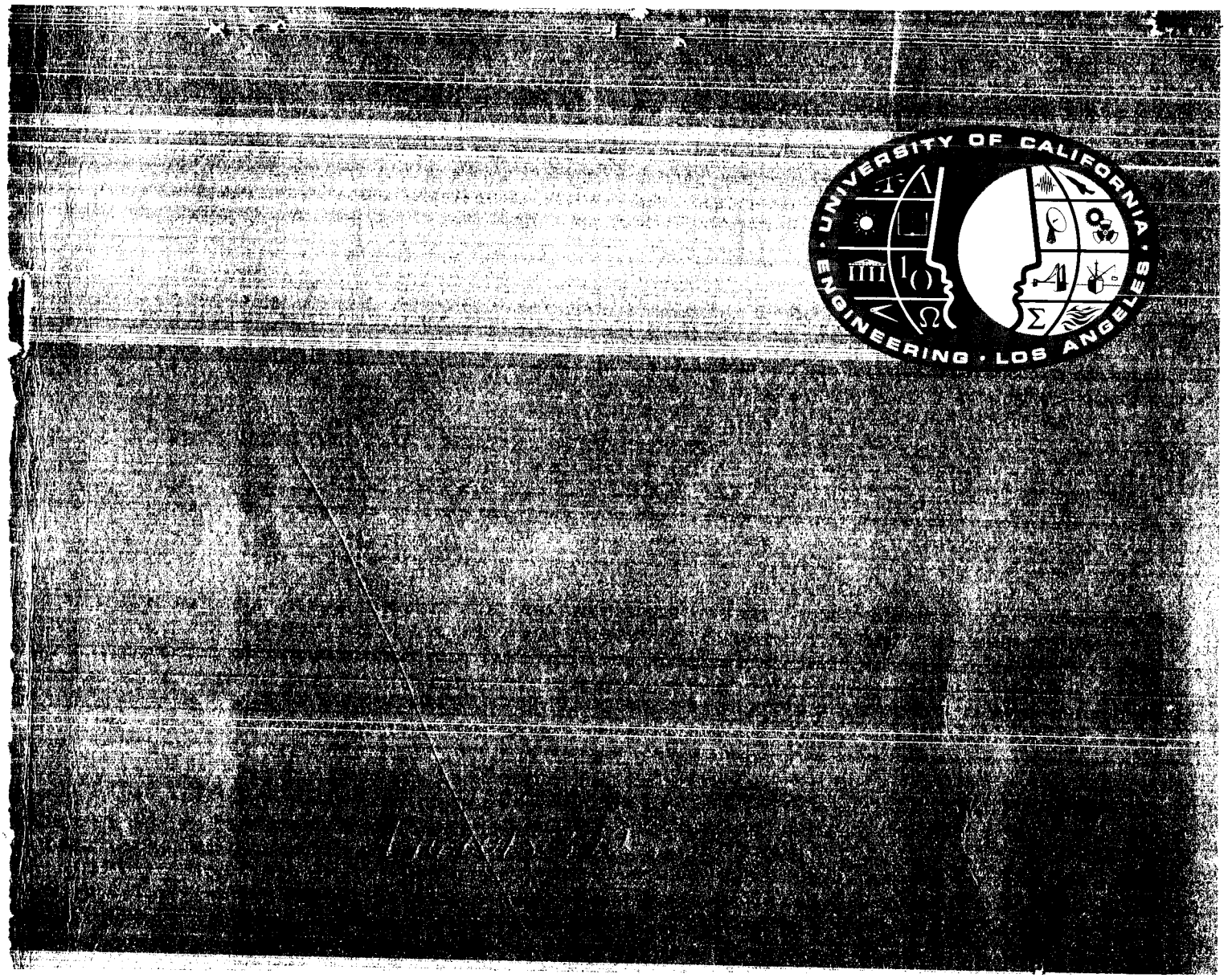


OPTIMAL CONTROL OF DISTRIBUTED PARAMETER SYSTEMS USING MULTILEVEL TECHNIQUES

D. A. WISMER, Jr.

April 1967
Report No. 66-55C

Report No. 66-55C
April 1967

OPTIMAL CONTROL OF DISTRIBUTED PARAMETER SYSTEMS USING MULTILEVEL TECHNIQUES

David Arthur Wismer, Jr.

Department of Engineering
University of California
Los Angeles, California

FOREWORD

The research described in this report, "Optimal Control of Distributed Parameter Systems Using Multilevel Techniques," Number 66-55C, by David Arthur Wismer, Jr., was carried out under the direction of C. T. Leondes, E. B. Stear, and A. R. Stubberud, in the Department of Engineering, University of California, Los Angeles.

This project is sponsored in part by the National Aeronautics and Space Administration under Grant NsG-237-62 to the Institute of Geophysics and Planetary Physics of the University.

This report was the basis for a dissertation submitted by the Author.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF ILLUSTRATIONS

CHAPTER 1
INTRODUCTION

1.1 Optimal Control Theory

In recent years, control systems theory has been largely concerned with problems of optimization. Most of this effort has dealt with systems described by ordinary differential equations and termed lumped parameter systems. More recently, considerable interest has emerged in the optimization of systems described by partial differential equations and termed distributed parameter systems. Still another class of optimization problems deals with static optimization or systems described by algebraic equations. This dissertation will treat the distributed parameter systems problem as an (approximately) equivalent problem for lumped parameter or static systems.

The particular problem considered is one of choosing a control variable(s) such that a given functional of independent, dependent, and/or control variables is maximized or minimized. This is to be accomplished while satisfying certain equality and/or inequality constraints called side conditions. Problems of this type involving differential equations can be handled by a number of methods; namely, the calculus of variations,[9] Pontryagin's maximum principle,[54] Bellman's dynamic programming,[7] functional analysis,[5,37] and gradient methods.[34,35] Recently Dantzig[19] has treated the optimization of linear dynamic systems by a generalized linear program. Certain of the above methods are also applicable to static system optimization.

Several authors have detailed the relationship between some of these methods. In particular, Dreyfus[21] has derived many of the theorems of the calculus of variations using dynamic programming,

and Hestenes[28] has demonstrated the relationship between the calculus of variations, the maximum principle, and dynamic programming.

Perhaps the greatest effort has been expended in solving linear problems with quadric cost functionals because these lend themselves nicely to analytical results. However, in many situations these properties do not adequately represent the physical situation and one is forced to deal with nonlinear systems which may be of high order and with complicated criterion functions. Since such problems are not amenable to analytical solutions, much research has been done on the computational aspects of obtaining numerical results. For example, the necessary conditions for optimality arising from the calculus of variations (or other) treatment give rise to various two point boundary value problems. These problems are of higher order than the original state equations and have initial values specified for some equations and final values specified for other (or the same) equations. Because of the nature of this problem, iterative methods are generally required for its solution. Particular techniques are (1) quasilinearization[30] which iterates on the solution trajectories, (2) Newton-Raphson[10] (a second variational technique) which iterates on certain initial conditions, and (3) gradient[34] or steepest descent which iterates on the control variable. The method of quasilinearization was used extensively in this investigation.

1.2 Distributed Parameter Systems

The pioneering work in optimal control theory for distributed parameter systems was done by Butkovskii and Lerner[15] in 1960 and Butkovskii has been a constant contributor since that time.[16, 17] The approach taken in much of this work is based on

extending Pontryagin's maximum principle. In this country, Wang[64, 65] has developed the necessary conditions for the optimal control of distributed parameter systems using the formalism of dynamic programming. Wang also discusses stability, controllability, observability, approximation methods, and instrumentation. Other investigators are Brogan,[11] Egorov,[22] Sakawa[57] and Axelband,[4] each of whom treats certain classes of problems by various methods.

For all of the theoretical work reported, computational results have been notably absent. Brogan[11] gives results for the linear one-dimensional diffusion equation with distributed control and Sakawa[56] treats a similar problem with boundary control. One of the reasons for the sparse number of examples is undoubtedly the computational difficulty involved in solving these problems. In fact, Wang[64] raises the question of the relative merit in discretizing the necessary conditions for optimality versus discretizing the original system partial differential equation since some approximation is generally required in obtaining a solution. The latter approach is taken in this dissertation by reducing partial differential equations to ordinary differential equations through spatial discretization. It is hoped that this approach will lead to numerical solutions for a broader class of problems than would otherwise be attained.

As mentioned above, two types of problems arise in the optimization of distributed parameter systems; namely, (1) distributed control and (2) boundary control. In the former case the control is distributed over the entire spatial domain, and in the latter, it is distributed only over the boundary domain. It is noteworthy that in most physical situations, true distributed controls are not present. This fact gives additional impetus to the discrete model approach. Some examples of distributed parameter systems

are continuous furnaces, electrical power transmission systems, and re-entry vehicles with ablative surfaces.[64]

1.3 Multilevel Control

The term multilevel control implies the decomposition of a (large) system into smaller (independent) subsystems and the coordination of these subsystem solutions by a "superior" controller which operates on several or all of the subsystems. The subsystem controllers might be called first-level, their "superiors," second-level, etc. This nomenclature is part of a general theory introduced by Mesarovic[48] for treating multilevel, multigoal systems where the term multigoal implies that the subsystems may have different goals or objective functions. Many papers have elaborated these concepts (see for instance the bibliography in Reference 49). The work reported here will refer to a two level, N goal system. As used to date, the term multilevel control has been used in connection with optimization problems and refers to an off-line type of control. A better term might be multilevel optimization; however, the former terminology is retained here.

In its present context, the idea of decomposition for solving optimal control problems seems to have originated with Dantzig and Wolfe[20] who, in 1960, adopted this procedure for solving large linear programming problems.† More recent work at Case Institute of Technology has considerably extended the theory of multilevel

†Elements of decomposition are also present in Kron's[38] method of "tearing" and Bellman's[7] dynamic programming for network analysis and optimization respectively.

control. Lasdon[41] treats the steady state optimization of nonlinear systems and shows, using the Kuhn-Tucker[39] theory, that a certain saddle-value problem results, and Macko[45] has extended these results to dynamic nonlinear systems. Takahara[59] treats linear dynamic systems in a somewhat different way using features inherent in the system linearity. Recently Bauman[6] has extended some of the above results to trajectory decomposition and has given some computational examples. References 41 and 45 provide the main background for the work reported here and they will be reviewed briefly in Chapter 2.

The multilevel control problem can also be approached from the point of view of mathematical programming. Dantzig[19] and Varaiya[63] have reported results in this area. The interesting question of duality also arises and is discussed by Pearson.[52, 53] Kulikowski[40] formulates a number of linear multilevel control problems using the theory of M. Krein[1] and also treats the related question of optimal aggregation. The latter question arises particularly in certain problems in operations research. Finally, Lefkowitz[43] discusses a multilevel approach to control system design and delineates four levels of control; namely, regulation, optimization, adaptation, and self-organization.

1.4 Scope of the Dissertation

Some of the objectives of this research are as follows:

1. To formulate distributed parameter systems in the context of multilevel control by spatial discretization.
2. To determine a decomposition approach which will tend to minimize coupling constraints between subsystems.

3. To determine the relative merits of various second-level controllers in handling the high dimensionality resulting from the spatial discretization.
4. To demonstrate the applicability of this approach to a wide class of problems including nonlinear problems and problems having irregular and/or higher order (greater than 1) spatial domains.
5. To solve a representative number of example problems using the multilevel approach.

This dissertation is intended to present the results of this research. The purpose of Chapter 1 is to lay a framework for the three major ideas with which the dissertation is concerned; namely, optimal control theory, distributed parameter systems, and multilevel control. A further purpose is to outline the goals of the research and to preview the subsequent material in the dissertation.

Chapter 2 describes some theoretical aspects of multilevel control and fixes the terminology to be used throughout the remainder of the dissertation. Much of this material was developed at Case Institute of Technology, Cleveland, Ohio.

Chapter 3 defines the classes of optimal control problems which can be attacked by a multilevel approach. A fairly general semidiscrete model is developed. The necessary conditions for the special case of a nonlinear diffusion equation with quadratic cost functional are then developed in order to fix ideas.

Chapter 4 discusses the pros and cons of using multilevel techniques for solving the problem proposed here and briefly treats some controllability questions. Also presented is a critique on various second-level controllers along with some miscellaneous topics such as state inequality constraints and time discretization.

Chapter 5 formulates a number of example problems using the ideas of multilevel control.

Chapter 6 presents numerical results for several of the examples formulated in Chapter 5. A brief description of the computer program and the method of quasilinearization used in obtaining subsystem solutions is also contained here.

Chapter 7 presents the conclusions reached in this research and details some areas where further study would be fruitful.

CHAPTER 2

MULTILEVEL CONTROL THEORY

2.1 Introduction

This chapter will review some multilevel control techniques for the optimization of nonlinear static and dynamic systems. Particular emphasis will be placed on the convergence properties of various forms of second-level controllers including a Gauss-Seidel type controller introduced here.

The theory of multilevel control has two significant features in solving an optimal control problem; namely, (1) a conceptual simplification for large systems and (2) a possible reduction in the computational burden involved in computing optimal controls. The first advantage is achieved by treating an n^{th} order system as N independent subsystems of order n_j where

$$n = \sum_{j=1}^{N} n_j$$

The subsystem independence is attained by relaxing one (or more) of the necessary conditions for optimality and then satisfying this condition with a second-level controller. This technique of solution requires an iteration between levels of control and thus no guarentee of computational time reduction can be made. However, in theory the reduced subsystem size may permit the solution of problems not otherwise possible.

