

1970

Applications of optimal control theory to space-time reactor kinetics

Steven Robert Specker
Iowa State University

Follow this and additional works at: <https://lib.dr.iastate.edu/rtd>

 Part of the [Nuclear Engineering Commons](#), and the [Oil, Gas, and Energy Commons](#)

Recommended Citation

Specker, Steven Robert, "Applications of optimal control theory to space-time reactor kinetics " (1970). *Retrospective Theses and Dissertations*. 4799.
<https://lib.dr.iastate.edu/rtd/4799>

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

71-7332

SPECKER, Steven Robert, 1945-
APPLICATIONS OF OPTIMAL CONTROL THEORY TO
SPACE-TIME REACTOR KINETICS.

Iowa State University, Ph.D., 1970
Engineering, nuclear

University Microfilms, Inc., Ann Arbor, Michigan

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED

APPLICATIONS OF OPTIMAL CONTROL THEORY TO
SPACE-TIME REACTOR KINETICS

by

Steven Robert Specker

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subject: Nuclear Engineering

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
Ames, Iowa

1970

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
II. TRAJECTORY OPTIMIZATION TECHNIQUES	13
A. Problem Formulation	14
B. Solution Methods	17
C. Penalty Function Approximation	26
III. THEORY	28
A. Natural Mode Approximation	28
B. Conjugate Gradient Method	38
IV. MATHEMATICAL MODEL AND ANALYSIS	50
A. Reactor Model	50
B. Modal Analysis	51
C. Formulation of the Optimal Control Problem	59
V. RESULTS AND CONCLUSIONS	70
VI. SUGGESTIONS FOR FURTHER STUDY	82
VII. ACKNOWLEDGMENTS	83
VIII. LITERATURE CITED	84
IX. APPENDIX: DERIVATION OF NECESSARY CONDITIONS FOR AN OPTIMAL CONTROL	89
A. Statement of Problem	89
B. Derivation of Necessary Conditions	90

I. INTRODUCTION

In recent years nuclear power has become economically attractive for many electric power systems throughout the world. With this increased use of nuclear power has come a corresponding sophistication in the design and control of nuclear reactors. The development of analytical and numerical methods for reactor analysis has reached the state where it is possible to compute accurately the effect of design or control changes on the reactor system. The question can now be asked as to which control scheme is "best" with respect to some predefined performance criteria.

In attempting to decide which control program is best, recourse is most often made to the theory and numerical methods of mathematical optimization. Of particular interest to the nuclear engineer is that branch of optimization usually referred to as optimal control (1). The optimal control problem can be stated imprecisely as the problem of selecting from a specified set of functions that control function which minimizes a given functional and which satisfies specified differential and algebraic constraints involving the problem or state variables (2). The necessary conditions for optimality of a control function have been derived by many authors. Berkovitz (3) adapts the classical calculus of variations to the control problem, while Pontryagin et al. (4) derive similar necessary conditions from a geometric view-

point. Although these necessary conditions rarely lead to an analytical determination of the optimal control, they form the theoretical foundation upon which the numerical solution techniques are built.

The methods of optimal control theory have been applied to "point" reactor control problems with some success. Lipinski (5) presents a comprehensive literature survey of previous applications of modern control theory to nuclear reactors, as well as a large number of selected nuclear and control engineering bibliographies.

When the reactor in question is small, the point reactor kinetics equations are usually adequate to analyze the response of the system to perturbations. However, for large power reactors, the various regions of the reactor core can become loosely coupled and any non-uniform changes in the properties of the reactor cause local distortions of the power distribution. In this case, the space-time separable assumption is no longer valid, and the use of the point model in the study of spatially dependent reactor systems is open to question.

Yasinsky and Henry (6) illustrated the importance of considering the spatial effect in reactor control. They made comparisons between exact and approximate solutions to the two-group space-time diffusion equations for two slab cores, one 240 cm. thick and the other 60 cm. thick.

Comparisons between space-time solutions and solutions obtained using the point kinetics equations clearly established the inadequacy of the point kinetics approximation in representing the neutron kinetic behavior for large reactor cores. It can be concluded from these results that if the point kinetics approximation is used to determine an optimal control scheme for a large reactor, the control may be optimal with respect to the point model, but it may not be optimal in the true sense, since it is only as good as the mathematical model employed. Therefore, it is not only desirable, but necessary to study the control of large reactor systems by taking into account the spatial distribution of the system responses.

A nuclear reactor in which spatial effects are important is an excellent example of a distributed parameter system (7). For control problems, such systems are characterized by the need to control a physical variable which is dependent not only on time but on its spatial distribution. This requires the direct consideration of distributed parameter mathematical models which are in the form of partial differential or integral equations. In a nuclear reactor the neutron flux is usually the distributed parameter of interest, and the time-dependent multi-group diffusion equations are the partial differential equations describing the system behavior.

In contrast to the wide scale application of optimal control theory to systems governed by ordinary differential equations, little has been done in the area of optimal control for distributed parameter systems. Certain general conditions and functional equations associated with the application of optimal control theory to distributed parameter systems have been derived for general systems (7), and for spatially dependent nuclear reactor systems (8, 9). However, in order to obtain solutions to practical problems it has been necessary to devise effective approximation schemes and computational procedures. These approximations are applied either to the distributed mathematical model of the system or to the equations derived from the optimization theory. The approximations applied to the distributed mathematical models can be grouped into two general classifications, discretizing methods and synthesis methods.

The discretizing methods generally take on one of the following forms (7):

1. spatial discretization - The discretized mathematical model consists of a finite dimensional system of continuous-time ordinary differential equations.
2. time discretization - The discretized model consists of a finite-dimensional system of spatially-continuous ordinary differential equations.

3. space-time discretization - The discretized model consists of a finite-dimensional system of difference equations.

It is interesting to compare these methods with the approximation techniques often applied to the time-dependent neutron diffusion equations. Spatial discretization is exemplified by the nodal analysis technique. In this method the reactor is divided into regions, or nodes, and a set of time-dependent equations is found involving the average fluxes at each node. The space part of the problem consists of the determination of the parameters governing the diffusion of neutrons from node to node (10). Time discretization is used quite often in fuel depletion studies in which a series of space dependent steady-state problems are solved. The digital computer code WIGLE (11) utilizes space-time discretization in constructing a useful algorithm for solving space-time problems.

Synthesis methods have been used widely in solving various problems in reactor theory. A review of the application of these methods to reactor analysis provides a comprehensive summary of the applicability of synthesis methods to any distributed parameter system. The unifying idea of synthesis methods is to reduce the number of independent variables by constructing an approximate solution in the form of a linear combination of known functions of one or more of the variables, with the coefficients of combination being func-

tions only of the remaining variables. In applying synthesis methods to time-dependent neutron diffusion problems the approximation most often employed is based on the expansion

$$\underline{\Phi}(\underline{r}, t) = \sum_{k=1}^K \psi_k(\underline{r}) \underline{T}_k(t) \quad (1)$$

where

$$\underline{\Phi}(\underline{r}, t) = \text{col}[\phi_1, \phi_2, \dots, \phi_G, C_1, C_2, \dots, C_P, \dots \text{etc.}]$$

$$\phi_g = \text{group } g \text{ neutron flux}$$

$$C_p = \text{group } p \text{ delayed neutron precursor}$$

$$G = \text{number of neutron flux groups}$$

$$P = \text{number of precursors}$$

$$\psi_k(\underline{r}) = \text{known space functions}$$

$$\text{etc.} = \text{other important variables (e.g. temp., xenon conc.)}$$

$$\underline{T}_k(t) = \text{coefficients of combination.}$$

The space part of the problem is concerned with the selection or construction of the space functions $\psi_k(\underline{r})$; the time part is the determination of the coefficients of combination. Based on the type of space functions employed, these methods can be divided into two classifications.

In the first type, commonly called modal analysis methods, the space functions are members of an orthogonal set of functions. The most common expansion of this type utilizes solutions of the Helmholtz equation. The Helmholtz modes

are attractive in that they are easily computed. The most serious defect of these modes is that a great number are necessary to describe the flux shape in a reactor with even the slightest geometrical complication.

The ω_d -modes or natural modes, which are the eigenvectors of a linear operator derived from the complete set of equations describing the reactor system at an initial reference condition, also form an orthogonal set. They have the advantage of being more readily tailored to the heterogeneities of a particular problem since those modes appropriate to the particular geometry in question can in principle be determined (12, 13).

A particular advantage of the orthogonal expansions is that expressions for the time coefficients (or coefficients of combination) can be obtained by using common orthogonality properties. In general the time coefficients are related by a set of coupled, nonlinear, ordinary differential equations. The main disadvantage of orthogonal expansions is the difficulty in extending the techniques to two and three dimensional problems. Foulke and Gyftopoulos (13) have attempted to overcome this difficulty by using space synthesis techniques.

The second method of constructing the space functions does not require orthogonality. One is free to use any set of space functions, the only criterion being the goodness

of the resulting approximation. In time-dependent neutron diffusion problems the space functions are usually chosen as known spatial flux shapes which are representative of reactor conditions at various times during a transient. A particular advantage of this method is the ability to consider two and three dimensional problems using space-time synthesis methods. The freedom to tailor the space functions to the problem is offset somewhat by the increased complexity involved in calculating the coefficients of combination. Orthogonality properties cannot be employed as in the modal analysis methods. Instead recourse is made to a variational or weighted residual argument in order to derive differential equations for the coefficients of combination (14, 15).

Instead of applying approximations directly to the distributed mathematical model of the system, an alternate approach has been to derive spatially dependent optimal control equations and then approximate the spatial dependence by a discretizing method. The principal motivation for this approach is the concern that the optimal control equations derived by the two methods will differ, and that the results based upon the spatially dependent derivation will in some sense be superior. Stacey (9) has used a spatially dependent derivation in formulating the problem of optimally controlling xenon spatial oscillations. The necessary conditions

for an optimal control are given by a system of space and time dependent differential equations. These equations are then approximated by a nodal representation to obtain a set of ordinary time-dependent differential equations with mixed boundary conditions. An iterative solution scheme, which utilizes a quasilinearization of the equations and a transformation matrix relating initial to final values of certain variables, is employed to obtain numerical results. In an appendix to this paper Stacey shows that identical equations for computing the optimal control are obtained regardless of whether the nodal approximation is applied to the distributed mathematical model of the system or to the spatially dependent optimal control equations.

The most common approximation scheme has been the use of a modal expansion of the state functions (e.g. flux, temp.) in order to obtain a system of time-dependent ordinary differential equations. Wiberg (16) utilized such an approximation in analyzing the problem of controlling xenon spatial oscillations. A general formulation of the problem was derived using the classical calculus of variations. However, in order to obtain equations for numerical computation, the equations resulting from the modal approximation were linearized. Since nonlinear effects are important in many problems, the practical application of these results to obtain quantitative results seems to be limited.

Suda (17) utilized a modal expansion of the flux distribution to investigate the controllability of the flux distribution of large reactors. Discussions were presented on the decomposition of the optimal control of flux distribution into those of the individual modes, on the controllability of reactors with feedback, and on the controllability when there is a constraint on the control rod speed.

Kliger (18) used a modal expansion in determining an optimal control law for a space-dependent nuclear reactor. A pseudo-control was defined which eliminated the coupling between modes.

Stacey (19) has attempted a different approach to the problem. The state functions (e.g. flux, temp.) are approximated by a modal expansion which, in conjunction with orthogonality conditions yields a non-coupled set of ordinary differential equations solvable in closed form. The non-coupling results from defining a pseudo-control function. This pseudo-control is expanded in a series of known time functions and unknown coefficients of combination. A variational argument is then used to obtain a set of coupled algebraic equations which can be solved for the unknown coefficients of combination.

Dynamic programming (20) is another technique that has been applied to the problem of computing the optimal control for spatially dependent reactors. In this method the optimal

control problem is considered to be a multistage decision process. Stacey (21) formulated the problem of controlling xenon-power spatial transients in a large thermal reactor in terms of the dynamic programming formalism. Wall and Fenech (22) applied dynamic programming to a fuel management optimization problem with the objective of minimizing power cost over the plant life. Unfortunately, because of the large number of state variables present in a realistic spatially dependent reactor model, it was necessary in both of these studies to limit severely the allowable control actions to be considered. Such limitations are necessitated by the enormous digital computer storage requirements encountered when more than two or three state variables are considered.

To date, most applications of optimal control theory to spatially dependent nuclear reactors have dealt with the problem of controlling spatial xenon oscillations in large thermal reactors. Of increasing concern, however, is the ability of nuclear power plants to make rapid changes in power under automatic control. Rapid changes will be required if nuclear plants are to be used to control power system frequency. Such control is now done by fossil-fuel-fired plants, but as the relative number of nuclear plants in a power system grows, it will become increasingly important for the nuclear plants to be able to assume a

considerable part of this frequency control (23). This need is recognized in the Liquid Metal Fast Breeder Reactor Program Plan (24) in which high priority is given to the development of methods of optimal control pertaining to rapid power level changes in spatially dependent reactors.

This dissertation concerns the development and application of an optimization technique from which optimal control schemes for rapid power level changes in a spatially dependent reactor may be computed. A modal approximation in conjunction with state variable techniques is used to formulate the problems in terms of a trajectory optimization problem. Various numerical optimization methods for solving trajectory optimization problems are discussed, with the good and bad points of each method outlined. The theory is applied to a bare slab reactor described by the two-group diffusion equations. Numerical results are obtained for a specific problem using a conjugate gradient algorithm. These results are compared with results obtained from point kinetics and space-time finite difference methods.

II. TRAJECTORY OPTIMIZATION TECHNIQUES

The optimal control problem to be studied in this thesis can be classified as a trajectory optimization problem. Interest in optimizing spacecraft trajectories has resulted in a large amount of recent literature devoted to both the mathematical theory of trajectory optimization problems and the methods for obtaining solutions to specific problems (25, 26, 27). The advent of large, high-speed digital computers has made possible the application of many numerical optimization techniques which were not feasible ten years ago. As a result there are a number of algorithms from which to choose when attempting to solve such a problem.

The purpose of this chapter is to review and summarize the pertinent theory and numerical methods applicable to trajectory optimization problems. In the first section a precise statement of the general problem is given. Several restrictions are then imposed which lead to a simplified formulation of the problem. The necessary conditions for optimality are discussed. In the following section the different numerical methods which can be used to solve for the optimal control are reviewed. The chapter concludes with a discussion of the penalty function method for converting constrained problems into unconstrained approximating problems.

A. Problem Formulation

In general, the trajectory optimization problem can be stated as follows:

Find the m -vector of control variables $\underline{u}^*(t)$ that minimizes a scalar performance index of the form

$$J = S(\underline{x}(t_f), t_f) + \int_{t_0}^{t_f} G(\underline{x}, \underline{u}, t) dt \quad (2)$$

subject to the differential constraints

$$\dot{\underline{x}} = \underline{f}(\underline{x}, \underline{u}, t) , \underline{x}(t_0) = \underline{x}_0 \quad (3)$$

and the terminal constraints

$$\underline{M}(\underline{x}(t_f), t_f) = 0 . \quad (4)$$

In the above, \underline{x} is an n -vector of state variables to be controlled, \underline{f} is an n -vector of nonlinear expressions defining the dynamical system to be controlled, \underline{u} is an m -vector of control functions on the interval $[t_0, t_f]$, \underline{M} is a p -vector of linear or nonlinear expressions constraining the terminal conditions of the dynamical system where $p \leq n+1$, G is a scalar function, and t_0 and t_f are the initial and final times which may or may not be specified. It is assumed that $\underline{f}(\underline{x}, \underline{u}, t)$ and all of its derivatives are continuous in the interval of interest. It is assumed also that the control variable $\underline{u}(t)$ is unbounded and that there are no constraints on the state history except at the initial and terminal

boundaries. The performance index of Equation 2 is said to be in the Bolza form. However, if $S = 0$ the performance index takes the Lagrange form, and if $G = 0$ it takes the Mayer form.

The question of existence and uniqueness of the optimal control is not considered here, since it is assumed that the optimal control problems of interest are "well posed" in the sense that they possess unique solutions.

For purposes of this dissertation the general trajectory optimization problem can be simplified by imposing several restrictions on the general formulation. They are:

1. the initial and final times are fixed
2. the performance index is in the Mayer form
3. only a single control function is considered
4. no terminal state constraints are allowed.

The last restriction does not eliminate from consideration problems with terminal state constraints. It will be shown later that the penalty function method can be used to formulate a constrained problem as an unconstrained approximating problem. Restriction 2 necessitates transforming a performance index of the Bolza or Lagrange form into the Mayer form. This poses no problems since simple methods exist for transforming any one of the three forms into either of the other two.

