \right\} .$$

Now assume that we have a fission source guess $\underline{\Psi}(0)$ and an approximation σ_0 for $\bar{\sigma}$ and we wish to perform the Chebyshev polynomial method (4.13) with inner iterations. If $\lambda(\ell)$ is a good approximation for the largest eigenvalue λ_1 of T, then from (4.13) and above we have

$$\hat{\underline{S}}(1) = T\underline{\Psi}(0) + \lambda_1\underline{E}(1)$$

so that

$$\hat{\underline{\Psi}}(1) = \underline{\Psi}(0) + \alpha_1\left[\frac{\hat{\underline{S}}(1)}{\lambda_1} - \underline{\Psi}(0)\right]$$

$$\hat{\underline{\Psi}}(1) - \underline{\Psi}(1) = \alpha_1\underline{E}(1) .$$

For the second outer iteration we have

$$\hat{\underline{S}}(2) = T\hat{\underline{\Psi}}(1) + \lambda_1\underline{E}(2) = \underline{S}(2) + \alpha_1 T\underline{E}(1) + \lambda_1\underline{E}(2)$$

$$\hat{\underline{\Psi}}(2) = \hat{\underline{\Psi}}(1) + \alpha_2\left[\frac{\hat{\underline{S}}(2)}{\lambda_1} - \hat{\underline{\Psi}}(1)\right] + \beta_2\left[\hat{\underline{\Psi}}(1) - \underline{\Psi}(0)\right]$$

$$\hat{\underline{\Psi}}(2) - \underline{\Psi}(2) = \alpha_1\left[1 - \alpha_2 + \beta_2 + \alpha_2\frac{T}{\lambda_1}\right]\underline{E}(1) + \alpha_2\underline{E}(2) .$$

Now using the fact that

$$(1 - \alpha_\ell + \beta_\ell) = -\frac{\sigma_0}{2}\alpha_\ell$$

we have

$$\hat{\underline{\Psi}}(2) - \underline{\Psi}(2) = \alpha_1\alpha_2 \left[-\frac{\sigma_0}{2} + \frac{T}{\lambda_1} \right] \underline{E}(1) + \alpha_2 \underline{E}(2)$$

which may be written in the form

$$\hat{\underline{\Psi}}(2) - \underline{\Psi}(2) = \frac{4}{\sigma_0} \frac{1}{C_2(\frac{2}{\sigma_0} - 1)} \left[C_1\left(\frac{2T}{\sigma_0\lambda_1}\right) - 1 \right] \underline{E}(1) + \alpha_2 \underline{E}(2) \ .$$

For the third outer iteration we obtain in similar fashion

$$\hat{\underline{\Psi}}(3) - \underline{\Psi}(3) = \frac{4}{\sigma_0} \frac{1}{C_3(\frac{2}{\sigma_0} - 1)} \left[C_2\left(\frac{2T}{\sigma_0\lambda_1} - 1\right) + \frac{1}{2} \right] \underline{E}(1)$$

$$+ \frac{8}{\sigma_0} \frac{C_1(\frac{2}{\sigma_0} - 1)}{C_3(\frac{2}{\sigma_0} - 1)} \left[C_1\left(\frac{2T}{\sigma_0\lambda_1} - 1\right) \right] \underline{E}(2) + \alpha_3 \underline{E}(3) \ ,$$

and in general for the ℓ-th outer iteration we have

$$\text{(B.5)} \qquad \hat{\underline{\Psi}}(\ell) - \underline{\Psi}(\ell) = \frac{4}{\sigma_0} \frac{1}{C_\ell(\frac{2}{\sigma_0} - 1)} \left\{ \gamma_\ell \underline{E}(1) + 2 \sum_{j=2}^{\ell} C_{j-1}(\frac{2}{\sigma_0} - 1) \gamma_{\ell-j+1} \underline{E}(j) \right\} \ ,$$

where

$$\gamma_0 = 0, \ \gamma_1 = \frac{1}{2} \ , \text{ and } \gamma_j = C_{j-1}\left(\frac{2T}{\sigma_0\lambda_1} - 1\right) + \gamma_{j-2} \text{ for } j \geq 2 \ .$$

The expression on the right side of Eq. (B.5) is thus the error introduced in the outer iterations by the inner iterations. But since the $\underline{E}(\ell)$'s are not known, this expression does not appear to be of any practical use to us. However, this expression does allow us to determine how the inner iteration errors are propagated by the Chebyshev polynomials.