2.2 Dynamic Systems

Consider the dynamic optimization problem of minimizing the functional

$$J(M) = \int_{o}^{t_1} F(U, M, t)\, dt \qquad (2.1)$$

subject to the side constraints given by

$$\dot{U} = G(U, M, t); \quad U(0) = U_o \tag{2.2}$$

$$R(U, M, t) \geq 0 \tag{2.3}$$

$$\psi_o(t_1) = 0 \tag{2.4}$$

$$\psi\left(U(t_1), t_1\right) = 0 \tag{2.5}$$

where

U = n dimensional state vector

M = m dimensional control vector

F = scalar function of class C^2

G = vector function of dimension n with components of class C^2

R = vector function of dimension r with components of class C^2

ψ = vector function of dimension $q(\leq n)$ with components of class C^2

In order to employ multilevel techniques, the system $\left((2.1)\text{-}(2.5)\right)$ is decomposed by partitioning the state vector U into N subvectors $U_1, \ldots, U_N$. In order to attain independent subsystems, a pseudo-control vector S_j is substituted for variables U_i, M_i $(i \neq j)$ (or functions thereof) appearing in the j^{th} subsystem. Assuming that (2.1), (2.3), and (2.5) are naturally separable into subsystems,† the optimization problem can be restated as minimizing

$$J(M) = \sum_{j=1}^{N} \int_o^{t_1} F_j(U_j, M_j, t)dt \tag{2.6}$$

†This assumption can be relaxed (see References 6 and 45) but this would not add to the present discussion.

with side constraints

$$\dot{U}_j = G_j(U_j, M_j, S_j, t) \; ; \quad U_j(0)=U_{jo} \tag{2.7}$$

$$R_j(U_j, M_j, t) \geq 0 \tag{2.8}$$

$$\psi_o(t_1) = 0 \tag{2.9}$$

$$\psi\left(U_j(t_1), t_1\right) = 0 \tag{2.10}$$

and coupling constraints

$$\theta_j(U_j, M_j, t) = S_i \qquad (i \neq j) \tag{2.11}$$

$$j = 1, \ldots, N.$$

where

θ_i = vector function of dimension h_i with components of class C^2

The elements of θ_i in (2.11) represent inputs to subsystems other than the i^{th}. A convenient notation for (2.11) is

$$\theta_j(U_j, M_j, t) = S^j \tag{2.12}$$

A representative subsystem is shown schematically in Figure 2.1.

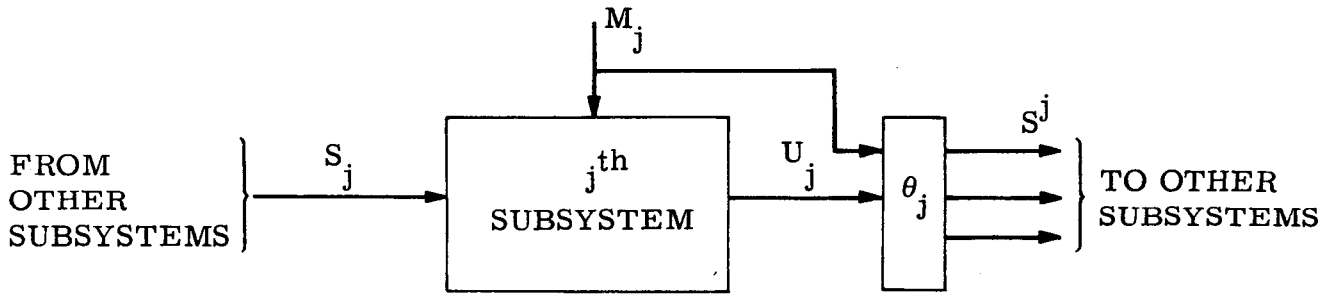
The Hamiltonian corresponding to $\left((2.6)\text{-}2.12)\right)$ can now be written as:

$$H = \sum_{j=1}^{N} \left\{ F_j + \lambda_j^T G_j + \rho_j^T \left(\theta_j - S^j\right) + \sum_{i=1}^{r_j} \mu_{ij} \left(R_{ij} - \xi_{ij}^2\right) \right\} \tag{2.13}$$

where $\lambda_j, \rho_j, \mu_{ij}$ are appropriate Lagrange multipliers of dimension n_j, h_j, and 1 respectively and are assumed to exist. The scalars ξ_{ij} correspond to the slack variables of Valentine.[60] The canonical Euler equations[9, 24] immediately yield the necessary conditions

$$\dot{U}_j = H_{\lambda_j} \tag{2.14}$$

$$\dot{\lambda}_j = - H_{U_j} \tag{2.15}$$



REPRESENTATIVE SUBSYSTEM

FIGURE 2.1

$$H_{M_j} = 0 \tag{2.16}$$

$$\mu_{ij} R_{ij} = 0 \qquad i = 1, \ldots, r_j \tag{2.17}$$

$$H_{\rho_j} = 0 \tag{2.18}$$

$$H_{S_j} = 0 \tag{2.19}$$

$$j = 1, \ldots, N$$

The transversality, Erdmann, Clebsch, and Weierstrass necessary conditions are likewise easily determined.[6, 45]

To realize the benefit from a reduction in dimensionality will require necessary conditions for the satisfaction of $\left(2.14\right)$-$(2.19)\big)$ which are dependent on individual subsystems only. Since this is not possible for all of $\big((2.14)$-$(2.19)\big)$, one or more of these conditions can be relaxed within the (first-level) subsystems and satisfied by a second-level control. The second-level variables are treated as parameters at the first-level. The multilevel approach is to guess these parameters and solve all first-level subsystems. The subsystem results are of course nonoptimal for the overall problem. These first-level results are then transferred to the second level where new parameter values are determined. These values are then passed to the first-level subsystems and the entire process repeated until all necessary conditions are satisfied simultaneously.

Defining the Hamiltonian for the j^{th} subsystem as

$$H_j = F_j + \lambda_j^T G_j + \rho_j^T \left(\theta_j - S^j\right) + \sum_{i=1}^{r_j} \mu_{ij} \left(R_{ij} - \xi_{ij}^2\right) \tag{2.20}$$

and treating the $S_j (j=1, \ldots, N)$ as parameters† yields the following first-level necessary conditions

$$\begin{aligned} \dot{U}_j &= (H_j)_{\lambda_j} \\ \dot{\lambda}_j &= -(H_j)_{U_j} \\ (H_j)_{M_j} &= 0 \\ \mu_{ij} R_{ij} &= 0 \qquad i = 1, \ldots r_j \\ (H_j)_{\rho_j} &= 0 \\ j &= 1, \ldots, N \end{aligned} \tag{2.21}$$

The condition remaining to be satisfied by a second-level control is (2.19). This method is termed "feasible" since the coupling constraints (2.12) are always satisfied by each subsystem. At any point in the iteration, the solution, while nonoptimal, does represent the actual overall system.

In order to put (2.13) in a form suitable for the determination of (2.19), rearrange the set of terms $\left\{ \rho_j^T S^j \,|\, j=1, \ldots, N \right\}$ into the form $\left(\rho^j\right)^T S_j$ $(j=1, \ldots, N)$. The second-level necessary condition is now

$$(H_j)_{S_j} - \rho^j = 0 \qquad j=1, \ldots, N \tag{2.22}$$

where the ρ^j are treated as parameters by the second-level controller. One method of satisfying (2.22) is by a gradient controller of the form

---

†The $S^j (j=1, \ldots N)$ are thereby parameters also since $\left\{ S_j \,|\, j=1, \ldots, N \right\} = \left\{ S^j \,|\, j=1, \ldots, N \right\}$

$$\frac{dS_j(t,\sigma)}{d\sigma} = a\left((H_j)_{S_j} - \rho^j\right) \tag{2.23}$$

where σ represents iteration time and a is a scalar to be determined. It can be shown[45] that if J is to be maximized, $a>0$; and if J is to be minimized, $a<0$. Furthermore (2.23) will converge to the desired solution (2.22) if $S_j(t,0)$ is suitably chosen. One restriction of this method is that each subsystem must have at least as many degrees of freedom (controls) as the number of coupling constraints (2.12) added to it.

Some added difficulty may be encountered in obtaining the subsystem solution when θ_j is a function of U_j only. In this case the coupling constraint is essentially a state equality constraint along the entire path and must therefore be differentiated until the control appears explicitly.[8, 13] The corresponding necessary conditions are different than (2.21) and are generally more difficult to solve.

Redefining the Hamiltonian for the j^{th} subsystem as

$$H_j = F_j + \lambda_j^T G_j + \rho_j^T \theta_j - (\rho^j)^T S_j + \sum_{i=1}^{r_j} \mu_{ij}\left(R_{ij} - \xi_{ij}^2\right) \tag{2.24}$$

where the subsystem criterion function is now

$$\tilde{J}(M_j) = \int_o^{t_1} \left[F_j + \rho_j^T \theta_j - (\rho^j)^T S_j\right] dt \tag{2.25}$$

and treating the ρ_j (j=1, . . . , N) (and therefore ρ^j) as parameters, yields the following first-level necessary conditions

$$
\begin{aligned}
\dot{U}_j &= \left(H_j\right)_{\lambda_j} \\
\dot{\lambda}_j &= \left(H_j\right)_{U_j} \\
\left(H_j\right)_{M_j} &= 0 \\
\mu_{ij} R_{ij} &= 0 \qquad i = 1, \ldots, r_j \\
\left(H_j\right)_{S_j} &= 0 \\
j &= 1, \ldots, N
\end{aligned}
\tag{2.26}
$$

The condition remaining to be satisfied by a second-level control is now (2.18). This method is termed "nonfeasible" since the coupling constraints are not satisfied during the course of the iteration but only when the second level has converged.

The second-level necessary conditions can now be stated from (2.20) as

$$
\theta_j - S^j = 0 \qquad j = 1, \ldots, N \tag{2.27}
$$

where the S^j are now treated as parameters. Again using the gradient controller at the second level yields

$$
\frac{d\rho_j(t, \sigma)}{d\sigma} = a\left(\theta_j - S^j\right) \qquad j = 1, \ldots, N \tag{2.28}
$$

In this case, it can be shown[45] that at the optimum a saddle value results between (2.26) and (2.27). Thus when J(M) is to be maximized, $a < 0$; and when J(M) is to be minimized, $a > 0$. Convergence of (2.28) to (2.27) is guaranteed if the $\rho_j(t, 0)$ are suitably chosen, and if each subsystem is formulated such that it has at least a local minimum with respect to both M_j and S_j. The latter condition is quite restrictive. To circumvent this difficulty, Bauman[6]

has proposed an alternative procedure for satisfying (2. 27) using the second-variational techniques introduced by Breakwell.[10]

An alternative to these methods is to assign both (2. 18) and (2. 19) to the second-level controller. In this case both ρ_j and S_j are treated as parameters in the subsystem and the second-level necessary conditions are given by (2. 22) and (2. 27). Takahara[59] has proposed this method for linear dynamic systems in which all subsystems are solved before determining new parameter values (ρ_j, S_j) at the second level. A variation of this approach as suggested here is to determine and use new values of the parameters ρ_j and S_j as soon as the required first-level data is available. This approach is analogous to the Gauss-Seidel technique[61] for solving linear algebraic equations and as such will be termed the Gauss-Seidel second-level controller. It is of interest to examine the conditions under which convergence is guaranteed by this method. This question has not been completely answered; however, some results are available. For static systems (for which the Gauss-Seidel iterative procedure is intended),[50, 61] the solution U to the linear system of equations

$$AU + F = M \tag{2.29}$$

converges for any initial guess U^o if all the eigenvalues of a matrix C have modulus less than unity where

A = n x n matrix

F, U, M = n vectors

$C = -\Lambda^{-1}\Delta$

Λ = lower triangular part of A

Δ = upper triangular part of A (above diagonal)

This condition can be shown to hold whenever A is Hermitian, positive definite.[50] Varga[62] has shown that the A matrix resulting from the discretization of the elliptic partial differential equation

$$
\begin{aligned}
&-\left(K_1(x_1, x_2)U_{x_1}\right)_{x_1} - \left(K_2(x_1, x_2)U_{x_2}\right)_{x_2} + \sigma(x_1, x_2) = m(x_1, x_2) \\
&u(x_1, x_2) = f(x_1, x_2) \qquad\qquad x_1, x_2 \in \Omega_b \qquad\qquad (2.30) \\
&K_1 > 0,\ K_2 > 0,\ \sigma \geq 0
\end{aligned}
$$

is real, symmetric, with positive diagonal entries and nonpositive off-diagonal entries. Moreover, if the vector of boundary mesh points F is written separately as in (2.29), A is irreducibly diagonally dominant so that A is positive definite.† Hence the state equation of the form (2.29) resulting from the discretization of elliptic partial differential equations will converge by the Gauss-Seidel method for given M and F. The optimization problem, however is to solve an augmented system of equations consisting of (2.29) plus the stationarity condition on the control M, and the state variables U (assuming no inequality constraints). For the case of a quadratic criterion function, this augmented system has the form

$$A'Z = B \qquad (2.31)$$

where

Z^T = [U, λ], a 2n dimensional vector

λ = Lagrange multiplier, n vector

A' = 2n x 2n matrix

B = 2n dimensional vector

†An identical result holds for the negative of the A matrix arising from the parabolic equations discussed in Chapter 3.

However, A' is no longer necessarily symmetric and therefore no longer positive definite although it may still contain some other properties of A; namely, real with positive diagonal entries and non-positive off-diagonal entries. Thus convergence is not guaranteed for (2.31). The fact that the Gauss-Seidel procedure as proposed for multilevel control utilizes subsystems composed of aggregated elements of U (rather than single elements) is not expected to change the convergence properties given above.

Antosiewicz[2] gives a convergence criterion for the iterative solution of nonlinear static systems of the form

$$Z = f(Z) \tag{2.32}$$

where f is a vector function defined over a normed linear space R^n and satisfying Lipschitz conditions with respect to Z. The notation corresponds to the augmented system (2.31) obtained for linear systems. Namely, the set of equations

$$Z^{k+1} = f(Z^k) \tag{2.33}$$

will converge to the solution of (2.32) for Z^o suitably chosen if

$$||L|| < 1 \tag{2.34}$$

where L is the matrix of Lipschitz constants. Naturally, convergence depends upon the norm chosen.

Kolmogorov[36] gives a convergence criterion for infinite dimensional function spaces based on the principle of contraction mapping. Consider the set of differential equations

$$\dot{Z} = f(Z, t) \tag{2.35}$$
$$Z(t_o) = Z_o$$

where f is a continuous vector function on the space R^{n+1} and satisfies a Lipschitz condition with respect to Z, namely,

$$|f(Z^1, t) - f_i(Z^2, t)| \leq L \max \left\{ |z_i^1 - z_i^2| \; ; \; 1 \leq i \leq n \right\} \qquad (2.36)$$

In the above $Z^T = [z_1, \ldots, z_n]$ and M is considered as a parameter. Letting T be the integral operator arising from (2.35), the set of iterative equations

$$Z^{k+1} = T(Z^k) \qquad (2.37)$$

will converge to the solution of (2.35) if

$$L(t - t_o) < 1 \qquad (2.38)$$

and Z^o is suitably chosen.

Takahara[59] has shown a sufficient convergence condition for the iterative solution of linear dynamic systems using decomposition and a second-level controller analogous to the Jacobi method[66] of solving discrete representations of elliptic partial differential equations.† This result, based on (2.38) requires that the norm of a complex function of inverse Laplace transforms be less than unity. In most cases this criterion is too complex to be of any practical value.