The simplified optimization problem to be considered in

this study can be stated as follows:

$$\text{minimize } J = S(\underline{x}(t_f)) \quad (5)$$

$$\text{subject to } \dot{\underline{x}} = \underline{f}(\underline{x}, u, t) \quad (6)$$

$$\underline{x}(t_0) = \underline{x}_0 . \quad (7)$$

The conditions that are necessary for the optimality of a control $u(t)$ are

$$\dot{\underline{x}} = \underline{f}(\underline{x}, u^*, t) \quad (8)$$

$$\underline{x}(t_0) = \underline{x}_0 \quad (9)$$

$$\dot{\lambda}_i(t) = - \frac{\partial H}{\partial x_i}(\underline{x}, u^*, t) \quad i = 1, n \quad (10)$$

$$\lambda_i(t_f) = \left. \frac{\partial S}{\partial x_i} \right|_{t=t_f} \quad i = 1, n \quad (11)$$

$$g(u^*) = \left. \frac{\partial H}{\partial u} \right|_{u=u^*} = 0 \quad (12)$$

where the Hamiltonian function H is defined as

$$\begin{aligned} H(\underline{x}, u, \underline{\lambda}, t) &= G + \underline{\lambda}^T \underline{f} \\ &= \underline{\lambda}^T \underline{f} \quad \text{when } G = 0 \end{aligned} \quad (13)$$

and where $\underline{\lambda}(t)$ is an n -vector of adjoint or costate variables on $[t_0, t_f]$. The superscript $*$ indicates the value of the variable at the minimum of the functional. A derivation of the necessary conditions is given in the Appendix.

The condition given by Equation 12 holds only at the

minimum of $J(u)$, i.e. when $J(u) = J(u^*)$. The expression $g(u) = \partial H / \partial u$ is the gradient to the Hamiltonian and points in the direction of increasing J (2). The optimality condition of Equation 12 is not given by Pontryagin et al. (4) as a necessary condition since Pontryagin's maximum principle is derived for a closed control space. However, for problems where the Hamiltonian has continuous first partial derivatives with respect to its arguments and the control space is unbounded, relative minima of J will occur when $u^*(t)$ satisfies

$$\left. \frac{\partial H}{\partial u} \right|_{u=u^*} = 0$$

and the other optimality conditions (Equations 8-11).

B. Solution Methods

Equations 8 to 11 form a nonlinear two-point boundary value problem. That is, for the system of $2n$ -equations given by Equation 8 and 10, n initial conditions are specified for \underline{x} and n terminal conditions can be calculated for $\underline{\lambda}$ using Equation 11. As a result, solution of the system of equations by direct numerical integration is not possible. Recourse must be made to iterative numerical methods.

In general, the methods used can be classified as either direct or indirect. The indirect methods use the necessary conditions for optimality, Equations 8-13, as a

starting point and attempt by various iterative techniques to satisfy these conditions. In the direct methods the constraining system of differential equations is satisfied and an iteration made on the control function such that each new iterate improves the performance index to be minimized.

The general computational procedure for the indirect methods is as follows:

1. Use the condition $\partial H/\partial u = 0$ to solve for $u^*(t)$.
 $u^*(t)$ is a function of $\underline{x}(t)$, $\underline{\lambda}(t)$, and t .
2. Substitute $u^*(\underline{x}, \underline{\lambda}, t)$ into Equations 8 and 10 to eliminate the control variable in these equations.
 The result can be expressed as

$$\begin{aligned}\dot{\underline{x}} &= \partial H/\partial \underline{\lambda} \\ \dot{\underline{\lambda}} &= -\partial H/\partial \underline{x}\end{aligned}\quad (14)$$

where H is now a function of \underline{x} , $\underline{\lambda}$, and t .

3. Assume initial values for the adjoint variables, $\underline{\lambda}(t)$, and integrate the system given by Equation 14 from t_0 to t_f . Initial conditions for \underline{x} are given by Equation 9.
4. Compute the difference between the resultant terminal values of $\underline{\lambda}$ and the values calculated from Equation 11.
5. Use this terminal error to generate a correction in the assumed initial values of the adjoint variables. The particular scheme used to compute this

correction distinguishes the different indirect methods from one another. Tapley and Lewallen (25) present a good discussion of these techniques.

6. Apply the corrections and generate a new trajectory by integrating Equation 14 from t_0 to t_f . The terminal error in the adjoint variables is reduced.
7. Iterate using steps 4, 5, and 6 until a given convergence criteria is satisfied.
8. Solve for $u^*(t)$ using the relationship obtained in step 1.

McGill and Kenneth (28) have developed an indirect technique called quasilinearization. This technique differs from the other indirect methods in that it solves the two-point boundary value problem of Equation 14 by choosing iterates that satisfy the boundary conditions exactly and that approach satisfaction of the differential equations as the iteration proceeds. A particular advantage of quasilinearization is that it is only necessary to solve the linearized form of Equation 14.

The main disadvantage associated with the indirect methods is that the variation of the terminal boundary conditions with variations in the initial values of the adjoint variables is often so sensitive that practical numerical application is impossible (26).

The direct methods can be grouped into two classifica-

tions, gradient techniques and second-variation methods. The basic difference between the two lies in the accuracy of the approximation used to relate control variations $\Delta u(t)$ to changes in the performance index, ΔJ . The gradient methods use a linear approximation, thereby ignoring second order and higher terms. Second-variation techniques, as their name implies, include second-order terms in the approximation for ΔJ . Since the second-order methods normally employ a gradient process in their initial phase of computation, a discussion of the gradient methods will be presented first.

In common with all gradient techniques is the use of a descent method to decrease the value of a performance index continuously from one step to the next. To demonstrate with a simple example, assume that a minimum of the functional $J(\underline{x})$ is sought and that an initial point \underline{x}_0 is given. The iterations are constructed according to an equation of the form

$$\underline{x}_{n+1} = \underline{x}_n + \alpha_n \underline{s}_n \quad (15)$$

where α_n is a scalar and \underline{s}_n is a direction vector. The manner in which these successive direction vectors \underline{s}_n are generated distinguishes the different gradient methods from one another. Once the direction vector is chosen, the scalar α_n is selected to minimize $J(\underline{x}_n + \alpha \underline{s}_n)$, regarded as a function

of the scalar α . Generally, the relationships are arranged (by multiplying \underline{s}_n by -1 if necessary) so that $J(\underline{x}_n + \alpha \underline{s}_n) < J(\underline{x}_n)$ for small positive α . The appropriate value of α is usually determined by an iterative search or a suitable approximation.

The descent process can be visualized in function space as shown in Figure 1 where the performance index J is represented by its contours. Starting from a point \underline{x}_0 , one moves along the direction vector \underline{s}_0 calculated by the particular method until reaching the first point where the line $\underline{x} + \alpha \underline{s}_0$ is tangent to a contour of J . From \underline{x}_1 one moves in the direction \underline{s}_1 until the line $\underline{x}_1 + \alpha \underline{s}_1$ is tangent to a contour of J . If J is bounded below, it is clear that a continuation of this process defines a bounded decreasing

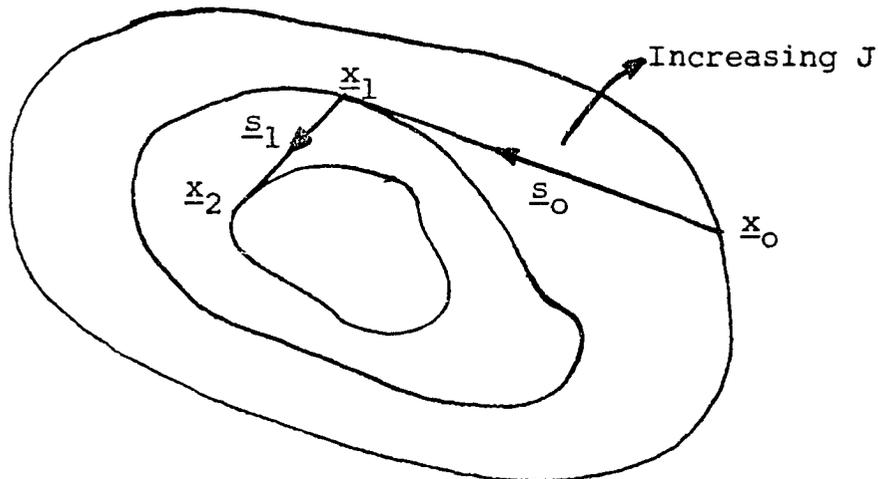


Figure 1. The descent process

sequence of functional values which tend toward a minimum (29).

Steepest descent is the gradient method most often used in solving optimal control problems, mainly because of its computational simplicity. Although considered as early as 1847 by Cauchy for the minimizing of an objective function of several variables, it was not until the early 1960's that the steepest descent technique was extended to function space. This extension was done by Bryson and Denham (30) and by Kelley (31). In addition, these authors and others have derived methods of applying steepest descent to control problems involving terminal state constraints, state-space constraints, and control variable constraints. They have utilized both a penalty function approach and Rosen's gradient projection method (32, 33) when dealing with constrained problems. The steepest descent technique is based on the principle of choosing a new direction vector that lies along the direction of maximum decrease of the functional J . For the general trajectory optimization problem, this direction is given by the negative of the gradient to the Hamiltonian, $g(u) = \partial H / \partial u$ (2).

The advantages of steepest descent are that convergence does not depend upon the availability of a good initial estimate of the optimum as a starting point, and that it seeks out relative minima instead of points which are merely

stationary as in the indirect methods. However, there are three principal disadvantages. First, the convergence, although usually relatively good in the beginning of the iterative sequence, often slows considerably as the optimum trajectory is approached. Second, the penalty function method used to solve problems with terminal constraints introduces a trial and error process in the calculation (see Section C). Third, regions of severe irregularity sometimes develop in the control functions.

The conjugate direction methods are a class of gradient techniques whose convergence does not slow appreciably as the optimum trajectory is approached. They owe their superior convergence to the fact that new directions of search are computed using information about the performance index J at all previous search points. This accumulation of information is in direct contrast to the steepest descent method in which the new direction of search is computed using information about the behavior of J at the current search point only. The conjugate direction methods will be discussed further in Chapter III where a detailed discussion of the conjugate gradient method is presented.

In an attempt to improve upon the shortcomings of the gradient methods, second-order methods of solving optimal control problems have been developed by Breakwell et al. (34), and Kelley et al. (35), and others. As with gradient

methods, the initial trajectory estimate is not required to be a good estimate of the optimal trajectory. In the initial phase of computation, the penalty function approximation of terminal constraints is employed and the behavior of the process strongly resembles that of a gradient process. In this stage of the process constraints are imposed on the step-size in order to limit the amount of improvement sought during each cycle. These constraints are progressively relaxed, finally dropped, and the terminal penalty scheme is replaced by exact terminal conditions that are ultimately satisfied if a solution exists. Thus, the basic philosophy of these methods is to use a gradient method to reach close to the optimal trajectory and then to switch to the second-order method for refinement.

There are three primary advantages of the second-variation methods when compared to the gradient methods (26):

1. In the final stages of the computational procedure the penalty function technique is no longer required and each successive approximation attempts to satisfy the boundary conditions exactly. Thus the undetermined constants associated with the penalty function terms are eliminated.
2. In both phases of the second-variation method, the step-size is automatically determined, thus

eliminating the independent procedure needed in the gradient methods.

3. In the final stages of the computational procedure the method exhibits quadratic convergence.

The main disadvantage associated with the second-variation methods is the additional programming effort needed to formulate the computing algorithm.

A method of numerical optimization that is not easily classified as direct or indirect has been developed by Bellman (20). In this method, known as dynamic programming, the optimal control problem is considered to be a multi-stage decision process. The main advantage of dynamic programming is that by the use of the principle of optimality the problem is reduced to a sequence of single-stage decision processes. Of further importance is the fact that constraints on either the control or state variables actually simplify the solution process. Unfortunately the systematic simplicity of the method is often offset by the enormous storage requirements. Many optimal control problems, when cast in the dynamic programming formalism, require so much computer storage that solution by that method is not feasible.

From the preceding discussion it should be apparent that there is not a categorically "best" method for solving optimization problems. Each method should be evaluated with respect to the particular problem to be solved and the

computer facilities available. There were several reasons for choosing the conjugate gradient method to solve the trajectory optimization problems considered in this study. First, the problem to be solved could be formulated in a manner well suited for solution by the conjugate gradient method. Secondly, a computer program utilizing a conjugate gradient algorithm, and written for the available computer facilities was available. Extensive refinement and adaptation of this original program yielded an effective tool for the solution of the trajectory optimization problems considered in this study.

C. Penalty Function Approximation

When employing a direct method it is often necessary to reformulate the problem so that all of the state variables are unconstrained at the final time. This can be done in a straightforward manner by using a penalty function approximation (2, 29, 36). If the terminal constraints are given as in Equation 4

$$\underline{M}(\underline{x}(t_f), t_f) = 0 \quad (16)$$

where \underline{M} is a p -vector of nonlinear or linear expressions, then a new performance index can be defined as

$$J' = S(\underline{x}(t_f), t_f) + 1/2 \sum_{i=1}^p K_i M_i^2 \quad (17)$$

where the K_i are constants called penalty coefficients. A minimum of J' is now sought without requiring the terminal values to satisfy Equation 16 exactly, but rather paying a "penalty" for deviations. As the K_i are made large, the trajectory that minimizes J' is in some sense close to the trajectory that minimizes J with Equation 16 satisfied.

A useful algorithm is obtained by successively solving the optimization problem with the K_i increasing for each iteration. However, there does exist one serious problem in attempting to implement this procedure. This is the question of determining how fast the penalty coefficients K_i should be increased. If the coefficients are increased too fast, the new control will tighten up the terminal constraints but leave the unaugmented part of J' , $S(x(t_f), t_f)$ essentially unchanged. By contrast, if the coefficients are not increased fast enough, the specified terminal constraints will be approached quite slowly. As a result, a certain amount of trial-and-error may be required to achieve a logical sequence of the penalty coefficients.

III. THEORY

When applying optimal control theory to a spatially dependent reactor control problem, two main steps are required in the analysis. It is first necessary to formulate the problem in a manner suitable for solution by an optimization technique. The Natural Mode Approximation (13) is utilized in this dissertation. Secondly, a suitable optimization technique must be chosen from among the many that are available. For the analysis here, the conjugate gradient method is used.

The purpose of this chapter is to discuss the theory underlying the Natural Mode Approximation method and the conjugate gradient method in order to form a basis for the applications considered in Chapter IV.

A. Natural Mode Approximation

As discussed in Chapter I, a modal approximation seeks to represent the neutron flux and other dependent variables, e.g. precursor concentration, by a finite linear combination of space-dependent functions with time-dependent coefficients of combination. The space-dependent functions, or space modes, are chosen to be members of an orthogonal set of functions in order to facilitate the derivation of expressions for the time-dependent coefficients. The particular set of space modes which are employed distinguishes the

different modal approximations from one another. The so-called "Natural Modes" of the reactor will be utilized as the space modes for the analysis here (12, 13).

Multi-group diffusion theory approximations enable the space and time dependent behavior of a reactor without feedback or external sources to be described by the matrix equation

$$L(\underline{r}, t)\underline{\Phi}(\underline{r}, t) = \frac{\partial \underline{\Phi}(\underline{r}, t)}{\partial t} \quad (18)$$

where

$$\underline{\Phi}(\underline{r}, t) = \text{col}[\phi_1, \phi_2, \dots, \phi_G, C_1, C_2, \dots, C_P]$$

$$\phi_g \quad = \text{group } g \text{ neutron flux}$$

$$C_p \quad = \text{group } p \text{ delayed neutron precursor}$$

$$G \quad = \text{number of neutron flux groups}$$

$$P \quad = \text{number of precursors}$$

$$K \quad = G + P .$$

The $K \times K$ matrix operator $L(\underline{r}, t)$ consists of all the production and destruction operators.

The Natural Mode Approximation (NMA) method seeks an approximate solution of Equation 18 according to the following procedure:

Consider the critical or steady-state reference condition defined by

$$L_c \underline{\Phi} = 0 \quad (19)$$

where L_0 represents the steady-state matrix operator. Define the eigenvectors $\psi_{mk}(\underline{r})$ of the eigenvalue problem

$$L_0 \psi_{mk}(\underline{r}) = \omega_{mk} \psi_{mk}(\underline{r}) \quad (20)$$

as the natural modes of Equation 19. It is assumed that the ψ_{mk} 's satisfy the same homogeneous boundary conditions as $\underline{\Phi}$.

Expand the solution vector $\underline{\Phi}(\underline{r}, t)$ into a finite series of the form

$$\underline{\Phi}(\underline{r}, t) \approx \sum_{m=1}^M \sum_{k=1}^K T_{mk}(t) \psi_{mk}(\underline{r}) . \quad (21)$$

The use of double subscript notation is prompted by an important property of the solutions of Equation 20. This property can be stated as follows (37): For a reactor described by P delayed precursor groups and G neutron energy groups, the eigenvectors (natural modes) of Equation 20 come in sets of K members. Each of the K eigenvectors of the m^{th} set has components of similar, but not necessarily identical, spatial shape. In general, the eigenvectors become more oscillatory in space as the set index m increases. Thus the m^{th} set of eigenvectors may be considered as representing the m^{th} spatial harmonic.

