MIT-3903-2
MITNE-100

FINITE DIFFERENCE TECHNIQUES FOR THE SOLUTION OF THE REACTOR KINETICS EQUATIONS

by

William H. Reed, K. F. Hansen

May 1969

Department of Nuclear Engineering
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Contract AT(30-1)-3903
United States Atomic Energy Commission

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

DEPARTMENT OF NUCLEAR ENGINEERING

Cambridge 39, Massachusetts

FINITE DIFFERENCE TECHNIQUES FOR THE SOLUTION OF THE REACTOR KINETICS EQUATIONS

by

William H. Reed, K. F. Hansen

May 1969

MIT - 3903 - 2

MITNE - 100

AEC Research and Development Report

Contract AT(30-1)3903

U.S. Atomic Energy Commission

FINITE DIFFERENCE TECHNIQUES FOR THE SOLUTION OF THE REACTOR KINETICS EQUATIONS

by

William Hudmon Reed

Submitted to the Department of Nuclear Engineering on May 16, 1969 in partial fulfillment of the requirements for the degree of Doctor of Science.

ABSTRACT

A class of finite difference methods called splitting techniques are presented for the solution of the multigroup diffusion theory reactor kinetics equations in two space dimensions. A subset of the above class is shown to be consistent with the differential equations and numerically stable. An exponential transformation of the semi-discrete equations is shown to reduce the truncation error of the above methods so that they beoome practical methods for two-dimensional problems. A variety of numerical experiments are presented which illustrate the truncation error, convergence rates, and stability of a particular member of the above class.

Thesis Supervisor: Kent F. Hansen
Title: Associate Professor of Nuclear Engineering

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

The author wishes to express his sincere appreciation to Prof. Kent F. Hansen for the support, guidance, and enthusiasm provided during the course of this work.

The author would also like to thank Mr. William T. McCormick, Jr. for his cooperation during joint projects and to thank Mr. McCormick and Dr. John B. Yasinsky for numerical results provided.

This work was performed under a USAEC contract AT(30-1)-3903. The computation was performed on the IBM 360/65 computer at the M. I. T. Information Processing Center.

The author would like to thank his parents and wife for their support throughout this work.

BIOGRAPHICAL NOTE

William Hudmon Reed was born on January 29, 1943, in Florence, Alabama. He received his elementary and secondary education in Sylvania, Georgia, and was graduated from Screven County High School in June, 1961.

In September, 1961, he enrolled at the Georgia Institute of Technology. While an undergraduate, he was a member of Beta Theta Pi social fraternity, and was elected to membership in Tau Beta Pi engineering honorary society. He received his Bachelor of Science degree in physics in June, 1965.

Mr. Reed is married to the former Sheila Marlene Dickie of Wilmington, North Carolina.

Chapter 1

INTRODUCTION

1.1 The Reactor Kinetics Equations

Knowledge of the kinetic behavior of a nuclear reactor subjected to a perturbation from a critical configuration is important for the safe design of that reactor. Most naturally occurring perturbations affecting the state of a reactor, whether from external or internal origin, are localized in space and are not distributed uniformly throughout the reactor. Examples are sodium voiding, burnout, and control rod motions. The transient behavior of the neutron population following such a perturbation will exhibit changes in both the energy spectrum and spatial shape. This behavior is described approximately by the multigroup diffusion theory reactor kinetics equations. These equations are obtained from the more precise mathematical models of transport theory; it is generally held that a solution to the time-dependent transport equations would be prohibitively expensive for practical use in more than one dimension.

This thesis will be concerned with the numerical solution of the linear multigroup diffusion theory reactor kinetics equations. Reactor parameters will be allowed to depend upon time as well as space, which would be the case with externally manipulated control rods. However, no attempt will be made to handle the nonlinear problems arising from the consideration of feedback effects such as temperature coefficients and related thermal effects. It is anticipated, however, that methods developed here will have applicability to the nonlinear problems excluded

above. In some of the future discussions it will be necessary to consider only constant coefficient problems. In no case should such a restriction be taken as a practical one.

The reactor kinetics equations may be written in the following form:

$$\frac{1}{v_g}\frac{d\phi_g}{dt}(\vec{r},t) = \vec{\nabla}\cdot D_g(\vec{r},t)\,\vec{\nabla}\phi_g(\vec{r},t) + \sum_{g'=1}^{G}\Sigma_{gg'}(\vec{r},t)\,\phi_{g'}(\vec{r},t) + \sum_{i=1}^{I} f_{gi}C_i(\vec{r},t) \qquad (1\leq g\leq G)$$

$$\frac{dC_i}{dt}(\vec{r},t) = -\lambda_i C_i(\vec{r},t) + \sum_{g'=1}^{G} p_{ig'}(\vec{r},t)\,\phi_g(\vec{r},t) \qquad (1\leq i\leq I) \qquad (1.1)$$

Parameters appearing in the above equations have the following meanings:

ϕ_g = neutron flux $\left(\frac{\text{neutrons}}{\text{cm}^2\text{-sec}}\right)$ in g'^{th} energy group

v_g = characteristic velocity $\left(\frac{\text{cm}}{\text{sec}}\right)$ of g'^{th} energy group

D_g = diffusion coefficient (cm) in g'^{th} energy group

C_i = concentration $\left(\frac{\text{atoms}}{\text{cm}^3}\right)$ of i'^{th} precursor

λ_i = decay constant (sec^{-1}) of i'^{th} precursor

$\Sigma_{gg'}$ = intergroup transfer cross section (cm^{-1}) $g \neq g'$

Σ_{gg} = negative of group removal cross section in g'^{th} group

f_{gi} = probability (sec^{-1}) that the i'^{th} precursor will produce a neutron in g'^{th} energy group

p_{ig} = cross section (cm^{-1}) for production of i'^{th} precursor by fission in g'^{th} group

G = total number of energy groups

I = total number of precursor groups.

Boundary conditions for the above equations will be homogeneous Neumann or Dirichlet, and an initial condition consisting of the flux and precursor distributions in space and energy must be specified.

Three broad classes of methods have been used to obtain approximate solutions to Eqs. (1. 1).

Modal methods,[1] whereby the unknowns $\phi_g(x, y, z, t)$ are expanded in sum of spatial modes with time-varying coefficients, have been used with a great deal of success, especially in two and three space dimensions. Synthesis techniques,[2] a subset of the above class, are particularly popular. Nodal methods[3] result from the division of the reactor into subregions or "nodes" and the specification of coupling between nodes. To date nodal methods have not been widely used except for rather particular and limited kinetics studies. A discussion of a representative nodal method is given in Ref. 4.

Finite difference techniques have until recently been limited to one space dimension. Well known examples are the GAKIN[5] method and the WIGLE[6] method. Each of these methods has now been extended to two space dimensions, resulting in the LUMAC[7] and TWIGL[8] methods, respectively. The GAKIN and LUMAC methods represent the most advanced finite difference kinetics methods available to date.

It is the subject of this thesis to explore the feasibility of finite difference techniques in higher spatial dimensions and to develop a particular method for two-dimensional kinetics which is "economical" enough for practical computations.

With the definition of a vector $\vec{\theta}$ defined in the following manner:

$$\vec{\theta} = \begin{bmatrix} \phi_1(\vec{r},t) \\ \phi_2(\vec{r},t) \\ \vdots \\ \phi_G(\vec{r},t) \\ C_1(\vec{r},t) \\ \vdots \\ C_I(\vec{r},t) \end{bmatrix} \tag{1.2}$$

Eqs. (1.1) may be written in the shorthand notation

$$\frac{d\vec{\theta}}{dt} = \mathbf{M}\,\vec{\theta} \tag{1.3}$$

where the (G+I) by (G+I) matrix operator $\mathbf{M}$ is given by

$$\mathbf{M} = \left[\begin{array}{cccc|cccc} \nabla\cdot D_1\nabla+\Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1G} & f_{11} & f_{12} & \cdots & f_{1I} \\ \Sigma_{21} & \nabla\cdot D_2\nabla+\Sigma_{22} & \cdots & \Sigma_{2G} & f_{21} & f_{22} & \cdots & f_{2I} \\ & \cdots & & & & \cdots & & \\ \Sigma_{G1} & \Sigma_{G2} & \cdots & \nabla\cdot D_G\nabla+\Sigma_{GG} & f_{G1} & f_{G2} & \cdots & f_{GI} \\ \hline p_{11} & p_{12} & \cdots & p_{1G} & -\lambda_1 & 0 & & 0 \\ p_{21} & p_{22} & \cdots & p_{2G} & 0 & -\lambda_2 & & 0 \\ & \cdots & & & & \cdots & & \\ p_{I1} & p_{I2} & \cdots & p_{IG} & 0 & 0 & & -\lambda_I \end{array}\right] \tag{1.4}$$

In the most general case, excluding feedback, the parameters p_{ig}, $\Sigma_{gg'}$, and D_g appearing in the above matrix may depend on time. The constants f_{gi} and λ_i are always independent of time. The form of the reactor kinetics equation represented by Eq. (1.3) will be convenient for later reference.

In section 1.2 the semi-discrete form of the reactor kinetics equations is developed. These equations may be written in a form similar to Eq. (1.3). They are called "semi-discrete" because the spatial variables have been discretized with the time dependence remaining continuous.

The semi-discrete form of the kinetics equations is the common base on which all presently used finite difference techniques are built. Each of the methods in use today differs from all the others only in its treatment of the time derivative of the semi-discrete equation.

1.2 The Semi-Discrete Equations

The discretization of the spatial variables is considered in this section. In order to maintain some degree of simplicity, the following discussion will be limited to rectangular geometries on a Cartesian coordinate system with the two space variables x and y. It is expected, however, that the methods developed in this thesis will be applicable to more complicated two-dimensional geometries.

In order to reduce the reactor kinetics equations to a two-dimensional problem, there will be assumed to be no variations in the z direction. Consider the single equation

$$\frac{1}{v_g}\frac{d\phi_g}{dt}(x,y,t) = \vec{\nabla}\cdot D_g\vec{\nabla}\phi_g(x,y,t) + \sum_{g'=1}^{G}\Sigma_{gg'}(x,y)\,\phi_{g'}(x,y,t)$$

$$+\sum_{i=1}^{I} f_{gi}C_i(x,y,t) \qquad (1.5)$$

and the grid with constant mesh spacings Δx and Δy

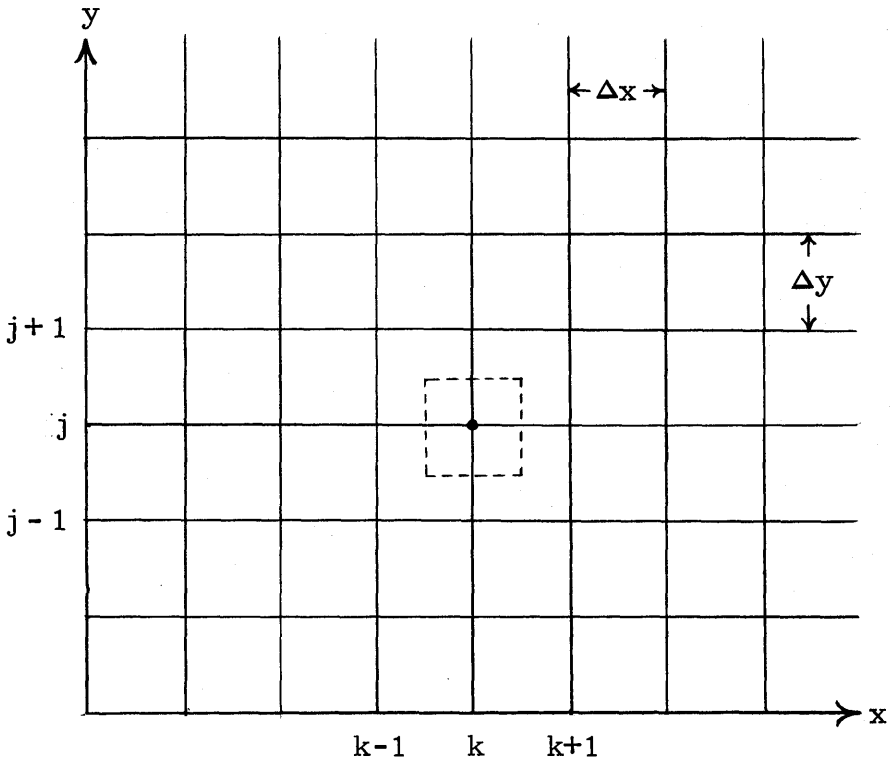


Fig. 1.1. Two-dimensional grid.

Equation (1.5) is integrated over a volume element ($\Delta x\Delta y$) about point (k, j) having unit extent in the z direction. This volume element is illustrated by dashed lines in Fig. 1.1. The equation is then divided

by the volume of the element, giving the following relation:

$$\frac{1}{v_g}\frac{d}{dt}\int\frac{\phi_g(x,y,t)}{\Delta x\Delta y}\,d\vec{r} = \frac{1}{\Delta x\Delta y}\int\vec{\nabla}\cdot D_g(x,y)\,\vec{\nabla}\phi_g(x,y,t)\,d\vec{r}$$

$$+\sum_{g'=1}^{G}\int\frac{\Sigma_{gg'}(x,y)\,\phi_{g'}(x,y,t)}{\Delta x\Delta y}\,d\vec{r}$$

$$+\sum_{i=1}^{I} f_{gi}\quad\frac{C_i(x,y,t)}{\Delta x\Delta y}\,d\vec{r}. \tag{1.6}$$

With the following definitions,

$$\phi_{g,k,j} = \int\frac{\phi_g(x,y)}{\Delta x\Delta y}\,d\vec{r} \tag{1.7a}$$

$$C_{i,k,j} = \int\frac{C_i(x,y)}{\Delta x\Delta y}\,d\vec{r} \tag{1.7b}$$

$$\Sigma_{gg',k,j}\phi_{g',k,j} = \int\frac{\Sigma_{gg'}(x,y)\,\phi_{g'}(x,y,t)}{\Delta x\Delta y}\,d\vec{r}, \tag{1.7c}$$

where all integrals are taken over the volume element $\Delta x\Delta y$ about (k,j); and using Gauss' theorem, we may write

$$\frac{1}{v_g}\frac{d}{dt}\phi_{g,k,j} = \frac{1}{\Delta x\Delta y}\int_S D_g(x,y)\,\vec{\nabla}\phi_g(x,y,t)\cdot d\vec{s}$$

$$+\sum_{g'=1}^{G}\{\Sigma_{gg',k,j}\phi_{g',k,j}\}+\sum_{i=1}^{I} f_{gi}C_{i,k,j} \tag{1.8}$$

where $d\vec{s}$ is an outward normal surface element and S is the surface of the volume element.