If we again assume that the eigenvectors $\underline{\psi}_i$ of T span the associated vector space, then we may express the $\underline{E}(\ell)$ as

$$\underline{E}(\ell) = \sum_{i=1}^{MN} a_{\ell,i}\underline{\psi}_i \quad .$$

Remembering that we are seeking the eigenvector $\underline{\psi}_1$ corresponding to λ_1, we are interested in seeing how the terms $a_{\ell,i}\underline{\psi}_i$ for $i \neq 1$ are propagated in the outer iterations by the Chebyshev polynomials. The following tables show how the term $a_{\ell,2}\underline{\psi}_2$ in the expansion of $\underline{E}(\ell)$ is propagated for various values of σ_0. The symbol $S_{j,\ell}$ represents the coefficient of $a_{j,2}\underline{\psi}_2$ in the expansion of $\underline{E}(j)$ at the end of the ℓ-th outer iteration.

TABLE I: $\sigma_0 = .5$; $\bar{\sigma} = .5$

ℓ	$S_{1,\ell}$	$S_{2,\ell}$	$S_{3,\ell}$	$S_{4,\ell}$	$S_{5,\ell}$	$S_{6,\ell}$	$S_{7,\ell}$	$S_{8,\ell}$	$S_{9,\ell}$	$S_{10,\ell}$
1	1.33									
2	.47	1.411								
3	.12	.485	1.37							
4	.028	.125	.471	1.37						
5	.006	.028	.121	.471	1.37					
6	.001	.006	.028	.121	.471	1.37				
7	.0002	.001	.006	.028	.121	.471	1.37			
8	---	.0002	.001	.006	.028	.121	.471	1.37		
9	---	---	.0002	.001	.006	.028	.121	.471	1.37	
10	---	---	---	.0002	.001	.006	.028	.121	.471	1.37

TABLE II: $\sigma_0 = .8$; $\bar{\sigma} = .8$

ℓ	$S_{1,\ell}$	$S_{2,\ell}$	$S_{3,\ell}$	$S_{4,\ell}$	$S_{5,\ell}$	$S_{6,\ell}$	$S_{7,\ell}$	$S_{8,\ell}$	$S_{9,\ell}$	$S_{10,\ell}$
1	1.66									
2	1.43	2.14								
3	.83	1.67	1.944							
4	.425	.957	1.49	1.915						
5	.203	.489	.853	1.46	1.91					
6	.093	.233	.435	.838	1.46	1.91				
7	.042	.107	.208	.427	.836	1.46	1.91			
8	.018	.048	.095	.204	.426	.836	1.46	1.91		
9	.0078	.021	.042	.094	.203	.426	.836	1.46	1.91	
10	.0033	.009	.019	.041	.093	.203	.426	.836	1.46	1.91

TABLE III: $\sigma_0 = .9$; $\bar{\sigma} = .9$

ℓ	$S_{1,\ell}$	$S_{2,\ell}$	$S_{3,\ell}$	$S_{4,\ell}$	$S_{5,\ell}$	$S_{6,\ell}$	$S_{7,\ell}$	$S_{8,\ell}$	$S_{9,\ell}$	$S_{10,\ell}$
1	1.82									
2	2.24	2.73								
3	1.83	2.99	2.43							
4	1.29	2.36	2.56	2.34						
5	.84	1.64	2.00	2.44	2.32					
6	.524	1.07	1.39	1.91	2.41	2.31				
7	.318	.67	.90	1.32	1.88	2.40	2.31			
8	.189	.40	.56	.86	1.30	1.87	2.40	2.31		
9	.110	.24	.34	.53	.85	1.30	1.87	2.40	2.31	
10	.063	.14	.20	.32	.53	.84	1.30	1.87	2.40	2.31