Various sufficiency conditions relating to the convergence of the Gauss-Seidel second-level controller have been reviewed above. That these conditions (in particular (2.38)) are overly restrictive when applied to the augmented system of equations arising in linear dynamic optimization problems is shown by example in Appendix A. In practice, the Gauss-Seidel second-level controller was found to have excellent convergence properties when applied to both linear and nonlinear problems of the type arising from semidiscrete approximations to parabolic partial differential equations. These convergence properties are further discussed in Chapters 4 and 6.

†The Gauss-Seidel method for treating these problems is known to converge exactly twice as fast as the Jacobi method.[64]

2.3 Static Systems

Consider the static optimization (minimization) problem given below

$$\min_{M} \; F(U, M) \tag{2.39}$$

such that

$$G(U, M) = 0 \tag{2.40}$$

and

$$R(U, M) \geqq 0 \tag{2.41}$$

where

U = state vector in E^n

M = control vector in E^m

F = scalar function in C^2

G = n vector of functions in C^2

R = r vector of functions in C^2

In order to treat $\big((2.39)\text{-}(2.41)\big)$ by multilevel techniques, it is again necessary to formulate independent subsystems by substituting pseudo-control variables S_j for all variables (or functions) entering the j^{th} subsystem from other subsystems. Assuming that (2.39) and (2.41) can be written in separable form, the above optimization problem can be stated as

$$\min_{M_j} \sum_{j=1}^{N} F_j(U_j, M_j) \tag{2.42}$$

such that

$$G_j(U_j, M_j, S_j) = 0 \tag{2.43}$$

$$R_j(U_j, M_j) \geqq 0 \tag{2.44}$$

with the coupling constraints

$$\theta_j(U_j, M_j) = S_i \qquad (i \neq j) \tag{2.45}$$

where θ_j is a vector function of dimension h_j having elements which are inputs to other subsystems. A convenient notation for (2.45) is

$$\theta_j(U_j, M_j) = S^j \tag{2.46}$$

The Lagrangian L associated with $\left((2.42)\text{-}(2.46)\right)$ can be written as

$$L = \sum_{j=1}^{N} \left\{ F_j + \lambda_j^T G_j + \rho_j^T \left(\theta_j - S^j\right) + \sum_{i=1}^{r_j} \mu_{ij} \left(R_{ij} - \xi_{ij}^2\right) \right\} \tag{2.47}$$

where $\lambda_j, \rho_j, \mu_{ij}$ are Lagrange multipliers of appropriate dimension and the ξ_{ij} are slack variables. The necessary conditions for a minimum are

$$L_{U_j} = L_{\lambda_j} = L_{M_j} = L_{S_j} = L_{\rho_j} = 0 \tag{2.48}$$

$$\mu_{ij} R_{ij} = 0 \qquad i = 1, \ldots, r_j$$

$$j = 1, \ldots, N$$

The remaining development of multilevel control for static systems is analogous to the discussion in Section 2.2 and will therefore not be given here. For a detailed discussion see References 6, 12, 41, and 42.

CHAPTER 3

DISCRETIZATION AND DECOMPOSITION OF DISTRIBUTED PARAMETER SYSTEMS

3.1 Introduction

The optimal control of distributed parameter systems has in recent years received considerable attention as a subject for research. The pioneering work in the field has been done by Butkowski[15,16] and Wang[65] beginning in 1960. These researchers have extended the theory of optimal control to include distributed fields by using Pontryagin's maximum principle and Bellman's dynamic programming respectively. However, this theory is still incomplete and the computational problems severe. Perhaps for these reasons the only applications which have appeared in the literature treat linear partial differential equations in one space dimension and with some form of quadratic cost functional.[11,56]

In solving these problems some form of approximation must inevitably be made. In the rare cases where a closed form solution is possible, it takes the form of an infinite series. Otherwise, the (numerical) determination of certain Green's functions and the solution of a two-point boundary value problem is required.

The approach proposed for this dissertation is to treat a lumped approximation in the spatial domain of the distributed parameter system. Although this approach admits the application of a larger body of control theory, several questions arise; namely, (1) How can boundary conditions and particularly boundary control be treated, and (2) Can computational techniques be devised to handle the (in general) large number of interacting differential equations required to obtain a sufficiently accurate approximation to the distributed

system? If these two questions can be answered satisfactorily, the present class of optimal control problems for distributed parameter systems which can be treated effectively can be considerably enlarged. These questions and the important related question concerning the convergence of the solution of the approximate system to the distributed system will be treated in this chapter.

3.2 Problem Statement

The optimal control problem for the case of distributed control considered here is to minimize the functional

$$J(m) = \int_{\bar{\Omega}} P_o\left(u(X,t_1), X, t\right)d\Omega + \int_{\bar{\Omega}} \int_{t_o}^{t_1} P_1\left(u(X,t),\ m(X,t)\right)dt\ d\Omega \qquad (3.1)$$

subject to the scalar partial differential equation side constraint

$$\frac{\gamma u}{\gamma t}(X,t) = \mathcal{G}\left(u(X,t), m(X,t), X, t\right) \qquad X \in \Omega,\ t \geq t_o \qquad (3.2)$$

with boundary conditions

$$\alpha(X,t)u(X,t) + \beta(X,t)\frac{\gamma u}{\gamma n}(X,t) = f(X,t) \qquad X \in \Omega_b,\ t \geq t_o \qquad (3.3)$$

and initial conditions

$$u(X,t_o) = u_o(X) \qquad X \in \Omega \qquad (3.4$$

In (3.3), n indicates a direction normal to the boundary. The control $m(X,t)$ may also be required to satisfy inequality constraints of the form

$$R\left(u(X,t),\ m(X,t),\ X,t\right) \geq 0 \qquad X \in \Omega, t \geq t_o \qquad (3.5)$$

and terminal constraints of the form

$$\psi_o(t_1) = 0$$

$$\psi\left(u(X,t_1), X\right) = 0 \qquad X \in \Omega \qquad (3.6)$$

In the above equations the following symbols are defined

$u(X,t) = u(x_1, x_2, \ldots, x_n, t)$, a scalar state variable

$m(X,t) = m(x_1, x_2, \ldots, x_n, t)$, a scalar control variable

Ω = a given finite (connected) region in Euclidean n-space and Ω_b is the boundary of Ω

$\mathcal{G}$ = spatially varying differential operator on u (up to second order) which may include parameters which are functions of u, m, X, or t

$\bar{\Omega}$ = closure of Ω

P_o, P_1 = real-valued functions of class C^2 on t and piecewise C^1 on Ω

α, β, f = real-valued functions, piecewise C^1 on Ω_b and C^2 on t which satisfy

$$\alpha(X,t) \geq 0 \qquad X \in \Omega_b$$
$$\beta(X,t) \geq 0 \qquad t \geq t_o$$
$$\alpha(X,t) + \beta(X,t) > 0 \qquad (3.7)$$

R = a vector-valued function of dimension r with components R_i which are of class C^2 on t and and piecewise C^1 on Ω

ψ = a vector-valued function of dimension q with components ψ_i which are of class C^2 on t and piecewise C^1 on Ω

It is assumed that the functions R and ψ are consistent with the boundary conditions in (3.3).

Consideration of a scalar state variable u excludes the class of problems defined by hyberbolic and biharmonic partial differential equations. The reason for their exclusion will be discussed in a later section. Note that the first of Equations (3.6) restricts attention to the class of problems having a fixed terminal time.

The optimal control problem for the case of boundary control (control variables contained only in boundary conditions) is not easily treated in the generality shown above. In this case consider the following criterion functional

$$J(f) = \int_{\Omega} \int_{t_o}^{t_1} P_1\left(u(X, t), X, t\right) dt\, d\Omega \tag{3.8}$$

subject to the partial differential side constraints

$$\frac{\partial u}{\partial t}(X, t) = \mathcal{G}\left(u(X, t), X, t\right) \qquad X \in \Omega,\ t \geq t_o \tag{3.9}$$

and boundary and initial conditions given by (3.3) and (3.4). The inequality constraints become

$$R\left(f(X, t), X, t\right) \qquad X \in \Omega_b,\ t \geq t_o \tag{3.10}$$

and the terminal conditions are given by (3.6). Cases where the boundary control f appears explicitly in the criterion functional, can also be treated if the spatial integration of f is taken only over the boundary domain. Several such examples are given in Chapter 5.

3.3 A Semidiscrete Approximation

The optimal control problems posed above can be solved by the theory of Chapter 2 if suitable approximations can be found. The approach taken here is to discretize the spatial variables by defining a vector

$$X_{\underline{i}} = \left(i_1(\Delta x_1), i_2(\Delta x_2), \ldots, i_j(\Delta x_j), \ldots, i_n(\Delta x_n)\right)^T \tag{3.11}$$

which in effect places a grid on the region Ω. Here the elements of $\underline{i} = \left[i_1, i_2, \ldots, i_j, \ldots, i_n\right]^N$ are intergers defined by

$$i_j = 0, 1, \ldots, \frac{(x_j)_{max} - (x_j)_{min}}{\Delta x_j} = N_j.$$

Denoting the set of points defined by (3.11) by # and following Young[66] and Varga[62] the following terms can be defined: mesh point-a point in # ; interior mesh point-a point in $\# \cap \Omega$; boundary mesh point-a point in $\# \cap \Omega_b$; exterior mesh point-a point not belonging to $\# \cap \overline{\Omega}$; regular point-a point belonging to $\# \cap \overline{\Omega}$ and such that all adjacent points also belong to $\# \cap \overline{\Omega}$; irregular point-a point belonging to # which is not a regular point.

Young[64] shows for the case $\beta = 0$ (see (3.3)) that irregular mesh points may be treated as regular mesh points by defining a boundary (pseudo) mesh point at the intersection of the boundary and the line segment connecting the irregular point with each exterior point. (See Figure 3.1.) These points can be denoted by

$$X_i = \left(i_1(\Delta x_1'), i_2(\Delta x_2'), \ldots, i_n(\Delta x_n')\right)^T$$

where

$$\Delta x_j' = e_j \Delta x_j \qquad j = 1, \ldots, n$$

and

$$0 \leq e_j \leq 1$$

Thus when $e_j = 1$, $(j = 1, \ldots, n)$ the point is a regular mesh point.

Varga[62] treats the case $\beta > 0$ for a symmetric linear operator $\mathcal{G}$ by approximating the boundary by line segments connecting boundary (pseudo) mesh points and using the identity

$$\frac{\partial u}{\partial n} = u_x \cos\theta + u_y \sin\theta \tag{3.12}$$

to obtain the required approximation (θ = angle between the linear boundary approximation and the positive x_1 axis). It is easily seen that a nonsymmetric operator $\mathcal{G}$ can be treated if the boundary is composed of orthogonal line segments, i.e., $\sin 2\theta = 0$. The

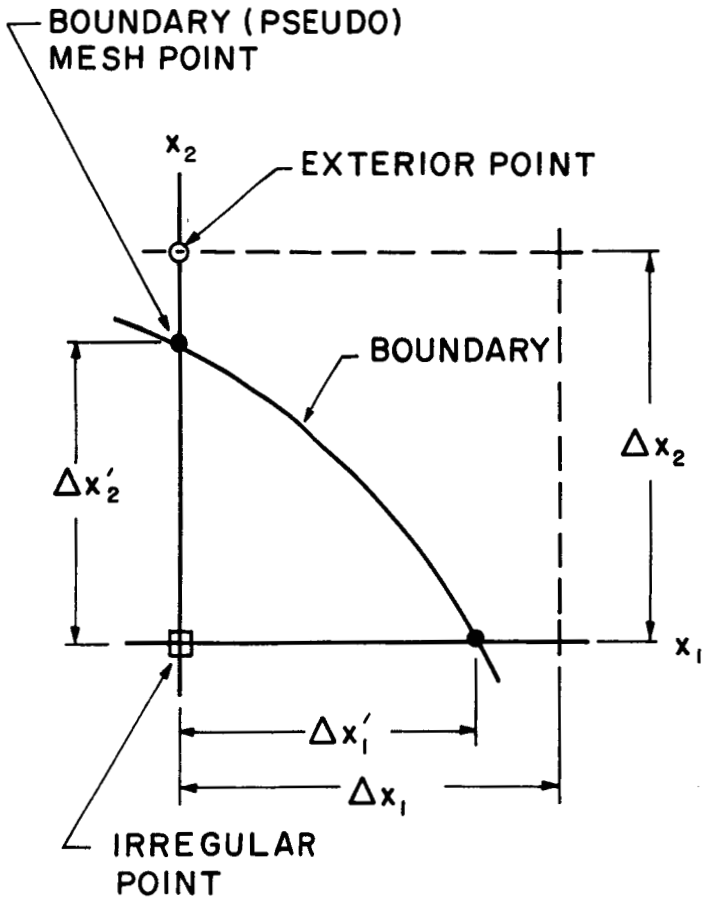


ILLUSTRATION OF AN IRREGULAR POINT

FIGURE 3.1

following development then requires either (a) $\beta = 0$, b) $\sin 2\theta = 0$, or c) $\mathcal{G}$ symmetric with respect to all x_i, $i = 1, \ldots, n_j$; for simplicity consider the case $\beta = 0$.