The surface integral in the above expression is approximated in the following manner:

$$\frac{1}{\Delta x \Delta y}\int_S D_g(x,y)\,\vec{\nabla}\phi_g(x,y,t)\cdot d\vec{s} \approx \frac{1}{\Delta x \Delta y}\; D_{g,i,j+\frac{1}{2}}\;\frac{\phi_{g,i,j+1}-\phi_{g,i,j}}{\Delta y}\;(\Delta x)$$

$$+ D_{g,i,j-\frac{1}{2}}\;\frac{\phi_{g,i,j}-\phi_{g,i,j-1}}{\Delta y}\;(-\Delta x)$$

$$+ D_{g,i,j+\frac{1}{2}}\;\frac{\phi_{g,i+1,j}-\phi_{g,i,j}}{\Delta x}\;(\Delta y)$$

$$+ D_{g,i-\frac{1}{2},j}\;\frac{\phi_{g,i,j}-\phi_{g,i-1,j}}{\Delta x}\;(-\Delta y), \tag{1.9}$$

where $D_{g,i,j+\frac{1}{2}} = D_g\left(x_i, y_j+\frac{\Delta y}{2}\right)$, etc.

For the case of a spatially constant diffusion coefficient D_g, the above treatment is exactly the five-point central difference approximation to the ∇^2 operator in two dimensions:

$$D_g\nabla^2\phi_g \approx D_g\;\frac{\delta_x^2\phi_g}{(\Delta x)^2}+\frac{\delta_y^2\phi_g}{(\Delta y)^2}$$

where δ^2 is the central difference operator.

An important feature of the above treatment of the spatial derivatives

is illustrated by the case of spatially dependent diffusion coefficients. Continuity of the neutron current $D\vec{\nabla}\phi$ across the surfaces of the volume elements has been preserved by the evaluation of the diffusion coefficients at these surfaces. Thus, if volume elements A and B share a common surface ds, then the integral term approximated by Eq. (1.9) representing neutron current across this surface ds is approximated in the same way for both elements. As a result, no neutrons have been artificially created or lost at these interfaces.

Equation (1.8) with the approximation of Eq. (1.9) may now be written in matrix notation with the definition of the vectors $\vec{\phi}_g$ and $\vec{C}_i$ containing the flux and precursor values $\phi_{g,k,j}$ and $C_{i,k,j}$ at all of the mesh points. The matrix equation is

$$\frac{d\vec{\phi}_g}{dt} = (H_g + V_g)\,\vec{\phi}_g + \sum_{g'} T_{gg'}\vec{\phi}_{g'} + \sum_i F_{gi}\vec{C}_i. \tag{1.10}$$

The matrices H_g and V_g each contain only three nonzero elements per row, and each may be made tridiagonal with the proper ordering of the unknowns $\phi_{g,k,j}$ within the vector $\vec{\phi}_g$. These matrices are given by

$$H_g\phi_{h,k,j} = \frac{v_g}{(\Delta x)^2}\left\{D_{g,k+\frac{1}{2},j}\phi_{g,k+1,j} - \left(D_{g,k+\frac{1}{2},j} + D_{g,k-\frac{1}{2},j}\right)\phi_{g,k,j} + D_{g,k-\frac{1}{2},j}\phi_{g,k-1,j}\right\} \tag{1.11a}$$

$$V_g\phi_{g,k,j} = \frac{v_g}{(\Delta y)^2}\left\{D_{g,k,j+\frac{1}{2}}\phi_{g,k,j+1} - \left(D_{g,k,j+\frac{1}{2}} + D_{g,k,j-\frac{1}{2}}\right)\phi_{g,k,j} + D_{g,k,j-\frac{1}{2}}\phi_{g,k,j-1}\right\}. \tag{1.11b}$$

The matrices $T_{gg'}$ and F_{gi} are diagonal and are given by

$$T_{gg'}\vec{\phi}_{g,k,j} = \Sigma_{gg',k,j}\vec{\phi}_{g,k,j} \tag{1.12}$$

$$F_{gi}C_{i,k,j} = f_{gi}C_{i,k,j}. \tag{1.13}$$

The matrices H_g, V_g, $T_{gg'}$, and F_{gi} are N by N matrices where N is the total number of space points on the grid.

The spatially discrete precursor equations are much simpler to derive and are given by

$$\frac{d\vec{C}_i}{dt} = -\Lambda_i\vec{C}_i + \sum_{g'} P_{ig'}\vec{\phi}_{g'}, \tag{1.14}$$

where Λ_i and $P_{ig'}$ are diagonal matrices given by

$$\Lambda_i C_{i,k,j} = \lambda_i C_{i,k,j} \tag{1.15}$$

$$P_{ig'}\vec{\phi}_{g',k,j} = p_{ig'}\vec{\phi}_{g',k,j} \tag{1.16}$$

With the further definition of a vector $\vec{\psi}$ as

$$\vec{\psi} = \begin{bmatrix} \vec{\phi}_1 \\ \vec{\phi}_2 \\ \vdots \\ \vec{\phi}_G \\ \vec{C}_1 \\ \vdots \\ \vec{C}_I \end{bmatrix} \tag{1.17}$$

the set of Eqs. (1.10) and (1.14) may be combined using the shorthand notation

$$\frac{d\vec{\psi}}{dt} = A\vec{\psi}. \tag{1.18}$$

The above equation is the semi-discrete form of the reactor kinetics equations.

The matrix A is given in terms of previously defined matrices as

$$A = \left[\begin{array}{cccc|ccc} H_1+V_1+T_{11} & T_{12} & \cdots & T_{1G} & F_{11} & \cdots & F_{1I} \\ T_{21} & H_2+V_2+T_{22} & \cdots & T_{2G} & F_{21} & \cdots & F_{2I} \\ & \cdots & & & \cdots & & \\ T_{G1} & T_{G2} & \cdots & T_{GG} & F_{G1} \; F_{G2} & \cdots & F_{GI} \\ \hline P_{11} & P_{12} & \cdots & P_{1G} & -\Lambda_1 & & \\ P_{21} & P_{22} & \cdots & P_{2G} & & -\Lambda_2 & \\ & \cdots & & & \cdots & & \\ P_{I1} & P_{I2} & \cdots & P_{IG} & & & -\Lambda_I \end{array}\right]. \tag{1.19}$$

For later use, the following definitions of the matrices D and E are made.

$$D = \left[\begin{array}{cccc|ccc} H_1+V_1 & 0 & \cdots & 0 & & & \\ 0 & H_2+V_2 & \cdots & 0 & & & \\ & \cdots & & & & 0 & \\ 0 & 0 & \cdots & H_G+V_G & & & \\ \hline & & & & -\Lambda_1 & & \\ & 0 & & & & -\Lambda_2 & \\ & & & & & \ddots & \\ & & & & & & -\Lambda_I \end{array}\right] \tag{1.20}$$

$$E = A - D. \tag{1.21}$$

Under the assumption of homogeneous Dirichlet boundary conditions, the matrix (-D) is a Stieltjes matrix. That is to say, it is symmetric and positive definite. Therefore the matrix D is symmetric and negative definite. This fact will be useful in the stability analysis of Chapter 2.

The above treatment of the $\vec{\nabla} \cdot D\vec{\nabla}$ operator is well known to be consistent and accurate to order $(\Delta x)^2$ and $(\Delta y)^2$.[9] That is to say, if $\vec{\theta}$ is a genuine solution of the differential equation

$$\frac{d\vec{\theta}}{dt} = M\vec{\theta},$$

then $A\vec{\theta} = M\vec{\theta} + 0(\Delta x^2) + 0(\Delta y^2)$. The above fact will be useful for the consistency analysis of fully discrete equations to be considered in Chapter 2.

Certain properties of the matrix A are of interest. It is nonsymmetric and has been observed to have complex eigenvalues in some instances (usually these appear only in problems with many energy groups). All the elements of A are positive or zero, except for diagonal elements which are strictly negative for problems of physical interest. A particular feature of A which should be noted is the wide disparity between the magnitudes of its elements. The parameters λ, for instance, are on the order of unity, while elements of the H and V matrices may be quite large. Typical elements are $v_g D_g / h_x^2$, and it may be pointed out that these elements depend inversely upon the square of the mesh spacing and will be unbounded in the limit as this spacing is decreased to zero.

It is for this reason that the matrices H_g and V_g are called the "principal part" of the matrix A. In Chapter 2 of this thesis it will be shown that it is the principal part of A which determines the all-important numerical property called stability.

For the theoretically important case of a step change in the properties of the reactor, the matrix A is independent of time. The solution operator for Eq. (1.18) is formally seen to be e^{At}, so that

$$\vec{\psi}(t) = e^{At}\vec{\psi}_o. \tag{1.22}$$

The computation of the operator $e^{At} = I + At + \frac{(At)^2}{2} + \ldots$ is so exceedingly difficult, however, that approximate methods for solving Eq. (1.18) are sought even for the case of constant coefficients.

It is possible to divide existing methods for the solution of Eq. (1.18) into three rather arbitrary classes: explicit, implicit, and alternating semi-implicit. The simplest explicit method results from the approximation of e^{At} by the first two terms of a Taylor's series expansion, so that

$$\vec{\psi}^{J+1} = (I+hA)\,\vec{\psi}^{J}, \tag{1.23}$$

where h is a time increment. This method is characterized by an absolute minimum computational effort for each time step, but it suffers a drawback known as numerical instability. In order to achieve a stable method, the time step size h must be extremely small, so that a large number of steps are required in any practical computation.

Fully implicit methods have numerical properties which are just the reverse of explicit methods. They require the inversion of the entire

A matrix at each time step, a necessarily iterative process that is very costly in computation time. An example of a fully implicit method is the TWIGL method with the theta weighting parameter equal to 0.5, i.e.,

$$\vec{\psi}^{J+1} = \left(I - \frac{hA}{2}\right)^{-1} \left(I + \frac{hA}{2}\right) \vec{\psi}^{J}. \tag{1.24}$$

Implicit methods are unconditionally stable, however, and therefore the time step size characteristic of these methods is quite large.

Semi-implicit methods represent a compromise between the explicit and implicit methods which, if properly devised, may have the advantages of both. A defining characteristic of this class of methods is that they require the inversion of only a segment of the matrix A at each time step. This segment is usually chosen so that the inversion process is simple and requires no iteration. Properly formulated, these methods may possess unconditional stability. An example of this type of method is the rather famous "Alternating Direction Implicit" method.[10] Unfortunately for the kinetics equations, these methods possess unacceptably large truncation error, which limits the time step size to the order of that necessary for an explicit method.

A primary result of this thesis is the development of a technique whereby the truncation error of alternating semi-implicit methods may be vastly decreased. This technique is presented in Chapter 2. The consistency and stability of these alternating semi-implicit methods, as well as their asymptotic behavior, are discussed in other sections of Chapter 2.

A particular member of the above class of methods is chosen for detailed examination in Chapter 3. The results of a number of numerical

experiments performed with this method, called MITKIN, are also presented in that chapter. Chapter 4 contains a summary discussion of the properties of the MITKIN method, including a discussion of computer requirements, and Chapter 5 presents recommendations concerning the utilization of the method and future improvements which might be made.

Chapter 2

THEORY

This chapter will be concerned with methods for obtaining an approximate solution to the semi-discrete equation

$$\frac{d\vec{\psi}}{dt} = A\vec{\psi} \tag{2.1}$$

which was developed in section 1.2 of this thesis. Methods to be considered are members of the class of alternating semi-implicit methods and are frequently referred to in this work as "splitting" techniques. Some rather general ideas concerning the consistency and stability of these methods may be formulated.

In section 2.1 an exponential transformation is introduced which has been experimentally observed to reduce significantly the truncation error of some splitting methods. Section 2.2 demonstrates the consistency and stability of a subset of this class of methods when applied to the transformed equation derived in section 2.1. This transformed equation involves a diagonal matrix of free parameters, called frequencies. The selection of these parameters is discussed in section 2.3.

A great deal of matrix algebra is involved in the ensuing sections of this chapter. Specifically, the concepts of vector and matrix norms are crucial to an understanding of the sections on consistency and stability. Reference 11 contains an excellent treatment of the properties of these norms.

2.1 The Exponential Transformation

The reactor kinetics equations are, from a mathematical point of view, "stiff" equations for reactivities less than prompt critical. Stiffness in this connotation will mean that the equations have time constants which span a wide range of values. This is due to the physical fact that the decay constants of the precursors are much smaller than the characteristic neutron diffusion time constants in the higher energy groups. These presursors retard the response of the reactor to a perturbation from a critical state, a well-known fact which facilitates the control of a nuclear reactor. The prediction of this response from numerical solutions of the kinetics equations is not facilitated but is actually made more difficult by the presence of these precursors. The time step size of many stable methods is limited by the neutron equations, but the presence of the precursors greatly lengthens the time interval during which the solution readjusts to a new shape. An unacceptably large number of time steps is necessary to predict this most interesting region of the transient. This is entirely a truncation error problem and has nothing in common with numerical instability. A means by which this difficulty may be surmounted is presented in this section.