TABLE IV: $\sigma_0 = .98$; $\bar{\sigma} = .98$

ℓ	$S_{1,\ell}$	$S_{2,\ell}$	$S_{3,\ell}$	$S_{4,\ell}$	$S_{5,\ell}$	$S_{6,\ell}$	$S_{7,\ell}$	$S_{8,\ell}$	$S_{9,\ell}$	$S_{10,\ell}$
1	1.96									
2	3.50	3.64								
3	4.41	6.12	3.43							
4	4.74	7.40	5.53	3.25						
5	4.65	7.74	6.50	5.16	3.20					
6	4.30	7.45	6.68	5.96	4.93	3.15				
7	3.82	6.82	6.37	6.06	5.64	4.80	3.11			
8	3.31	6.04	5.80	5.75	5.71	5.46	4.72	3.10		
9	2.82	5.21	5.11	5.21	5.39	5.50	5.35	4.68	3.08	
10	2.36	4.42	4.41	4.59	4.88	5.18	5.38	5.29	4.65	3.08

TABLE V: $\sigma_0 = .8$; $\bar{\sigma} = .98$

ℓ	$S_{1,\ell}$	$S_{2,\ell}$	$S_{3,\ell}$	$S_{4,\ell}$	$S_{5,\ell}$	$S_{6,\ell}$	$S_{7,\ell}$	$S_{8,\ell}$	$S_{9,\ell}$	$S_{10,\ell}$
1	1.66									
2	2.07	2.14								
3	2.05	2.42	1.944							
4	1.97	2.36	2.16	1.915						
5	1.89	2.27	2.107	2.12	1.91					
6	1.81	2.17	2.02	2.07	2.12	1.91				
7	1.74	2.07	1.93	1.98	2.07	2.12	1.91			
8	1.64	1.98	1.84	1.90	1.98	2.07	2.12	1.91		
9	1.57	1.91	1.76	1.81	1.89	1.98	2.07	2.12	1.91	
10	1.50	1.82	1.72	1.73	1.80	1.89	1.98	2.07	2.12	1.91

Even if inner iterations are not required, the above tables may be used to determine the propagation of round-off errors in the generation of Chebyshev polynomials using the three-term recurrence relation (4.12).

For the straight power method (4.1), one may easily show that the analog of Eq. (B.5) is

$$\hat{\underline{\Psi}}(\ell) - \underline{\Psi}(\ell) = \sum_{j=1}^{\ell} \left(\frac{T}{\lambda_1}\right)^{\ell-j} \underline{E}(j) \quad . \tag{B.6}$$

In the next section we shall take into account the fact that inner iterations are being performed and obtain a new eigenvalue problem which under certain conditions has the same fundamental solution as (B.1).

B. A New Eigenvalue Problem

From Eq. (3.17) it is clear that the straight power method applied to the eigenvalue problem (B.1) may be carried out by solving successively the system of group equations

$$\left\{ A_g \underline{\varphi}_g(\ell) = R_{g-1} \underline{\varphi}_{g-1}(\ell) + \frac{X_2}{\lambda(\ell-1)} \sum_{k=1}^{G} F_k \underline{\varphi}_k(\ell-1) \right\}_{g=1}^{g=G} \quad . \tag{B.7}$$

For two and three dimensional problems, the A_g's generally cannot be directly inverted so that inner iterations must be used to obtain approximations for the $\underline{\varphi}_g(\ell)$'s.

Let A_g be written as

$$A_g = H_g^{-1}(I - P_g) \quad , \quad \text{where } \rho(P_g) < 1 \, , \tag{B.8}$$

so that a stationary iterative method to solve $A_g \underline{\varphi}_g = \underline{S}$ may be written as

(B.9) $$\underline{\varphi}_g^{(m)} = P_g \underline{\varphi}_g^{(m-1)} + H_g \underline{S} \quad ,$$

where m is the inner iteration index. Equation (B.9) may be written in terms of the initial guess $\underline{\varphi}_g^{(0)}$ as

(B.10) $$\underline{\varphi}_g^{(m)} = P_g^m \underline{\varphi}_g^{(0)} + (I + P_g + \cdots + P_g^{m-1}) H_g \underline{S} \quad .$$

For the rest of this section we shall assume that a fixed number m_g of inner iterations are performed in group g for every outer iteration. If we now define $\tilde{A}_g^{-1}$ as

$$\tilde{A}_g^{-1} \equiv (I + P_g + P_g^2 + \cdots + P_g^{m_g-1}) H_g \quad ,$$

then Eq. (B.10) may be expressed as

$$\underline{\varphi}^{(m_g)} = P_g^{m_g} \underline{\varphi}_g^{(0)} + \tilde{A}_g^{-1} \underline{S} \quad .$$