Since the operator $\mathcal{G}$ is at most second order, it can be approximated as follows

$$\mathcal{G}\left(u(X_{\underline{i}}, t), m(X_{\underline{i}}, t), X_{\underline{i}}, t\right) = \tag{3.13}$$
$$G_{\underline{i}}\left(u_{\underline{i}}(t), m_{\underline{i}}(t), t, u_{\underline{i}\pm I_1}(t), u_{\underline{i}\pm I_2}(t), \ldots, u_{\underline{i}\pm I_n}(t)\right)$$

where $I_k = \left\{\underline{i} \,|\, \underline{i} = 0 \text{ except for the } k^{th} \text{ element which } = 1\right\}$, $\underline{i}$ ranges over all regular points, and the functions $G_{\underline{i}}$ are assumed to be real valued and of class C^2. Applying the definition (3.11) to Equations (3.1) to (3.6) for the case of distributed control results in the following set of equations for the discretized system:

$$J(m_{\underline{i}}) = \sum_{\underline{i} \in \Omega} P_{oi}\left(u_{\underline{i}}(t_1), t_1\right) \Delta\Omega + \sum_{\underline{i} \in \Omega} \int_{t_o}^{t_1} P_{1\underline{i}}\left(u_{\underline{i}}(t), m_{\underline{i}}(t), t\right) dt\, \Delta\Omega + \sum_{i \in \Omega} \tau_i \tag{3.14}$$

where $\Delta\Omega = \Delta x_1 \cdot \Delta x_2 \cdot \ldots \cdot \Delta x_n$

$\tau_{\underline{i}}$ = truncation error at $X_{\underline{i}}$

$$\frac{du_{\underline{i}}}{dt} = G_{\underline{i}}\left(u_{\underline{i}}(t), m_{\underline{i}}(t), t, u_{\underline{i}\pm I_1}(t), u_{\underline{i}\pm I_2}(t), \ldots, u_{\underline{i}\pm I_n}(t)\right) + v_{\underline{i}} \tag{3.15}$$
$$X_{\underline{i}} \in \Omega, \; t \geq t_o$$

where $v_{\underline{i}}$ = truncation error at $X_{\underline{i}}$

$$\alpha_{\underline{i}}(t) u_{\underline{i}}(t) = f_{\underline{i}}(t) \qquad X_{\underline{i}} \in \Omega_b, \; t \geq t_o \tag{3.16}$$

$$u_{\underline{i}}(t_o) = u_{o\underline{i}} \qquad X_{\underline{i}} \in \Omega \tag{3.17}$$

$$R_{\underline{i}}\left(u_{\underline{i}}(t), m_{\underline{i}}(t), t\right) \geq 0 \qquad X_{\underline{i}} \in \overline{\Omega}, \; t \geq t_o \tag{3.18}$$

$$\psi_o(t_1) = 0$$
$$\psi_{\underline{i}}\, u_{\underline{i}}(t_1) = 0 \qquad X_{\underline{i}} \in \Omega \tag{3.19}$$

Minimizing (3.14) is equivalent to minimizing

$$\frac{J(m_{\underline{i}})}{\Delta\Omega} = \sum_{\underline{i}\epsilon\Omega} \left\{ P_{o\underline{i}}\left(u_{\underline{i}}(t_1), t_1\right) + \int_{t_o}^{t_1} P_{1\underline{i}}\left(u_{\underline{i}}(t), m_{\underline{i}}(t), t\right) dt + \tau_{\underline{i}} \right\} + \sum_{\underline{i}\epsilon\Omega_b} \left\{ P_{oi}\left(u_{\underline{i}}(t_1), t_1\right) + \int_{t_o}^{t_1} P_{1\underline{i}}\left(u_{\underline{i}}(t), m_{\underline{i}}(t), t\right) dt + \tau_{\underline{i}} \right\} \tag{3.20}$$

It is easily seen that the side constraints (3.16) and (3.18) completely specify $m_{\underline{i}}$ on the boundary. Thus if $R_i = 0$, m_i is determined by

$$R_{\underline{i}}\left(\frac{f_{\underline{i}}(t)}{\alpha_{\underline{i}}(t)}, m_{\underline{i}}(t), t\right) = 0 \qquad X_{\underline{i}} \epsilon \Omega_b \tag{3.21}$$

and if $R_i > 0$ a necessary condition for a minimum of (3.20) is that

$$\frac{\partial P_{1\underline{i}}}{\partial m_{\underline{i}}}\left(\frac{f_{\underline{i}}(t)}{\partial_{\underline{i}}(t)}, m_{\underline{i}}(t), t\right) = 0 \qquad X_{\underline{i}} \epsilon \Omega_b \tag{3.22}$$

where $\frac{\partial P_{1\underline{i}}}{\partial m_{\underline{i}}}$ is the Frechet derivative[44] of $P_{1\underline{i}}$ at $m_{\underline{i}}$

Thus the optimal value of $m_{\underline{i}}$ is completely determined on the boundary and the criterion function can be written as

$$J'(m_{\underline{i}}) = \sum_{\underline{i}\epsilon\Omega} \left\{ P_{o\underline{i}}\left(u_{\underline{i}}(t_1), t_1\right) + \int_{t_o}^{t_1} P_{1\underline{i}}\left(u_{\underline{i}}(t), m_{\underline{i}}(t), t\right) dt + \tau_{\underline{i}} \right\} \tag{3.23}$$

It is not intended that $\tau_{\underline{i}}$ should enter into the optimization, but only that it can be determined to evaluate the worth of the approximation. Note that only in (3.15) do cross coupling terms between spatial mesh points appear and they appear there in a very special way. The arguments in (3.15) corresponding to boundary points can be evaluated using (3.16).

The system of Equations (3.15), (3.17) to (3.19), (3.23) is now in a form quite suitable for decomposition and multilevel

solution as seen in Chapter 2. However, further treatment in the general notation considered here is unwieldy and contributes little to the development. Hence further consideration of the decomposition method is postponed until Section 3.4 where an important special case of a nonlinear parabolic partial differential equation is discussed.

Applying the definition (3.11) to (3.8) by (3.10) for the case of boundary control results in the following set of equations for the discretized system:

$$\frac{J(f_{\underline{i}})}{\Delta\Omega} = \sum_{\underline{i}\in\Omega} \left\{ \int_{t_o}^{t_1} P_{1\underline{i}}\left(u_{\underline{i}}(t), t\right)dt + \tau_{\underline{i}} \right\} \tag{3.24}$$

$$\frac{du_{\underline{i}}(t)}{dt} = G_{\underline{i}}\left(u_{\underline{i}}(t), t, u_{\underline{i}\pm I_1}(t), u_{\underline{i}\pm I_2}(t), \ldots, u_{\underline{i}\pm I_n}(t)\right) + v_{\underline{i}} \tag{3.25}$$

$$X_{\underline{i}}\in\Omega,\ t \geq t_o$$

$$R_{\underline{i}}\left(f_{\underline{i}}(t), t\right) \geq 0 \qquad X \in \Omega_b,\ t \geq 0 \tag{3.26}$$

The boundary condition (3.16) can be used to get the control as an explicit argument in (3.25). In general the optimal control $f_{\underline{i}}^{*}$ may not be unique unless it is assumed (as does Wang[64]) that the boundary control function does not vary along the boundary.

3.4 Discretization and Decomposition-a Special Case

Consider the minimization of the functional

$$J_1\left[m(x,t)\right] = \int_o^1 \int_o^1 \int_{t_o}^{t_1} \left[q(X)\left(u_d(X) - u(X,t)\right)^2 + c(X)m^2(X,t) \right] dt\, dX \tag{3.27}$$

subject to the nonlinear partial differential equation side constraints

$$\nabla_t u(X,t) = K_1(X,u,t)\nabla_{x_1}^2 u + K_2(X,u,t)\nabla_{x_2}^2 u - \sigma(X,t)u + b(X,t)m(X,t) \tag{3.28}$$

where

$$K_1 > 0,\ K_2 > 0,\ \sigma \geq 0,\ X = [x_1 x_2]^T$$

$$0 \leq x_1 \leq 1, \quad 0 \leq x_2 \leq 1, t, \text{ is fixed} \tag{3.28a}$$

with boundary conditions

$$u(X, t) = f(X, t) \qquad X \in \Omega_b,\ t \geq t_o \tag{3.29}$$

and initial conditions

$$u(X, 0) = u_o(X) \tag{3.30}$$

Equations (3.27) to (3.30) represent a fairly general class of problems which can be treated by decomposition. (Inequality constraints can be handled equally well using a technique due to Valentine[60] but are omitted here for clarity of presentation. Inequalities are considered in Chapter 5.) Variants of (3.27) which have appeared in the literature are (a) $c = 0$, minimum deviation from a desired trajectory; (b) $u(X,t) = u(X,t_1)$, minimum terminal error; (c) $q = 0$, $u(X,t_1) = u_1(X)$, minimum effort with fixed terminal conditions; (d),(e), $m(X,t) = 0$ $(X \in \Omega)$, $m(X,t) = f(X,t)$ $(X \in \Omega_b)$, boundary control corresponding to (a) and (b).

To handle (3.28) in its more general form

$$\nabla_t u = \sum_{i=1}^{n} \left(K_i(X, u, t)u_{x_i}\right)_{x_i} - \sigma(X, t)u + b(X, t)m(X, t) \tag{3.31}$$

$$K_i > 0,\ \sigma \geq 0, \quad X = [x_1, \ldots, x_n]^T \tag{3.31a}$$

is straightforward but cumbersome. In fact, a general treatment of the nonlinearity is impossible and K_1, K_2 will be specialized to linear functions once their effect upon the decomposition is made clear.

Using the usual approximation[66] of the second partial derivative, the discrete form of (3.28) is

$$\frac{du_{ij}(t)}{dt} = \frac{1}{h^2}\left\{\left[-2K_{1ij}-2K_{2ij}-h^2\sigma_{ij}\right]u_{ij}+K_{1ij}\left(u_{i+1,j}+u_{i-1,j}\right)\right.$$
$$\left. + K_{2ij}\left(u_{i,j+1}-u_{i,j-1}\right)\right\}+b_{ij}m_{ij}+v_{ij} \qquad (3.32)$$

where

$$K_{1ij} = K_1\left(u_{ij}, X_{ij}, t\right)$$
$$K_{2ij} = K_2\left(u_{ij}, X_{ij}, t\right)$$
$$\sigma_{ij} = \sigma\left(X_{ij}, t\right)$$
$$h = \Delta x_1 = \Delta x_2$$

For the case of a linear system, the semidiscrete equation (3.32) can be shown to be consistent with and to converge to the corresponding partial differential equation (3.28) for h sufficiently small. The definition of these terms, along with this proof for the case of a linear system of the form (3.31) having one space dimension, is given in Appendix B.

Define a natural ordering of interior mesh points over the square region Ω as the sequence of mesh points taken from left to right and bottom to top. Then a vector U having the natural ordering is

$$U = \left(u_1, u_2, \ldots, u_k, u_{k+1}, \ldots u_{2k}, \ldots, u_{k^2}\right)^T \qquad (3.33)$$

where

$$k = \frac{1}{h} - 1$$

and A, the matrix of coefficients, for (3.32) is of the general form

$$\begin{bmatrix} d_{i-1} & \cdots & e_{i-1} & -a_{i-1} & e_{i-1} & 0 & \cdots & d_{i-1} & & \\ 0 & d_i & \cdots & e_i & -a_i & e_i & \cdots & 0 & d_i & \\ & 0 & d_{i+1} & \cdots & 0 & e_{i+1} & -a_{i+1} & e_{i+1} & \cdots & 0 & d_{i+1} \end{bmatrix} \begin{bmatrix} u_{i-1} \\ u_i \\ u_{i+1} \end{bmatrix}$$

where

$$a_i = \frac{1}{h^2}\left(2K_{1j} + 2K_{2i} + h^2\sigma_i\right)$$

$$e_i = \frac{1}{h^2}\left(K_{1i}\right)$$

$$d_i = \frac{1}{h^2}\left(K_{2i}\right)$$

Note that the elements of each row of A, say the i^{th} are functions of only u_i (as well as X_i and t).

Defining a vector

$$F = \left(f_1, f_2 \ldots, f_i, \ldots, f_\gamma\right)^T$$

composed of all boundary mesh points taken in the natural ordering, (3.32) can be written

$$\dot{U}(t) = A(U, t)\,U(t) + B(t)M(t) + G(U, t)\,F(t) + V(t) \quad (3.35)$$

$$U(0) = U_o$$

where

$A = k^2 \times k^2$ matrix of form (3.34)

$B = k^2 \times k^2$ diagonal matrix with elements b_i

$G = k^2 \times \gamma$ matrix

$M = k^2$ vector with elements m_i

The general element of the truncation error vector can be determined from a Taylor Series expansion as

$$v_i = -\frac{1}{12} h^2 \left[K_{1i} \nabla^4_{x_{1i}} u_i + K_{2i} \nabla^4_{x_{2i}} u_i \right]$$

Having determined this expression for evaluating the truncation error, V(t) will not be considered further.

By using the natural ordering defined above the criterion function (3.37) can be discretized as

$$J_1[m_i(t)] = \sum_{i=1}^{k^2} \left\{ h^2 \int_{t_o}^{t_1} \left[q_i \left(u_{di} - u_i(t) \right)^2 + c_i m_i^2(t) \right] dt + \tau_i \right\} \qquad (3.37)$$

The truncation error τ_i can be evaluated as

$$\tau_i = h^3 \int_{t_o}^{t_1} \left[\left. \frac{\partial P_i}{\partial x_1} \right|_i (t) + \left. \frac{\partial P_i}{\partial x_2} \right|_i (t) \right] dt \qquad (3.38)$$

where P_i is the expression in brackets in (3.37). In order to evaluate τ_i and v_i as the computation progresses, the derivative terms must be suitably approximated.

Since no coupling or nonlinear terms appear in (3.37) this equation may be written in more compact notation as

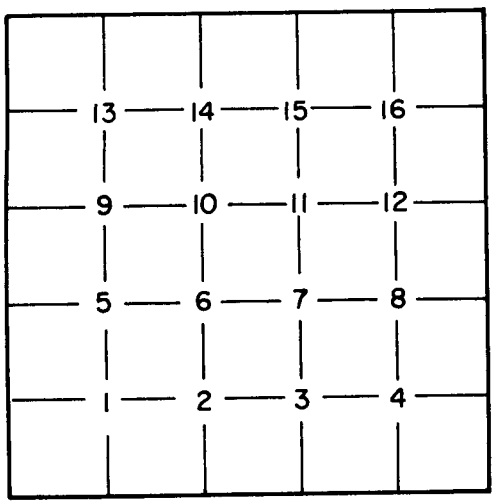
$$J[M] = \frac{1}{h^2} J_1[M(t)] = \int_{t_o}^{t_1} \left[(U_d - U)^T Q (U_d - U) + M^T C M \right] dt \qquad (3.39)$$

where U_d, U, and M are vectors of dimension k^2 and Q, C are $k^2 \times k^2$ diagonal matrices.