Consider the following change of variables. Let

$$\vec{\psi}(t) = e^{\omega t}\,\vec{\phi}(t), \tag{2.2}$$

where ω is a diagonal matrix and the function $\vec{\phi}(t)$ represents what is hopefully a small modulation on the assumed behavior $e^{\omega t}$. It is emphasized that, since ω is a diagonal matrix, the matrix $e^{\omega t}$ is readily

computed; and therefore the transformation between the variables $\vec{\psi}$ and $\vec{\phi}$ is easily made. An equation for the new variable $\vec{\phi}(t)$ may be derived in the following manner. Differentiating Eq. (2.2),

$$\frac{d\vec{\psi}}{dt} = e^{\omega t} \frac{d\vec{\phi}}{dt} + \omega\, e^{\omega t}\, \vec{\phi} \tag{2.3}$$

The substitution of Eqs. (2.2) and (2.3) in Eq. (2.1) gives

$$e^{\omega t} \frac{d\vec{\phi}}{dt} + \omega\, e^{\omega t}\, \vec{\phi} = A\, e^{\omega t}\, \vec{\phi}$$

or

$$\frac{d\vec{\phi}}{dt} = e^{-\omega t}\, (A-\omega)\, e^{\omega t}\, \vec{\phi} \tag{2.4}$$

It is contended that Eq. (2.4) with a proper selection of the free parameters ω is less difficult to solve than Eq. (2.1). By "less difficult to solve" it is intended to mean that, given any method, the truncation error of that method will be less when applied to Eq. (2.4) than when applied to Eq. (2.1). It is not intended to mean that the stability properties of the Eqs. (2.1) are altered by the transformation. Equation (2.4) is still a rather formidable one from the viewpoint of numerical stability.

If the matrix ω is chosen so that

$$\omega\, \vec{\psi}(t_1) = A\, \vec{\psi}(t_1), \tag{2.5}$$

then it is easily seen that

$$\left.\frac{d\vec{\phi}}{dt}\right|_{t_1} = 0. \tag{2.6}$$

Thus in some interval about time t_1 the function $\vec{\phi}$ does indeed represent a small modulation on the assumed behavior. Unfortunately, these time intervals are rather small. Because of the stiffness of the equations, however, the intervals are much longer than the time step size mentioned in the first paragraph of this section. It is for this reason that this transformation increases the time step size which can be taken by many methods for reasonably small truncation error.

There is one difficulty with the above analysis that must be noted. The exact solution $\vec{\psi}(t)$ which appears in Eq. (2.5) is, of course, unavailable, so that the frequencies of Eq. (2.5) can not be determined. Methods for selecting these free parameters are discussed in section 2.3.

2.2 Splitting Methods

A general class of splitting techniques is attractive for the solution of Eq. (2.4) over a time interval 2h. These splitting methods encompass such well known methods as ADI and ADE (alternating direction implicit and explicit, respectively). In order to obtain a general stability criterion for all methods of this type, the following analysis will be completed in terms of arbitrary splittings of the matrices involved. The matrices D and E which were defined by Eqs. (1.20) and (1.21) are split in the following manner:

$$\begin{aligned} D &= \Lambda_1 + \Lambda_2 \\ E &= E_1 + E_2 \\ E &= E_3 + E_4 \end{aligned} \tag{2.7}$$

so that

$$A = \Lambda_1 + \Lambda_2 + E_1 + E_2 \tag{2.8a}$$

$$A = \Lambda_1 + \Lambda_2 + E_3 + E_4 \tag{2.8b}$$

With the evaluation of the matrices $e^{\omega t}$ and $e^{-\omega t}$ at the midpoint of the time interval the following difference equations may be written:

$$\frac{\vec{\phi}(h) - \vec{\phi}(0)}{h} = e^{-\omega h}[\Lambda_1+E_1-\omega]\, e^{\omega h}\vec{\phi}(h) - e^{-\omega h}[\Lambda_2+E_2]\, e^{\omega h}\vec{\phi}(0)$$

$$\frac{\vec{\phi}(2h) - \vec{\phi}(h)}{h} = e^{-\omega h}[\Lambda_2+E_3-\omega]\, e^{\omega h}\vec{\phi}(2h) - e^{-\omega h}[\Lambda_1+E_4]\, e^{\omega h}\vec{\phi}(h). \tag{2.9}$$

The above equations may be solved for $\vec{\phi}(2h)$ in terms of $\vec{\phi}(0)$, which is equal to $\vec{\psi}(0)$. The equations give

$$\vec{\phi}(2h) = e^{-\omega h}[I-h(\Lambda_1+E_1-\omega)]^{-1}[I+h(\Lambda_2+E_2)][I-h(\Lambda_2+E_3-\omega)]^{-1}$$

$$[I+h(\Lambda_1+E_4)]\, e^{\omega h}\vec{\phi}(0). \tag{2.10}$$

But since $\vec{\psi}(2h) = e^{2\omega h}\vec{\phi}(2h)$, we have

$$\vec{\psi}(2h) = B(\omega, h)\,\vec{\psi}(0), \tag{2.11}$$

where the advancement matrix B is given by

$$B(\omega, h) = e^{\omega h}[I-h(\Lambda_1+E_1-\omega)]^{-1}[I+h(\Lambda_2+E_2)][I-h(\Lambda_2+E_3-\omega)]^{-1}$$

$$[I+h(\Lambda_1+E_4)]\, e^{\omega h}. \tag{2.12}$$

If ω and h are held constant with time, Eq. (2.11) becomes

$$\vec{\psi}^{J+2} = B(\omega, h)\, \vec{\psi}^{J}, \tag{2.13}$$

where $\vec{\psi}^{J}$ is a shorthand notation for $\vec{\psi}(Jh)$. Given the initial condition $\vec{\psi}(0) = \vec{\psi}_o$, we may write

$$\vec{\psi}^{2N} = B(\omega, h)^N\, \vec{\psi}_o$$

with the additional assumption that reactor properties do not change with time. This assumption is necessary for the discussions of consistency and stability of the following two sections.

2.2.1 Consistency

It is, of course, hoped that $\vec{\psi}^{2N}$ is some approximation to the exact solution $\vec{\theta}(2Nh)$ of the differential equations. If this is to be true, it is necessary that the difference Eqs. (2.13) be consistent over a single time step. Since $\dfrac{\vec{\psi}^{J+2} - \vec{\psi}^{J}}{2h}$ is to be an approximation to the time derivative, the ratio

$$\frac{B(\omega, h)\, \vec{\theta} - \vec{\theta}}{2h}$$

must be an approximation to $M\vec{\theta}$ in some sense, where M is given by Eq. (1.4). It is not necessary that this be true for all $\vec{\theta}$, since M, for example, is not even defined for discontinuous $\vec{\theta}$. It must be true for $\vec{\theta}$ a genuine solution of the differential equations for some arbitrary initial condition. The consistency condition may be written formally as[12]

$$\left\| \left[\frac{B(\omega, h) - I}{2h} - M \right] \theta(t) \right\| \to 0 \qquad \text{as } h \to 0 \qquad \text{for } 0 \leq t \leq T.$$

Since $\vec{\theta}$ is a genuine solution of $\frac{d\vec{\theta}}{dt} = M\vec{\theta}$, the above condition may be expressed in the following form which is more convenient in most cases.

$$\left\| \frac{\vec{\theta}(t+2h) - B(\omega, h)\,\vec{\theta}(t)}{h} \right\| \to 0 \qquad \text{as } h \to 0 \qquad \text{for } 0 \leq t \leq T.$$

Implicit in these conditions is the requirement that the spatial mesh spacings be reduced as the time step size is decreased. A relation of the form $h = g(\Delta x, \Delta y)$ must be specified to precisely determine the manner in which these increments are taken to zero.

The following theorem will be useful for the consistency analysis of Eq. (2.13).

Theorem 1. If each of the matrices $C_1(h)$ and $C_2(h)$ are consistent, then the matrix $B = C_1C_2$ is also consistent.

Proof: If C_1 and C_2 are consistent, then

$$\left\| \frac{\vec{\theta}(t+h) - C_2\vec{\theta}(t)}{h} \right\| \to 0 \qquad \text{as } h \to 0$$

$$\left\| \frac{\vec{\theta}(t+2h) - C_1\vec{\theta}(t+h)}{h} \right\| \to 0 \qquad \text{as } h \to 0$$

But since C_1 is consistent, its norm is bounded as $h \to 0$, and

$$\|C_1\| \left\| \frac{\vec{\theta}(t+h) - C_2\vec{\theta}(t)}{h} \right\| \to 0 \qquad \text{as } h \to 0.$$

Using the fact that $\|Ax\| \leq \|A\|\,\|x\|$,

$$\left\| \frac{C_1\vec{\theta}(t+h) - C_1C_2\vec{\theta}(t)}{h} \right\| \to 0 \qquad \text{as } h \to 0.$$

The triangle inequality $\|x+y\| \leq \|x\| + \|y\|$ gives

$$\left\| \frac{C_1\vec{\theta}(t+h) - C_1C_2\vec{\theta}(t)}{h} + \frac{\vec{\theta}(t+2h) - C_1\vec{\theta}(t+h)}{h} \right\| \to 0 \qquad \text{as } h \to 0$$

or, finally,

$$\left\| \frac{\vec{\theta}(t+2h) - C_1C_2\vec{\theta}(t)}{h} \right\| \to 0 \qquad \text{as } h \to 0.$$

But the above equation is the consistency condition for the matrix C_1C_2, and the theorem is proved.

The matrix B of Eq. (2.13) is factored into the product of the two matrices C_1 and C_2, so that

$$B = C_1C_2. \tag{2.15}$$

The matrices C_1 and C_2 are defined by

$$C_1 = e^{\omega h}[I-h(\Lambda_1+E_1-\omega)]^{-1}\,[I+h(\Lambda_2+E_2)] \tag{2.16a}$$

$$C_2 = [I-h(\Lambda_2+E_3-\omega)]^{-1}\,[I+h(\Lambda_2+E_4)]\,e^{\omega h}. \tag{2.16b}$$

Each of the matrices C_1 and C_2 will now be shown to be consistent. Consider first the matrix C_2.

$$\left\| \frac{\vec{\theta}(t+h) - C_2\vec{\theta}(t)}{h} \right\| = \left\| \frac{\vec{\theta}(t+h) - [I-h(\Lambda_2+E_3-\omega)]^{-1}\,[I+h(\Lambda_1+E_4)]e^{\omega h}\vec{\theta}(t)}{h} \right\|$$

$\vec{\theta}(t+h)$ is expanded in a Taylor's series about time t:

$$\vec{\theta}(t+h) = \vec{\theta}(t) + h\,\frac{\partial\vec{\theta}}{\partial t} + \frac{h^2}{2}\,\frac{\partial^2\vec{\theta}}{\partial t^2} + \ldots .$$

Substituting,

$$\left\| \frac{\vec{\theta}(t+h) - C_2\vec{\theta}(t)}{h} \right\| = \left\| \vec{\theta}(t) + h\,\frac{\partial\vec{\theta}}{\partial t} + \frac{h^2}{2}\,\frac{\partial^2\vec{\theta}}{\partial t^2} - C_2\vec{\theta}(t) \right\|$$

$$\leq \|[I-h(\Lambda_2+E_3-\omega)]^{-1}\| \left\| [I-h(\Lambda_2+E_3-\omega)] \right.$$

$$\left. \left[\vec{\theta}(t) + h\,\frac{\partial\vec{\theta}}{\partial t} + \frac{h^2}{2}\,\frac{\partial^2\vec{\theta}}{\partial t^2} + \ldots \right] - [I+h(\Lambda_1+E_4)]\,e^{\omega h}\vec{\theta}(t) \right\|.$$

Let β be a bound for $\|[I-h(\Lambda_2+E_3-\omega)]^{-1}\|$. This matrix may easily be shown to have a bounded norm as $h \to 0$, under the restriction that the time step size h and the mesh spacings Δx and Δy are related by expressions of the form

$$h = r_1(\Delta x)^2 = r_2(\Delta y)^2$$

where r_1 and r_2 are constants. This boundedness is demonstrated during the stability analysis of the next section. Therefore,

$$\left\| \frac{\vec{\theta}(t+h) - C_2\vec{\theta}(t)}{h} \right\| \leq \beta \left\| [I-h\Lambda_2-hE_3+\omega h]\left[\vec{\theta}(t)+h\frac{\partial\vec{\theta}}{\partial t}+0(h^2)\right] \right.$$

$$\left. - [I+h\Lambda_1+hE_4+\omega h+0(h^2)]\ \theta(t) \right\|$$

$$\leq \beta \left\| \frac{-hA\,\vec{\theta}(t) + h\frac{\partial\vec{\theta}}{\partial t} + 0(h^2)\ \theta(t)}{h} \right\|$$

$$\left\| \frac{\vec{\theta}(t+h) - C_2\vec{\theta}(t)}{h} \right\| \leq \beta \left\| \frac{\partial\vec{\theta}}{\partial t} - A\,\vec{\theta} + 0(h) \right\|.$$

But since $\vec{\theta}$ is a genuine solution of $\frac{d\vec{\theta}}{dt} = M\,\vec{\theta}$, and since $A\,\vec{\theta} = M\,\vec{\theta} + 0(\Delta x^2) + 0(\Delta y^2)$,

$$\left\| \frac{\vec{\theta}(t+h) - C_2\vec{\theta}(t)}{h} \right\| \leq \beta \left\| 0(h) + 0(\Delta x^2) + 0(\Delta y^2) \right\|.$$

Recalling the requirement that the time increment is to be decreased with the square of the mesh spacings Δx and Δy, we have

$$\left\| \frac{\vec{\theta}(t+h) - C_2\vec{\theta}(t)}{h} \right\| \leq \beta \left\| 0(h) \right\| \to 0 \qquad \text{as } h \to 0.$$

The matrix operator C_2 is therefore consistent with the differential equations. An exactly analogous treatment of C_1 will show that it, too, is consistent, and the consistency of the matrix $B = C_1C_2$ follows from Theorem 1.