In uniform reactors each of the K eigenvectors of the m^{th} set has components of identical shape; and all of these components become more oscillatory in space as the index m

increases. In nonuniform reactors the classification of the natural modes into sets is not as straightforward as in the uniform reactor. In general it is still possible to make the classifications, but ambiguities arise because of appreciable differences in the spatial shapes of the eigenvectors of a set. Such differences were noted by Betancourt (38) when applying the NMA to a coupled core reactor using a two energy group, one delayed group model.

For the optimal control problems considered in this study a one-dimensional, bare, homogeneous reactor model is utilized. It is worthwhile to discuss in some detail the characteristics of the natural modes of such a reactor configuration.

From the First Fundamental Theorem of Reactor Theory (39), which is valid for a bare, homogeneous reactor, it can be concluded that the first G components of each of the eigenvectors of Equation 20 are solutions of the wave equation

$$\nabla^2 \psi_{mk}^g(r) + B^2 \psi_{mk}^g(r) = 0 \quad (22)$$

where

$$\psi_{mk} = \text{col}[\psi_{mk}^1, \psi_{mk}^2, \dots, \psi_{mk}^G, C_{mk}^1, C_{mk}^2, \dots, C_{mk}^P] .$$

The basic assumption underlying this theorem is that the neutron flux, for neutrons of all energy, vanishes at a fixed boundary. For a large bare reactor the neutron flux

can be assumed to vanish at the boundaries of the reactor core. Equation 22 follows directly from the previous statement that the eigenvectors satisfy the same homogeneous boundary conditions as Φ .

The vanishing of ψ_{mk} at the boundaries determines both B^2 and ψ_{mk} , except that ψ_{mk} can be multiplied by any positive constant. For positive B^2 the general solution of Equation 22 is

$$\psi_{mk}^g = e_k^g \sin(Br) + h_k^g \cos(Br) \quad g = 1, G .$$

For a one-dimensional reactor of width a , the boundary conditions are

$$\psi_{mk}^g(0) = 0 , \quad \psi_{mk}^g(a) = 0 \quad g = 1, G .$$

Satisfaction of the first condition requires that h_k^g equals zero. In order to satisfy the second condition it is required that

$$\sin(Ba) = 0$$

or

$$B_m = \frac{m\pi}{a} \quad m = 1, 2, \dots$$

and

$$\psi_{mk}^g = e_k^g \sin\left(\frac{m\pi r}{a}\right) \quad g = 1, G . \quad (23)$$

The subscript m can now be identified with the m^{th} spatial harmonic of the eigenvectors. Since the spatial distribution of the delayed neutron precursors is proportional to the neutron flux, the eigenvector $\psi_{mk}(r)$ can be represented as the product of an amplitude vector \underline{e}_{mk} , and a space-dependent scalar function $S_m(r)$.

$$\begin{aligned} \psi_{mk}(r) &= \text{col}[\psi_{mk}^1, \psi_{mk}^2, \dots, \psi_{mk}^K] \\ &= \underline{e}_{mk} S_m(r) \\ &= S_m(r) \begin{bmatrix} e_{mk}^1 \\ e_{mk}^2 \\ \cdot \\ \cdot \\ \cdot \\ e_{mk}^K \end{bmatrix} \end{aligned} \quad (24)$$

where

$$S_m(r) = \sin\left(\frac{m\pi r}{a}\right) .$$

For each mode specified by m and each eigenvalue specified by index k , there is an eigenvector \underline{e}_{mk} with K elements. The space function $S_m(r)$ depends only on the mode and is independent of the eigenvalues corresponding to that mode.

Substitution of Equation 24 into Equation 20 yields

$$L_{o-mk} e_{mk} S_m(r) = \omega_{mk} e_{mk} S_m(r) \quad . \quad (25)$$

Canceling the factor $S_m(r)$ from both sides of Equation 25 results in the conventional eigenvalue problem

$$L_{o-mk} e_{mk} = \omega_{mk} e_{mk} \quad (26)$$

where, for the bare slab reactor, the ∇^2 operator is replaced with $-B_m^2$ in the L_o matrix.

Thus the L_o matrix yields the eigenvalues, ω_{mk} ; and the corresponding eigenvectors represent the relative amplitudes of the components of the space functions $\psi_{mk}(r)$.

It is now necessary to determine relationships for the time dependent coefficients $T_{mk}(t)$. This is accomplished by application of an orthogonality relationship.

The eigenvectors of Equation 20 and those of the adjoint equation

$$L_{o-nj}^T \psi_{nj}^*(r) = \omega_{nj} \psi_{nj}^*(r) \quad (27)$$

where L_o^T is the transpose of L_o , have a very useful orthogonality property if ψ_{mk} and ψ_{nj} satisfy the same homogeneous boundary conditions. This orthogonality property is (40)

$$\langle \psi_{nj}^*, \psi_{mk} \rangle = 0 \quad \omega_{mk} \neq \omega_{nj} \quad (28)$$

where

$\langle \underline{X}, \underline{Y} \rangle =$ inner product

$$= \int_{\text{reactor}} \underline{X}^T \underline{Y} \, dr . \quad (29)$$

Basic to this result is the assumption that the eigenvalues, ω_{mk} , are distinct. It will also be assumed that

$$\langle \psi_{mk}^*, \psi_{mk} \rangle \neq 0 \quad (30)$$

which implies that the eigenvalues ω_{nj} are the same as ω_{mk} .

Substituting Equation 21, for $\underline{\Phi}(r,t)$, into Equation 18 yields

$$\frac{\partial}{\partial t} \sum_{m=1}^M \sum_{k=1}^K T_{mk}(t) \psi_{mk}(r) = L \sum_{m=1}^M \sum_{k=1}^K T_{mk}(t) \psi_{mk}(r) . \quad (31)$$

Now multiply this equation by ψ_{nj}^* and integrate over the reactor to obtain on the L.H.S.

$$\int_{\text{reactor}} \left\{ \frac{\partial}{\partial t} \sum_{m=1}^M \sum_{k=1}^K T_{mk}(t) \psi_{nj}^* \psi_{mk} \right\} dr . \quad (32)$$

Since the integration is over space, and not time, this can be rewritten as

$$\begin{aligned} & \frac{\partial}{\partial t} \sum_{m=1}^M \sum_{k=1}^K T_{mk}(t) \int_{\text{reactor}} \psi_{nj}^* \psi_{mk} \, dr \\ &= \frac{\partial}{\partial t} \sum_m \sum_k T_{mk}(t) \langle \psi_{nj}^*, \psi_{mk} \rangle . \end{aligned} \quad (33)$$

From the orthogonality relationship of Equation 28 one can see that the inner product term in the double summation will be zero except when $m = n$ and $k = j$; thus the L.H.S. of Equation 31 becomes

$$\frac{\partial}{\partial t} T_{nj}(t) \langle \psi_{nj}^*, \psi_{nj} \rangle \quad (34)$$

Now on the R.H.S. of Equation 31 add and subtract

$$\sum_{m=1}^M \sum_{k=1}^K L_0 T_{mk} \psi_{mk}$$

before multiplying by ψ_{nj}^* and integrating. That is, first write

$$\begin{aligned} \sum_m \sum_k L T_{mk}(t) \psi_{mk}(r) &= \sum_m \sum_k (L - L_0) T_{mk}(t) \psi_{mk} \\ &+ \sum_m \sum_k L_0 T_{mk} \psi_{mk} \quad (35) \end{aligned}$$

Then multiply by ψ_{nj}^* and integrate over the reactor to obtain

$$\sum_m \sum_k \langle \psi_{nj}^* (L - L_0) T_{mk}(t) \psi_{mk} \rangle + \sum_m \sum_k \langle \psi_{nj}^*, L_0 T_{mk} \psi_{mk} \rangle \quad (36)$$

The second term can be simplified by noting that L_0 does not operate on $T_{mk}(t)$, and from Equation 20 that $L_0 \psi_{mk} =$

$\omega_{mk} \psi_{mk}$. Thus

$$\begin{aligned}
& \sum \sum \langle \psi_{nj}^*, L_0 T_{mk} \psi_{mk} \rangle \\
&= \sum \sum \langle \psi_{nj}^*, L_0 \psi_{mk} \rangle T_{mk} \\
&= \sum \sum \langle \psi_{nj}^*, \omega_{mk} \psi_{mk} \rangle T_{mk} \\
&= \sum \sum \omega_{mk} \langle \psi_{nj}^*, \psi_{mk} \rangle T_{mk} \quad . \quad (37)
\end{aligned}$$

Applying the orthogonality property once more leaves simply

$$\omega_{nj} \langle \psi_{nj}^*, \psi_{nj} \rangle T_{nj} \quad (38)$$

as the second term in Equation 36. Combining the results of Equations 34, 36, and 38 it is seen that the relationship of Equation 31 has become

$$\begin{aligned}
\frac{d}{dt} T_{nj} \langle \psi_{nj}^*, \psi_{nj} \rangle &= \omega_{nj} \langle \psi_{nj}^*, \psi_{nj} \rangle T_{nj} \\
&+ \sum \sum \langle \psi_{nj}^*, (L-L_0) T_{mk} \psi_{mk} \rangle \quad (39)
\end{aligned}$$

or

$$\begin{aligned}
\frac{d}{dt} T_{nj} &= \omega_{nj} T_{nj} \\
&+ \sum_m \sum_k \langle \psi_{nj}^*, (L-L_0) \psi_{mk} \rangle T_{mk} / \langle \psi_{nj}^*, \psi_{nj} \rangle \quad . \quad (40)
\end{aligned}$$

If this is repeated for every ψ_{mk}^* , a coupled set of differential equations for the time coefficients is obtained. The system of equations can be written in the form

$$\frac{dT}{dt} = \text{diag}[\omega]T + PT \quad (41)$$

where

$$\underline{T}(t) = \text{col}[T_{11}, T_{12}, \dots, T_{MK}]$$

$$\text{diag}[\omega] = \begin{bmatrix} \omega_{11} & & & 0 \\ & \omega_{12} & & \\ & & \ddots & \\ 0 & & & \omega_{mk} \end{bmatrix}$$

and P is the perturbation matrix

$$P = \begin{bmatrix} \frac{\langle \psi_{11}^*, (L-L_0)\psi_{11} \rangle}{\langle \psi_{11}^*, \psi_{11} \rangle} & \frac{\langle \psi_{11}^*, (L-L_0)\psi_{MK} \rangle}{\langle \psi_{11}^*, \psi_{11} \rangle} \\ \cdot & \cdot \\ \cdot & \cdot \\ \frac{\langle \psi_{MK}^*, (L-L_0)\psi_{11} \rangle}{\langle \psi_{MK}^*, \psi_{MK} \rangle} & \frac{\langle \psi_{MK}^*, (L-L_0)\psi_{MK} \rangle}{\langle \psi_{MK}^*, \psi_{MK} \rangle} \\ \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$$

B. Conjugate Gradient Method

The conjugate gradient method is a member of the family of techniques known as conjugate direction methods which combine the computational simplicity of the gradient techniques with the rapid convergence properties of second-order

techniques. The improved directions of search result from the assumption that the performance index can be approximated by a quadratic expression in the neighborhood of the current search point. The properties of the quadratic function are used implicitly in the derivation of the methods to produce directions of search that are superior to the negative gradient directions.

The conjugate gradient (CG) technique was first developed by Hestenes and Stiefel (41) as a means of solving a system of linear algebraic equations. In 1964, the technique was used by Fletcher and Reeves (42) to minimize a function of several variables. The first extension of the method to a function space was presented by Hayes (43) in 1954. Other treatments of the extension have been given by Sinnott and Luenberger (44), Lasdon et al. (45), and Tripathi and Narendra (46). The treatment of the CG method presented here follows closely that given by Willoughby (2).

The theoretical basis and the computational procedure of the CG method are most readily understood by first examining the finite-dimensional version from which the function space extensions have been derived. Thus, initially the method is discussed in the context of minimizing a function f of n real variables which are elements of a real Euclidean vector space E_n . It is assumed for simplicity that only one minimum of f exists over E_n , with the minimum denoted by $f(\underline{x}^*)$, i.e.,

$$f(\underline{x}^*) \leq f(\underline{x}) \quad \forall \underline{x} \in E_n . \quad (42)$$

The solution procedure involves choosing a new trial vector \underline{x}_{i+1} using the relation

$$\underline{x}_{i+1} = \underline{x}_i + \alpha_i \underline{s}_i \quad (43)$$

where the subscript i represents the iteration number, α_i is a scalar called the stepsize, and \underline{s}_i is an n -vector called the direction of search. Specifically, the CG procedure is as follows:

1. For $i = 0$, guess an initial state vector \underline{x}_0 .
2. Calculate the gradient vector \underline{g}_i at \underline{x}_i

$$\underline{g}_i = \underline{g}(\underline{x}_i) = \nabla f(\underline{x}_i) . \quad (44)$$

3. Calculate the CG parameter β_i

$$\beta_i = \frac{(\underline{g}_i, N\underline{s}_{i-1})}{(\underline{s}_{i-1}, N\underline{s}_{i-1})} . \quad (45)$$

$(\underline{y}, \underline{z})$ denotes the Euclidean inner product defined to be

$$(\underline{y}, \underline{z}) = \sum_{j=1}^n y_j z_j = \underline{y}^T \underline{z} , \quad (46)$$

and N is the Hessian matrix defined by

$$\left[\frac{\partial^2 f}{\partial \underline{x}^2} \right]_{\underline{x}=\underline{x}_i} .$$

If $i = 0$, $\beta_0 = 0$.

4. Calculate the direction of search \underline{s}_i

$$\underline{s}_i = -\underline{g}_i + \beta_i \underline{s}_{i-1} \quad . \quad (47)$$

5. Perform a one-dimensional minimization to determine \underline{x}_{i+1} , i.e.

$$\underline{x}_{i+1} = \underline{x}_i + \alpha_i \underline{s}_i \quad (48)$$

where α_i is such that

$$f(\underline{x}_i + \alpha_i \underline{s}_i) \leq f(\underline{x}_i + \gamma \underline{s}_i) \quad \forall \gamma > 0 \quad .$$

6. Increase i and repeat from step 2 until the minimum is reached.

The above procedure is quadratically convergent meaning that it will find the minimum of any quadratic function in a finite number of steps. In particular, the CG method will minimize a quadratic function of n variables in at most n steps (45).

The derivation of the method requires the notion of conjugacy between vectors. Two vectors \underline{y} and \underline{w} are said to be conjugate, N -conjugate, or N -orthogonal with respect to the matrix N if

$$(\underline{y}, N\underline{w}) = 0 \quad . \quad (49)$$

If the objective function is quadratic so that

$$f(\underline{x}) = f(\underline{x}^*) + \frac{1}{2} ((\underline{x} - \underline{x}^*), N(\underline{x} - \underline{x}^*)) \quad (50)$$

and N is a positive definite matrix with constant elements corresponding to the second partial derivatives of f , then the directions of search given by Equation 47 form a mutually conjugate set with respect to N , i.e.

$$(\underline{s}_i, N\underline{s}_j) = 0, \quad i \neq j \quad . \quad (51)$$

It follows that the \underline{s}_i are linearly independent vectors which span E_n . Therefore

$$\underline{x}^* = \sum_{k=0}^{n-1} c_k \underline{s}_k \quad . \quad (52)$$

The objective is to determine the coefficients c_k in Equation 52. Forming the inner product $(N\underline{x}^*, \underline{s}_k)$,

$$(N\underline{x}^*, \underline{s}_k) = c_k (\underline{s}_k, N\underline{s}_k) \quad k = 0, 1, 2, \dots, n-1 \quad (53)$$

since Equation 51 eliminates all the terms with mixed subscripts. Thus

$$c_k = \frac{(N\underline{x}^*, \underline{s}_k)}{(\underline{s}_k, N\underline{s}_k)} \quad (54)$$

but

$$\nabla f(\underline{x}_i) = N(\underline{x}_i - \underline{x}^*) = \underline{g}_i \quad (55)$$

so that

$$c_k = \frac{(N\underline{x}_i, \underline{s}_k) - (\underline{g}_i, \underline{s}_k)}{(\underline{s}_k, N\underline{s}_k)} \quad . \quad (56)$$

Using Equation 48 repetitively results in

$$\underline{x}_i = \underline{x}_j + \sum_{k=j}^{i-1} \alpha_k \underline{s}_k, \quad 0 \leq j \leq i. \quad (57)$$

From Equations 55 and 57,

$$\begin{aligned} \underline{g}_i &= N\underline{x}_j + \sum_{k=j}^{i-1} N\alpha_k \underline{s}_k - N\underline{x}^* \\ &= \underline{g}_j + \sum_{k=j}^{i-1} N\alpha_k \underline{s}_k, \quad 0 \leq j \leq i. \end{aligned} \quad (58)$$

Also

$$(\underline{g}_j, \underline{s}_{j-1}) = 0, \quad (59)$$

as a result of the one-dimensional minimization in Equation 48. Therefore, if the inner product $(\underline{g}_i, \underline{s}_{j-1})$ is calculated from Equation 58,

$$(\underline{g}_i, \underline{s}_{j-1}) = (\underline{g}_j, \underline{s}_{j-1}) + \sum_{k=j}^{i-1} \alpha_k (N\underline{s}_k, \underline{s}_{j-1}) \quad 1 \leq j \leq i. \quad (60)$$

The first term is zero from Equation 59 and the last term is zero from Equation 51. Therefore

$$(\underline{g}_i, \underline{s}_{j-1}) = 0 \quad 1 \leq j \leq i \quad (61)$$

i.e., the gradient at the i^{th} iteration of the search is orthogonal to all the previous directions of search. Returning to Equation 56, if $k = i-1$ then

$$c_i = \frac{(N\underline{x}_{i+1}, \underline{s}_i)}{(\underline{s}_i, N\underline{s}_i)} = \frac{(N(\underline{x}_i + \alpha_i \underline{s}_i), \underline{s}_i)}{(\underline{s}_i, N\underline{s}_i)} \quad i = 1, 2, \dots, n-1 \quad .$$

(62)

Thus if \underline{s}_i is chosen N-conjugate to all previous directions of search, and the stepsize α_i is found using a one-dimensional minimization, the value of c_i is determined from Equation 62. It is clear that after at most n steps, all of the coefficients in Equation 52 are determined and the minimum is located. For nonquadratic functions, the rate of convergence of the method depends upon the nature of f , and the location of \underline{x} .