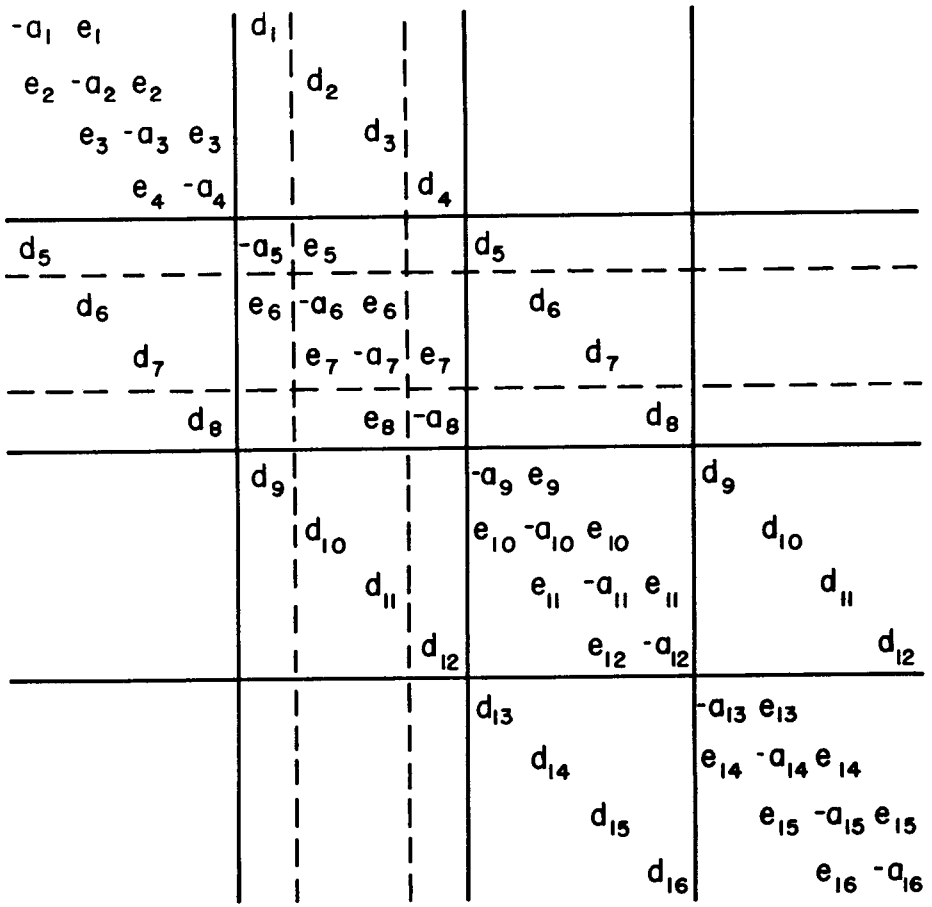
In order to minimize (3.39) subject to (3.35), some decomposition technique seems warranted, especially for large values of k. Various system decompositions will now be discussed with regard to the application of three of the second-level controllers presented in Chapter 2; namely, feasible, nonfeasible and Gauss-Seidel.

Decomposition merely involves partitioning the state vector U in some convenient way and then introducing pseudo-control vectors S to achieve subsystem independence. It is obviously desirable to make use of any "natural decomposition" which the system may afford by a careful examination of the A matrix. For definiteness in discussion, the A matrix is shown explicitly in Figure 3.2 for the case k = 4. One convenient decomposition is to consider each row of the region Ω as a subsystem. The subsystem matrices A_j are then given by the block diagonal matrices in Figure 3.2 and the coupling constraints by the diagonals d_i. If subsystems of this form are too large to be conveniently handled, a further partitioning can be made as shown by the dotted lines in Figure 3.2. In this case, the coupling constraints are more unwieldy, as will be shown The choice of an optimum subsystem size may depend on many factors for instance, (a) form of the A matrix, (b) accuracy desired, (c) computer capacity and speed (d) availability of programs capable of handling certain size problems. It should be noted that for problems having only one space dimension d_i equals zero and the A matrix is of tridiagonal form† (all zeros except the main diagonal and the

†Gelfand[24] calls this type of matrix a Jacobi matrix.



THE Ω REGION



THE A MATRIX CORRESPONDING TO Ω

THE A MATRIX

FIGURE 3.2

two adjacent diagonals). Correspondingly, if the problem has three space dimensions, the A matrix contains four symmetrically spaced diagonals in addition to the tridiagonal band.

First consider the decomposition first mentioned above in which each subsystem corresponds to a row in the region Ω. This partitioning of U yields

$$U^T = \left[U_1^T, U_2^T, \ldots, U_j^T, \ldots, U_N^T\right] \tag{3.40}$$

where

$$U_j^T = \left(u_{p_j}, u_{p_j+1}, \ldots, u_{q_j}\right) \qquad p_j = (j-1)k+1,\ q_j = jk$$

Then

$$\dot{U}_j = A_j(U_j, t)U_j + B_j M_j + D_{1j}(U_j, t)S_j + G_j(U_j, t)F_j \tag{3.41}$$
$$U_j(0) = U_{oj}$$

where $D_{1j} = \left[D_j \mid D_j\right]$, a $k \times 2k$ matrix

S_j = pseudo-control vector of dimension 2k

D_j = $k \times k$ diagonal matrix with elements $\left(d_{p_j}, \ldots, d_{q_j}\right)$

Because of the form of D_{1j}, (3.41) can be written equivalently

$$\dot{U}_j = A_j(U_j, t)U_j + B_j M_j + D_j(U_j, t)(S_j^u + S_j^l) + G_j(U_j, t)F_j \tag{3.42}$$
$$U_j(0) = U_{oj}$$

where S_j^u = k-dimensional vector, the upper half of S_j

S_j^l = k-dimensional vector, the lower half of S_j

The coupling constraint required by this decomposition is

$$S_j^T = \left[U_{j-1}^T, U_{j+1}^T \right] \tag{3.43}$$

The set of constraints (3.43) can also be written

$$\begin{aligned} U_j &= S_{j-1}^l \\ U_j &= S_{j+1}^u \end{aligned} \tag{3.44}$$

From (3.39) the subsystem criterion function becomes

$$J_j(M_j) = \int_{t_o}^{t_1} \left[\left(U_{dj} - U_j \right)^T Q_j \left(U_{dj} - U_j \right) + M_j^T C_j M_j \right] dt \tag{3.45}$$

where

$$J(M) = \sum_{j=1}^{N} J_j(M_j)$$

In the event that cross coupling terms entered into the criterion function, pseudo-controls could be introduced there also to maintain subsystem independence.

By adding (3.44) with vector Lagrange multipliers ρ_j, ν_j of dimension k, the Hamiltonian becomes

$$\begin{aligned} H = \sum_{j=1}^{N} H_j = \sum_{j=1}^{N} \Big\{ & \left(U_{dj} - U_j \right)^T Q_j \left(U_{dj} - U_j \right) + M_j^T C_j M \\ & + \lambda_j^T (A_j U_j + B_j M_j + D_{1j} S_j + G_j F_j) + \rho_j^T \left(U_j - S_{j-1}^l \right) \\ & + \nu_j^T \left(u_j - S_{j+1}^u \right) \Big| \end{aligned} \tag{3.46}$$

where

$$\rho_1 = \nu_N = S_1^u = S_N^l = 0$$

In its present form (3.46) is suitable for solution only with a Gauss-Seidel second-level controller. The feasible method is excluded

since 2k interconnecting constraints were added to each subsystem (except 1 and n) when only k degrees of freedom were available. In order to insure convergence of the nonfeasible controller, at least local subsystem minima must exist. A necessary condition (the Clebsch condition) for this is that a certain Hessian matrix be non-negative definite, i. e.,

$$\Pi^T \nabla^2_{\xi_j} H_j \Pi \geq 0 \tag{3.47}$$

for every $\Pi \neq 0$

where $\xi_j^T = \left[M_j^T, S_j^T \right]$

Equation (3. 47) is clearly not satisfied since S_j appears linearly in (3. 46). In addition, the latter condition leads to singular subsystem control in S_j which considerably complicates the solution of the necessary conditions (see Reference 6). However, both of these drawbacks can be overcome by adding coupling constraints in which each term is squared. With this modification, the nonfeasible controller can be used and will be discussed subsequently.

For the Gauss-Seidel second-level controller, the necessary conditions to be satisfied by the first-level control are

$$\begin{aligned} \dot{U}_j &= \frac{\partial H_j}{\partial \lambda_j} \\ \dot{\lambda}_j &= \frac{\partial H_j}{\partial U_j} \\ \frac{\partial H_j}{\partial M_j} &= 0 \end{aligned} \tag{3.48}$$

Assuming (3. 28) is linear, (3. 48) would become

$$\dot{U}_j = A_j U_j - \frac{1}{2} B_j C_j^{-1} B_j^T \lambda_j + D_{1j} S_j + G_j F_j$$
$$\dot{\lambda}_j = 2Q_j \left(U_{dj} - U_j\right) - A_j^T \lambda_j - \rho_j - \nu_j \qquad (3.49)$$

with boundary conditions

$$U_j(0) = U_{jo}$$
$$\lambda_j(t_1) = 0$$

The necessary conditions to be satisfied by the second-level control are

$$\frac{\partial H_j}{\partial \rho_j} = \frac{\partial H_j}{\partial \nu_j} = \frac{\partial H}{\partial S_j} = 0$$

In order for S_j to appear explicitly in the coupling constraint of (3.46), subscripts can be rearranged, i.e.,

$$\nu_{j-1}^T \left(U_{j-1} - S_j^u\right) + \rho_{j+1}^T \left(U_{j+1} - S_j^l\right) \qquad (3.51)$$

Equations (3.50) can then be written

$$U_j - S_{j-1}^l = 0 \qquad \rho_j = D_{j-1}^T \lambda_{j-1} \qquad (3.52)$$
$$U_j - S_{j+1}^u = 0 \qquad \nu_j = D_{j+1}^T \lambda_{j+1}$$

or in more useful form

$$S_j^u = U_{j-1} \qquad \rho_j = D_{j-1}^T \lambda_{j-1}$$
$$S_j^l = U_{j+1} \qquad \nu_j = D_{j+1}^T \lambda_{j+1} \qquad (3.53)$$

The operation of the Gauss-Seidel controller is as follows:

1. Guess S_j^l, ν_j $(j = 1, \ldots, N)$ and set $j = 0$
2. Set $j = j + 1$
3. Solve subsystem j for U_j, λ_j

4. Substitute $S^u_{j+1} = U_j$, $\rho_{j+1} = D_j^T \lambda_j$
5. Is j = N?
   No - Go to 2
   Yes - Are (3.52) satisfied to the desired accuracy for
   j=1, . . . , N?
   No - Go to 7
   Yes - Stop
6. Substitute $S^l_{j-1} = U_j, \nu_{j-1} = D_j^T \lambda_j$
7. Set j = j-1
8. Solve subsystem j for U_j, λ_j
9. Is j = 1?
   No - Go to 6
   Yes - Go to 2

This controller has been found to have excellent convergence properties for both linear and nonlinear problems as will be shown in Chapter 6.

To satisfy (3.47) for the nonfeasible second-level controller, the Hamiltonian (3.46) can be written

$$H = \sum_{j=1}^{N} H_j = \sum_{j=1}^{N} \Big\{ \left(U_{dj}-U_j\right)^T Q_j \left(U_{dj}-U_j\right) + M_j^T C_j M_j + \lambda_j^T \left(A_j U_j + B_j M_j + D_{1j} S_j + G_j F_j\right) + U_j^T \bar{\rho}_j U_j - \left(S^l_{j-1}\right)^T \bar{\rho}_j S^l_{j-1} + U_j^T \bar{\nu}_j U_j - \left(S^u_{j+1}\right)^T \bar{\nu}_j S^u_{j+1} \Big\} \tag{3.54}$$

where $\bar{\rho}_j$ = diagonal k x k matrix with elements

$$\left(\rho_{p_j}, \ldots, \rho_{q_j}\right)$$

$\bar{\nu}_j$ = diagonal k x k matrix with elements

$$\left(\nu_{p_j}, \ldots, \nu_{q_j}\right)$$

and $\quad \bar{\rho}_1 = \bar{\nu}_N = 0, \quad S_1^u = S_N^l = 0$

Rearranging (3.54) in a form suitable for the nonfeasible method H_j becomes

$$H_j = \left(U_{dj} - U_j\right)^T Q_j \left(U_{dj} - U_j\right) + M_j^T C_j M_j + \lambda_j^T \left(A_j U_j + B_j M_j + D_j (S_j^u + S_j^l) + G_j F_j\right)$$

$$+ U_j^T (\bar{\rho}_j + \bar{\nu}_j - \left(S_j^l\right)^T \bar{\rho}_{j+1} S_j^l - (S_j^u) \bar{\nu}_{j-1} S_j^u \qquad (3.55)$$

The first-level necessary conditions are

$$\dot{U}_j = \frac{\partial H_j}{\partial U_j} \qquad \frac{\partial H_j}{\partial M_j} = 0$$

$$\dot{\lambda}_j = -\frac{\partial H_j}{\partial U_j} \qquad \frac{\partial H_j}{\partial S_j} = 0 \qquad (3.56)$$

Assuming (3.28) is linear, the top row of equations in (3.56) result in exactly the first equation of (3.49). The lower right equation in (3.56) yields

$$S_j^u = \frac{1}{2} \left(\bar{\nu}_{j-1}\right)^{-1} D_j^T \lambda_j \qquad j = 2, \ldots, N$$

$$S_j^l = \frac{1}{2} \left(\bar{\rho}_{j+1}\right)^{-1} D_j^T \bar{\lambda}_j \qquad j = 1, \ldots, N-1 \qquad (3.57)$$

where the inverses exist since $\bar{\nu}_j$ and $\bar{\rho}_j$ are of full rank. Taken together (3.56) can then be written

$$\dot{U}_j = A_j U_j - \frac{1}{2}\left[B_j C_j^{-1} B_j^T - D_j\left((\bar{\nu}_{j-1})^{-1} + (\bar{\rho}_{j+1})^{-1}\right) D_j^T\right] \lambda_j + G_j F_j$$

$$\dot{\lambda}_j = 2Q_j\left(U_{dj} - U_j\right) - A_j^T \lambda_j - 2\left(\bar{\rho}_j + \bar{\nu}_j\right) U_j \qquad (3.58)$$

with boundary conditions

$$U_j(0) = U_{jo} \qquad \lambda_j(t_1) = 0$$

The necessary conditions to be satisfied by the second-level control are

$$\frac{\partial H}{\partial \bar{\rho}_j} = \frac{\partial H}{\partial \bar{\nu}_j} = 0 \qquad (3.59)$$

where

$$\frac{\partial}{\partial \bar{\rho}_j} \triangleq \text{diag.}\left[\frac{\partial}{\partial \rho_{p_j}}, \ldots, \frac{\partial}{\partial \rho_{q_j}}\right]$$

$$\frac{\partial}{\partial \bar{\nu}_j} \triangleq \text{diag.}\left[\frac{\partial}{\partial \nu_{p_j}}, \ldots, \frac{\partial}{\partial \nu_{q_j}}\right]$$

In this method (3.59) is satisfied using a gradient controller defined by

$$\frac{d\bar{\rho}_j}{d\sigma} = a_1 \frac{\partial H}{\partial \rho_j} = a_1\left[U_j U_j^T - S_{j-1}^l \left(S_{j-1}^l\right)^T\right] I$$

$$\frac{d\bar{\nu}_j}{d\sigma} = a_2 \frac{\partial H}{\partial \bar{\nu}_j} = a_2\left[U_j U_j^T - S_{j+1}^u \left(S_{j+1}^u\right)^T\right] I \qquad (3.60)$$

$$a_1 > 0 \,, \qquad a_2 > 0$$

where I is the identity matrix and σ denotes iteration time. Because of the squaring of the constraints, (3.59) is satisfied by either

$$U_j = S_{j+1}^l = S_{j+1}^u$$

or

$$U_j = -S_{j-1}^l = -S_{j+1}^u \qquad (3.61)$$

and hence this controller may converge to a false solution as shown by Bauman.[6] The choice of the scalars a_1 and a_2 greatly affects the rate of convergence; however, the lack of a completely general procedure for making this choice constitues one of the disadvantages of the method.