Thus, if we let

$$\tilde{\underline{\varphi}}_g(\ell) \equiv \underline{\varphi}_g^{(m_g)}(\ell) \quad \text{and} \quad \underline{\varphi}_g^{(0)}(\ell) \equiv \tilde{\underline{\varphi}}_g(\ell-1) \quad ,$$

then when inner iterations are performed we actually obtain instead of (B.7) the pseudo system of group equations

(B.11) $$\left\{ \begin{array}{l} \tilde{\underline{\varphi}}_g(\ell) = P_g^{m_g} \tilde{\underline{\varphi}}_g(\ell-1) + \tilde{A}_g^{-1} R_{g-1} \tilde{\underline{\varphi}}_{g-1}(\ell) \\ \qquad\qquad + \tilde{A}_g^{-1} \dfrac{X_2}{\lambda(\ell-1)} \sum_{k=1}^{G} F_k \tilde{\underline{\varphi}}_k(\ell-1) \end{array} \right\}_{g=1}^{g=G} \quad .$$

Since $\rho(P_g) < 1$, we may express A_g^{-1} as

$$A_g^{-1} = (I + P_g + \cdots + P_g^{m_g} + \cdots)H_g$$

so that

$$\tilde{A}_g^{-1} = (I - P_g^{m_g})A_g^{-1} \quad . \tag{B.12}$$

Thus, $\tilde{A}_g^{-1}$ is nonsingular and we may write Eq. (B.11) as

$$\left\{ \tilde{A}_g \underline{\tilde{\varphi}}_g(\ell) - R_{g-1}\underline{\tilde{\varphi}}_{g-1}(\ell) = \tilde{A}_g P_g^{m_g} \underline{\tilde{\varphi}}_g(\ell-1) + \frac{X_g}{\lambda(\ell-1)} \sum_{k=1}^{G} F_k \underline{\tilde{\varphi}}_k(\ell-1) \right\}_{g=1}^{g=G} \tag{B.13}$$

Thus, if we let

$$\tilde{\Phi}(\ell) \equiv \begin{pmatrix} \underline{\tilde{\varphi}}_1(\ell) \\ \underline{\tilde{\varphi}}_2(\ell) \\ \vdots \\ \underline{\tilde{\varphi}}_G(\ell) \end{pmatrix} \quad , \qquad \tilde{E} \equiv \begin{pmatrix} \tilde{A}_1 & & & 0 \\ -R_1 & \tilde{A}_2 & & \\ & \ddots & \ddots & \\ 0 & & -R_{G-1} & \tilde{A}_G \end{pmatrix}$$

$$\tilde{D} \equiv \begin{pmatrix} \tilde{A}_1 & & & 0 \\ & \tilde{A}_2 & & \\ & & \ddots & \\ 0 & & & \tilde{A}_G \end{pmatrix} \quad , \quad \text{and} \quad P \equiv \begin{pmatrix} P_1^{m_1} & & & 0 \\ & P_2^{m_2} & & \\ & & \ddots & \\ 0 & & & P_G^{m_G} \end{pmatrix}$$

then (B.13) may be expressed in the matrix form

$$\tilde{E}\,\underline{\Phi}\,(\ell) = \frac{XF}{\lambda(\ell-1)}\,\underline{\tilde{\Phi}}\,(\ell-1) + \tilde{DP}\,\underline{\tilde{\Phi}}\,(\ell-1)$$

or equivalently

$$\underline{\tilde{\Phi}}\,(\ell) = \tilde{E}^{-1}\left[\frac{XF}{\lambda(\ell-1)} + \tilde{DP}\right]\underline{\tilde{\Phi}}\,(\ell-1) \qquad (B.14)$$

From Eq. (B.12) we have

$$\tilde{A}_g = A_g(I - P_g^{m_g})^{-1}$$

which may be expressed in the form

$$\tilde{A}_g = A_g\left[I + (I - P_g^{m_g})^{-1} P_g^{m_g}\right] = A_g + \tilde{A}_g P_g^{m_g} \quad .$$