The extension of the CG method to function space is best illustrated by considering the simplified optimization problem presented in Chapter II. The problem is stated as follows:

$$\text{minimize } J = S(\underline{x}(t_f)) \quad (63)$$

$$\text{subject to } \dot{\underline{x}} = \underline{f}(\underline{x}, u, t) \quad (64)$$

$$\underline{x}(t_0) = \underline{x}_0 \quad (65)$$

where \underline{x} is an n -vector, u is a single control function, and t_0, t_f are fixed. It is assumed that given a control, u , Equations 64 and 65 can be solved for a unique $\underline{x} = \underline{x}(u)$, and thus $J = J(u)$ is a function of u alone.

The important connection between the function space extension and the finite-dimensional analysis is obtained

by examining the expression

$$g(u) = \frac{\partial H}{\partial u} . \quad (66)$$

As discussed in Chapter II, $g(u)$ is the gradient to the Hamiltonian and points in the direction of increasing J .

The first variation in J is given by

$$\delta J = \frac{\partial S}{\partial \underline{x}} \Big|_{t_f} \delta \underline{x}_f . \quad (67)$$

The notation δJ represents the first-order approximation to $J(u) - J(\hat{u})$ where \hat{u} is a given nominal control function. A new expression for δJ can be formulated using the adjoint system defined by the system of differential equations

$$\dot{\lambda}_i(t) = - \frac{\partial H(\underline{x}, u, t)}{\partial x_i} \quad i = 1, n \quad (68)$$

and the Hamiltonian function

$$H(\underline{x}, u, \underline{\lambda}, t) = \underline{\lambda}^T \underline{f} . \quad (69)$$

The first variation of $\dot{\underline{x}}$ is given by

$$\delta \dot{\underline{x}}_i = \sum_{j=1}^n \frac{\partial f_i}{\partial x_j} \delta x_j + \frac{\partial f_i}{\partial u} \delta u . \quad (70)$$

Form the expression

$$\frac{d}{dt} \sum_{i=1}^n \lambda_i \delta x_i = \sum_{i=1}^n \dot{\lambda}_i \delta x_i + \sum_{i=1}^n \lambda_i \delta \dot{x}_i \quad (71)$$

and evaluate it using Equations 68, 69, and 70. Integrating the result with respect to time with $\underline{x}(t_0) = 0$ gives

$$\sum_{i=1}^n \lambda_i(t_f) \delta x(t_f) = \int_{t_0}^{t_f} \left[\frac{\partial H}{\partial u} \delta u \right] dt \quad . \quad (72)$$

Using Equation 11, Equation 72 can be written as

$$\delta J = \frac{\partial S}{\partial \underline{x}} \Big|_{t_f} \delta \underline{x}_f = \int_{t_0}^{t_f} \left[\frac{\partial H}{\partial u} \delta u \right] dt \quad . \quad (73)$$

If the variation of the control u is along a direction of search s , i.e.,

$$\delta u = s \delta \alpha \quad (74)$$

where α is a scalar, then the derivative of J along s is

$$\frac{dJ}{d\alpha} = \int_{t_0}^{t_f} \left[\frac{\partial H}{\partial u} s \right] dt \quad . \quad (75)$$

Equation 75 is the inner product of the direction of search and $\frac{\partial H}{\partial u}$. A comparison of Equation 75 with the comparable expression in the finite-dimensional case indicates that $\frac{\partial H}{\partial u}$ is analogous to the gradient vector in the finite-dimensional analysis.

In order to apply the CG method to the general functional, J , in Equation 63, it is necessary to approximate J by a quadratic functional in the neighborhood of the

current search point. Thus

$$J(u) = 1/2(u - \hat{u}, A(u - \hat{u})) + J(\hat{u}) \quad (76)$$

where A is a positive definite linear operator analogous to the N-matrix in the finite-dimensional analysis.

The CG algorithm for the function space extension is as follows:

1. For $i = 0$ guess an initial control function $u_0(t)$.
2. Integrate the state system, Equations 64 and 65, from t_0 to t_f .
3. Integrate the costate system, Equations 10 and 11, from t_f to t_0 .
4. Calculate

$$\frac{\partial H}{\partial u} = g[u_i(t)] = g_i \quad (77)$$

5. Calculate β_i using

$$\beta_i = \frac{(g_i, As_{i-1})}{(s_{i-1}, As_{i-1})} \quad (78)$$

If $i = 0$, $\beta_0 = 0$.

6. Calculate the direction of search

$$s_i(t) = -g_i(t) + \beta_i s_{i-1}(t) \quad (79)$$

7. Let

$$u_{i+1}(t) = u_i(t) + \alpha_i s_i(t) \quad (80)$$

and determine α_i by performing a one-dimensional minimization, i.e.

$$J(u_i + \alpha_i s_i) \leq J(u_i + \gamma s_i) \quad \forall \gamma > 0 \quad .$$

8. Increase i and repeat from step 2 until the minimum is reached.

The principal difficulty encountered in applying this algorithm is the calculation of the matrix A . Although the A -matrix can be identified as the second-order expansion of the performance index J , it is usually impractical to calculate the matrix directly. In practice, it is easier to calculate the functional As_{i-1} . Sinnott and Luenberger (44) and Tripathi and Narendra (46) have derived methods for calculating As_{i-1} which necessitate integrating two sets of auxiliary equations. For the optimal control problems of interest here the auxiliary equations are:

$$\dot{\underline{y}}(t) = \underline{f}_x(t)\underline{y}(t) + \underline{f}_u(t)s(t) \quad (81)$$

$$\underline{y}(t_0) = 0 \quad (82)$$

and

$$\dot{\underline{\eta}}(t) = -\underline{f}_x^T \underline{\eta}(t) \quad (83)$$

$$\underline{\eta}(t_f) = J_{\underline{x}_f \underline{x}_f} \underline{y}(t_f) \quad (84)$$

where $\underline{y}(t)$ and $\underline{\eta}(t)$ are n -vectors and

$$\underline{f}_x = \begin{bmatrix} \partial f_1 / \partial x_1 & \partial f_1 / \partial x_2 & \dots & \partial f_1 / \partial x_n \\ \partial f_2 / \partial x_1 & & & \vdots \\ \partial f_n / \partial x_1 & & & \partial f_n / \partial x_n \end{bmatrix}$$

$$\underline{f}_u = \text{col}[\partial f_1/\partial u, \partial f_2/\partial u, \dots, \partial f_n/\partial u]$$

$$\underline{J}_{\underline{xx}} = \begin{bmatrix} \partial J/\partial x_1 \partial x_1 & \partial J/\partial x_1 \partial x_2 & \dots & \partial J/\partial x_1 \partial x_n \\ \partial J/\partial x_2 \partial x_1 & & & \\ \vdots & & & \\ \partial J/\partial x_n \partial x_1 & \dots\dots\dots & & \partial J/\partial x_n \partial x_n \end{bmatrix} .$$

A derivation of these equations is given by Willoughby (2).

The function As_{i-1} is then given by the expression

$$As_{i-1} = \underline{f}_u(t) \underline{\eta}(t) \quad . \quad (85)$$

IV. MATHEMATICAL MODEL AND ANALYSIS

A. Reactor Model

The reactor model for this investigation is a bare homogeneous slab reactor (see Figure 2). The reactor parameters are typical of a large uniform, light-water-moderated assembly.

The two energy group one delayed neutron group diffusion theory equations used to describe the reactor dynamics are:

$$\begin{aligned} D_1 \nabla^2 \phi_1 - \Sigma_1 \phi_1 - D_1 B_T^2 \phi_1 + (1-\beta)[v_1 \Sigma_{f1} \phi_1 + v_2 \Sigma_{f2} \phi_2] + \lambda C \\ = 1/v_1 \partial \phi_1 / \partial t \end{aligned} \quad (86)$$

$$\begin{aligned} D_2 \nabla^2 \phi_2 - \Sigma_a^0 \phi_2 + U(r, t) - D_2 B_T^2 \phi_2 + \Sigma_{R1} \phi_1 \\ = 1/v_2 \partial \phi_2 / \partial t \end{aligned} \quad (87)$$

$$\beta [v_1 \Sigma_{f1} \phi_1 + v_2 \Sigma_{f2} \phi_2] - \lambda C = \partial C / \partial t \quad . \quad (88)$$

The nuclear parameters are defined as:

ϕ_i = group i neutron flux

D_i = diffusion coefficient for group i

Σ_1 = macroscopic removal plus absorption cross section for group 1

Σ_{R1} = macroscopic removal cross section for group 1

Σ_{a2}^0 = initial macroscopic absorption cross section
for group 2

Σ_{fi} = group i macroscopic fission cross section

ν_i = average number of neutrons produced per fission
occurring in group i

v_i = group i neutron velocity

B_T^2 = transverse buckling

β = effective delayed neutron fraction

λ = delayed neutron group decay constant

C = concentration of delayed neutron precursors

$U(r,t)$ = control function $\equiv [\Sigma_{a2}^0 - \Sigma_{a2}(t)]\phi_2$
 $= \Delta\Sigma_{a2}(r,t)\phi_2$

∇^2 = Laplacian operator.

The subscripts $i = 1$ and 2 refer to the fast and thermal group respectively.

The steady state or critical reactor parameters are given in Table 1.

B. Modal Analysis

For the problems considered in this dissertation, the rate of change of the neutron flux is relatively slow compared to that of an excursion. Thus, it is a good approximation to assume that

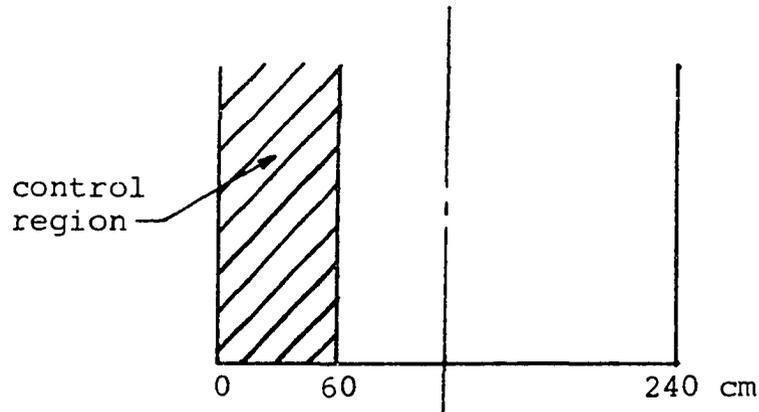


Figure 2. Reactor configuration

Table 1. Steady state reactor parameters

Parameter	Units	Value
D_1	cm	1.69531
D_2	cm	0.40972
Σ_1	cm^{-1}	0.049959
Σ_{R1}	cm^{-1}	0.016444
Σ_{a2}^0	cm^{-1}	0.26614
$v_1 \Sigma_{f1}$	cm^{-1}	0.019496
$v_2 \Sigma_{f2}$	cm^{-1}	0.49786
v_1	cm/sec	$4.06 (10)^6$
v_2	cm/sec	$2.20 (10)^5$
B_T^2	cm^{-2}	0.009440
β		0.0064
λ		0.08
a	cm	240.0

$$\frac{1}{v_1} \frac{\partial \phi_1}{\partial t} = 0 \quad \frac{1}{v_2} \frac{\partial \phi_2}{\partial t} = 0 \quad . \quad (89)$$

The two group diffusion equations can then be written as

$$D_1 \nabla^2 \phi_1 - \Sigma_1 \phi_1 - D_1 B_T^2 \phi_1 + (1-\beta)[v_1 \Sigma_{f1} \phi_1 + v_2 \Sigma_{f2} \phi_2] + \lambda C = 0 \quad (90)$$

$$D_2 \nabla^2 \phi_2 - \Sigma_{a2}^0 \phi_2 + U(r,t) - D_2 B_T^2 \phi_2 + \Sigma_{R1} \phi_1 = 0 \quad (91)$$

$$\beta[v_1 \Sigma_{f1} \phi_1 + v_2 \Sigma_{f2} \phi_2] - \lambda C = \partial C / \partial t \quad . \quad (92)$$

As discussed previously, the spatial distribution of the neutron flux in a bare reactor is a solution of the wave equation

$$\nabla^2 \phi_g + B_m^2 \phi_g = 0 \quad (93)$$

where

$$B_m^2 = \left(\frac{m\pi}{a}\right)^2$$

for a slab reactor of width a . Making the substitution

$\nabla^2 \phi_g = -B_m^2 \phi_g$ in Equations 90, 91, and 92 gives

$$[-D_1 B_m^2 - \Sigma_1 - D_1 B_T^2 + (1-\beta)v_1 \Sigma_{f1}] \phi_1 + (1-\beta)v_2 \Sigma_{f2} \phi_2 + \lambda C = 0 \quad (94)$$

$$[-D_2 B_m^2 - \Sigma_{a2}^0 - D_2 B_T^2] \phi_2 + \Sigma_{R1} \phi_1 + U(r,t) = 0 \quad (95)$$

$$\beta[v_1 \Sigma_{f1} \phi_1 + v_2 \Sigma_{f2} \phi_2] - \lambda C = \partial C / \partial t \quad . \quad (96)$$

Using the expression $U(r,t) = \Delta \Sigma_{a2}(t) \phi_2$, Equations 94 and 95 can be solved for ϕ_1 and ϕ_2 in terms of the precursor concentration C to give

$$\phi_1 = (Q_m S_m + R_m) C \quad (97)$$

$$\phi_2 = S_m C \quad (98)$$

where

$$Q_m = \frac{-(1-\beta) v_2 \Sigma_{f2}}{D_1 B_m^2 - \Sigma_1 - D_1 B_T^2 + (1-\beta) v_1 \Sigma_{f1}}$$

$$R_m = \frac{-\lambda}{D_1 B_m^2 - \Sigma_1 - D_1 B_T^2 + (1-\beta) v_1 \Sigma_{f1}}$$

$$S_m = \frac{\Sigma_{R1} R_m}{D_2 B_m^2 - \Sigma_{a2}^0 - D_2 B_T^2 + \Sigma_{R1} Q_m + \Delta \Sigma_{a2}(r,t)}$$

Substituting Equations 97 and 98 for ϕ_1 and ϕ_2 allows Equation 96 to be written as

$$L_m C = \frac{\partial C}{\partial t} \quad (99)$$

where L_m is a scalar function dependent on the mode and the change in the thermal absorption cross section $\Delta \Sigma_{a2}(r,t)$.

The steady state scalar operator L_0 is formed by setting $\Delta \Sigma_{a2}(r,t)$ equal to zero in the expression for L_m .

Utilizing the Natural Mode Approximation, $C(r,t)$ is expressed as

$$C(r,t) = \sum_{m=1}^M C_m(t)\psi_m(r) \quad (100)$$

where the $\psi_m(r)$ are the eigenfunctions of the eigenvalue problem

$$L_0 \psi_m = \omega_m \psi_m \quad (101)$$

It should be noted that double subscript notation is not needed because there is only one eigenvalue for each mode. From the discussion in Chapter III it is readily seen that the eigenfunctions are given by

$$\psi_m = \sin(B_m r) \quad (102)$$

where

$$B_m = \frac{m\pi}{a} \quad .$$

The eigenvalues ω_m for the first three modes of the reactor model are listed in Table 2.