The necessary condition (3.47) for the existence of local subsystem minima results in

$$\begin{aligned} \rho_i &\leq 0 \qquad i = p_{j+1}, \ldots, q_{j+1} \\ \nu_k &\leq 0 \qquad k = p_{j-1}, \ldots, q_{j-1} \end{aligned} \tag{3.62}$$

However, if the equality in (3.62) holds for any element, the corresponding coupling constraint would not be satisfied over that period of time. Hence strict inequality must hold and (3.62) becomes

$$\begin{aligned} \rho_i(t) &< 0 \qquad i = p_{j+1}, \ldots, q_{j+1} \\ \nu_k(t) &< 0 \qquad k = p_{j-1}, \ldots, q_{j-1} \end{aligned} \tag{3.63}$$

In the case where a different subsystem decomposition is performed (e.g., each row of Ω may contain many subsystems) or where the spatial region Ω is of higher or lower dimension, the above results remain valid. A systematic procedure for writing general coupling constraints can be stated but, although precise, the notation is somewhat cumbersome. For example, consider the mesh points 6 and 7 of Figure 3.2 to be a subsystem, say the j^{th}. The subsystem matrix A_j is then the block diagonal matrix containing a_6 and a_7. The matrix D_{1j} in (3.41) is obtained by consecutively writing all elements appearing in a horizontal band through A_j where the ordering is from left to right and A_j is excluded. For this example

$$D_{1j} = \begin{bmatrix} d_6 & 0 & e_6 & 0 & d_6 & 0 \\ 0 & d_7 & 0 & e_7 & 0 & d_7 \end{bmatrix} \tag{3.64}$$

The A_j matrix has dimension $n_j \times n_j$ and D_{1j} has dimension $n_j \times m_j$ where m_j is also the dimension of S_j.

Defining a vector S^j as an ordered set of inputs to all subsystems which are outputs from the j^{th} subsystem, the coupling constraint in (3.46) can be written

$$\rho_j^T \left(\theta_j U_j - S^j\right) \tag{3.65}$$

where $\rho_j = m_j$ vector Lagrange multiplier

$\theta_j = m_j \times n_j$ matrix

$S^j = m_j$ vector

The matrix θ_j is obtained by scanning a vertical band through A_j from top to bottom and consecutively writing 1's wherever a nonzero element appears (excluding A_j). For this example

$$\theta_j = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{3.66}$$

Note that θ_j has elements in the same position as D_{1j}^T. This is a consequence of the symmetry of A. Defining

$$S_j = \left[s_j^1, s_j^2, \ldots, s_j^{\xi_j}\right]^T \tag{3.67}$$

the vector S^j for this example would be

$$S^j = \left[s_{j-2}^2, s_{j-2}^3, s_{j-1}^2, s_{j+1}^2, s_{j+2}^2, s_{j+2}^3\right]^T \tag{3.68}$$

Upon rearranging in the form $-\left(\rho^j\right)^T S$; , the ρ^j vector becomes by symmetry

$$\rho^j = \left[\rho^2_{j-2}, \rho^3_{j-2}, \rho^2_{j-1}, \rho^2_{j+1}, \rho^2_{j+2}, \rho^3_{j+2}\right]^T \tag{3.69}$$

Once mastered, this notation permits a mechanistic approach toward decomposition regardless of the problem size. Of course some special attention must be given if the spatial region is non-rectangular since in this case the A matrix is generally nonsymmetric. This notation will be further employed in describing the examples in Chapter 5.

The above discussion has considered coupling constraints which are one to one, i.e., each element s_{ij} of S_j corresponds to one and only one state element, say u_k. Now consider the case where (3.41) is written

$$\dot{U}_j = A_j(U_j, t)U_j + B_j M_j + S_j + G_j(U_j, t)F_j \tag{3.70}$$

$$U_j(0) = U_{oj}$$

and the coupling constraints become

$$S_j = D_j(U_j, t)\left(U_{j-1} + U_{j+1}\right) \tag{3.71}$$

There are now only k coupling constraints; however, these are now coupled as seen in (3.71). For the nonlinear problem posed, this coupling constraint cannot be handled by any of the three controllers previously mentioned. If the problem were linear, only the Gauss-Seidel controller could be employed; however, no advantage is gained over its previous application. Takahara[59] employs coupling constraints of the form (3.71) in solving linear problems. Note that the feasible method is not excluded because of the number of coupling constraints as it was previously, but because the right hand side of (3.71) contains variables from more than one subsystem.

3.5 Elliptic, Hyberbolic, and Biharmonic Equations

Parabolic partial differential equations constitute an important class of equations governing transient physical phenomena. In preceding sections, the distributed parameter systems considered were of this class. This distinction was clearly made in defining the scalar equation (3.2). However, many types of physical phenomena are described by other types of partial differential equations and it is of interest to examine the optimization of these systems by decomposition and multilevel techniques.

In particular, elliptic partial differential equations describe the steady-state behavior of those systems whose transient states are described by parabolic equations. Thus the question of steady-state optimization arises. In terms of a general notation following from Section 3.2, such a problem could be posed as minimizing the functional

$$J(m) = \int_{\bar{\Omega}} P\Big(u(X), m(X), X\Big)\, d\Omega \tag{3.72}$$

subject to the scalar partial differential equation side constraint

$$\mathcal{G}\Big(u(X), m(X), X\Big) = 0 \qquad X \in \Omega \tag{3.73}$$

with boundary conditions

$$\alpha(X)u(X)+\beta(X)\frac{\partial u(X)}{\partial n} = f(X) \qquad X \in \Omega_b \tag{3.74}$$

and possibly inequality constraints

$$R\Big(u(X), m(X), X\Big) \geq 0 \qquad X \in \Omega \tag{3.75}$$

The problem (3.72) to (3.75) corresponds to the case of distributed control as described earlier. Elliptic equations with control only on the boundary can be formulated in a similar way. The most familiar elliptic equations corresponding to these two problems are

of course Poisson's equation and Laplace's equation respectively. By discretizing (3.72) to (3.75) over the spatial domain Ω, the resulting equations are in the form considered by Lasdon[41] and discussed in Chapter 2. Although of considerable interest, steady-state problems are not considered further in this dissertation. It is interesting to note that elliptic equations do not fit into the mathematical machinery developed for determining exact optimal solutions because they are not well-posed as discussed by Brogan.[11]

Optimization problems for the class of distributed parameter systems described by hyperbolic and biharmonic partial differential equations are not readily treated by the methods presented in this dissertation. The reason stems from the fact that successful decomposition requires independent subsystems with as few coupling constraints as possible. In fact the success achieved in applying this method to distributed parameter systems is largely due to the strongly diagonal nature of the A matrix shown above. In the case of hyperbolic and biharmonic equations exactly the opposite is true; in fact, all elements appear off of the main diagonal. To illustrate, consider the simplest hyperbolic equation

$$\nabla_t^2 u = k\nabla_x^2 u \tag{3.76}$$

which when written in normal form is

$$\nabla_t \begin{bmatrix} u \\ v \end{bmatrix} = k^{\frac{1}{2}} \begin{bmatrix} 0 & \nabla_x \\ \nabla_x & 0 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \tag{3.77}$$

The A matrix is now composed entirely of cross coupling terms with no entries on the main diagonal. Similarly for the biharmonic equation

$$\nabla_t^2 u = -k\nabla_x^4 u \tag{3.78}$$

which in normal form is

$$\nabla_t \begin{bmatrix} u \\ v \end{bmatrix} = k^{\frac{1}{2}} \begin{bmatrix} 0 & -\nabla_x^2 \\ \nabla_x^2 & 0 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \tag{3.79}$$

which suffers from the same difficulties.

CHAPTER 4

VARIATIONS OF THE OPTIMAL CONTROL PROBLEM FOR DISTRIBUTED PARAMETER SYSTEMS AND THEIR EFFECT ON MULTILEVEL CONTROL

4.1 Variations in Problem Statement

The multiplicity of possible control problems that can be conceived for distributed parameter systems is many orders of magnitude greater than for lumped parameter systems. The reason stems from the higher dimensionality introduced by considering spatial as well as time dependent variables. This higher dimensionality leads to: (a) boundary control, which has no direct equivalent for lumped parameter systems, (b) greater flexibility in specifying admissible controls, (c) greater flexibility in specifying terminal and inequality constraints, (d) a wider choice of meaningful criterion functions. However, it was shown in Chapter 3 that a lumped parameter approximation can be formulated for any of these problems. In particular, by discretizing the spatial variations, an n dimensional set of ordinary differential equations is attained. Obviously n increases as the desired accuracy of approximation increases and approaches infinity in the limit. This type of approximation is important because it admits a much larger body of theory with which to attack the problem. In particular, only a few optimal control problems in distributed parameter systems currently admit to analytical solutions while for their lumped counterparts only a few do not. Of course, care must be taken to interpret the lumped solutions as approximations to the exact solution only when that approximation is valid. A case where it is not valid will be discussed subsequently. The main problem in this discussion then seems to be how to extend

present optimization techniques to handle the large systems of equations arising in this application. One possible solution is by multilevel control.

In the case of the distributed control problem, the approximation yields a single control for each mesh point with the only cross coupling terms arising from the approximations of spatial partial derivatives. The fact that no such partial derivatives appear in the criterion functional yields an independent functional for each mesh point, a convenient outcome. Moreover, the approximation of the integral over the spatial domain always yields a summable criterion functional for the integrated problem as required for the application of multilevel control.

In order to treat a large number of coupling constraints some convenient notation and a systematic procedure are essential. In general no such luxury is available when treating nonlinear equations. However, in most distributed parameter systems having physical significance, the nonlinearity appears only in a subsystem and not in the coupling constraints when treated as a lumped approximation. Thus the aspects of the fairly general nonlinear system (3.69) given for treating the coupling constraints systematically, (opposed to subsystem optimization) could be treated by the vector-matrix notation developed therein. The general scheme (3.64) to (3.69) is given for treating the coupling constraints systematically, while somewhat cumbersome, has proven useful in practice.

As mentioned above, the problem of boundary control has no analog in lumped parameter systems. A few problems of this type have been attacked using the extended definition of an operator.[11, 23] With this technique, the optimal control problem for the system

$$\frac{\partial u}{\partial t} = \alpha \frac{\partial^2 u}{\partial x^2} \tag{4.1}$$

$$u(x, 0) = 0$$

with boundary control f(t) related by

$$u(0, t) = 0 \qquad u(1, t) = f(t) \tag{4.2}$$

is written as

$$\frac{\partial u}{\partial t} = \alpha \frac{\partial^2 u}{\partial x^2} + \alpha\, \delta'(x-1) f(t) \tag{4.3}$$

$$u(x, 0) = 0$$

with homogeneous boundary conditions

$$u(0, t) = 0 \qquad u(1, t) = 0 \tag{4.4}$$

where

$$\delta' = \frac{d}{dx}\delta$$

Apparently the generality of problems which can be treated in this manner is severely limited. Although this method readily yields the optimal form f(t) of the boundary control in terms of the adjoint variables, numerical results are not easily obtained especially when the control is constrained. The only numerical results for this problem known to the author are given by Sakawa[56] (constrained) and Brogan[11] (unconstrained) and the latter's numerical efforts were unsuccessful. By considering the discrete approximation to the boundary control problem, a larger class of problems can be formulated than by the extended operator technique. However, care must be taken to assure that the lumped system yields an optimal solution which indeed approximates the actual necessary conditions for the distributed parameter system. In Chapter 5, examples are formulated in which this is, and is not, true.

Some of the difficulties involved in obtaining solutions to the boundary control problem may be intimately connected with the

concept of controllability of distributed parameter systems. The notion of controllability was first introduced by Kalman[31] for linear finite dimensional dynamical systems. These ideas have been extended by Wang[64] for distributed parameter systems. Gilbert[25] has defined controllability for n dimensional linear systems in terms of the system's structural decomposition. Gilbert's definition may be shown to be equivalent to Kalman's, at least for the case of constant coefficient linear systems having distinct eigenvalues.

Various qualifying adjectives often appear with definitions of controllability, for example, null-δ, completely-null-δ, etc. Consider the definition of complete controllability on $[t_o, t_1]$: a state $u(X, t_o) \epsilon \Gamma(\Omega)$ (e.g., a Hilbert space) is said to be completely controllable on $[t_o, t_1]$ if there exists an admissible control function which will transfer $u(X, t_o)$ to any desired final state $u_d(X, t_1) \epsilon \Gamma(\Omega)$ in a finite time t_1. It is easily shown that under certain conditions the discrete approximation will always be completely controllable on $[t_o, t_1]$ while the actual distributed parameter system may not be. For example, consider the discrete approximation of a one (space) dimensional constant coefficient distributed parameter system

$$\dot{U} = AU + M \tag{4.5}$$

where A = n x n tridiagonal symmetric matrix and assume that A has distinct eigenvalues. Defining normal coordinates

$$U = \rho Y$$

(4.5) can be written

$$\dot{Y} = \rho^{-1} A \rho Y + \rho^{-1} M \tag{4.6}$$

and Gilbert's controllability criterion requires that ρ^{-1} have no zero rows. The columns of ρ are the eigenvectors of A. Because A is tridiagonal and the desired eigenvectors are to be non-trivial,

it appears that each eigenvector must contain no zero element. Since A is real and symmetric, ρ is orthogonal[23] and

$$\rho^{-1} = \rho^{T}$$

Thus ρ^{T} not only contains no zero rows but indeed no zero elements. Hence, the semidiscrete distributed control problem is controllable and the boundary control problem likewise. Suppose however that the desired final state contained a finite discontinuity in $u_d(X, t_1)$. Then, from physical reasoning, it is clear that no control exists which could attain the desired final state in the distributed parameter system; however, such is not the case in the approximate system as seen above. Although this example is rather extreme, it serves to illustrate that care must be taken in the formation and interpretation of the semidiscrete approximations.