2.2.2 Stability

A fundamental theorem due to Lax assures the convergence of the solution of a consistent difference equation to the solution of a well-posed differential equation if and only if the difference equations are stable.[13] If the difference equations are written in the form

$$\psi^{J+1} = B(h)\, \psi^{J}$$

a sufficient condition for numerical stability is that

$$\| B(h)^{N} \| \text{ is bounded for } \quad 0 \leq h \leq \tau$$
$$0 \leq Nh \leq T.$$

This condition is equivalent to the restriction that

$$\| B(h) \| \leq 1 + 0(h).$$

In order to prove the stability of the difference system of Eq. (2.13), it will be necessary to make two assumptions. First, the matrix B must not depend on time; i.e., the consideration is limited to step changes in the reactor properties and to a constant frequency matrix ω. Second, any consideration of stability assumes a relation of the form $h = g(\Delta x, \Delta y)$ between the time step size and the spatial mesh spacings. It will shortly become evident that for the kinetics equations the above relationship must be

$$\frac{h}{(\Delta x)^2} = r_1 \tag{2.17a}$$

$$\frac{h}{(\Delta y)^2} = r_2, \tag{2.17b}$$

where r_1 and r_2 are constants.

For the following stability analysis it will be helpful to establish that the norms of certain matrices are bounded for $0 \leq h \leq \tau$. These matrices are $(I+h\Lambda_1)$, $(I+h\Lambda_2)$, $(I-h\Lambda_1)^{-1}$, and $(I-h\Lambda_2)^{-1}$. This is not a trivial issue, since the elements of the matrices Λ_1 and Λ_2 are inversely proportional to $(\Delta x)^2$ and $(\Delta y)^2$. With the restriction of Eq. (2.17), however, most of the elements of $h\Lambda$ are constant with decreasing h. The exceptions are the precursor decay constants appearing on the diagonal of the lower part of the matrix. For these elements, the multiplication by a small h produces an element of $h\Lambda$ tending to zero in the limit. In the L_∞ or maximum norm, where the norm of a matrix is the maximum of the row sums, the norms of the matrices $(I+h\Lambda_1)$ and $(I+h\Lambda_2)$ are bounded (they are actually constant with h) by observation, independent of the choice of the matrices Λ_1 and Λ_2. This is true even though the dimension of the matrices involved is increasing with decreasing h.

The situation is not so easy for the case of the matrices $(I-h\Lambda_1)^{-1}$ and $(I-h\Lambda_2)^{-1}$, since the inversions generally produce a full matrix. It is now necessary to place a condition on the choice of the splitting of D into Λ_1 and Λ_2 if the above matrices are to possess bounded norms. A simple argument shows that if the matrices $(I-h\Lambda_1)$ and $(I-h\Lambda_2)$ are strictly diagonally dominant the norms of their inverses are bounded. Let

$$I - h\Lambda = Q + R,$$

where Q is a diagonal matrix consisting of the diagonal elements of $I - h\Lambda$. These diagonal elements are all greater than 1. Then

$$\|(I-h\Lambda)^{-1}\| = \|(Q+R)^{-1}\| = \|(I+Q^{-1}R)^{-1}\, Q^{-1}\|$$

$$\leq \|(I+Q^{-1}R)^{-1}\|\, \|Q^{-1}\|.$$

But

$$\|Q^{-1}\| < 1,$$

thus

$$\|(I-h\Lambda)^{-1}\| < \|I - Q^{-1}R + (Q^{-1}R)^2 + \ldots\|$$

$$< 1 + \|Q^{-1}R\| + \|Q^{-1}R\|^2 + \ldots .$$

But since $(I-h\Lambda)^{-1}$ is to be strictly diagonally dominant, it follows that the L_∞ norm of $Q^{-1}R$ is less than one. The above series converges to a finite limit which may serve as a bound for the L_∞ norm of $(I-h\Lambda)^{-1}$.

The L_∞ norms of $(I-h\Lambda_1)^{-1}$ and $(I-h\Lambda_2)^{-1}$ have been shown to be bounded for $0 \leq h \leq \tau$ provided the matrices $(I-h\Lambda_1)$ and $(I-h\Lambda_2)$ are strictly diagonally dominant. This will be the case if D is split so that Λ_1 and Λ_2 are both diagonally dominant. Boundedness for other norms follows from the equivalence theorem, which states that for any matrix A and two matrix norms, $\|A\|_1$ and $\|A\|_2$, there exist positive constants K_1 and K_2 such that[14]

$$K_1\|A\|_1 \leq \|A\|_2 \leq K_2\|A\|_1.$$

The matrix B is factored into the product of two matrices, C_1 and C_2, defined by Eqs. (2.16). It will be shown next that each of these matrices can be separated into two terms, one being the difference approximation of the principal part of A and the other term being of order h.

$$C_1 = e^{\omega h}[I-h(\Lambda_1+E_1-\omega)]^{-1}\,[I+h(\Lambda_2+E_2)]$$

$$C_1 = [I+0(h)][I-h(I-h\Lambda_1)^{-1}(E_1-\omega)]^{-1}\,[I-h\Lambda_1]^{-1}\,[I+h(\Lambda_2+E_2)].$$

But since $\|(I-h\Lambda_1)^{-1}\|$ is bounded for $0 \leq h \leq \tau$, the second term of the right-hand side of the above expression is $[I-0(h)]^{-1}$ which is equal to $I + 0(h)$. Therefore

$$C_1 = [1+0(h)][1+0(h)](I-h\Lambda_1)^{-1}\,[I+h\Lambda_2+0(h)].$$

Finally,

$$C_1 = (I-h\Lambda_1)^{-1}\,(I+h\Lambda_2) + 0(h). \tag{2.18}$$

Similar algebra produces an equivalent result for the matrix C_2:

$$C_2 = (I-h\Lambda_2)^{-1}\,(I+h\Lambda_1) + 0(h). \tag{2.19}$$

The matrix B becomes

$$B = C_1C_2 = (I-h\Lambda_1)^{-1}\,(I+h\Lambda_2)(I-h\Lambda_2)^{-1}\,(I+h\Lambda_1) + 0(h). \tag{2.20}$$

From Eq. (2.20) it may be seen that the advancement matrix B has been written as the sum of its principal part, which will be called B', and terms of order h. The principal part of the matrix B is given by

$$B' = (I-h\Lambda_1)^{-1}\,(I+h\Lambda_2)(I-h\Lambda_2)^{-1}\,(I+h\Lambda_1).$$

In order to complete the proof of stability it is necessary to show that

the matrices $(I+h\Lambda_2)(I-h\Lambda_2)^{-1}$ and $(I+h\Lambda_1)(I-h\Lambda_1)^{-1}$ are norm reducing under some conditions on Λ_1 and Λ_2.

In the L_2 norm,

$$\left\| (I+h\Lambda)(I-h\Lambda)^{-1} \right\| = \max_{\vec{v}} \frac{\vec{v}^T(I-h\Lambda^T)^{-1}\,(I+h\Lambda^T)(I+h\Lambda)(I-h\Lambda)^{-1}\,\vec{v}}{\vec{v}^T\vec{v}}.$$

Let

$$\vec{u} = (I-h\Lambda)^{-1}\,\vec{v}.$$

$$\left\| (I+h\Lambda)\,(I+h\Lambda)^{-1} \right\| = \max_{\vec{u}} \frac{\vec{u}^T(I+h\Lambda^T)(I+h\Lambda)\,\vec{u}}{\vec{u}^T(I-h\Lambda^T)(I-h\Lambda)\,\vec{u}}$$

$$\left\| (I+h\Lambda)(I-h\Lambda)^{-1} \right\| = \max_{\vec{u}} \frac{\vec{u}^T[I+h(\Lambda^T+\Lambda) + h^2\Lambda^T\Lambda]\,\vec{u}}{\vec{u}^T[I-h(\Lambda^T+\Lambda) + h^2\Lambda^T\Lambda]\,\vec{u}}.$$

Since a matrix of the form $\Lambda^T\Lambda$ is always positive definite, Λ having real entries, the above norm is clearly less than one if the matrix $\Lambda^T+\Lambda$ is negative definite. Therefore, if a particular splitting method is chosen so that $\Lambda_1 + \Lambda_1^T$ and $\Lambda_2 + \Lambda_2^T$ are negative definite matrices, then the matrices $(I+h\Lambda_1)(I-h\Lambda_1)^{-1}$ and $(I+h\Lambda_2)(I-h\Lambda_2)^{-1}$ are norm-reducing.

Since the advancement matrix B is given by

$$B = B' + 0(h)$$

the stability of B is evident provided that it can be shown that its principal part B' is bounded less than $1 + 0(h)$. This boundedness will be established in the L_2 norm. Let

$$B' = L\,M\,R,$$

where $L = (I-h\Lambda_1)^{-1}$, $M = (I+h\Lambda_2)(I-h\Lambda_2)^{-1}$, and $R = (I+h\Lambda_1)$. It has been shown that $\|M\| < 1$ and $\|R\| < 1$ if Λ_1 and Λ_2 are chosen properly. But

$$B'^N = LMR\,LMR\,LMR \ldots LMR$$

$$\|B'^N\| \leq \|L\|\,\|R\|.$$

However, $\|L\| = \|(I-h\Lambda_1)^{-1}\|$ and $\|R\| = \|I+h\Lambda_1\|$ have been previously shown to be bounded for $0 \leq h \leq \tau$. Thus $\|B'^N\|$ is bounded for large N. This implies that

$$\|B'\| < 1 + 0(h) \qquad \text{for } 0 < h < \tau$$
$$0 \leq Nh \leq T.$$

Therefore

$$\|B\| < 1 + 0(h) \qquad \text{for } 0 < h < \tau$$
$$0 \leq Nh \leq T,$$

and the difference equations are stable.

It should be emphasized that several assumptions and restrictions were made in the preceding stability proof. The splitting of D must be done in such a way that Λ_1 and Λ_2 are diagonally dominant and $\Lambda_1+\Lambda_1^T$ and $\Lambda_2+\Lambda_2^T$ are negative definite. The ratio of the time step size to the square of the spatial mesh spacings was assumed to be fixed, although no limit was placed on the size of this ratio. It is in this sense that unconditional stability has been achieved.

The final assumption was that the frequency matrix ω is constant independent of time. Generally these frequencies will be allowed to vary with time. The larger question of the stability of the resulting nonlinear

advancement has not yet been answered by means other than numerical experiment. Conclusions concerning the experimental stability of methods in which the frequencies are changed at each time step are presented in section 4. 2.

2. 2. 3 Asymptotic Behavior

A particularly attractive feature of the exponential transformation is that the splitting methods just discussed may be forced to yield the correct asymptotic behavior. If the largest eigenvalue ω_o of A corresponds to the eigenvector $\vec{\mu}_o$, i. e.,

$$A\vec{\mu}_o = \omega_o\vec{\mu}_o, \tag{2. 21}$$

then the asymptotic behavior of the function $\vec{\psi}(t)$ is

$$\vec{\psi}_{asy}(t) = a_o e^{\omega_o t}\vec{\mu}_o.$$

The eigenvector $\vec{\mu}_o$ may be shown to be an eigenvector of the advancement matrix B with eigenvalue $e^{2\omega_o h}$, provided the frequencies are chosen equal to the eigenvalue ω_o of A.

$$B(\omega_o, h)\vec{\mu}_o = e^{\omega_o h}[I-h(\Lambda_1+E_1-\omega_o)]^{-1}[I+h(\Lambda_2+E_2)]$$

$$[I-h(\Lambda_2+E_3-\omega_o]^{-1}[I+h(\Lambda_1+E_4)]\, e^{\omega_o h}\vec{\mu}_o.$$

But

$$(\Lambda_1+E_4)\vec{\mu}_o = (\omega_o-\Lambda_2-E_3)\vec{\mu}_o$$

$$(\Lambda_2+E_2)\vec{\mu}_o = (\omega_o-\Lambda_1-E_1)\vec{\mu}_o.$$

Substituting,

$$B(\omega_o, h)\,\vec{\mu}_o = e^{2\omega_o h}\,[I-h(\Lambda_1+E_1-\omega_o)]^{-1}\,[I+h(\Lambda_2+E_2)]$$

$$[I-h(\Lambda_2+E_3-\omega_o)]^{-1}\,[I-h(\Lambda_2+E_3-\omega_o)]\,\vec{\mu}_o$$

$$B(\omega_o, h)\,\vec{\mu}_o = e^{2\omega_o h}\,[I-h(\Lambda_1+E_1-\omega_o)]^{-1}\,[I-h(\Lambda_1+E_1-\omega_o)]\,\vec{\mu}_o.$$

Finally,

$$B(\omega_o, h)\,\vec{\mu}_o = e^{2\omega_o h}\,\vec{\mu}_o. \tag{2.23}$$

From the above equation it may be seen that the matrix $B(\omega_o, h)$ operating on the asymptotic solution, which is a multiple of $\vec{\mu}_o$, produces the exact growth of $e^{2\omega_o h}$ over the time step 2h.

2.3 Frequency Selection

It has been previously stated that the frequencies ω must be updated frequently in order that the function $\vec{\phi}$ be a small modulation on the assumed behavior $e^{\omega t}$. In the interest of accuracy, it would seem most reasonable to alter the frequencies after each time step, provided that a stable method for the selection of these frequencies may be found.

A cursory examination of the problem might lead one to select "instantaneous" frequencies at each time step, using the approximate solution $\vec{\psi}^J$ as

$$\omega^J \vec{\psi}^J = A\,\vec{\psi}^J. \tag{2.24}$$

The choice of these "instantaneous" frequencies leads to a zero derivative

of $\vec{\phi}(t)$ at the beginning of the time step. Small perturbations in $\vec{\psi}^J$ produce enormous instantaneous frequencies, however, which are in serious error over most of the time step. The above method of frequency selection has been experimentally observed to induce instability in difference systems of the type of Eq. (2.12).