At this point it is necessary to depart from the standard method of obtaining equations for the time coefficients as given in Chapter III. Such a departure is necessary in order to formulate the problem in terms of a trajectory optimization problem. Since ϕ_1 and ϕ_2 were solved for in terms of $C(r,t)$ it is reasonable to approximate the group

fluxes by

$$\phi_1(r, t) = \sum_{m=1}^M A_m^{(1)}(t) \psi_m(r) \quad (103)$$

$$\phi_2(r, t) = \sum_{m=1}^M A_m^{(2)}(t) \psi_m(r) \quad (104)$$

where $A_m^{(1)}(t)$ and $A_m^{(2)}(t)$ are the time dependent coefficients of combination. Expressions for these coefficients and those of Equation 100 are obtained by substituting Equations 100, 103, and 104 into Equations 94, 95, and 96 and using the orthogonality property of the sine functions

$$\int_{\text{reactor}} \sin(B_m r) \sin(B_n r) \, dr = 0 \quad m \neq n$$

or

$$\langle \psi_m(r), \psi_n(r) \rangle = 0 \quad m \neq n \quad . \quad (105)$$

The resulting expressions for $A_m^{(2)}$ and C_m can be written as

$$\frac{dC_m}{dt} = \omega_m C_m + J_m \frac{\langle \psi_m, U(r, t) \rangle}{\langle \psi_m, \psi_m \rangle} \quad (106)$$

$$0 = V_m A_m^{(2)} - W_m C_m - J_m \frac{\langle \psi_m, U(r, t) \rangle}{\langle \psi_m, \psi_m \rangle} \quad (107)$$

where V_m and W_m are functions of the reactor parameters and the mode. An expression for the $A_m^{(1)}$ is not needed at this

time.

The motivation for this approach has been to isolate the control function $U(r,t)$ in a single term. The standard derivation of the equations for the time coefficients would have yielded expressions with $\Delta\Sigma_{a2}(t)$ as one of a sum of terms in a denominator. A practical formulation of a trajectory optimization problem is not feasible in such a case.

Having isolated $U(r,t)$, it is now possible to simplify the term in which it appears. Substituting Equation 104 for $\phi_2(r,t)$, $U(r,t)$ can be written as

$$U(r,t) = \Delta\Sigma_{a2}(r,t) \sum_{m=1}^M A_m^{(2)} \psi_m \quad . \quad (108)$$

Assuming that the thermal absorption cross section is changed uniformly over a specified region of the core, one can write

$$\langle \psi_m, U \rangle = \Delta\Sigma_{a2}(t) \left[\langle \psi_m, \sum_{n=1}^M A_n^{(2)} \psi_n \rangle \right] \quad . \quad (109)$$

This can be further simplified by noting that

$$\begin{aligned} \langle \psi_m, \sum_{n=1}^M A_n^{(2)} \psi_n \rangle &= \langle \psi_m, \psi_1 \rangle A_1^{(2)} + \langle \psi_m, \psi_2 \rangle A_2^{(2)} \\ &+ \dots + \langle \psi_m, \psi_M \rangle A_M^{(2)} \quad . \end{aligned} \quad (110)$$

Now define

$$P_{mn} = -J \frac{\langle \psi_m, \psi_n \rangle}{m \langle \psi_m, \psi_m \rangle} \quad (111)$$

Assuming that three modes are used in the expansion, Equations 106 and 107 can be written as

$$\dot{\underline{c}} = \begin{bmatrix} \omega_1 & & \\ & \omega_2 & \\ & & \omega_3 \end{bmatrix} \underline{c} + \Delta \Sigma_{a2}(t) P \underline{A}^{(2)} \quad (112)$$

$$\begin{bmatrix} v_1 & & \\ & v_2 & \\ & & v_3 \end{bmatrix} \underline{A}^{(2)} - \begin{bmatrix} w_1 & & \\ & w_2 & \\ & & w_3 \end{bmatrix} \underline{c} - \Delta \Sigma_{a2}(t) P \underline{A}^{(2)} = 0 \quad (113)$$

where

$$\underline{c} = \text{col}[c_1, c_2, c_3]$$

$$\underline{A}^{(2)} = \text{col}[A_1^{(2)}, A_2^{(2)}, A_3^{(2)}]$$

$$P = \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix} \quad .$$

Equations 112 and 113 in conjunction with the expansions of Equations 100 and 104 describe the space and time dependent response of the thermal neutron flux and the delayed precursor concentration to a variation in the thermal absorption cross section in a region of the core.

C. Formulation of the Optimal Control Problem

1. Problem statement

The problem to be studied can be stated as follows: Determine the variation of the thermal absorption cross section $\Sigma_{a2}(t)$, in the control region (see Figure 2), for a fixed time interval $[t_0, t_f]$ which results in

1. a doubling of the thermal neutron flux in the center of the reactor,
2. a steady-state thermal neutron flux at the terminal time t_f ,

and which minimizes the functional

$$\int_{t_0}^{t_f} [\Delta \dot{\Sigma}_{a2}(t)]^2 dt \quad .$$

Assuming that $\Delta \dot{\Sigma}_{a2}$ is proportional to the control rod velocity, minimization of this functional minimizes the energy requirements of the control system (47). The asymmetric control region was chosen so as to accentuate space-time effects in the reactor.

2. State variable formulation

In order to formulate the optimal control problem in the Mayer form, it is necessary to introduce three new state variables. These variables are defined by the following expressions:

$$\dot{x}_6(t) = \dot{\Delta \Sigma}_{a2}(t) \equiv u(t) \quad (114)$$

$$\dot{x}_4(t) = x_6(t) = \Delta \dot{\Sigma}_{a2}(t) \Rightarrow x_4(t) = \Delta \Sigma_{a2}(t) \quad (115)$$

$$\dot{x}_5(t) = [x_6(t)]^2 \quad . \quad (116)$$

The value of $x_5(t)$ at the terminal time is

$$x_5(t_f) = \int_{t_0}^{t_f} [x_6]^2 dt = \int_{t_0}^{t_f} [\Delta \dot{\Sigma}_{a2}(t)]^2 dt \quad . \quad (117)$$

Therefore, minimization of $x_5(t_f)$ results in a minimization of the control energy.

In order to facilitate the analysis, the following new notation is adopted:

$$\begin{aligned} x_1(t) &= C_1(t) & x_7(t) &= A_1^{(2)}(t) \\ x_2(t) &= C_2(t) & x_8(t) &= A_2^{(2)}(t) \\ x_3(t) &= C_3(t) & x_9(t) &= A_3^{(2)}(t) \end{aligned} \quad (118)$$

3. System equations

Using the new notation and Equations 114-116 in conjunction with Equations 112 and 113 yields the following system equations:

$$\begin{aligned}\dot{x}_1 &= f_1(\underline{x}, u, t) \\ &= \omega_1 x_1 + (p_{11} x_7 + p_{12} x_8 + p_{13} x_9) x_4\end{aligned}\quad (119)$$

$$\begin{aligned}\dot{x}_2 &= f_2(\underline{x}, u, t) \\ &= \omega_2 x_2 + (p_{21} x_7 + p_{22} x_8 + p_{23} x_9) x_4\end{aligned}\quad (120)$$

$$\begin{aligned}\dot{x}_3 &= f_3(\underline{x}, u, t) \\ &= \omega_3 x_3 + (p_{31} x_7 + p_{32} x_8 + p_{33} x_9) x_4\end{aligned}\quad (121)$$

$$\dot{x}_4 = f_4(\underline{x}, u, t) = x_6 \quad (122)$$

$$\dot{x}_5 = f_5(\underline{x}, u, t) = [x_6]^2 \quad (123)$$

$$\dot{x}_6 = f_6(\underline{x}, u, t) = u(t) \quad (124)$$

$$\begin{aligned}f_7(\underline{x}, u, t) &= -(p_{11} x_7 + p_{12} x_8 + p_{13} x_9) x_4 \\ &\quad + v_1 x_7 - w_1 x_1 = 0\end{aligned}\quad (125)$$

$$\begin{aligned}f_8(\underline{x}, u, t) &= -(p_{21} x_7 + p_{22} x_8 + p_{23} x_9) x_4 \\ &\quad + v_2 x_8 - w_2 x_2 = 0\end{aligned}\quad (126)$$

$$\begin{aligned}f_9(\underline{x}, u, t) &= -(p_{31} x_7 + p_{32} x_8 + p_{33} x_9) x_4 \\ &\quad + v_3 x_9 - w_3 x_3 = 0\end{aligned}\quad (127)$$

where

$$\underline{x} = \text{col}[x_1, x_2, \dots, x_9] \quad .$$

The calculated values of the constant parameters in the system equations are listed in Table 2.

Table 2. Calculated values for system equation parameters

$\omega_1 = 0$	$V_1 = 0.0052141$	$W_1 = 0.079672$
$\omega_2 = -.05900$	$V_2 = 0.0051577$	$W_2 = 0.020681$
$\omega_3 = -.07058$	$V_3 = 0.0050703$	$W_3 = 0.009110$
$P = \begin{bmatrix} 0.1674 & 0.2766 & 0.2933 \\ 0.07385 & 0.1230 & 0.1551 \\ 0.03611 & 0.07150 & 0.06877 \end{bmatrix}$		

4. Initial conditions

Since $\omega_1 = 0$, and setting $x_6(t_0) = 0$, the following initial conditions can be imposed which are consistent with the requirement that the system is initially at steady-state:

$$\begin{aligned}
 x_1(t_0) &= 1.0 & x_6(t_0) &= 0 \\
 x_2(t_0) &= 0 & x_7(t_0) &= 15.28 \\
 x_3(t_0) &= 0 & x_8(t_0) &= 0 \\
 x_4(t_0) &= 0 & x_9(t_0) &= 0 \\
 x_5(t_0) &= 0 & &
 \end{aligned} \tag{128}$$

5. Performance index

At the final time t_f , the system must satisfy the two constraints imposed by the statement of the problem. Expressions for each of these constraints will be obtained in a form suitable for use in a penalty function approximation.

The first constraint requires the thermal neutron flux at the center of the reactor at the final time to be double its initial value. That is:

$$\phi_2(120, t_f) = 2[\phi_2(120, 0)] \quad (129)$$

or

$$\sum_{m=1}^M A_m^{(2)}(t_f) \psi_m(120) = 2 \sum_{m=1}^M A_m^{(2)}(0) \psi_m(120) \quad (130)$$

Equation 130 can be expanded to obtain

$$x_7(t_f) - x_9(t_f) - 30.56 = 0 \quad (131)$$

In order to use Equation 131 in the computational algorithm it is necessary to rewrite the equation to include the state variable x_4 . Using Equations 125 and 127, Equation 131 can be written in the form

$$\begin{aligned} Z_1 x_1(t_f) - Z_3 x_3(t_f) + (E_1 x_7 + E_2 x_8 + E_3 x_9) x_4 \\ - 30.56 = 0 \end{aligned} \quad (132)$$

where

$$Z_i = W_i/V_i$$

$$E_i = p_{1i}/V_1 - p_{3i}/V_3 \quad .$$

The second constraint requires that the rate of change of the thermal neutron flux be approximately zero at the final time. In order to obtain a suitable expression for this constraint, the time derivative of Equation 125 is taken:

$$\begin{aligned} -\dot{x}_4(p_{11}x_7 + p_{12}x_8 + p_{13}x_9) - x_4p_{11}\dot{x}_7 - x_4p_{12}\dot{x}_8 \\ - x_4p_{13}\dot{x}_9 + v_1\dot{x}_7 - w_1\dot{x}_1 = 0 \quad . \end{aligned} \quad (133)$$

Substituting Equation 119 for \dot{x}_1 , letting $\omega_1 = 0$, and solving for \dot{x}_7 lets Equation 133 be written as

$$\dot{x}_7 = \frac{(p_{11}x_7 + p_{12}x_8 + p_{13}x_9)(x_6 + w_1x_4) + x_4(p_{12}\dot{x}_8 + p_{13}\dot{x}_9)}{(v_1 - x_4p_{11})} \quad . \quad (134)$$

Assuming that \dot{x}_8 and \dot{x}_9 are small compared to \dot{x}_7 , a good approximation is obtained by requiring that

$$x_6(t_f) + w_1x_4(t_f) = 0 \quad . \quad (135)$$

Equation 135 is exactly the expression which would have been obtained using a point kinetics derivation.

The performance index for this optimal control problem can now be constructed using a penalty function approximation. Following the procedure discussed in Chapter II, the

performance index can be written as

$$J = x_5(t_f) + \frac{1}{2} K [(Z_1 x_1 - Z_3 x_3 + x_4 (E_1 x_7 + E_2 x_8 + E_3 x_9) - 30.56)^2 + (x_6 + W_1 x_4)^2] \quad (136)$$

where the K_i of Equation 16 are assumed to be equal.

As discussed in Chapter III, the conjugate gradient algorithm requires the integration of three additional sets of equations in order to determine the gradient to the Hamiltonian g_i and the parameter β_i . These relations will now be discussed.

6. Adjoint equations

The adjoint equations defined by Equation 10 were derived for the system equations of Equation 8. The system equations of the current problem differ from those of Equation 8 due to the presence of the algebraic equations (Equations 125-127). From an examination of the derivation of the adjoint equations it can be shown that the modified adjoint equations for this problem are given by

$$\dot{\lambda}_i = -\partial H / \partial x_i \quad i = 1, 6 \quad (137)$$

$$\partial H / \partial x_i = 0 \quad i = 7, 9 \quad (138)$$

where the Hamiltonian function H is defined as

$$\begin{aligned}
H = & f_1\lambda_1 + f_2\lambda_2 + f_3\lambda_3 + f_4\lambda_4 + f_5\lambda_5 + f_6\lambda_6 + f_7\lambda_7 \\
& + f_8\lambda_8 + f_9\lambda_9 \quad .
\end{aligned} \tag{139}$$

The specific equations for this problem are

$$\dot{\lambda}_1 = -\omega_1\lambda_1 + W_1\lambda_7 \tag{140}$$

$$\dot{\lambda}_2 = -\omega_2\lambda_2 + W_2\lambda_8 \tag{141}$$

$$\dot{\lambda}_3 = -\omega_3\lambda_3 + W_3\lambda_9 \tag{142}$$

$$\begin{aligned}
\dot{\lambda}_4 = & -(p_{11}x_7 + p_{12}x_8 + p_{13}x_9)\lambda_1 - (p_{21}x_7 + p_{22}x_8 + p_{23}x_9)\lambda_2 \\
& - (p_{31}x_7 + p_{32}x_8 + p_{33}x_9)\lambda_3 + (p_{11}x_7 + p_{12}x_8 + p_{13}x_9)\lambda_7 \\
& + (p_{21}x_7 + p_{22}x_8 + p_{23}x_9)\lambda_8 + (p_{31}x_7 + p_{32}x_8 + p_{33}x_9)\lambda_9
\end{aligned} \tag{143}$$

$$\dot{\lambda}_5 = 0 \tag{144}$$

$$\dot{\lambda}_6 = -\lambda_4 - 2x_6\lambda_5 \tag{145}$$

$$\begin{aligned}
& -(p_{11}\lambda_1 + p_{21}\lambda_2 + p_{31}\lambda_3)x_4 + (p_{11}\lambda_7 + p_{21}\lambda_8 + p_{31}\lambda_9)x_4 \\
& - V_1\lambda_7 = 0
\end{aligned} \tag{146}$$

$$\begin{aligned}
& -(p_{12}\lambda_1 + p_{22}\lambda_2 + p_{32}\lambda_3)x_4 + (p_{12}\lambda_7 + p_{22}\lambda_8 + p_{32}\lambda_9)x_4 \\
& - V_2\lambda_8 = 0
\end{aligned} \tag{147}$$

$$\begin{aligned}
 & -(p_{13}\lambda_1 + p_{23}\lambda_2 + p_{33}\lambda_3)x_4 + (p_{13}\lambda_7 + p_{23}\lambda_8 + p_{33}\lambda_9)x_4 \\
 & - v_3\lambda_9 = 0 \quad . \quad (148)
 \end{aligned}$$

The terminal conditions for the first six adjoint variables are calculated by Equation 11 i.e.

$$\lambda_i(t_f) = \left. \partial S / \partial x_i \right|_{t=t_f} \quad i = 1, 6 \quad . \quad (149)$$

Having computed the terminal values of the first six adjoint variables, the values of λ_7 , λ_8 , and λ_9 can be found by solving the system of simultaneous equations consisting of Equations 146, 147, and 148.

7. Gradient to the Hamiltonian function

The gradient to the Hamiltonian, $\partial H / \partial u$, for this problem is given by

$$g = \partial H / \partial u = \lambda_6(t) \quad . \quad (150)$$

8. Auxiliary equations

The auxiliary equations given by Equation 81 must be modified to conform to the system equations of this problem. From the derivation of the auxiliary equations it can be shown that the modified auxiliary equations corresponding to Equation 81 are

$$\dot{y}_i(t) = (f_i)_{\underline{x}} y(t) + (f_i)_u s(t) \quad i = 1,6 \quad (151)$$

$$0 = (f_i)_{\underline{x}} y(t) + (f_i)_u s(t) \quad i = 7,9 \quad (152)$$

The specific equations for this problem are

$$\begin{aligned} \dot{y}_1 = & \omega_1 y_1 + (p_{11} x_7 + p_{12} x_8 + p_{13} x_9) y_4 \\ & + (p_{11} y_7 + p_{12} y_8 + p_{13} y_9) x_4 \end{aligned} \quad (153)$$

$$\begin{aligned} \dot{y}_2 = & \omega_2 y_2 + (p_{21} x_7 + p_{22} x_8 + p_{23} x_9) y_4 \\ & + (p_{21} y_7 + p_{22} y_8 + p_{23} y_9) x_4 \end{aligned} \quad (154)$$

$$\begin{aligned} \dot{y}_3 = & \omega_3 y_3 + (p_{31} x_7 + p_{32} x_8 + p_{33} x_9) y_4 \\ & + (p_{31} y_7 + p_{32} y_8 + p_{33} y_9) x_4 \end{aligned} \quad (155)$$

$$\dot{y}_4 = y_6 \quad (156)$$

$$\dot{y}_5 = 2x_6 y_6 \quad (157)$$

$$\dot{y}_6 = s(t) \quad (158)$$

$$\begin{aligned} -w_1 y_1 - (p_{11} x_7 + p_{12} x_8 + p_{13} x_9) y_4 + (v_1 - p_{11} x_4) y_7 \\ - p_{12} x_4 y_8 - p_{13} x_4 y_9 = 0 \end{aligned} \quad (159)$$

$$\begin{aligned} -w_2 y_2 - (p_{21} x_7 + p_{22} x_8 + p_{23} x_9) y_4 + p_{21} x_4 y_7 \\ + (v_2 - p_{22} x_4) y_8 - p_{23} x_4 y_9 = 0 \end{aligned} \quad (160)$$

$$\begin{aligned}
& -W_3 Y_3 - (P_{31} x_7 + P_{32} x_8 + P_{33} x_9) Y_4 + P_{31} x_4 Y_7 \\
& + P_{32} x_4 Y_8 + (V_3 - P_{33} x_4) Y_9 = 0
\end{aligned} \tag{161}$$

The initial value of all of the y-variables is zero as given by Equation 82.