By considering problems where the criterion function is to minimize the norm of some terminal error, the question of controllability can be avoided. This type of criterion function seems to prevail in the literature when boundary controls are employed. A notable exception is the example by Brogan[11] cited above where reasonable results were not obtained.

4.2 Multilevel Control Considerations

The multilevel control techniques discussed here are not of universal applicability. In this section various advantages and limitations of several second-level controllers will be pointed out, particularly as they relate to the solution of problems arising from partial differential equations. In addition, there are three basic limitations of the multilevel technique which are implied in Chapter 3 by (3.5) and (3.6); namely, that (1) inequality and (2) terminal constraints are separable between subsystems and (3) that the terminal time is

fixed. Seemingly, the first two restrictions might be removed by further decomposition (a topic for future research). However, the limitation of fixed terminal time seems inescapable since, without it, each subsystem could satisfy its terminal conditions at a different time and thus confound the second-level controller.

As pointed out earlier, the nonfeasible gradient controller requires unique subsystem minima with respect to M_j and S_j as a sufficient condition for convergence. This requirement with respect to M_j is expected and is usually accounted for in the problem definition (e.g., by minimizing a convex function). However, the requirement of local minima with respect to S_j, i.e., arbitrary state variables, is extreme and is often not satisfied. Indeed there may be physical reasons why the stationary points with respect to S_j are maxima or saddle points. One reason for the latter requirement is to enable the correct sign to be chosen for the constant a_i (3.60) where

$$\rho_i^{(k+1)} = \rho_i^{(k)} + a_i\left(u_i^2 - (s^i)^2\right)^{(k)} \tag{4.7}$$

and k is the iteration number. Because of the saddle value properties arising in the nonfeasible gradient method, it can be shown[6, 45] that a_i is positive if local subsystem minima occur and negative if local maxima occur. In this as in all gradient techniques, the convergence depends rather heavily on the magnitude of a_i (step size). Unfortunately nothing more than a few general guidelines are available for choosing these values[14] and experience must be heavily relied upon. In case a scattering of maxima and minima occurred in various subsystems, it would seem possible to make an exploration over both signs of a_i and to choose the one which minimizes J(M). However, this approach would quickly become unwieldy as the number of

coupling constraints increased beyond one. A relaxation of the local minima condition can be obtained by using the Newton-Raphson type second-level controller derived by Bauman.[6] However, this controller involves considerably more computation and requires the nonsingularity of certain matrices to insure its convergence. It is noteworthy that for most problems the sufficiency conditions for convergence in each of these methods cannot be guaranteed apriori because of difficulty in evaluating the conditions. An additional requirement for convergence of the nonfeasible gradient controller is that the initial guess ρ^{o} of the Lagrange multipliers be sufficiently close to the optimum value ρ^{*} which satisfies the coupling constraint. However, one has no physical intuition regarding Lagrange multipliers and hence this choice is often difficult to make. In particular the nonfeasible method was attempted for the minimum effort problem where the system was described by the one-dimensional diffusion equation. Two four-dimensional subsystems were employed which resulted in two coupling constraints. The dominant roots corresponding to the state variables which appeared in the coupling constraints were described by the characteristic equation

$$s^{2} + f(\rho;h) = 0 \tag{4.8}$$

where $h = \Delta x$ is a parameter. For negative values of ρ (as required by (3.63)) between zero and some minimum value, $\rho_{min}(h)$, the roots of (4.8) were real and the response was monotonic as expected. However, for values of ρ less than $\rho_{min}(h)$ the roots were imaginary and the response was oscillatory, an unacceptable result for the diffusion equation. Note that $\rho_{min}(h)$ depends explicitly on h and therefore changes with the number of mesh points employed. Acceptable convergence was never obtained for this simple example and the nonfeasible gradient controller is therefore not considered acceptable for this application.

In deriving the feasible gradient controller, the coupling constraints are attached to each subsystem and are actually satisfied by the first-level necessary conditions (2.21). According to Macko,[45] the only limiting condition in applying this second-level controller is that each subsystem have at least as many degrees of freedom as the number of coupling constraints attached to it. It is easily seen that this restriction cannot be satisfied for the partial differential equation application discussed here, except possibly for systems having only one space dimension. Consider a one dimensional system having distributed control and no inequality constraints. Let the space domain be broken into n internal mesh points and N subsystems where the j^{th} subsystem has dimension n_j and

$$n = \sum_{j=1}^{N} n_j \tag{4.9}$$

For the distributed control problem, the j^{th} subsystem has n_j controls and therefore n_j degrees of freedom. Naturally if active inequality constraints are present, n_j is reduced appropriately. Using the decomposition (partitioning) shown in Chapter 3, it is clear that the total number of coupling constraints is $2(N-1)$ where a general subsystem contains at most 2 such constraints. Hence the requirement for using a feasible second-level controller in this case is that

$$n_j \geq 2 \tag{4.10}$$

For a system having two space dimensions and again using the decomposition shown in Chapter 3, any subsystem, say the j^{th}, has at most $2(1+n_j)$ coupling constraints and n_j degrees of freedom. Similarly for 3 space dimensions, the number of coupling constraints in the j^{th} subsystem is given by $2(1+2n_j)$. Hence the feasible method

cannot be used for partial differential equations having more than one space dimension when the decomposition is as stated. Further, it is considered unlikely that any other form of decomposition will greatly alter this conclusion. Because of this lack of generality, and because of the added difficulty, discussed in Chapter 2, in solving the necessary conditions since the control does not appear explicitly in the coupling constraint, the feasible gradient controller is not recommended for these applications to distributed parameter systems. It should be noted, however, that, when applicable, this controller requires initial guesses of coupled state variables only. This is a considerable advantage since one can usually estimate these variables from physical considerations.

The Gauss-Seidel type second-level controller is extremely simple and does not suffer from the chief difficulties of the gradient techniques; namely, choosing the magnitude and sign of the step size a_i. Various sufficiency conditions for the convergence of general iterative techniques and the Gauss-Seidel method in particular were discussed in Chapter 2 and were found to be rather restrictive when applied to linear dynamic systems. In practice, however, the convergence of the Gauss-Seidel second-level controller was extremely good; in fact the convergence of this method was not the limiting factor on any of the problems attempted. However, one potential limitation arises from the fact that the spectral radius for the static Gauss-Seidel procedure is inversely proportional to the square of Δx, the spatial discritization interval.[66] Thus as Δx is decreased, the convergence becomes slower. This phenomena was also observed with the Gauss-Seidel second-level controller and is reported in Section 6.3.

It is conjectured that additional improvement in the convergence rate could be obtained by employing the obvious generalization of the successive overrelaxation scheme for solving the discrete systems of equations arising from elleptic partial differential equations. It is known that this method speeds convergence over the Gauss-Seidel method for solving similar problems.[66]

In order to use multilevel techniques on the boundary control problem posed earlier further restrictions are required. Since in this case there is no longer a control at each mesh point, it is quite possible that a subsystem may arise which contains no control variable. The feasible gradient controller is thus immediately eliminated. The nonfeasible gradient controller can be used if the coupling constraint is altered to satisfy the necessary conditions (3.47) for local subsystem minima. The Gauss-Seidel technique can be used only if the state terminal conditions are free. Since no control is present, the two-point boundary value problem could otherwise not be solved. In this case, however, the state variables can be integrated forward and the adjoint variables integrated backward. This example is one of the few where the Gauss-Seidel controller is not always applicable. This problem with the state terminal conditions free is further considered in Chapters 5 and 6.

Recently, optimal control problems involving inequality constraints which are functions of the state variables only have received considerable attention.[8, 13, 21] In such problems it has been shown[8,13] that the adjoint variables possess discontinuities at points where they enter onto and/or exit from the constraint boundary. The numerical evaluation of this discontinuity usually involves considerable effort. An alternative approach to problems of this type which

yields approximate results when on the constraint boundary is the penalty function approach discussed by Kelly.[34, 35]

In treating a distributed parameter system having a state inequality of the form

$$u_{min}(X) \leq u(X, t) \leq u_{max}(X) \qquad X \in \Gamma_c(\Omega) \qquad (4.7)$$

by decomposition, it is especially convenient to consider the inequality constrained variables $u(X_{\underline{i}}, t)\left(X_{\underline{i}} \in \Gamma_c(\Omega)\right)$ as pseudo-control variables. The problem can now be handled by the (simpler) theory applying to inequality constrained control variables. Since the adjoint variables are known to be continuous when the inequality constraints contain control variables explictly, the treatment of state inequalities as pseudo-controls is not expected to yield exact results; in particular, discontinuous adjoint variables. This is also the case for the penalty function approach and the two methods are indeed similar. A comparison of the two methods is given in Chapter 5 where a particular example is formulated and solved.

The possibility of discretizing the time variable in addition to the space variables has not yet been mentioned. The resulting set of algebraic equations, albeit very large, can then be treated as a static system. If the system and the criterion function are linear, the technique of linear programming can be applied along with its own decomposition theory as developed by Dantzig.[20] If either the system or the criterion function or both are nonlinear, the static optimization method discussed earlier can be applied. An example of this theory applied to a simple linear dynamic system discretized in time is given by Lasdon.[42] Bauman[6] discusses the time decomposition of an optimal trajectory problem containing discontinuities along the trajectory; however, in this case, the independent time

segments are treated continuously. Although a time discretization may be feasible for distributed parameter systems having a short time domain, no particular advantage is anticipated and hence this method will not be pursued further.

One of the potential arguments for employing decomposition and multilevel control is that each subsystem can be solved by a technique which is best suited to it. Particularly in the case of linear subsystems, it may be possible to write the optimal solution in closed form[18] and then just perform the integration rather than solving a two-point boundary value problem during each iteration. For nonlinear subsystems, an iterative procedure for solving the two-point boundary value problem is inevitable; possible methods include (a) the gradient technique which iterates on the control, (b) the Newton-Raphson technique[58] which iterates on the adjoint initial conditions, and (c) quasilinearization[30] which iterates on the state and adjoint solutions themselves. In any case, when large problems having many subsystems are involved, it is important that the subsystems converge readily from arbitrary initial guesses since otherwise it is necessary to interrupt the iteration between the subsystems and the second-level control. When this occurs it is inconvenient to resume at the point where the cyclic process was interrupted. The convergence of the subsystems for any of the methods mentioned above depends on the initial guesses and the degree of the subsystem nonlinearity.

For the computational work done here, quasilinearization was used for solving all subsystems. In all cases the initial guesses were linear and satisfied the boundary conditions for the state variables and were zero for the adjoint variables. No convergence difficulties were encountered in either linear or mildly nonlinear

problems when sufficient accuracy in the integration process could be maintained. This point is discussed further in Chapter 6. Quasi-linearization has the disadvantages of requiring more storage than some of the other methods and of being rather inflexible toward changes in the integration step size.

CHAPTER 5

FORMULATION OF SOME EXAMPLE PROBLEMS

5.1 Introduction

In this chapter some examples of optimal control problems will be formulated using multilevel techniques. Selected problems from this group have been solved numerically and these results are presented in Chapter 6. In Section 5.4 a problem involving inequality constraints on the state variables is formulated and compared with the penalty function approach. In Section 5.5 the optimal control law for several boundary control problems is formulated analytically and these results are compared with those obtained from solving the lumped parameter approximation. In these examples the subsystems will be formulated in terms of fourth-order systems.

5.2 Minimum Effort, Fixed End, Linear Problems

Consider the minimization of

$$J(m) = \int_0^l \int_0^{t_1} m^2(x, t)\, dt, dx \tag{5.1}$$

subject to a side constraint given by the one-dimensional diffusion equation (with a forcing function m)

$$\frac{\partial u}{\partial x}(x, t) = \alpha \frac{\partial^2 u}{\partial x^2}(x, t) + m(x, t) \tag{5.2}$$

with boundary conditions

$$u(0, t) = u(l, t) = 0 \tag{5.3}$$

and initial conditions

$$u(x, 0) = u_o(x) \tag{5.4}$$

It is desired to attain the terminal condition

$$u(x, t_1) = u_1(x) \tag{5.5}$$

at a given time t_1.