Thus,

$$\tilde{E} = E + \tilde{DP}$$

and (B.14) may be written as

$$\underline{\tilde{\Phi}}\,(\ell) = (E + \tilde{DP})^{-1}\left[\frac{XF}{\lambda(\ell-1)} + \tilde{DP}\right]\underline{\tilde{\Phi}}\,(\ell-1)$$

or equivalently

$$\underline{\tilde{\Phi}}\,(\ell) = (I + E^{-1}\,\tilde{DP})^{-1}\left[\frac{E^{-1}XF}{\lambda(\ell-1)} + E^{-1}\,\tilde{DP}\right]\underline{\tilde{\Phi}}\,(\ell-1) \qquad (B.15)$$

If the iterative process (B.15) is convergent, i.e., $\underline{\tilde{\Phi}}(\ell) = \underline{\tilde{\Phi}}(\ell-1) \equiv \underline{\tilde{\Phi}}$ and $\lambda(\ell) = \lambda(\ell-1) \equiv \lambda$, then from (B.15) we have

$$\underline{\tilde{\Phi}} = \frac{E^{-1}XF}{\lambda}\,\underline{\tilde{\Phi}}$$

so that $\left\{\underline{\tilde{\Phi}},\ \lambda\right\}$ is also an eigen-pair of the matrix $E^{-1}XF$. Moreover, from

Thm. 3.2 we know that any positive eigenvector of $E^{-1}XF$ is just a scalar multiple of the fundamental eigenvector $\underline{\Phi}_1$. Thus, if the iterative process (B.15) converges and if $\tilde{\underline{\Phi}}$ is a positive vector, then $\lambda = \lambda_1$ and $\tilde{\underline{\Phi}}$ is a scalar multiple of $\underline{\Phi}_1$. Even though we do not know what conditions are needed on the matrix P to insure the convergence of (B.15), it is of some consolation to know that if the process (B.15) does converge and has a positive solution vector, then this solution is also the desired answer to the eigenvalue problem (B.1).

Assuming that the process (B.15) is convergent and that $\lambda(\ell)$ is a good approximation to λ_1, then when inner iterations are performed the power method is actually being applied to the eigenvalue problem

$$\gamma \tilde{\underline{\Phi}} = (I + E^{-1} \tilde{D}P)^{-1} \left[\frac{E^{-1}XF}{\lambda_1} + E^{-1} \tilde{D}P \right] \tilde{\underline{\Phi}} \tag{B.16}$$

and thus the rate of convergence of (B.15) is determined by the eigenvalues γ of the matrix $(I + E^{-1} \tilde{D}P)^{-1} \left[\frac{E^{-1}XF}{\lambda_1} + E^{-1} \tilde{D}P \right]$. We note that $\gamma = 1$ and $\tilde{\underline{\Phi}} = \underline{\Phi}_1$ is a solution of (B.16).

The eigenvalue problem (B.16) cannot be restated in terms of the fission source as was done in Chapter III for the system (B.1). Thus, when inner iterations are performed, the Chebyshev extrapolation of the fission source vectors as described in Chapter IV are not associated with any well-defined eigenvalue problem. The eigenvalue problem (B.16) is well-defined, however, so might it not be better to apply the Chebyshev polynomial method to this problem?

From Chapter IV, sufficient conditions for the rigorous application of Chebyshev polynomials to (B.16) are that the matrix $(I + E^{-1} \tilde{D}P)^{-1} \left[\frac{E^{-1}XF}{\lambda_1} + E^{-1} \tilde{D}P \right]$ have real eigenvalues and a complete set of

eigenvectors. The range of the eigenvalues is also needed. An experimental and analytical investigation is currently being undertaken to determine how nearly these conditions can be met by the matrix $(I + E^{-1}\tilde{D}P)^{-1}\left[\frac{E^{-1}XF}{\lambda_1} + E^{-1}\tilde{D}P\right]$.

Other approaches to the general inner-outer iteration problem are also being investigated. For example, E. L. Wachspress of the Knolls Atomic Power Laboratory is currently investigating the possibility of using Wielandt's method of fractional iterations to accelerate the outer iterations.

The author would like to express his appreciation to Professor R. S. Varga for his aid in the development of this appendix.

REFERENCES

1. Arms, R. J., Gates, L. D., and Zondek, B. [1956] "A Method of Block Iteration". Journal of the Soc. for Indust. and Appl. Math. Vol. 4, pp. 220-229.

2. Bilodeau, G. and Hageman, L. [1957] "A Survey of Numerical Methods in the Solution of Diffusion Problems". Bettis Atomic Power Laboratory Report WAPD-TM-64.