The other set of auxiliary equations is identical to the adjoint equations with the dependent variables λ_i replaced by the variables η_i . However, the terminal value of $\underline{\eta}$ is different from that for the adjoint vector $\underline{\lambda}$. The terminal values for the η_i are given by a modified form of Equation 84, i.e.,

$$\eta_i(t_f) = J_{x_i \underline{x}} \Big|_{t=t_f} \underline{y}(t_f) \quad i = 1, 6 \quad . \tag{162}$$

Having computed these terminal values, the terminal values of η_7 , η_8 , and η_9 can be found by solving the system of simultaneous equations consisting of Equations 146, 147, and 148 with the λ_i replaced by η_i .

9. Calculation of the parameter β_i

In order to calculate β_i by Equation 78 it is necessary to calculate the term As_{i-1} . This term is calculated by Equation 85 which, for this problem, is given by

$$As = \eta_6(t) \quad . \tag{163}$$

V. RESULTS AND CONCLUSIONS

The results presented here are the products of digital computer calculations based on the conjugate gradient algorithm described in Chapter IV. All numerical calculations were performed on the IBM 360/65 digital computer using the FORTRAN IV language.

All integrations were performed using the subroutine DHPCG (48). This subroutine uses Hamming's modified predictor-corrector method for the solution of a general system of first-order differential equations. It is a stable fourth-order integration procedure that requires the evaluation of the right-hand side of the system only two times per step. The procedure also generates an estimate of the local truncation error at each time step and automatically adjusts the size of the next step. Information from the integrations was stored at fixed time intervals so that trajectories computed from forward integrations would be stored at the same time points as those obtained from backward integration.

The general procedure used to solve the combined system of differential and algebraic equations was, at each time step, to solve the differential system using subroutine DHPCG and use the resultant values to solve the system of algebraic equations. The systems of simultaneous algebraic equations were solved using the subroutine DGELG (48).

The one-dimensional minimization required to determine

the stepsize α_i was based upon a cubic polynomial approximation to the contour of the functional along the direction of search. Both function values and derivative values were used to determine the polynomial. After a satisfactory approximation was made, the minimum of the polynomial was chosen as the optimum stepsize.

Computation times required to obtain the optimal control varied considerably due to the arbitrariness involved in choosing a sequence of increasing penalty coefficients. As one gains experience in choosing these values for a particular problem, the computation time decreases substantially. In general, the computation times were of the order of 50 to 100 seconds.

As a convenient reference, the problem to be solved is restated here.

Determine the variation of $\Sigma_{a2}(t)$, in the control region (see Figure 2), for a fixed time interval, which results in

1. a doubling of the thermal neutron flux at the center of the reactor,
2. a steady-state thermal neutron flux at the terminal time t_f ,

and which minimizes the functional

$$\int_{t_0}^{t_f} [\Delta \dot{\Sigma}_{a2}(t)]^2 dt .$$

Optimal control programs were obtained for three different transition time intervals, 10 seconds, 30 seconds, and 60 seconds. These optimal controls are presented in Figures 3, 4, and 5. The curves labeled space-time were obtained using an expansion in three modes as discussed in Chapter IV. Also presented are the optimal control programs obtained using a one mode, or point kinetics, formulation. The point kinetics approximation assumes that the spatial distribution of the neutron flux is adequately described by the first or fundamental mode of the solution of the wave equation $\nabla^2\phi + B^2\phi = 0$. Thus, for a bare slab reactor, the neutron flux maintains a spatial distribution described by $\phi(r) = \sin(\pi r/a)$ throughout the duration of the transition.

As can be seen from the figures, the space-time optimal control requires smaller changes in the thermal absorption cross section than does the point kinetics optimal control. The difference between the control programs decreases as the transition time interval increases. The reasons for this behavior can be readily seen by examining Figures 6, 7, and 8 in which the spatial distribution of the thermal neutron flux, as obtained by the space-time model, is plotted for various times during a transition. The 10 second transition exhibits considerable flux tilting as a result of the removal of thermal neutron absorber from the control region. The effectiveness of a neutron absorber is dependent upon

the product of the absorption cross section and the neutron flux, i.e. $\Sigma_{a2}(t)\phi_2(r,t)$. It is evident, therefore, that the removal of a given amount of thermal neutron absorber from the control region will cause a greater increase in flux in the space-time model than in the "point" model, where the flux is not allowed to tilt. As the transition time interval is increased the flux tilting is not as pronounced, and consequently the discrepancy between the space-time and point kinetics optimal controls is diminished. It is interesting to note, however, that even the very slight flux tilting evident in the 60 second transition causes an appreciable effect on the optimal control program.

In order to determine the accuracy of the mathematical model used in the analysis, the resultant optimal control programs were input to a modified version of the WIGLE-40 (11) finite difference kinetics code. The flux profiles calculated by this code are often labeled as "exact" solutions by many authors. The WIGLE solutions for the thermal flux profile at the final time are denoted by crosses in Figures 6, 7, and 8. For the 30 second and 60 second transitions the agreement is very good. The discrepancy evident for the 10 second transition is most probably due to the truncation of the modal expansion after three modes. A more accurate approximation could be obtained by using higher modes, with a resultant increase in the computational time

needed to obtain a solution.

The WIGLE solutions verified the accuracy of the approximations used to obtain expressions for the constraint on the rate of change of the thermal neutron flux at the final time. For all three cases the rate of change of ϕ_2 was approximately zero at the final time.

In summary, for the problem studied in this dissertation, the conjugate gradient algorithm in conjunction with the Natural Mode Approximation provides an effective computational scheme for the determination of optimal control programs for a spatially dependent nuclear reactor. Contrary to the opinions of several authors, it was shown that the general nonlinear equations describing the reactor dynamics can be handled directly without the necessity of linearization. The superiority of a space-time analysis over a simpler point kinetics analysis was also demonstrated.

For a more complex reactor model involving heterogeneities, and two or three dimensions, the Natural Mode Approximation would not be a practical approximation. For these complex configurations the synthesis techniques which utilize a variational method appear to be the most promising. Of particular significance is that in nearly all of the synthesis techniques the time dependent coefficients of combination are related by a coupled set of ordinary, nonlinear differential equations. Thus, the conjugate gradient

technique should be applicable to almost all of the available synthesis methods.

Perhaps the most significant contribution of this dissertation is that it provides a basic procedural framework for determining spatially dependent optimal controls, upon which the more intricate mathematical models needed to describe complex systems can be constructed.