The most desirable frequencies are those accurate over some finite interval of time, i.e., those which produce as small as possible total change in $\vec{\phi}(t)$ over that interval. These frequencies are, of course, theoretically given by the relation

$$e^{\omega^J h}\,\vec{\psi}^J = e^{Ah}\,\vec{\psi}^J. \tag{2.25}$$

As a practical relation for obtaining frequencies Eq. (2.25) is useless. If, however, instead of the exact solution operator e^{Ah} the difference approximation B is used, the following relation is obtained:

$$e^{2\omega h}\,\vec{\psi}^J = B(\omega, h)\,\vec{\psi}^J.$$

The above relation may be put in a much simpler form.

$$e^{2\omega h}\,\vec{\psi}^J = e^{\omega h}[I-h(\Lambda_1+E_1-\omega)]\,[I+h(\Lambda_2+E_2)]$$

$$[I-h(\Lambda_2+E_3-\omega)]^{-1}\,[I+h(\Lambda_1+E_4)]\;e^{\omega h}\,\vec{\psi}^J. \tag{2.27}$$

Assume that ω satisfies the equation

$$(A-\omega)\;e^{\omega h}\,\vec{\psi}^J = 0. \tag{2.28}$$

Then

$$(\Lambda_1+E_4)\ e^{\omega h}\ \vec{\psi}^J = [\omega-(\Lambda_2+E_3)]\ e^{\omega h}\ \vec{\psi}^J.$$

Substituting in Eq. (2.27),

$$e^{2\omega h}\ \vec{\psi}^J = e^{\omega h}[I-h(\Lambda_1+E_1-\omega)]^{-1}\ [I+h(\Lambda_2+E_2)]$$

$$[I-h(\Lambda_2+E_3-\omega)]^{-1}\ [I-h(\Lambda_2+E_3-\omega)]\ e^{\omega h}\ \vec{\psi}^J$$

$$e^{2\omega h}\ \vec{\psi}^J = e^{\omega h}[I-h(\Lambda_1+E_1-\omega)]^{-1}\ [I-h(\Lambda_1+E_1-\omega)]\ e^{\omega h}\ \vec{\psi}^J$$

or

$$e^{2\omega h}\ \vec{\psi}^J = e^{2\omega h}\ \vec{\psi}^J.$$

Therefore the relations (2.27) and (2.28) are equivalent. If the frequencies satisfy one of these relations, then the solution at the next time step may be found from

$$\vec{\psi}^{J+2} = e^{2\omega h}\ \vec{\psi}^J. \tag{2.29}$$

The frequencies determined by Eq. (2.28) represent an instantaneous frequency evaluated from the approximate solution in the middle of the time step. Unfortunately, the solution of Eq. (2.28) must be obtained by use of time-consuming iterative methods.

A third method for selecting the frequencies rests upon the assumption that the change in the frequencies over a time step is small. In this case, the frequencies computed from the relation

$$\omega^J = \frac{1}{2h}\ \ln\left[\frac{\vec{\psi}^J}{\vec{\psi}^{J-2}}\right] \tag{2.30}$$

may be a reasonable approximation to those of Eq. (2.28). Since the above formula is explicit ($\vec{\psi}^J$ is calculated from $\vec{\psi}^{J-2}$ using the frequencies ω^{J-2}), the saving of computation is quite large for each time step, although it is likely that the use of the above frequencies will require a greater number of time steps.

The particular method for selecting frequencies that is finally chosen for use with a splitting technqiue should be the one most effective for that technique.

Chapter 3

RESULTS

Certain theoretical properties of a broad class of finite difference techniques, the splitting methods, were developed within the previous chapter. A particular member of this class of methods is chosen for detailed study in this chapter. The method, called MITKIN, will be defined in section 3.1 and be shown to possess desirable numerical characteristics. In section 3.2 the results of many numerical experiments with the MITKIN method are presented.

3.1 The MITKIN Method

A numerical method may be specified uniquely by the definition of the matrices Λ_1, Λ_2, E_1, E_2, E_3, and E_4 appearing in Eq. (2.12) and by the choosing of a method for the selection of the frequencies ω. A particular method called MITKIN that has been thoroughly studied and for which results are quoted at the end of this chapter results from the following choice. Let Λ_1 be lower triangular containing half the diagonal of D and Λ_2 be upper triangular containing the other half of the diagonal of D. Let the matrix E_1 be lower triangular containing the full diagonal of E and E_2 be strictly upper triangular. Then let $E_3 = E_1$ and $E_4 = E_2$.

The primary advantage of this method is the rapidity with which $\vec{\psi}^{J+2}$ may be computed from $\vec{\psi}^{J}$. The amount of computation necessary to carry out a time step is only slightly greater than that required by a fully explicit method. This easily may be seen by an examination of the

matrices $[I - h(\Lambda_1 + E_1 - \omega)]^{-1}$ and $[I - h(\Lambda_2 + E_1 - \omega)]^{-1}$. The matrix $[I - h(\Lambda_1 + E_1 - \omega)]$ is lower triangular, and the inversion process is a simple back-substitution. The matrix $[I - h(\Lambda_2 + E_1 - \omega)]$ is not triangular; it may be easily put in lower triangular form, however, by a simple reordering of the unknowns. The inversion again becomes a simple back-substitution.

The final realization of the MITKIN method is a double sweep of the spatial mesh. The first sweep begins at one corner and ends up at the opposite corner. The second sweep reverses the order in which the space points are solved. For each sweep of the spatial mesh the group structure is solved starting with neutron group one and ending with the precursor groups. This sweeping of the spatial mesh is precisely the same as that employed by the so-called "alternating direction explicit" method.

The above method easily can be shown to satisfy the stability conditions of the previous chapter. The matrices Λ_1 and Λ_2 are diagonally dominant by observation, and, further,

$$\Lambda_1 + \Lambda_1^T = \Lambda_2 + \Lambda_2^T = D.$$

Since D is negative definite the second stability condition is also satisfied.

The selection of the frequencies at each time step for the MITKIN method was done in accordance with Eq. (2.30), with one exception. Because the truncation error of the precursor equations is small without the use of frequencies, zero frequencies are assumed for these equations. Equation (2.30) fails to specify the frequencies to be used over

the first time step. These frequencies are taken as zero.

The final section of this chapter contains the results of many numerical experiments performed with the MITKIN method.

3.2 Numerical Results

A set of numerical experiments was designed in order to test the stability and the truncation error of the MITKIN method. Although the stability of the method with time-independent frequencies has been conclusively established on theoretical grounds, instability due to the changing of the frequencies at each time step must be admitted as a serious possibility. Such instability has in fact been seen for some splitting methods and will be discussed in section 4.2. Truncation error in this thesis means the percentage discrepancy between the solution of the difference equations and the exact solution of the semi-discrete equations. It will be treated as an experimentally determined quantity. This is done because the emphasis in this work is on taking the maximum possible time step, whereas mathematical concepts of truncation error are valid as the time step size approaches zero.

All of the cases which have been examined represent perturbations from one of five critical configurations. These five configurations are described in detail in Appendix A. Three of the configurations are homogeneous; the other two represent true space-dependent problems. There are both two-group and four-group problems with one precursor group, and a two neutron group problem with six precursor groups.

3.2.1 Homogeneous Problems

Spatially homogeneous problems, frequently referred to as "model" problems, represent a highly valued tool to the numerical analyst, since it is for these problems that exact solutions may be easy to obtain. If the spatial dependence of the initial condition can be expressed as a sum of a small number of the spatial modes, then solutions of both the differential equations and the semi-discrete equations can be obtained for the case of a uniform step change in the reactor properties. The emphasis in this thesis is upon the treatment of the time derivative, and the truncation error produced by the discretization of the spatial variables is not of interest here. Accordingly, the solution of the semi-discrete equations will be used in this section as the standard for comparison.

The following method is used to obtain the exact solution of the semi-discrete equations. If $\vec{\psi}$ is in the fundamental spatial mode, then $\nabla^2\vec{\phi}$ may be replaced by $B^2\vec{\phi}$ where B^2 is the difference buckling[15] given by

$$B^2 = \frac{2}{(\Delta x)^2}\left[1 - \cos\frac{\pi}{K}\right] + \frac{2}{(\Delta y)^2}\left[1 - \cos\frac{\pi}{J}\right],$$

K and J being the number of mesh intervals in the x and y directions. Since the spatial shape is known for all time, the problem is reduced to the solution of the set of equations

$$\frac{d\vec{S}(t)}{dt} = A'\vec{S}(t),$$

where A' is a (G+I) by (G+I) matrix. The above equations are the point kinetics equations where $\vec{S}$ is a vector containing the flux and precursor

concentrations in all groups. The vector $\vec{S}$ may be thought to represent the reactor "spectrum."

The initial condition $\vec{S}_o$ is expanded in a sum of the eigenvectors of A', which must be computed along with their respective eigenvalues.

$$\vec{S}_o = \sum_i a_i \vec{e}_i,$$

where $\vec{e}_i$ is an eigenvector of A' corresponding to the eigenvalue λ_i. The exact solution then becomes

$$\vec{S}(t) = \sum_i a_i e^{\lambda_i t} \vec{e}_i.$$

MITKIN and exact solutions were obtained for four test cases. In each of these test cases the initial condition $\vec{\psi}_o$ is in the fundamental spatial mode, which is a cosine shape for rectangular geometries. The test cases represent two and four group, one and six precursor group problems with positive and negative reactivities. Each test case represents a particular perturbation from one of the critical configurations of Appendix A. The precise perturbation is specified for each test case in the following way. If P_o represents a critical reactor parameter, then

$$P(t) = P_o + \Delta P(t).$$

TEST CASE 1

Critical Configuration: 1

Perturbation: Step change, $\Delta\Sigma_c$ (group 2) = $-.369 \times 10^{-4}$

This test case represents a two neutron group, one precursor group problem with ten mesh intervals in each direction. The perturbation from a critical configuration is made in a step fashion at a time 0.0 seconds by a uniform change in the thermal capture cross section. A series of computer runs were made with different time step sizes in order to illustrate convergence rates both with and without the exponential transformation. For the second case, the frequencies ω were set to zero. The values tabulated below are the thermal flux at the center of the reactor.

Table 3.1. MITKIN results for test case 1.

TIME (sec)	MITKIN				EXACT
	h=.004	h=.002	h=.001	h=.0005	
0.00	.382	.382	.382	.382	.382
0.08	.526	.596	.610	.612	.613
0.16	.773	.815	.816	.816	.816
0.24	1.008	1.004	.999	.997	.997
0.32	1.205	1.170	1.161	1.159	1.158
0.40	1.370	1.318	1.307	1.304	1.303

Table 3.2. MITKIN ($\omega = 0$) results for test case 1.

TIME (sec)	MITKIN ($\omega = 0$)				EXACT
	h=.004	h=.002	h=.001	h=.0005	
0.00	.382	.382	.382	.382	.382
0.08	.420	.447	.482	.520	.613
0.16	.457	.509	.576	.649	.816
0.24	.495	.569	.666	.768	.997
0.32	.529	.627	.752	.880	1.158
0.40	.564	.684	.833	.985	1.303

Table 3.1 demonstrates that the convergence of the MITKIN method to the exact solution is rather fast with decreasing time step size. In section 4.1 this convergence is illustrated to be of order h^2. Additionally, the truncation error for a time step size of .001 is less than half of one per cent, which seems more than adequate for most cases of interest.

By contrast, the results of Table 3.2 are disappointing. The truncation error at a time step size of .001 sec is about 40%, which is completely unacceptable. Perhaps the most distressing result illustrated by this table, however, is that at a time step size of .0005 second the error is still approximately 20%. For this range of time step sizes, the convergence rate is of order h. This slow convergence rate, in addition to the large truncation error, makes this a most undesirable method.

TEST CASE 2

Critical Configuration: 1

Perturbation: Step change, $\Delta\Sigma_c$ (group 2) = $+.231 \times 10^{-4}$

This case is similar to test case 1 but with a negative reactivity insertion. The following results illustrate the feasibility of increasing the time step size as the solution changes less rapidly. Values tabulated are the thermal flux at the center of the reactor.

The results in table 3.3 show that the MITKIN method is capable of producing accurate results for negative reactivity insertions. A second feature of the method is demonstrated by these results. This is the ability to increase the time step size in time zones where the solution is not changing too rapidly.

Table 3.3. Results for test case 2.

TIME	h	MITKIN	EXACT
0.00	—	.3823	.3823
0.04	.001	.3263	.3243
0.08	.001	.2922	.2920
0.12	.001	.2735	.2739
0.16	.001	.2633	.2638
0.20	.001	.2576	.2580
0.60	.01	.2441	.2485
2.20	.01	.2377	.2379

TEST CASE 3

Critical Configuration: 3

Perturbation: Step change, $\Delta\Sigma_c$ (group 2) = $-.369 \times 10^{-4}$

This case is similar to test case 1 but with six delayed neutron groups. The values tabulated are the neutron flux in group 2 at the reactor center.

Table 3.4. Results for test case 3.

TIME	h	MITKIN	EXACT
0.00	.001	.3823	.3823
0.08	.001	.6102	.6136
0.16	.001	.8216	.8220
0.24	.001	1.0160	1.0147
0.32	.001	1.1995	1.1973
0.40	.001	1.3766	1.3738

The above results show that the method is not affected by the additional number of precursor groups.

TEST CASE 4

Critical Configuration: 2

Perturbation: Step change, $\Delta\nu = +.01172$

This is a spatially homogeneous, four neutron group, one precursor group problem with ten mesh intervals in each coordinate direction. The problem is intended to demonstrate the feasibility of treating fast reactor kinetics with the MITKIN method. The perturbation from a critical configuration is made by changing ν in all groups by the above amount. Values tabulated are the flux in group 4 at the reactor center. The flux in groups 1, 2, and 3 shows a similar behavior.

Table 3.5. Results for test case 4.

TIME (sec)	h	MITKIN	EXACT
0.00	.000002	.004475	.004475
0.00016	.000002	.005473	.005481
0.00032	.000002	.006378	.006381
0.00048	.000002	.007148	.007149
0.00064	.000002	.007805	.007804
0.00080	.000002	.008364	.008362
0.00180	.00001	.01048	.01040

The extremely small time step size taken for this problem was necessary because of the very fast initial transient. The flux approximately doubles in .0008 second. Test case 8 will illustrate that such a small time step size is not inherent for a four group problem.

The results of Table 3.5 show excellent agreement of the MITKIN results with the exact answer.