3. Birkhoff, Garrett and Varga, R. S. [1958] "Reactor Criticality and Nonnegative Matrices". Journal of the Soc. for Indust. and Appl. Math. Vol. 6, pp. 354-377.

4. Cadwell, W. R. [1963] "PDQ-5 -- A FORTRAN Program for the Solution of the Two-Dimensional Neutron-Diffusion Problem -- Part I: Steady-State Version". Bettis Atomic Power Laboratory Report WAPD-TM-363.

5. Cuthill, E. H. and Varga, R. S. [1959] "A Method of Normalized Block Iteration". J. Assoc. Comput. Mach. Vol. 6, pp. 236-244.

6. Faddeeva, V. N. [1959] Computational Methods of Linear Algebra. (Translated from Russian by Curtis D. Benster.) New York: Dover Publications, Inc.

7. Flanders, D. A. and Shortley, G. [1950] "Numerical Determination of Fundamental Modes". J. of Appl. Phys. Vol. 21, pp. 1326-1332.

8. Frankel, S. P. [1950] "Convergence Rates of Iterative Treatments of Partial Differential Equations". Math. Tables Aids Comput. Vol. 4, pp. 65-75.

9. Geiringer, H., "On the Solution of Systems of Linear Equations by Certain Iterative Methods". Reissner Anniversary Volume. Ann Arbor: University of Michigan, pp. 365-393.

10. Habetler, G. J. and Martino, M. A. [1961] "Existence Theorems and Spectral Theory for the Multigroup Diffusion Model". Proceedings of Symposia in Applied Mathematics. Vol. XI, pp. 127-139.

11. Hageman, L. [1962] "Block Iterative Methods for Two-Cyclic Matrix Equations with Special Application to the Numerical Solution of the Second-Order Self-Adjoint Elliptic Partial Differential Equation in Two Dimensions". Bettis Atomic Power Laboratory Report WAPD-TM-327.

12. Kahan, W. [1958] "Gauss-Seidel Methods of Solving Large Systems of Linear Equations". Doctoral Thesis, University of Toronto.

13. Keller, H. B. [1958] "On Some Iterative Methods for Solving Elliptic Difference Equations". Quart. Appl. Math. Vol. 16, pp. 209-226.

14. Marlowe, O. J. and Suggs, M. C. [1960] "WANDA-5 -- A One-Dimensional Neutron Diffusion Equation Program for the Philco-2000 Computer". Bettis Atomic Power Laboratory Report WAPD-TM-241.

15. Marlowe, O. J. [1963] "Bettis FORTRAN Programming: Auxiliary Subroutines". Bettis Atomic Power Laboratory Report WAPD-TM-365.

16. Ostrowski, A. M. [1959] "On the Convergence of the Rayleigh Quotient Iteration for the Computation of the Characteristic Roots and Vectors VI". Archive for Rational Mechanics and Analysis. Vol. 4, pp. 153-165.

17. Varga, R. S. [1957a] "Numerical Solution of the Two-Group Diffusion Equation in x-y Geometry". IRE Trans. of the Professional Group on Nuclear Science. Vol. NS-4, pp. 52-62.

18. Varga, R. S. [1957b] "On Estimating Rates of Convergence in Multigroup Diffusion Problems". Bettis Atomic Power Laboratory Report WAPD-TM-41.

19. Varga, R. S. [1960] "Factorization and Normalized Iterative Methods". Boundary Problems in Differential Equations. Madison: The University of Wisconsin Press. pp. 121-142.

20. Varga, R. S. [1961] "Numerical Methods for Solving Multi-Dimensional Multigroup Diffusion Equations". Proceedings of Symposia in Applied Mathematics. Vol. XI, pp. 164-189.

21. Varga, R. S. [1962] Matrix Iterative Analysis. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

22. Wachspress, E. L. [1957] "CURE: A Generalized Two-Space Dimension Multigroup Coding for the IBM-704". Knolls Atomic Power Laboratory Report KAPL-1724.

23. Wilkinson, J. H. [1961] "Error Analysis of Direct Methods of Matrix Inversion". J. Assoc. Comput. Mach. Vol. 8, pp. 281-330.

24. Young, David [1954] "Iterative Methods for Solving Partial Difference Equations of Elliptic Type". Trans. Amer. Math. Soc. Vol. 76, pp. 92-111.