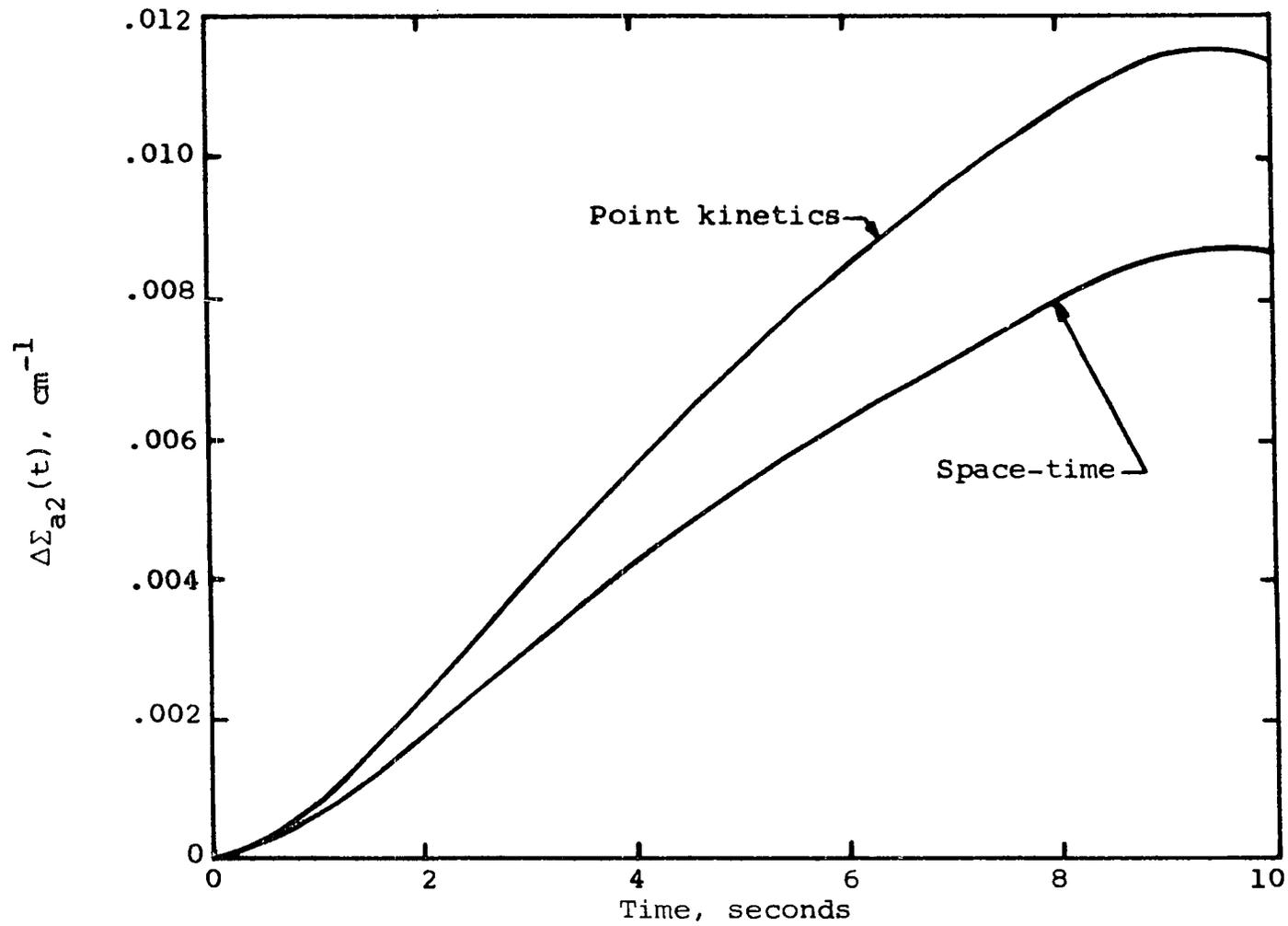


Figure 3. Optimal control program for 10 second transition

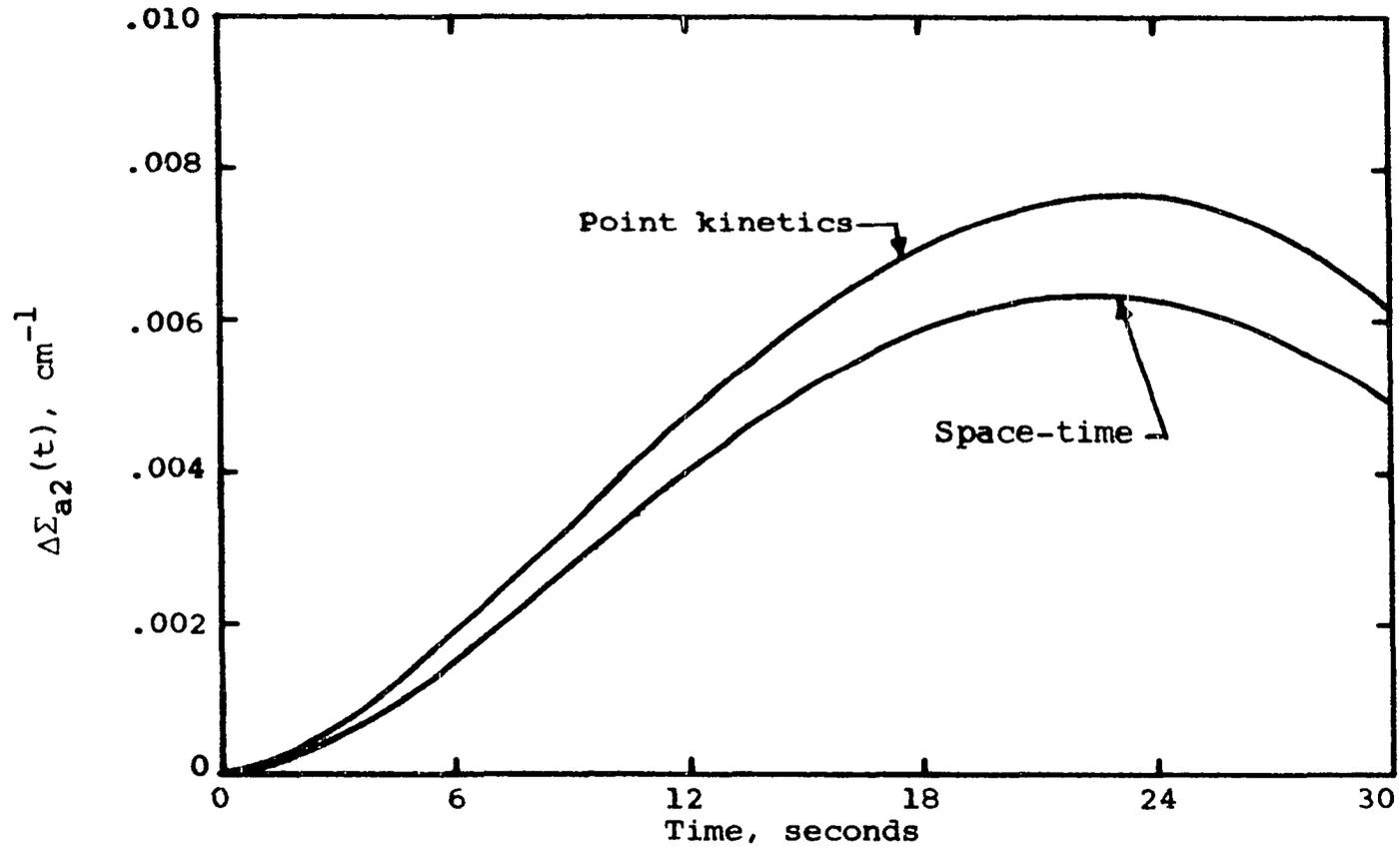


Figure 4. Optimal control program for 30 second transition

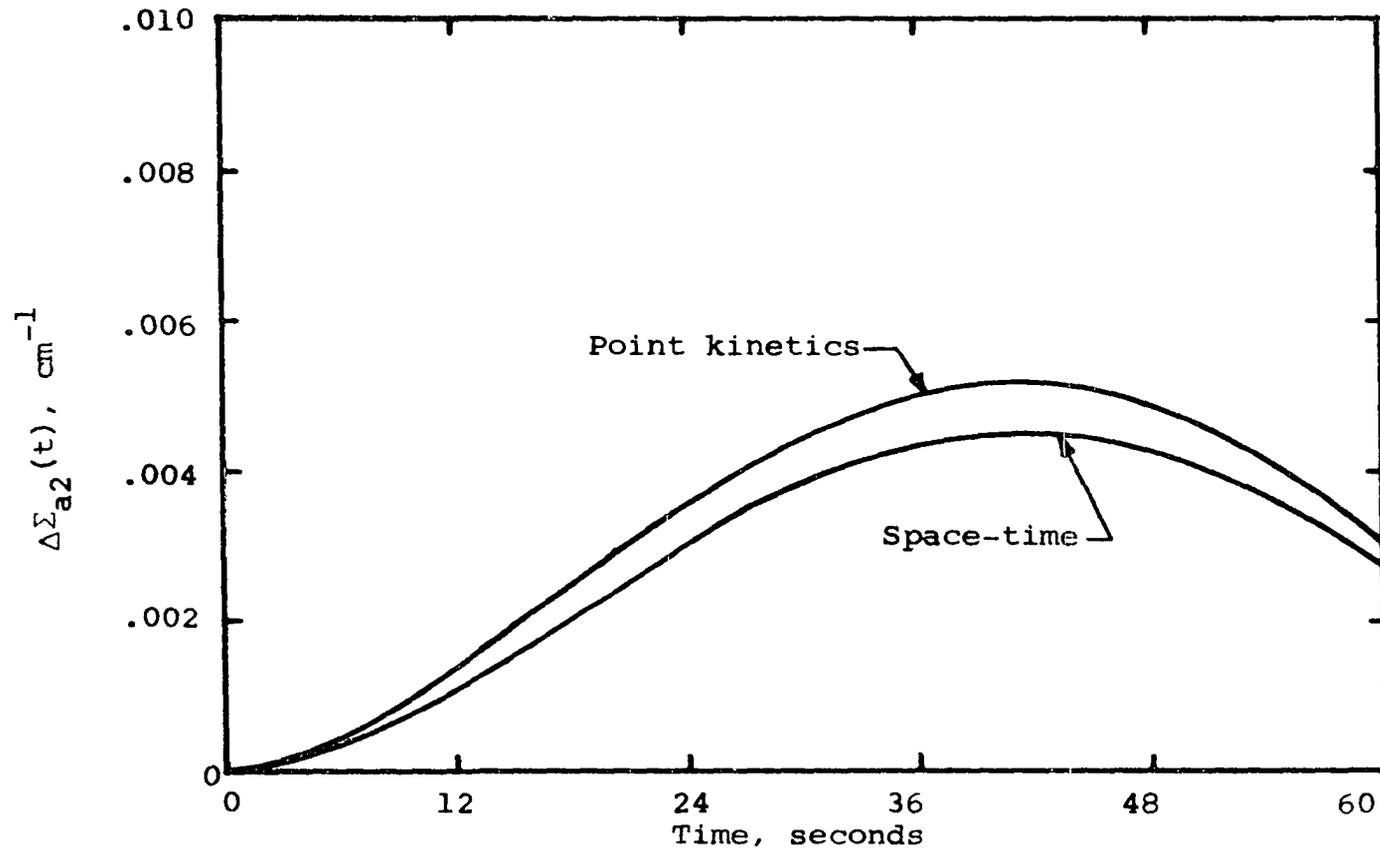


Figure 5. Optimal control program for 60 second transition

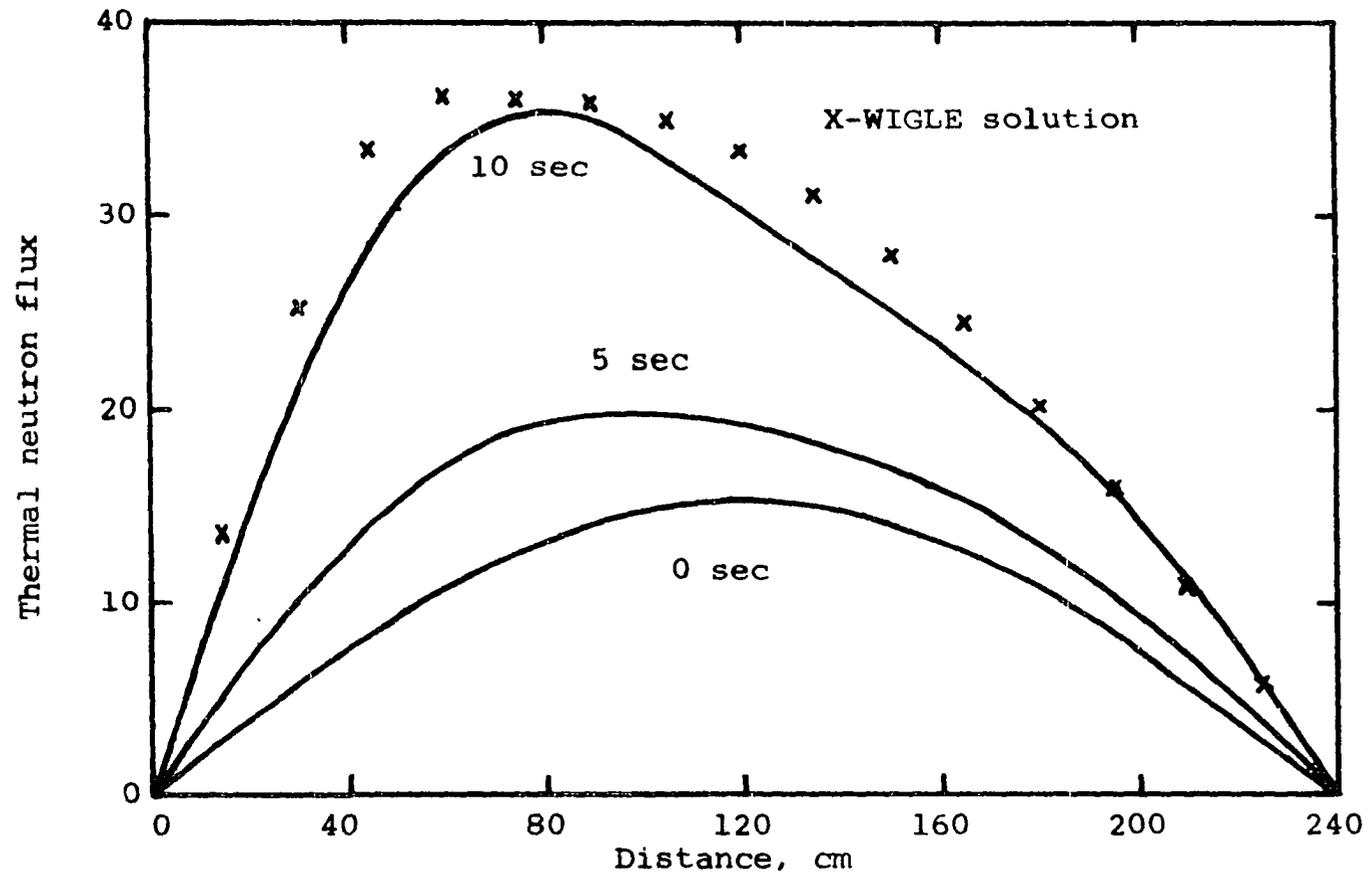


Figure 6. Thermal neutron flux profile during the 10 second transition

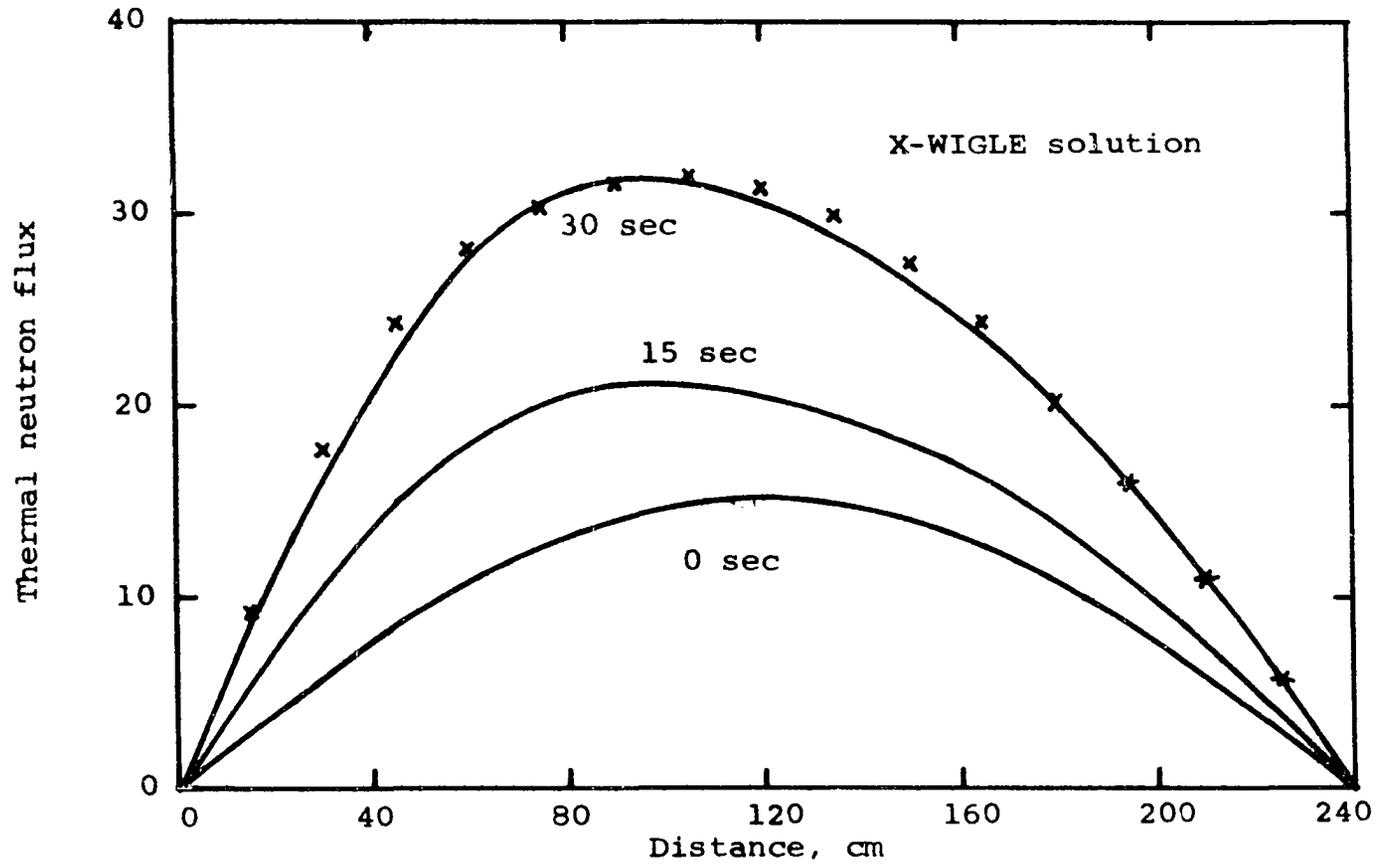


Figure 7. Thermal neutron flux profile during the 30 second transition

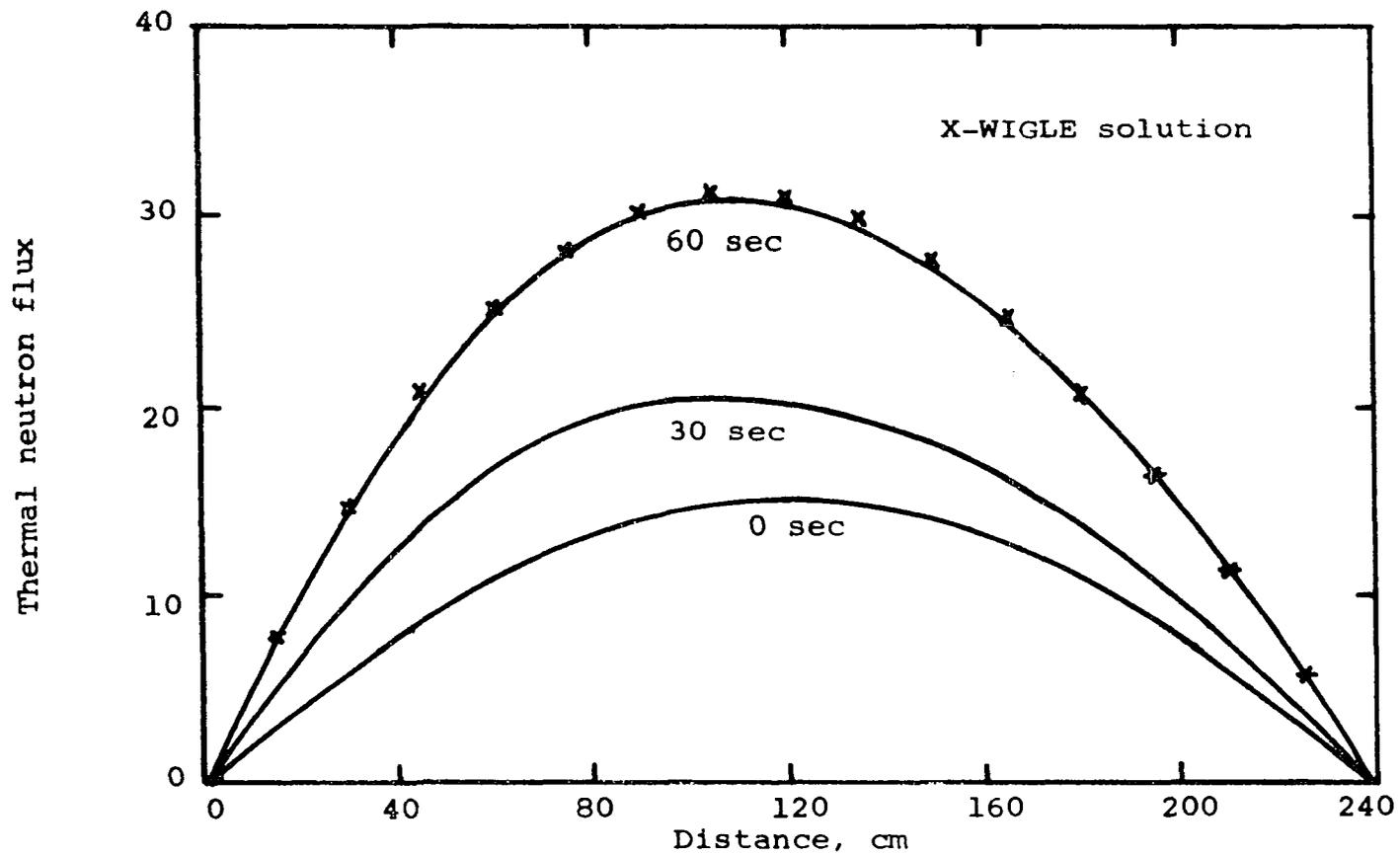


Figure 8. Thermal neutron flux profile for the 60 second transition

VI. SUGGESTIONS FOR FURTHER STUDY

There is a wealth of opportunities for additional investigations based on extensions of this work. The replacement of the Natural Mode Approximation by a synthesis method capable of handling more complex problems merits considerable attention. It would also be interesting to apply other trajectory optimization techniques to the problem studied in this dissertation in order to determine the best method to use for such a problem. An extension of the conjugate gradient method to handle several control variables would also be very useful.

The extensions of this work to fuel management problems appears to be quite feasible. In particular, applications to the problem of determining optimum fuel loading patterns should be attempted. Methods of minimizing or maximizing the production of certain isotopes could also be determined.

It would also be worthwhile to attempt the determination of a closed-loop control law for the problem studied in this thesis. Such an extension would involve a linearization about the optimal reference trajectory determined in this study.

VII. ACKNOWLEDGMENTS

Many individuals are responsible for the successful completion of the graduate program culminated by this dissertation. The author wishes to express this gratitude in particular to Dr. Richard Danofsky who has contributed much appreciated advice and encouragement during the course of the author's doctoral program. Greatly appreciated also is the thoughtful counseling and guidance given by Dr. Glenn Murphy throughout the author's college career.

The author gratefully acknowledges support from a United States Atomic Energy Commission Special Fellowship in Nuclear Science and Engineering.

Special thanks are due to the author's wife and children for their many sacrifices, constant encouragement and understanding during the course of the author's college career.

VIII. LITERATURE CITED

1. Athans, M. and P. L. Falb. Optimal control: an introduction to the theory and its applications. New York, New York, The McGraw-Hill Book Co. 1966.
2. Willoughby, J. K. Adaptations of the conjugate gradient method to optimal control problems with terminal state constraints. Iowa State University Engineering Research Institute Technical Report ERI-62500 (Iowa State University, Ames, Iowa). 1970.
3. Berkovitz, L. D. Variational methods in problems of control and programming. Journal of Mathematical Analysis and Applications 8: 145-169. 1961.
4. Pontryagin, L. S., V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko. The mathematical theory of optimal processes. New York, New York, John Wiley and Sons, Inc. 1962.
5. Lipinski, W. C. Optimal digital computer control of nuclear reactors. United States Atomic Energy Commission Report ANL-7530 (Argonne National Laboratory, Argonne, Ill.). 1969.
6. Yasinsky, J. B. and A. F. Henry. Some numerical experiments concerning space-time reactor kinetics behavior. Nuclear Science and Engineering 22: 171-181. 1965.
7. Wang, P. K. C. Control of distributed parameter systems. In Leondes, C. T., ed. Advances in control systems. Vol. 1. Pp. 75-172. New York, New York, Academic Press Inc. 1964.
8. Hsu, C. Control and stability analysis of spatially dependent nuclear reactor systems. United States Atomic Energy Commission Report ANL-7322 (Argonne National Laboratory, Argonne, Ill.). 1967.
9. Stacey, W. M., Jr. Control of xenon spatial oscillations. Nuclear Science and Engineering 38: 229-243. 1969.
10. Avery, R. Theory of coupled reactors. Second International Conference on the Peaceful Uses of Atomic Energy Proceedings 12: 182-191. 1958.