The four test cases which have been presented in this chapter give important data concerning the MITKIN method. The exact truncation error may be seen and convergence rates determined. No evidence of instability appeared.

It is apparent that the ability of the method to accurately predict a spatial transient is not illustrated by these results. Since this is the fundamental purpose of a space-dependent method such as MITKIN, the next section contains results for four problems in which there are spatial effects.

3.2.2 Space-dependent Problems

The four test cases for which results are quoted in this section are multiregion problems with flux shapes deviating significantly from a cosine. More important is the fact that the perturbation from the initial critical configuration is made in only a few of the spatial regions. The resulting transient contains changes in the spatial shape of the flux.

Exact solutions are not available for these problems. Approximate solutions for test cases 5, 6, and 7 are available from the TWIGL code,[16] and similar solutions for the four-group test case 8 are available from the LUMAC code.[17] The proposed method MITKIN is compared with these approximate solutions.

TEST CASE 5

Critical Configuration: 4

Perturbation: Step change, $\Delta\Sigma_c$ (material 1, group 2) = -.0035

This is a two neutron group, one presursor group problem with a spatially dependent step insertion of positive reactivity. A TWIGL

solution is available for comparison. The thermal flux value is quoted at two points, the reactor center (11,11) and at a mesh point (6,6) in the center of a driven region. These points are illustrated in the following figure which shows the problem geometry. For material properties see Appendix A.

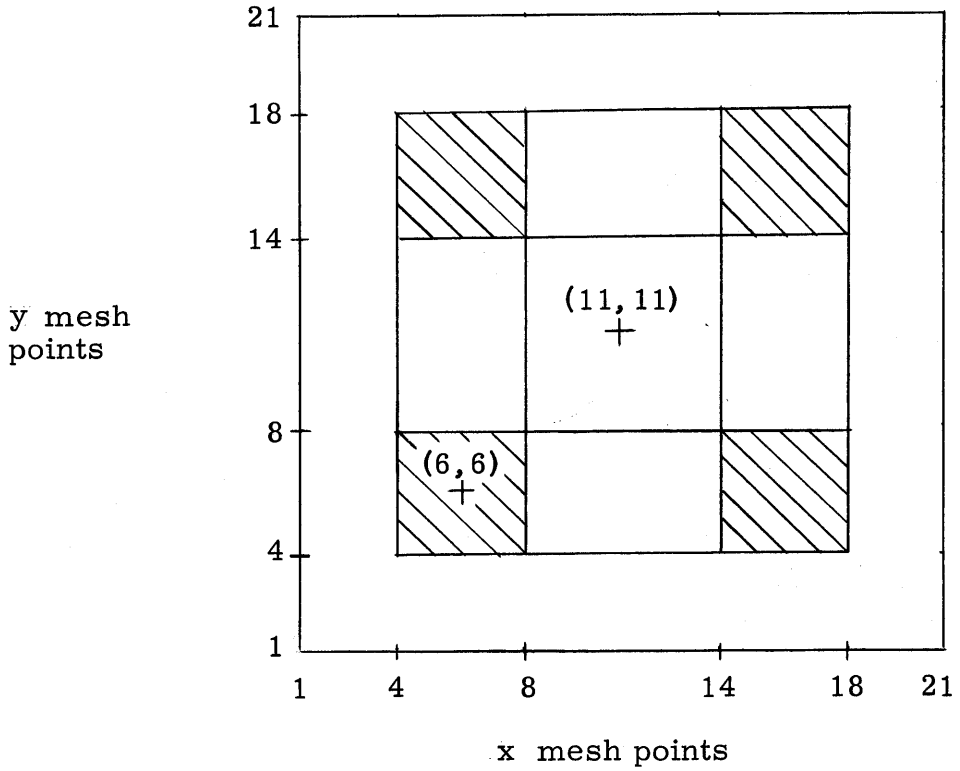


Fig. 3.1. Geometry for test case 5.

The cross-hatched areas of the above figure represent regions in which the perturbation is made.

The results of Table 3.6 show a discrepancy of about 3% between the MITKIN and the TWIGL solutions. This discrepancy, although small, is significant. Runs made with smaller time step sizes indicate that the

Table 3.6. Results for test case 5.

Time	h	Point (11, 11)		Point (6, 6)	
		MITKIN	TWIGL	MITKIN	TWIGL
0.0	.0001	16.75	16.75	4.390	5.390
0.01	.0001	27.69	27.69	9.488	9.273
0.02	.0001	31.50	30.78	10.907	10.658
0.03	.0002	33.13	32.40	11.460	11.209
0.04	.001	33.97	33.15	11.752	11.467
0.05	.001	34.63	33.53	11.976	11.597
0.06	.001	34.99	33.73	12.100	11.666
0.07	.001	35.06	33.85	12.122	11.705

MITKIN results are accurate to better than 1%. The disagreement between the two codes is thought to be due to truncation error in the TWIGL solutions caused by an excessive time step size for the problem.

TEST CASE 6

Critical Configuration: 4

Perturbation: Ramp change, $\Delta\Sigma_c$ (material 1, group 2) = $-.0035(t/0.2)$

for $0 \leq t \leq 0.2$ sec

$\Delta\Sigma_c$ (material 1, group 2) = $-.0035$

for $t > 0.2$ sec

This is a ramp version of test case 5. Values tabulated are the thermal flux.

The results of Table 3.7 show extremely close agreement between the MITKIN and TWIGL methods for this problem. The MITKIN method is observed to be stable for this problem, which represents the first

Table 3.7. Results for test case 6.

Time	h	Point (11,11)		Point (6,6)	
		MITKIN	TWIGL	MITKIN	TWIGL
0.0	.0005	16.75	16.75	5.39	5.39
0.05	.0005	18.79	18.76	6.16	6.15
0.10	.0005	21.75	21.74	7.25	7.25
0.15	.0005	25.95	25.96	8.82	8.82
0.20	.0005	32.31	32.37	11.18	11.20
0.25	.0005	34.12	34.05	11.80	11.77

case in which reactor parameters are allowed to vary with time.

TEST CASE 7

Critical Configuration: 4

Perturbation: Ramp change, $\Delta\Sigma_c$ (material 1, group 2) = +.03(t/.02)

for $0 \leq t \leq .02$ sec

$\Delta\Sigma_c$ (material 1, group 2) = +.03

for $t > 0.02$ sec

This is a negative ramp version of test case 5. Values tabulated are the thermal flux.

Both the MITKIN and TWIGL methods are guaranteed to produce the exact asymptotic shape and spectrum, and this behavior is clearly illustrated by this problem. During the early part of the transient these approximate solutions disagree by about 2%. But at 0.04 second, 0.02 second after the end of the ramp change in the cross section, the solutions agree to better than 0.025%. This is due to the fact that the precursor equations are treated with great accuracy by both methods, and

Table 3.8. Results for test case 7.

Time	h	Point (11, 11)		Point (6, 6)	
		MITKIN	TWIGL	MITKIN	TWIGL
0.0	.0001	16.750	16.750	5.390	5.390
0.004	.0001	13.880	13.700	3.910	3.865
0.008	.0001	9.000	9.949	2.275	2.296
0.012	.0001	6.421	6.506	1.489	1.512
0.016	.0001	5.246	5.288	1.120	1.131
0.020	.0001	4.573	4.594	0.902	0.907
0.040	.0001	4.385	4.385	0.871	0.871

since the asymptotic behavior for negative reactivities is solely determined by the precursors, the methods produce extremely accurate asymptotic results for negative reactivities.

TEST CASE 8

Critical Configuration: 5

Perturbation: Ramp change, $\Delta\Sigma_c$ (material 4, group 4) = $-.003(t/.2)$

for $0 \leq t \leq .2$ sec

$\Delta\Sigma_c$ (material 4, group 4) = $-.003$

for $t > .2$ sec

This is a four neutron group, one precursor group problem having no symmetry along either coordinate. LUMAC approximate solutions are available for comparison with the MITKIN results. The reactor is perturbed from a critical configuration by a ramp change in the thermal cross section of one region. Since this region contains a material having a zero fission cross section, the lower energy groups in this region are coupled to the higher ones (excluding downscattering) only by

diffusion to and from surrounding regions, where thermal neutrons may cause fissions giving rise to fast neutrons. The result is a problem in which large spatial and spectrum shape changes occur over the transient. It is for this reason that this problem is considered the most severe test of the proposed method.

Results are tabulated for groups 1 and 4 at space points (12, 3) and (3, 9) in order to illustrate the magnitude of these shape changes. These two points are shown in the following figure illustrating the problem geometry. See Appendix A for further detail of material properties in the various regions.

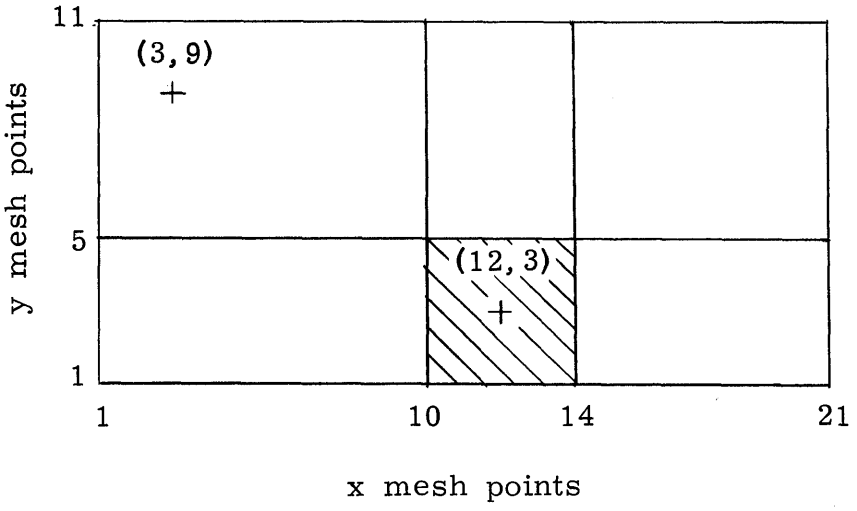


Fig. 3.2. Geometry for test case 8.

The cross-hatched area in the above figure represents the region in which the perturbation is made.

The MITKIN results in the following tables are quoted for two time step sizes.

Table 3.9. Group 1 flux at point (12, 3).

Time (sec)	MITKIN		LUMAC
	h = .001	h = .0005	
0.0	.1314	.1314	.1314
0.05	.1385	.1385	.1385
0.10	.1434	.1433	.1453
0.15	.1485	.1488	.1499
0.20	.1556	.1553	.1551
0.30	.1550	.1553	.1605

Table 3.10. Group 4 flux at point (12, 3).

Time (sec)	MITKIN		LUMAC
	h = .001	h = .0005	
0.00	.968	.968	.968
0.05	1.055	1.054	1.056
0.10	1.155	1.153	1.166
0.15	1.268	1.270	1.278
0.20	1.412	1.408	1.410
0.30	1.407	1.409	1.451

Table 3.11. Group 1 flux at point (3, 9).

Time (sec)	MITKIN		LUMAC
	h = .001	h = .0005	
0.00	.4463	.4463	.4463
0.05	.4567	.4566	.4569
0.10	.4681	.4679	.4730
0.15	.4796	.4807	.4830
0.20	.4964	.4955	.4943
0.30	.4948	.4957	.5123

Table 3.12. Group 4 flux at point (3, 9).

Time (sec)	MITKIN		LUMAC
	h = .001	h = .0005	
0.0	.03594	.03594	.03594
0.05	.03677	.03677	.0368
0.10	.03768	.03767	.0381
0.15	.03861	.03869	.0389
0.20	.03995	.03988	.0398
0.30	.03982	.03989	.0412

Examination of the results of Tables 3.9-3.12 shows that over a period of 0.2 second the flux in groups 1 and 4 at point (3, 9) increases by about 10%, whereas the growth at point (12, 3) is 46% for group 4 and 18% for group 1. The proximity of the MITKIN solutions for the two time step sizes shown implies that the results are quite accurate, and the close agreement with the results of the LUMAC code is strong evidence that a good approximation to the exact solution has been obtained.

Chapter 4

CONCLUSIONS

Certain conclusions may be drawn from the numerical results of Chapter 3. These conclusions fall roughly into three categories: truncation error, stability, and computer requirements. Special properties of the MITKIN method pertinent to each of these categories will be discussed in the following three sections.