11. Rad, M. E. WIGLE-40, a two-group, time-dependent diffusion theory program for the IBM 7040 computer. United States Atomic Energy Commission Report IDO-17125 (Phillips Petroleum Corp. Atomic Energy Division, Idaho Falls, Idaho). 1965.
12. Kaplan, S. The property of finality and the analysis of problems in reactor space-time kinetics by various modal expansions. Nuclear Science and Engineering 9: 357-361. 1961.
13. Foulke, L. R. and E. P. Gyftopoulos. Application of the natural mode approximation to space-time reactor problems. Nuclear Science and Engineering 30: 419-435. 1967.
14. Kaplan, S. Synthesis methods in reactor analysis. In Greebler, P. and E. H. Henley, ed. Advances in nuclear science and technology. Vol. 3. Pp. 233-266. New York, New York, Academic Press Inc. 1966.
15. Kaplan, S. Variational methods in nuclear engineering. In Henley, E. J. and J. Lewins, eds. Advances in nuclear science and technology. Vol. 5. Pp. 185-221. New York, New York, Academic Press Inc. 1969.
16. Wiberg, D. M. Optimal control of nuclear reactor systems. In Leondes, C. T., ed. Advances in control systems. Vol. 5. Pp. 301-388. New York, New York, Academic Press Inc. 1967.
17. Suda, N. On controllability of neutron flux distribution. Journal of Nuclear Science and Technology 5: 271-278. 1968.
18. Kliger, I. Optimal control of space-dependent nuclear reactors. American Nuclear Society Transactions 8: 233-324. 1965.
19. Stacey, W. M., Jr. Application of variational synthesis to the optimal control of spatially dependent reactor models. Nuclear Science and Engineering 39: 225-230. 1970.
20. Bellman, R. E. Dynamic programming. Princeton, New Jersey, Princeton University Press. 1957.

21. Stacey, W. M., Jr. Optimal control of xenon-power spatial transients. Nuclear Science and Engineering 33: 162-168. 1968.
22. Wall, I. and H. Fenech. The application of dynamic programming to fuel management optimization. Nuclear Science and Engineering 22: 285-297. 1965.
23. Jungnell, D. and T. Spanne. Studies of the frequency and power regulation of a power system containing nuclear and water power stations. Third International Conference of the Peaceful Uses of Atomic Energy Proceedings 1: 222-232 1964.
24. LMFBR Program Office Argonne National Laboratory. Liquid metal fast breeder reactor program, instrumentation and control. United States Atomic Energy Commission Report WASH-1105 (Atomic Energy Commission, Washington, D.C.). 1968.
25. Tapley, B. D. and J. M. Lewallen. Comparison of several numerical optimization methods. Journal of Optimization Theory and Applications 1: 1-32. 1967.
26. Kopp, R. E. and H. G. Moyer. Trajectory optimization techniques. In Leondes, C. T., ed. Advances in control systems. Vol. 4. Pp. 103-155. New York, New York, Academic Press Inc. 1967.
27. Balakrishnan, A. V. and L. W. Neustadt. Computing methods in optimization problems. New York, New York, Academic Press Inc. 1964.
28. McGill, R. and P. Kenneth. Solution of variational problems by means of a generalized Newton-Raphson operator. American Institute of Aeronautics and Astronautics Journal 2: 1761-1766. 1964.
29. Luenberger, D. G. Optimization by vector space methods. New York, New York, John Wiley and Sons, Inc. 1969.
30. Bryson, A. E. and W. F. Denham. A steepest-ascent method for solving optimum programming problems. Journal of Applied Mechanics 29: 247-257. 1962.
31. Kelley, H. J. Gradient theory of optimal flight paths. American Rocket Society Journal 30: 947-954. 1960.

32. Rosen, J. B. The gradient projection method for non-linear programming: part I. Linear Constraints. J. Soc. Industrial Appl. Math. 8: 181-217. 1967.
33. Rosen, J. B. The gradient projection method for non-linear programming: part II. Nonlinear Constraints. J. Soc. Industrial Appl. Math. 9: 514-532. 1961.
34. Breakwell, J. V., J. L. Speyer, and A. E. Bryson. Optimization and control of nonlinear systems using the second variation. Industrial Appl. Math. J. Control Soc. Series A, 1: 193-223. 1963.
35. Kelley, H. J., R. E. Kopp, and H. G. Moyer. A trajectory optimization technique based upon the theory of the second variation. In Szebehely, V. G., ed. Progress in astronautics and astrodynamics. Vol. 14. Pp. 559-582. New York, New York, Academic Press, Inc. 1964.
36. Lapidus, L. and R. Luus. Optimal control of engineering processes. Waltham, Mass., Blaisdell Publishing Co. 1967.
37. Foulke, L. R. A modal expansion technique for space-time reactor kinetics. Unpublished Ph.D. thesis. Cambridge, Mass., Department of Nuclear Engineering, Massachusetts Institute of Technology. 1966.
38. Betancourt, J. M. Analysis of coupled core reactors using the natural mode approximation. Unpublished Ph.D. thesis. Ames, Iowa, Library, Iowa State University of Science and Technology. 1968.
39. Weinberg, A. and E. P. Wigner. The physical theory of neutron chain reactors. 2nd ed. Chicago, Ill., The University of Chicago Press. 1959.
40. Henry, A. F. The application of inhour modes to the description of nonseparable reactor transients. Nuclear Science and Engineering 20: 338-351. 1964.
41. Hestenes, M. R. and E. Stiefel. Methods of conjugate gradients for solving linear systems. J. Res. Natl. Bur. Std. 49: 409-436. 1952.
42. Fletcher, R. and C. M. Reeves. Function minimization by conjugate gradients. Computer Journal 7: 149-154. 1964.

43. Hayes, R. M. Iterative methods of solving linear problems in hilbert space. Natl. Bur. Std. Appl. Math. Series 39: 71-104. 1954.
44. Sinnott, J. F., Jr. and D. G. Luenberger. Solution of optimal control problems by the method of conjugate gradients. Joint Automatic Control Conference Proceedings 1967: 566-574. 1967.
45. Lasdon, L. S., S. Mitter, and A. D. Warren. The method of conjugate gradients for optimal control problems. Institute of Electrical and Electronics Engineers Transactions on Automatic Control AC-12: 132-138. 1967.
46. Tripathi, S. S. and K. S. Narendra. Conjugate direction methods for nonlinear optimization problems. National Electronics Conference Proceedings 1968: 125-129. 1968.
47. Weaver, L. E. Reactor dynamics and control. New York, New York, American Elsevier Publishing Co., Inc. 1968.
48. International Business Machines Corporation, Technical Publications Department. System/360 scientific subroutine package (360A-CM-03X) version III programmers manual. 4th ed. White Plains, New York, Author. c1968.

IX. APPENDIX: DERIVATION OF NECESSARY
CONDITIONS FOR AN OPTIMAL CONTROL

The purpose of this Appendix is to present an understandable, but non-rigorous derivation of the conditions that are necessary for the optimality of a control $u(t)$. For the reader desiring a derivation with more rigor and formalism the work of Berkovitz (3) or Pontryagin (4) should be consulted.

A. Statement of Problem

The necessary conditions will be derived for the following simplified optimization problem:

$$\text{minimize } J = S(x(t_f)) \quad (\text{A.1})$$

$$\text{subject to } F(x, \dot{x}, u, t) = f(x, u, t) - \dot{x} = 0 \quad (\text{A.2})$$

$$x(t_0) = x_0 \quad (\text{A.3})$$

In these equations, x is a state variable to be controlled, u is a control function on the fixed time interval $[t_0, t_f]$, and f is a nonlinear expression defining the dynamical system to be controlled. It is assumed that $f(x, u, t)$ and all of its derivatives are continuous on the interval of interest. It is assumed also that the control variable $u(t)$ is unbounded and that there are no constraints on x except at the initial and terminal times.

B. Derivation of Necessary Conditions

Let

$$S(x(t_f)) = \int_{t_0}^{t_f} \frac{dS}{dt} dt$$

where

$$\frac{dS}{dt} = \frac{\partial S}{\partial x} \frac{dx}{dt} + \frac{\partial S}{\partial t} = \frac{\partial S}{\partial x} \dot{x} + \frac{\partial S}{\partial t} .$$

Rewrite J as

$$J = \int_{t_0}^{t_f} \left(\frac{\partial S}{\partial x} \dot{x} + \frac{\partial S}{\partial t} \right) dt = \int_{t_0}^{t_f} G(\dot{x}, x, t) dt \quad (A.4)$$

where

$$G(x, \dot{x}, t) = \frac{\partial S}{\partial x} \dot{x} + \frac{\partial S}{\partial t} .$$

Set the first variation of J equal to zero:

$$\delta J = \int_{t_0}^{t_f} \left[\frac{\partial G}{\partial x} \delta x + \frac{\partial G}{\partial \dot{x}} \delta \dot{x} \right] dt = 0 . \quad (A.5)$$

The first variation of F is given by

$$\delta F = \frac{\partial F}{\partial x} \delta x + \frac{\partial F}{\partial u} \delta u + \frac{\partial F}{\partial \dot{x}} \delta \dot{x} = 0 . \quad (A.6)$$

Multiply Equation A.6 by an as yet undefined variable $\lambda(t)$, and integrate from t_0 to t_f :

$$\int_{t_0}^{t_f} \left[\lambda \frac{\partial F}{\partial \mathbf{x}} \delta \mathbf{x} + \lambda \frac{\partial F}{\partial u} \delta u + \lambda \frac{\partial F}{\partial \dot{\mathbf{x}}} \delta \dot{\mathbf{x}} \right] dt = 0 \quad . \quad (\text{A.7})$$

Add Equations A.5 and A.7 to obtain

$$\int_{t_0}^{t_f} \left[\left(\frac{\partial G}{\partial \mathbf{x}} + \lambda \frac{\partial F}{\partial \mathbf{x}} \right) \delta \mathbf{x} + \lambda \frac{\partial F}{\partial u} \delta u + \left(\frac{\partial G}{\partial \dot{\mathbf{x}}} + \lambda \frac{\partial F}{\partial \dot{\mathbf{x}}} \right) \delta \dot{\mathbf{x}} \right] dt = 0 \quad . \quad (\text{A.8})$$

Integrate the third term of the integrand of Equation A.8 by parts to obtain

$$\begin{aligned} \int_{t_0}^{t_f} \left[\frac{\partial G}{\partial \dot{\mathbf{x}}} + \lambda \frac{\partial F}{\partial \dot{\mathbf{x}}} \right] \delta \dot{\mathbf{x}} dt &= \left[\frac{\partial G}{\partial \dot{\mathbf{x}}} + \lambda \frac{\partial F}{\partial \dot{\mathbf{x}}} \right] \delta \mathbf{x} \Bigg|_{t_0}^{t_f} \\ &- \int_{t_0}^{t_f} \frac{d}{dt} \left[\frac{\partial G}{\partial \dot{\mathbf{x}}} + \lambda \frac{\partial F}{\partial \dot{\mathbf{x}}} \right] \delta \mathbf{x} dt \quad . \end{aligned} \quad (\text{A.9})$$

Substitute Equation A.9 into Equation A.8 and collect terms:

$$\begin{aligned} \int_{t_0}^{t_f} \left\{ \left[\frac{\partial G}{\partial \mathbf{x}} + \lambda \frac{\partial F}{\partial \mathbf{x}} - \frac{d}{dt} \left(\frac{\partial G}{\partial \dot{\mathbf{x}}} + \lambda \frac{\partial F}{\partial \dot{\mathbf{x}}} \right) \right] \delta \mathbf{x} + \lambda \frac{\partial F}{\partial u} \delta u \right\} \\ + \left[\frac{\partial G}{\partial \dot{\mathbf{x}}} + \lambda \frac{\partial F}{\partial \dot{\mathbf{x}}} \right] \delta \mathbf{x} \Bigg|_{t_0}^{t_f} = 0 \quad . \end{aligned} \quad (\text{A.10})$$

It must follow that:

$$\frac{\partial}{\partial \mathbf{x}}[G + \lambda F] - \frac{d}{dt} \frac{\partial}{\partial \dot{\mathbf{x}}}[G + \lambda F] = 0 \quad (\text{A.11})$$

$$\frac{\partial}{\partial u}[\lambda F] = 0 \quad (\text{A.12})$$

$$\frac{\partial}{\partial \dot{\mathbf{x}}}[G + \lambda F] \delta \mathbf{x} \Big|_{t_0}^{t_f} = 0 \quad . \quad (\text{A.13})$$

For the optimal control problem of interest

$$F(\mathbf{x}, \dot{\mathbf{x}}, u, t) = f(\mathbf{x}, u, t) - \dot{\mathbf{x}} = 0 \quad .$$

Thus

$$\frac{\partial F}{\partial \dot{\mathbf{x}}} = -1$$

and

$$\frac{\partial G}{\partial \dot{\mathbf{x}}} = \frac{\partial}{\partial \dot{\mathbf{x}}} \left[\frac{\partial S}{\partial \mathbf{x}} \dot{\mathbf{x}} + \frac{\partial S}{\partial t} \right] = \frac{\partial S}{\partial \mathbf{x}} \quad .$$

It follows that

$$\frac{\partial}{\partial \dot{\mathbf{x}}}[G + \lambda F] = \frac{\partial S}{\partial \mathbf{x}} - \lambda \quad . \quad (\text{A.14})$$

Using the definition of G, Equation A.11 can be written as

$$\frac{\partial}{\partial \mathbf{x}} \left[\frac{\partial S}{\partial \mathbf{x}} \dot{\mathbf{x}} + \frac{\partial S}{\partial t} + \lambda F \right] = \frac{d}{dt} \left[\frac{\partial S}{\partial \mathbf{x}} - \lambda \right] \quad . \quad (\text{A.15})$$

However,

$$\frac{d}{dt} \left[\frac{\partial S}{\partial \mathbf{x}} \right] = \frac{\partial^2 S}{\partial \mathbf{x}^2} \frac{d\mathbf{x}}{dt} + \frac{\partial^2 S}{\partial \mathbf{x} \partial t} \quad (\text{A.16})$$

and

$$\frac{\partial}{\partial \mathbf{x}} \left[\frac{\partial S}{\partial \mathbf{x}} \dot{\mathbf{x}} + \frac{\partial S}{\partial t} \right] = \frac{\partial^2 S}{\partial \mathbf{x}^2} \dot{\mathbf{x}} + \frac{\partial^2 S}{\partial \mathbf{x} \partial t} \quad . \quad (\text{A.17})$$

Inserting Equations A.16 and A.17 into Equation A.15 gives

$$\frac{\partial}{\partial \mathbf{x}} [\lambda F] + \frac{\partial^2 S}{\partial \mathbf{x}^2} \dot{\mathbf{x}} + \frac{\partial^2 S}{\partial \mathbf{x} \partial t} = -\dot{\lambda} + \frac{\partial^2 S}{\partial \mathbf{x}^2} \frac{d\mathbf{x}}{dt} + \frac{\partial^2 S}{\partial \mathbf{x} \partial t} \quad .$$

Therefore,

$$\dot{\lambda} = - \frac{\partial}{\partial \mathbf{x}} [\lambda F] \quad . \quad (\text{A.18})$$

Inserting Equation A.14 into Equation A.13 gives

$$\left[\frac{\partial S}{\partial \mathbf{x}} - \lambda \right] \delta \mathbf{x} \Bigg|_{t_0}^{t_f} = 0 \quad . \quad (\text{A.19})$$

However, $\mathbf{x}(t_0) = \mathbf{x}_0$ and thus $\delta \mathbf{x}(t_0) = 0$. Equation A.19 can be written as

$$\left[\frac{\partial S(\mathbf{x}(t_f))}{\partial \mathbf{x}} - \lambda(t_f) \right] \delta \mathbf{x}(t_f) = 0 \quad .$$

Thus for arbitrary variations in $\delta \mathbf{x}(t_f)$ it must follow that

$$\frac{\partial S(\mathbf{x}(t_f))}{\partial \mathbf{x}} - \lambda(t_f) = 0 \quad .$$

Therefore,

$$\lambda(t_f) = \frac{\partial S(\mathbf{x}(t_f))}{\partial \mathbf{x}} \quad . \quad (\text{A.20})$$

Now define the Hamiltonian H to be

$$H = \lambda(t)f(\mathbf{x}, u, t) \quad . \quad (\text{A.21})$$

From Equation A. 18

$$\begin{aligned} \dot{\lambda} &= - \frac{\partial}{\partial \mathbf{x}} [\lambda (f(\mathbf{x}, u, t) - \dot{\mathbf{x}})] \\ &= - \frac{\partial}{\partial \mathbf{x}} [\lambda f(\mathbf{x}, u, t)] \quad . \end{aligned}$$

Thus,

$$\dot{\lambda} = - \frac{\partial H}{\partial \mathbf{x}} \quad . \quad (\text{A.22})$$

From Equation A.12,

$$\frac{\partial}{\partial u} [\lambda (f(\mathbf{x}, u, t) - \dot{\mathbf{x}})] = 0$$

and it follows that:

$$\frac{\partial H}{\partial u} = 0 \quad . \quad (\text{A.23})$$

The necessary conditions have been derived for a scalar system. The extension to a vector state-space is straightforward and results in vector expressions analogous to those derived here.