4.1 Truncation Error

Test case 1 gives supportive evidence for the contention of the closing paragraph of Chapter 1 of this thesis, where it was stated that the truncation error of semi-implicit methods is large for the reactor kinetics equations. The MITKIN method with $\omega \equiv 0$ is such a method, and the truncation error is indeed large. Several other types of semi-implicit methods were explored in the course of this work, and all appear to have truncation error of the magnitude of that exhibited for MITKIN ($\omega \equiv 0$) in test case 1. In particular, the ADI method suffers equivalently large truncation error.[18]

The exponential transformation appears to reduce the truncation error so drastically that most of these splitting methods become presentable candidates for the solution of the reactor kinetics equations. In order to illustrate this convergence acceleration, the truncation error of the solution at 0.4 second for test case 1 is plotted in Fig. 4.1. The precise definition of this truncation error is

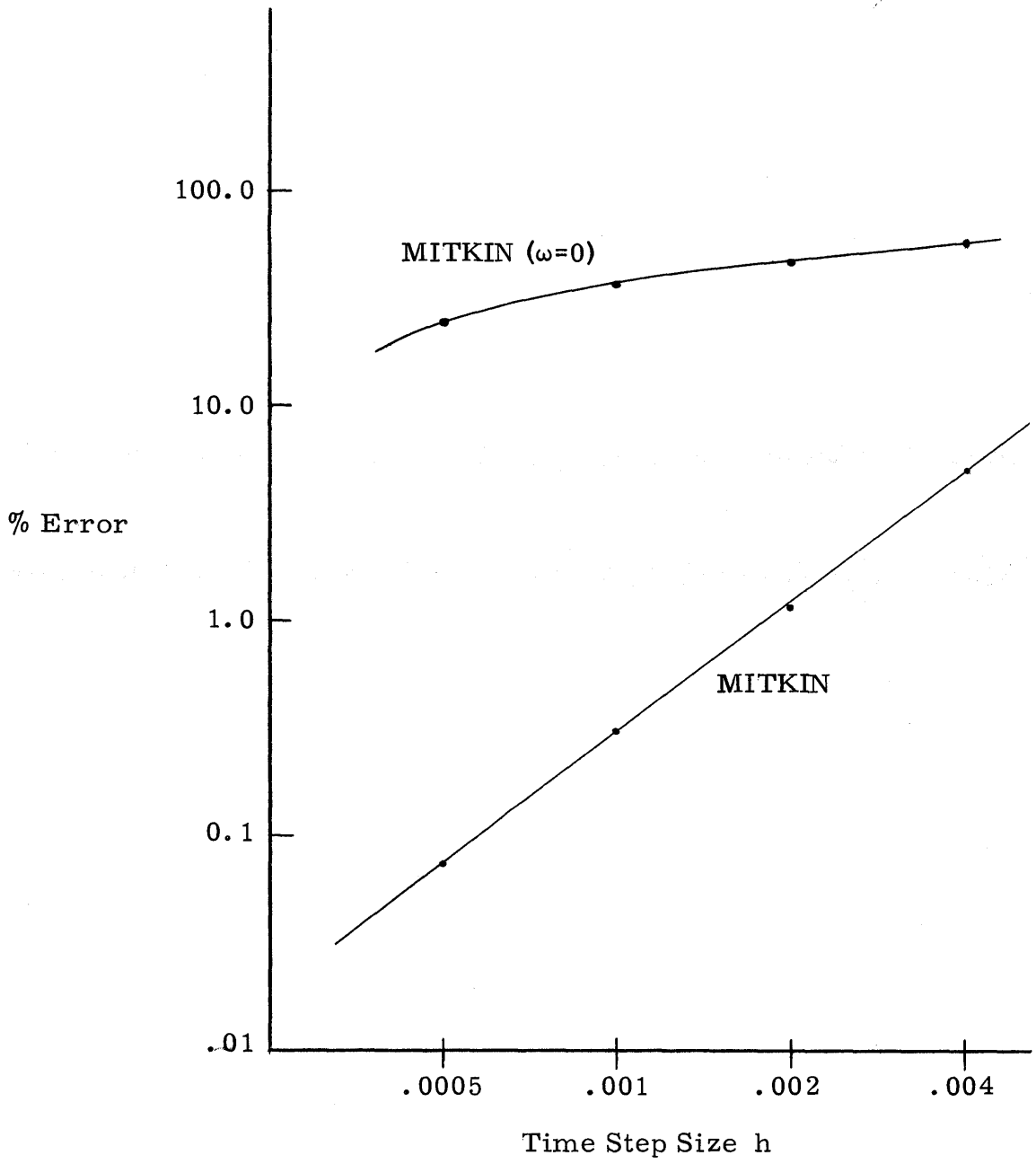


Fig. 4.1. Convergence rates.

$$\% \text{ Error} = 100 \times \left(\frac{\text{MITKIN-EXACT}}{\text{EXACT}}\right).$$

The percentage error for the zero frequency method is easily seen to be several hundred times as large as that for the MITKIN method with frequency correction. An examination of the dependence of the error for the MITKIN method upon the time step size h shows an almost exact relationship of the form Error = ah^2 is valid over this range of time step sizes. The method therefore behaves experimentally as if it were accurate to order h^2. Without the use of the frequencies, the convergence rate is much slower.

A large amount of experience with the MITKIN method gives a convenient rule of thumb relating the truncation error to the percentage change in the solution over one time step. A 1% change in the solution over each time step produces about 1% truncation error after 100 steps. This result is very significant. It implies that, regardless of the nature of the problem, the important time interval during which the flux doubles can be solved in 100 steps to an accuracy of about 1%.

4.2 Stability

The time step size for the MITKIN method does not appear to be limited by stability problems, but instead is limited by the accuracy desired by the user. If the time step size is increased, the computed solution will depart farther from the exact solution, but the wild oscillations characteristic of instability have never been observed.

This fact is not true for all splitting methods. In particular, the method similar to MITKIN but having E_1 lower triangular containing

half the diagonal of E and E_2 upper triangular with $E_3 = E_2$ and $E_4 = E_1$ has been observed to be strongly limited by stability problems.

4.3 Computer Requirements

The computer time necessary to generate a MITKIN approximate solution has been found to compare favorably with that required by other methods for comparable truncation error. The MITKIN time step size is roughly equivalent to that of the LUMAC method for most problems. The computer time necessary to carry out one time step, however, is about half that required by the LUMAC code. Comparison with running times for the TWIGL code is difficult, since the TWIGL runs were made on a different computer. It appears, however, that the MITKIN method is at least as fast as TWIGL. The following table gives experimental running times for the MITKIN method on the IBM 360/65 for various numbers of unknowns. The times quoted were computed by dividing the total execution time for a run by the number of time steps taken. It should be noted that one MITKIN time step will advance real time by 2h.

Table 4.1 Computer times.

Mesh points	Groups	Precursors	Seconds/Step
400	2	1	1.34
100	4	1	0.56
100	2	6	0.54
100	2	1	0.33
200	4	1	1.20

The computer time necessary to carry out a time step 2h will depend on the number of flux unknowns and the number of precursor unknowns. The precursors do not require as much computer time as the neutron groups, however, since the equations for the precursors are much simpler. If it is assumed that an expression of the following form is valid for the computer time necessary to carry out a time step, then the empirical constants α and β may be determined from the data of Table 4.1.

$$\text{Time/step} = \alpha N(G+\beta I)$$

where

N = number of mesh points

G = number of groups

I = number of precursors.

The above constants have been found to be approximately $\alpha = .00144$ sec and $\beta = 0.3$.

The above formula may be used to obtain a time estimate for larger problems. For example, a problem with 10 energy groups, 6 precursor groups, and 1000 mesh points would require an amount of computer time given by the following relation.

$$\text{Time/step} = (.00144 \text{ sec})(1000)[10+.3(6)]$$

$$\text{Time/step} \approx 17 \text{ seconds.}$$

The above time estimates are valid for the IBM 360/65 computer. A correction factor should be applied if time estimates are desired for another machine. The CDC 6600 is about four times as fast as the above-mentioned machine, for example, and, accordingly, the time per time

step on the CDC 6600 would be about 4 seconds for the above problem.

The MITKIN method as presently formulated requires the allocation of three storage locations to each of the neutron flux unknowns and one location to each precursor unknown. The three locations are required since it is necessary to simultaneously store the value of the flux at two time levels and a frequency. Clever programming can reduce the above requirement to two locations per flux unknown plus an additional N locations which hold the flux values in one energy group.

A primary advantage of the MITKIN method is that it is exceptionally easy to program. This logical simplicity is reflected in a compact program requiring a relatively small amount of core storage. The MITKIN code which is listed in Appendix C occupies approximately 52,000 bytes of core storage when compiled by the Fortran G compiler on the IBM 360/65.

Chapter 5

RECOMMENDATIONS

The MITKIN method possesses the following advantages:

a) low truncation error,

b) convergence rate $0(h^2)$,

c) experimentally stable,

d) minimal computation per time step,

e) simple programming logic.

Because of these advantages it is recommended for the solution of the reactor kinetics equations in two-dimensional rectangular geometries.

The extension of the MITKIN method to three-dimensional rectangular geometries is straightforward, and the method itself requires no modification. All of the stability and consistency proofs of Chapter 2 apply in three dimensions, thus the method may be expected to perform satisfactorily. The number of unknowns in a fine mesh three-dimensional computation is so enormous, however, that computer times for a reasonable number of steps becomes excessive. The MITKIN method does not represent a practical method in three space dimensions.

The method of the selection of the frequencies ω holds much promise for future refinement of the method. It would seem reasonable to use the MITKIN algorithm $B(\omega, h)$ to iteratively solve Eq. (2.28) for the frequencies ω. A particular method might be the successive substitution

$$\omega_k = g(\omega_{k-1}),$$

where

$$g(\omega) = \frac{1}{2h} \ln \left[\frac{B(\omega, h)\, \vec{\psi}^{J}}{\vec{\psi}^{J}} \right] .$$

If the time step size for the same truncation error is lengthened by more than the computer time necessary to obtain convergence of the above iteration, then this method would represent an improvement in the MITKIN technique.

BIBLIOGRAPHY

1. S. Kaplan, "The Property of Finality and the Analysis of Problems in Reactor Space-Time Kinetics by Various Modal Expansions," Nucl. Sci. Eng. 9, 357-361 (1961).

2. S. Kaplan, O. J. Marlowe, and J. Bewick, "Application of Synthesis Techniques to Problems Involving Time Dependence," Nucl. Sci. Eng. 18, 163-176 (1964).

3. C. G. Chezem and W. H. Köhler, Eds., "Coupled Reactor Kinetics," Proceedings of the National Topical Meeting of the American Nuclear Society, Texas, Jan., 1967.

4. W. H. Köhler, "Summary of Derivations of Coupled Point Reactor Kinetics Equations," Ibid., 192-213.

5. K. F. Hansen and S. R. Johnson, "GAKIN, A Program for the Solution of the One-Dimensional, Multigroup, Space-Time Dependent Diffusion Equations," USAEC Report GA-7543, General Atomic Division, General Dynamics Corporation (1967).

6. W. R. Cadwell, A. F. Henry, and A. J. Vigilotti, "WIGLE – A Program for the Solution of the Two-Group Space-Time Diffusion Equations in Slab Geometry," WAPD-TM-416 (1964).

7. W. T. McCormick, Jr., "Numerical Solution of the Two-Dimensional Time-Dependent Multigroup Equations," Ph. D. Thesis, Department of Nuclear Engineering, Massachusetts Institute of Technology, (1969).

8. J. B. Yasinsky, M. Natelson, and L. A. Hageman, "TWIGLE – A Program to Solve the Two-Dimensional, Two-Group, Space-Time Neutron Diffusion Equations with Temperature Feedback," WAPD-TM-743, Bettis Atomic Power Laboratories (Feb. 1968).

9. Eugene L. Wachspress, Iterative Solution of Elliptic Systems, Section 3.3.2, Prentice-Hall, Inc., Englewood Cliffs, N.J. (1966).

10. Richard S. Varga, Matrix Iterative Analysis, Chapter 8, Prentice-Hall, Inc., Englewood Cliffs, N.J. (1962).

11. Eugene Isaacson and Herbert Bishop Keller, Analysis of Numerical Methods, John Wiley and Sons, Inc., New York (1966).

12. Robert D. Richtmyer and K. W. Morton, Difference Methods for Initial-Value Problems, Chapter 3, John Wiley and Sons, New York (1967).

13. Ibid., Section 3.5.

14. Ibid., Section 3.8.

15. Melville Clark, Jr. and Kent F. Hansen, Numerical Methods of Reactor Analysis, Section 2.5, Academic Press, New York (1964).

16. J. B. Yasinsky, personal communication.

17. W. T. McCormick, Jr., op. cit., Chapter 3.

18. Alan L. Wight, personal communication.

APPENDICES

Appendix A

CRITICAL CONFIGURATIONS

This appendix contains the specification of reactor parameters defining five critical configurations. Symbols used in this appendix are defined below.

Δx = mesh spacing (cm) along x coordinate

Δy = mesh spacing (cm) along y coordinate

λ_i = decay constant (sec^{-1}) of i'^{th} precursor

β_i = delay fraction of i'^{th} precursor

f_{ij} = delay spectrum, i.e., the probability that precursor j produces a neutron in energy group i

v_i = velocity of i'^{th} neutron group (cm/sec)

x_i = prompt fission spectrum

D = diffusion coefficient

Σ_c = capture

ν = average number of neutrons per fission

Σ_f = fission cross section

$\Sigma_{J\rightarrow J+1}$ = scattering cross section (cm^{-1})

Critical Configuration 1

Number of neutron groups = 2

Number of precursor groups = 1

Geometry: Homogeneous square 200 cm on a side

$\Delta x = 20$ cm

$\Delta y = 20$ cm

Precursor Constants:

$\lambda_1 = .08, \quad \beta_1 = .0064, \quad f_{11} = 1.0, \quad f_{21} = 0.0$

	Group 1	Group 2
v	0.3×10^8	0.22×10^6
x	1.0	0.0

Material Properties:

	Group 1	Group 2
D	1.35	1.08
Σ_c	.00114	.0014069
ν	2.41	2.41
Σ_f	.000242	.00428
$\Sigma_{J \to J+1}$	.0023	0.0

Initial Conditions:

Spatial shape: Cosine

Spectrum: 1.0
.38234
.00034742

Critical Configuration 2

Number of neutron groups = 4

Number of precursor groups = 1

Geometry: Homogeneous square 150 cm on a side

$\Delta x = 15.0$ cm

$\Delta y = 15.0$ cm

Precursor Constants:

$\lambda_1 = .08$, $\beta_1 = .0074$, $f_{11} = 0.0$, $f_{21} = 1.0$, $f_{31} = 0.0$, $f_{41} = 0.0$

	Group 1	Group 2	Group 3	Group 4
v	$.25 \times 10^{10}$	$.5 \times 10^{9}$	$.43 \times 10^{7}$	$.25 \times 10^{6}$
x	0.575	0.425	0.0	0.0

Material Properties:

	Group 1	Group 2	Group 3	Group 4
D	2.0291	1.1609	.76965	.35676
Σ_c	.00237	.00438	.03266	.1339
ν	3.16578	3.16578	3.16578	3.16578
Σ_f	0.01316	0.00111	0.0182	0.38769
$\Sigma_{J \to J+1}$	0.06532	0.00481	0.00232	0.0

Initial Condition:

Spatial shape: Cosine

Spectrum: 1.0000000

11.2690000

1.0066000

0.0044746

0.0133890

Critical Configuration 3

Number of neutron groups = 2

Number of precursor groups = 6

Geometry: Homogeneous square 200 cm on a side

$\Delta x = 20.0$

$\Delta y = 20.0$

Delayed Constants:

$\lambda_1 = .0127, \quad \beta_1 = .000244$

$\lambda_2 = .0317, \quad \beta_2 = .001363$

$\lambda_3 = .1150, \quad \beta_3 = .001203$

$\lambda_4 = .3110, \quad \beta_4 = .002605$

$\lambda_5 = 1.40, \quad \beta_5 = .000819$

$\lambda_6 = 3.87, \quad \beta_6 = .000166$

$f_{1i} = 1.0, \quad f_{2i} = 0.0$

	Group 1	Group 2
v	$.3 \times 10^8$	$.22 \times 10^6$
x	1.0	0.0

Material Properties:

	Group 1	Group 2
D	1.35	1.08
Σ_c	.00114	.0014069
ν	2.41	2.41
Σ_f	.000242	.00408
$\Sigma_{J \to J+1}$	0.0023	0.0

Initial Condition:

Spatial shape: Cosine

Spectrum: 1.0

.38234

$.83435 \times 10^{-4}$

$.18672 \times 10^{-3}$

$.45429 \times 10^{-4}$

$.36376 \times 10^{-4}$

$.25405 \times 10^{-5}$

$.18628 \times 10^{-6}$

Critical Configuration 4

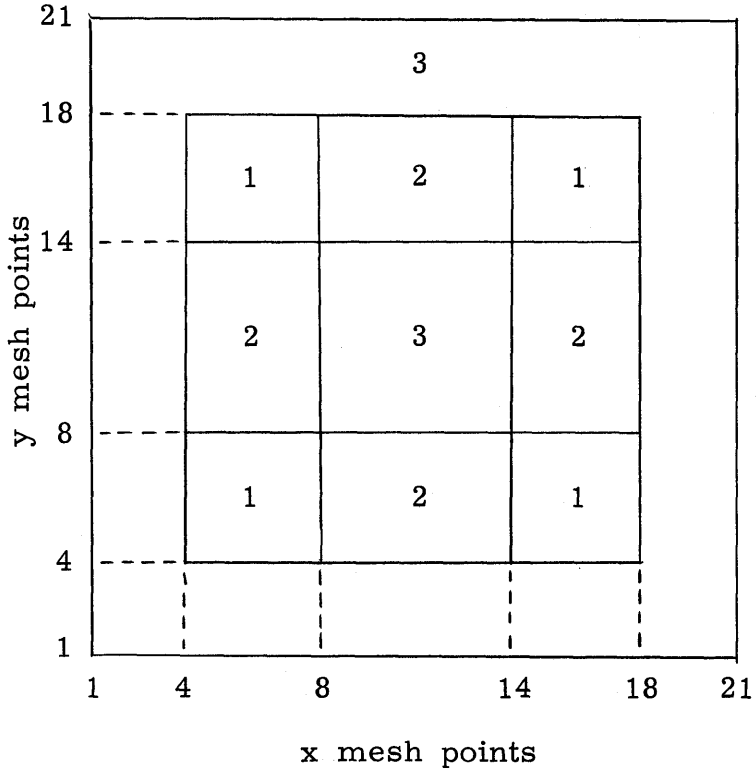
Number of neutron groups = 2

Number of precursor groups = 1

Geometry:

$\Delta x = 8.0$ cm

$\Delta y = 8.0$ cm



where the numbers indicate the material composition of that space region.

Delayed Constants:

$\lambda_1 = .08, \quad \beta_1 = .0075, \quad f_{11} = 1.0, \quad f_{21} = 0.0$

	Group 1	Group 2
v	$.1 \times 10^8$	$.2 \times 10^6$
x	1.0	0.0

Material Properties:

Material 1

	Group 1	Group 2
D	1.4	0.4
Σ_c	.0065	.05
ν	2.1877	2.1877
Σ_f	.0035	0.1
$\Sigma_{J \to J+1}$	0.01	0.0

Material 2

(same as material 1)

Material 3

	Group 1	Group 2
D	1.3	0.5
Σ_c	.0065	0.02
ν	2.1877	2.1877
Σ_f	.0015	.03
$\Sigma_{J \to J+1}$	.01	0.0

Critical Configuration 5

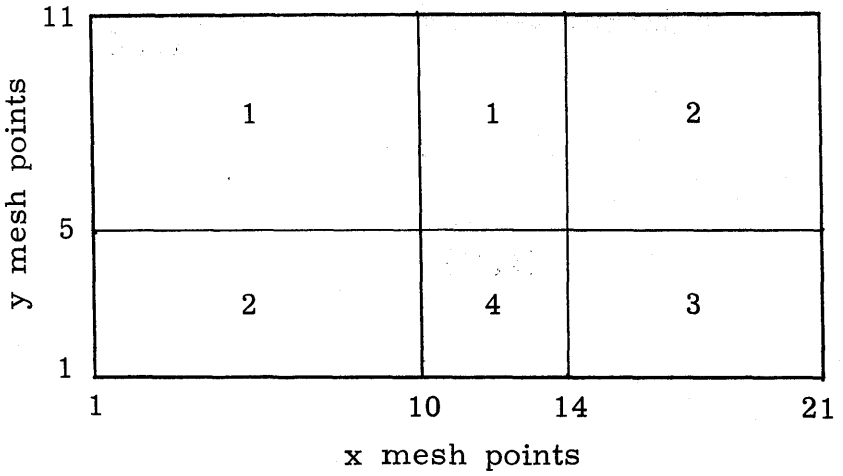
Number of neutron groups = 4

Number of precursor groups = 1

Geometry:

Δx = 8.0 cm

Δy = 8.0 cm



where the numbers indicate the material composition of that space region.

Delayed Constants:

$\lambda_1 = .08$, $\beta_1 = .0064$, $f_{11} = 0.0$, $f_{21} = 1.0$, $f_{31} = 0.0$, $f_{41} = 0.0$

	Group 1	Group 2	Group 3	Group 4
v	$.1 \times 10^{10}$	$.1 \times 10^{9}$	$.5 \times 10^{7}$	$.2 \times 10^{6}$
x	0.755	0.245	0.0	0.0

Material Properties:

Material 1

	Group 1	Group 2	Group 3	Group 4
D	2.7778	1.0753	.64103	.16260
Σ_c	.0013	.001	.0097	.115
ν	1.4507	1.4507	1.4507	1.4507
Σ_f	.00136	.00197	.0262	.54
$\Sigma_{J\to J+1}$	.0586	.00197	.085	0.0

Material 2

	Group 1	Group 2	Group 3	Group 4
D	3.3333	1.3889	.83333	2.0833
Σ_c	.00065	.0005	.0045	.058
ν	1.4507	1.4507	1.4507	1.4507
Σ_f	.0007	.0009	.0131	.274
$\Sigma_{J\to J+1}$	.0586	.0828	.0850	0.0

Material 3

	Group 1	Group 2	Group 3	Group 4
D	4.1667	2.0833	1.0753	.26247
Σ_c	.00077	.00072	.00051	.012
ν	0.0	0.0	0.0	0.0
Σ_f	0.0	0.0	0.0	0.0
$\Sigma_{J\to J+1}$	.0570	.0822	.0847	0.0

Material 4

(same as material 3)

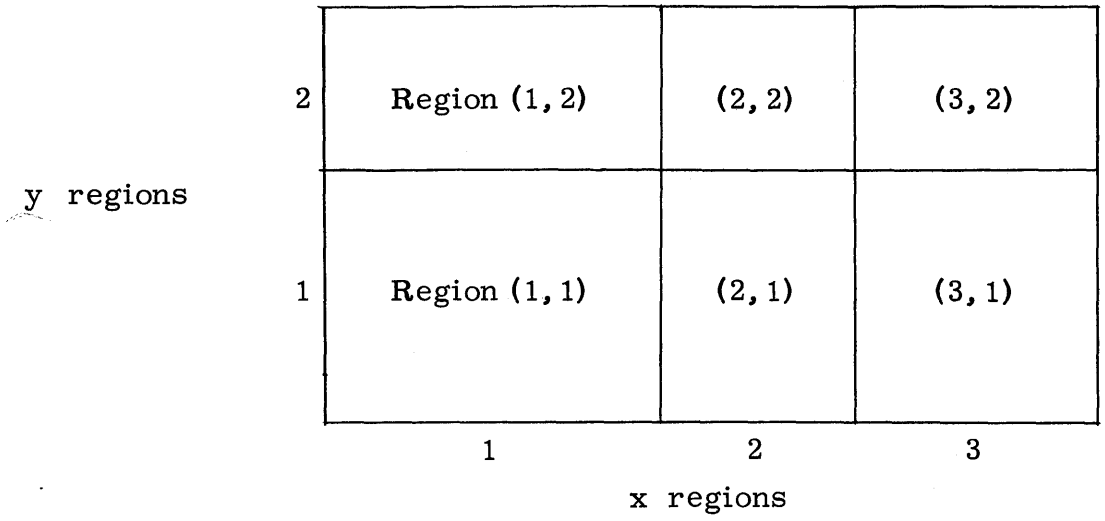
Appendix B

INPUT PREPARATION FOR MITKIN CODE

The MITKIN code was written in Fortran IV for the IBM 360 for the purpose of testing the MITKIN method with frequency correction in two dimensional rectangular geometries. The code is not intended to be universally general or for use as a production code, but it does possess certain flexibility. Only homogeneous Dirichlet boundary conditions can be handled, and the boundary zeros are stored and counted as mesh points. Maximum dimension sizes are 4 neutron groups, 6 precursor groups, 26 mesh points in each direction, 5 materials, and 10 spatial regions in each coordinate direction. Mesh spacings are allowed to vary from region to region.

The geometry is handled in a simple way. The reactor is divided into spatial regions by laying a grid across the reactor with lines parallel to the x and y axes. A particular region is identified by giving the set of numbers (I, J) specifying the X region and the Y region. A different material may be assigned to each region (I, J). The geometry of a sample problem is illustrated below.

The subroutine ALTER must be supplied by the user. It is called at each time step if IRAMP = 1 and is used to change the reactor properties during a time zone.



Card Type 1	FORMAT (20A4)
(TITLE(I), I = 1, 20)	Alphanumeric Title with 1 in column for page control
Card Type 2	FORMAT (16I5)
NNG	Number of neutron groups
NPG	Number of precursor groups
NXR	Number of X regions
NYR	Number of Y regions
NMAT	Number of materials
NXTP	Number of X test points
NYTP	Number of Y test points

IX1	IF IX1 = 0 cosine spatial shape IF IX1 = 1 initial condition read from cards
Card Type 3	FORMAT (5E15.5)
(THX(I), I = 1, NXR)	Thickness of I'th X region (in cm)
Card Type 4	FORMAT (5E15.5)
(THY(I), I = 1, NYR)	Thickness of I'th Y region (in cm)
Card Type 5	FORMAT (16I5)
(NIX(I), I = 1, NXR)	Number of mesh intervals in X region I
Card Type 6	FORMAT (16I5)
(NIY(I), I = 1, NYR)	Number of mesh intervals in Y region I
Card Type 7	FORMAT (16I5)
((MATC(I,J), J=1, NYR), I=1,NXR)	Material composition I. D. number of spatial region (I, J)
Card Type 8	FORMAT (16I5)
(ITP(I), I = 1, NXTP)	X coordinate of test points
(JTP(I), I = 1, NYTP)	Y coordinate of test points

Repeat Card Type 9 for I = 1, NPG

Card Type 9	FORMAT (5E15.5)
LAM(I)	Decay constant of I'th precursor (sec^{-1})

BETA(I)	Delayed fraction for I'th precursor
(F(I, K), K = 1, NNG)	Delay spectrum of I'th precursor

Card Type 10 — FORMAT (5E15. 5)

(V(I), I = 1, NNG)	Velocity for I'th neutron group (cm/sec)

Card Type 11 — FORMAT (5E15. 5)

(CHI(I), I = 1, NNG)	Fission spectrum

Repeat Card Types 12 through 17 for II = 1, NMAT

Card Type 12 — FORMAT (16I5)

I	I. D. number of material
INDF	IF INDF ≤ 0, $\nu = \Sigma_f = 0$
	IF INDF > 0, ν and Σ_f are read from cards

Card Type 13 — FORMAT (5E15. 5)

(DIFF(I, K), K = 1, NNG)	Diffusion coefficient material I, group K (cm)

Card Type 14 — FORMAT (5E15. 5)

(SIGC(I, K), K = 1, NNG)	Capture cross section material I, group K (cm^{-1})

Delete Card Types 15 and 16 if INDF ≤ 0

Card Type 15 — FORMAT (5E15. 5)

(XNU(I, K), K = 1, NNG)	ν for material I in group K

Card Type 16 — FORMAT (5+15. 5)

(SIGF(I, K), K = 1, NNG)	Fission cross section material I, group K (cm^{-1})

Card Type 17 — FORMAT (5E15. 5)

(SIGS(I,K,L), L=1, NNG), K=1, NNG)	Scattering matrix $\Sigma_{s_{K\rightarrow L}}$ (cm^{-1})
	Note: $\Sigma_{s_{J\rightarrow J}}$ should be set equal to zero.

Delete Card Type 18 if IX1 = 1

Card Type 18 — FORMAT (5E15. 5)

(EIGEN(I), I = 1, NNG + NPG)	Initial spectrum for homogeneous problems

Delete Card Type 19 if IX1 ≠ 1

Repeat Card Type 19 for K = 1, NNG + NPG

Card Type 19 — FORMAT (6E12. 6)

(A(K,I,J), I=1, ITOT), J=1, JTOT)	The initial condition where ITOT + JTOT are total number of mesh points in X + Y directions counting boundary points.

Repeat Card Type 20 as often as desired

Card Type 20 — FORMAT (3I5, E15. 5)

IPRIN	Printout will occur every IPRIN steps
MAX	Total number of steps to be taken in this time zone

IRAMP	IF IRAMP = 1 the subrouting "alter" is called. User must supply own routine.
H	Time step size to be used in this time interval.

Appendix C

CODE LISTING