The semidiscrete approximation to this problem written in the decomposed subsystem form given by (3.41) and (3.43) can be stated as minimizing

$$J'(M) = \frac{J(M)}{\Delta x} = \sum_{j=1}^{N} M_j^T(t)\, M_j(t)\, dt \tag{5.6}$$

subject to the side constraints

$$\dot{U}_j = A_j U_j + M_j + D_j S_j \tag{5.7}$$

$$U_j(0) = U_{oj} \tag{5.8}$$

$$U_j(t_1) = U_{1j} \tag{5.9}$$

$$j = 1, \ldots, N$$

Making the (arbitrary) choice of fourth-order subsystems for convenience, the A_j and D_j matrices can be defined explicitly as

$$A_j = k \begin{bmatrix} -2 & 1 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & -2 \end{bmatrix} \qquad k = \frac{\alpha}{h^2} \qquad h = \Delta x$$

$$D_j = k \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \qquad D_1 = k \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \qquad D_N = k \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

and $$S_j = \left[s_j^1, s_j^2\right]^T$$

Consider the use of a nonfeasible second-level controller where each term of the coupling constraints is squared in order to satisfy (3.47). The coupling constraints can then be written as

$$\left(\theta_j U_j\right)^T \bar{\rho}_j \theta_j U_j - \left(S^j\right)^T \bar{\rho}_j S^j = 0 \qquad (5.10)$$

$$(j = 1, \ldots, N)$$

where

$$\theta_j = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \bar{\rho}_j = \begin{bmatrix} \rho_j^1 & 0 \\ 0 & \rho_j^2 \end{bmatrix}$$

$$\theta_1 = [0 \quad 0 \quad 0 \quad 1]$$

$$\theta_N = [1 \quad 0 \quad 0 \quad 0]$$

The first-level necessary conditions follow from a development similar to $\left((3.54)\text{-}(3.58)\right)$ and can be written

$$\dot{Z}_j = B_j Z_j \qquad j = 1, \ldots, N \qquad (5.11)$$

where

$$Z_j = \left[U_j, \lambda_j\right]^T$$

and

$$B_j = k \begin{bmatrix} -2 & 1 & 0 & 0 & \left(\frac{k}{2\rho_{j-1}^2} - \frac{h}{2\alpha}\right) & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 & -\frac{h}{2\alpha} & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 & -\frac{h}{2\alpha} & 0 \\ 0 & 0 & 1 & -2 & 0 & 0 & 0 & \left(\frac{k}{2\rho_{j+1}^1} - \frac{h}{2\alpha}\right) \\ -\frac{2\rho_j^1}{k} & 0 & 0 & 0 & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -\frac{2\rho_j^2}{k} & 0 & 0 & -1 & 2 \end{bmatrix}$$

For $j=1$, $\rho_j^1 = 0$ and $\rho_{j-1}^1 = \infty$; and for $j = N$, $\rho_j^2 = 0$ and $\rho_{j+1}^1 = \infty$. Note that B_j is a time-varying matrix. The boundary conditions for (5.11) are

$$Z_j^u(0) = U_j(0) = U_{oj}$$
$$Z_j^u(t_1) = U_j(t_1) = U_{1j} \tag{5.12}$$

The necessary conditions $\left((3.59)-(3.60)\right)$ to be satisfied by the second-level controller are then

$$\frac{d\rho_j^1}{d\sigma} = a_1\left[\left(u_j^1\right)^2 - \left(s_{j-1}^2\right)^2\right] \qquad a_1 > 0 \tag{5.13a}$$

$$\frac{d\rho_j^2}{d\sigma} = a_2\left[\left(u_j^4\right)^2 - \left(s_{j+1}^1\right)^2\right] \qquad a_2 > 0 \tag{5.13b}$$

where for $j = 1$, (5.13a) disappears and for $j = N$, (5.13b) disappears. As in (3.60), ρ_j^1 and ρ_j^2 are required to be negative.

Treating the same problem, but using the Gauss-Seidel second-level controller, the coupling constraints can be written

$$\rho_j^T\left(\theta_j U_j - S^j\right) = 0 \qquad (j = 1, \ldots, N) \tag{5.14}$$

where $\rho_j^T = \left[\rho_j^1, \rho_j^2\right]$

and θ_j is the same as in (5.10). The necessary conditions to be solved at the first level now follow from $\left((3.48)-(3.49)\right)$ and can be written

$$\dot{Z}_j = B_j Z_j + P_j \qquad (j = 1, \ldots, N) \tag{5.15}$$

where $Z_j = \left[U_j, \lambda_j\right]^T$

$$B_j = k \begin{bmatrix} -2 & 1 & 0 & 0 & \frac{-1}{2k} & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 & \frac{-1}{2k} & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 & \frac{-1}{2k} & 0 \\ 0 & 0 & 1 & -2 & 0 & 0 & 0 & \frac{-1}{2k} \\ 0 & 0 & 0 & 0 & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 \end{bmatrix}$$

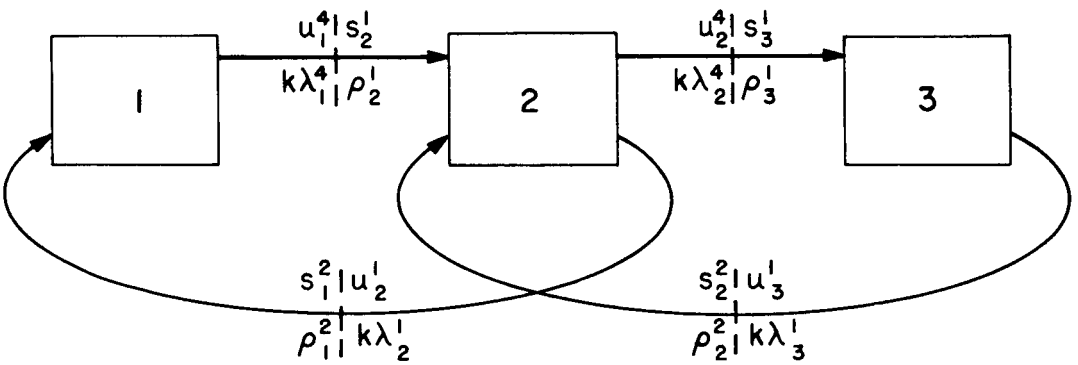
$$P_j = \left[ks_j^1,\ 0, 0, ks_j^2,\ -\rho_j^1,\ 0, 0,\ -\rho_j^2\right]^T$$

Note that P_j is a vector whose elements are parameters in (3.15) since both S_j and ρ_j are determined by the Gauss-Seidel second-level controller. In this case B_j is a constant matrix.

The second-level necessary conditions (3.50) can then be written explicitly as

$$\begin{aligned} s_j^1 &= u_{j-1}^4 & \rho_j^1 &= k\,\lambda_{j-1}^4 \\ s_j^2 &= u_{j+1}^1 & \rho_j^2 &= k\,\lambda_{j+1}^1 \end{aligned} \qquad (5.16)$$

where $s_1^1 = \rho_1^1 = 0$ and $s_N^2 = \rho_N^2 = 0$. The Gauss-Seidel procedure detailed in Chapter 3 is shown schematically in Figure 5.1 for three subsystems. The only initial guesses required are for s_1^2, ρ_1^2, s_2^2, and ρ_2^2.



GAUSS-SEIDEL PROCEDURE FOR N=3

FIGURE 5.1

5.3 A Nonlinear Problem

Consider the minimum effort problem of Section 5.2 with a side constraint given by the nonlinear diffusion equation

$$\frac{\partial u(x,t)}{\partial t} = \alpha\, u(x,t)\, \frac{\partial^2 u(x,t)}{\partial x^2} + m(x,t) \tag{5.17}$$

with boundary, initial, and terminal conditions given by ((5.3)-(5.5)). The semidiscrete approximation of (5.17) written in terms of fourth-order subsystems is then (suppressing the j subscript for notational convenience)

$$\begin{aligned}
\dot{u}_1 &= k\left(-2u_1^2 + u_1 u_2\right) + k\, u_1 s_1 + m_1 \\
\dot{u}_2 &= k\left(u_1 u_2 - 2u_2^2 + u_3 u_2\right) + m_2 \\
\dot{u}_3 &= k\left(u_2 u_3 - 2u_3^2 + u_4 u_3\right) + m_3 \\
\dot{u}_4 &= k\left(u_3 u_4 - 2u_4^2\right) + k\, u_4 s_2 + m_4
\end{aligned} \tag{5.18}$$

with boundary conditions given by ((5.8)-(5.9)). The necessary condition

$$\frac{\partial H_j}{\partial M_j} = 0$$

is satisfied by

$$M_j = -\frac{1}{2}\lambda_j \tag{5.19}$$

as was the case in (5.15). Writing the coupling constraint as (5.14), the final first-level necessary condition

$$\dot{\lambda}_j = -\frac{\partial H_j}{\partial U_j}$$

becomes (again suppressing the j subscript)

$$
\begin{aligned}
\dot{\lambda}_1 &= k\left[(4u_1 - u_2 + s_1)\lambda_1 - u_2\lambda_2\right] - \rho_1 \\
\dot{\lambda}_2 &= k\left[-u_1\lambda_1 + (4u_2 - u_1 - u_3)\lambda_2 - u_3\lambda_3\right] \\
\dot{\lambda}_3 &= k\left[-u_2\lambda_2 + (4u_3 - u_2 - u_4)\lambda_3 - u_4\lambda_4\right] \\
\dot{\lambda}_4 &= k\left[-u_3\lambda_3 + (4u_4 - u_3 - s_2)\lambda_4\right] - \rho_2
\end{aligned}
\tag{5.20}
$$

Equations $\left((5.18)-(5.20)\right)$ can be written in the form

$$
\dot{Z}_j = B_j(Z_j)\,Z_j + C_j(Z_j)\,P_j \tag{5.21}
$$

where Z_j and P_j are defined in (5.15) and B_j and C_j are determined from $\left((5.18)-(5.20)\right)$. Using the Gauss-Seidel controller, the second-level necessary conditions yield values for the parameter vector P_j; namely,

$$
\begin{aligned}
s_j^1 &= u_{j-1}^4 & \rho_j^1 &= k\,u_{j-1}^4\,\lambda_{j-1}^4 \\
s_1^2 &= u_{j+1}^1 & \rho_j^2 &= k\,u_{j+1}^1\,\lambda_{j+1}^1
\end{aligned}
\tag{5.22}
$$

where $s_1^1 = \rho_1^1 = 0$ and $s_N^2 = \rho_N^2 = 0$. The subsystem two-point boundary value problem given by $\left((5.18)-(5.20)\right)$ cannot be solved in closed form as was the case with linear subsystems. The iterative method of quasilinearization (described in Chapter 6) was used to solve this problem. Note that although the subsystem necessary conditions for this problem are considerably more complex than in the linear case, the second-level necessary conditions are only slightly different. Hence one expects the convergence properties of the Gauss-Seidel controller to be similar for linear and mildly nonlinear problems. For this example no degradation in the convergence rate was observed.

5.4 A State Inequality Constrained Problem

Consider the minimum effort problem of Section 5.2 subject to the state inequality constraint

$$u_j^4 \geq c(t) \tag{5.23}$$

where $c(t)$ is a known function of time. The second-level necessary conditions (5.16) require that

$$u_j^4 = s_{j+1}^1 \tag{5.24}$$

and hence (5.23) can be written

$$s_{j+1}^1 \geq c(t) \tag{5.25}$$

The inequality (5.25) is now on a (pseudo) control variable of subsystem (j_o+1) and can thus be conveniently handled by the method due to Valentine.[60] Consider the constrained pseudo-control to be in the k^{th} subsystem. Decomposing and using a Gauss-Seidel second-level controller, the Hamiltonian can be written

$$H = \sum_{j=1}^{N} \left\{ M_j^T M_j + \lambda_j^T (A_j U_j + M_j + D_j S_j) + \rho_j^T \left(\theta_j U_j - S^j \right) + \nu \left(\xi^2 - (\beta_k^T S_k - c) \right) \right\} \tag{5.26}$$

where $\beta_k^T = [1\ 0\ 0\ 0]$ and ξ is a real slack variable. Note that (5.25) has been converted to an equality constraint such that (5.25) is satisfied when ξ is real. The first-level necessary conditions for subsystem k now yield

$$\dot{U}_k = A_k U_k + M_k + D_k S_k$$

$$\dot{\lambda}_k = - A_k^T \lambda_k - \theta_k^T \rho_k$$

$$M_k = - \frac{1}{2} \lambda_k \qquad (5.27)$$

$$\xi^2 = \beta_k^T S_k - c$$

$$2\nu\xi = 0$$

The second-level necessary conditions are given by

$$D_k^T \lambda_k - \rho^k - \beta_k \nu = 0 \qquad (5.28a)$$

$$\theta_k U_k - S^k = 0 \qquad (5.28b)$$

The Clebsch necessary condition[9] required that $\nu \geq 0$. If $\nu = 0$, (5.25) is satisfied by strict inequality and the solution proceeds as in Section 5.2. However, if $\xi = 0$ ($\nu > 0$), (5.25) is satisfied by equality and (5.28a) requires the determination of ν. This can be accomplished by using a gradient method to determine ν when on the boundary. The variation of H with respect to ν is, from (5.26)

$$\delta H = \left(c - \beta_k^T S_k\right) \delta\nu \qquad (5.29)$$

By Macko's[45] saddle value proof, it is seen that minimizing J(M) with respect to M corresponds to maximizing $H(\nu, \rho, M, S)$ with respect to ν (and ρ). Hence $\partial\nu$ in (5.29) can be chosen as

$$\delta\nu = a\left(c - \beta_k^T S_k\right) \qquad a > 0 \qquad (5.30)$$

Since $\beta_k^T S_k$ does not equal u_{k-1}^4 until the second level has converged, it is convenient to write (5.30) as

$$\delta\nu = a\left(c - u_{k-1}^4\right) \qquad a > 0 \qquad (5.31)$$

The gradient controller is then defined by

$$\frac{d\nu}{d\sigma} = a\left(c - u_{k-1}^4\right) \qquad a > 0 \tag{5.32}$$

Solving (5.32) for ν and substituting into (5.28a) yields for the first element of ρ^k

$$\rho_1^k = \rho_{k-1}^2 = \frac{\alpha}{\Delta x^2}\,\lambda_k^1 - \sum_{i=1}^{m} a\left(c - u_{k-1}^4\right)^{(i)} + \nu^{(0)} \qquad a > 0 \tag{5.33}$$

where (i) is the iteration index for the second-level controller and m is the "current" iteration. By using (5.33) one iteration on ν is obtained during one iteration through the second-level controller. Substituting (5.33) into the appropriate equation in (5.21) gives

$$\dot{\lambda}_{k-1}^4 = \frac{\alpha}{\Delta x^2}\left[-\lambda_{k-1}^3 + 2\lambda_{k-1}^4 - \lambda_k^1\right] + \sum_{i=1}^{m} a\left(c - u_{k-1}^4\right)^{(i)} + \nu^{(0)} \tag{5.34}$$

In solving this problem by the penalty function approach[34] a new state variable P is defined by

$$\begin{aligned} \dot{P} &= K\left(c - u_{k-1}^4\right)^2 && \text{if} \quad u_{k-1}^4 \leqq c \\ \dot{P} &= 0 && \text{if} \quad u_{k-1}^4 > c \end{aligned} \tag{5.35}$$

$$P(0) = 0$$

where K is an arbitrary constant. The inequality (5.23) will be satisfied along the entire path only if $P(t_1) = 0$. Using this approach the adjoint equation corresponding to (5.34) follows from a straightforward application of the maximum principle[54] as

$$\dot{\lambda}_{k-1}^4 = \frac{\alpha}{\Delta x^2}\left[-\lambda_{k-1}^3 + 2\lambda_{k-1}^4 + \lambda_k^1\right] - 2K\lambda_P\left(c - u_{k-1}^4\right) \tag{5.36}$$

where λ_P is the constant adjoint variable corresponding to (5.35) and must be determined iteratively by trying to drive $P(t_1)$ to zero.[46]

In general either (5.34) or (5.36) will bring u^4_{k-1} arbitrarily close to the boundary but will not attain it exactly. This is easily seen from these equations since the error term $(c - u_{k-1})$ is zero when going onto or off of the boundary, and thus the discontinuities in λ^4_{k-1} indicated by the general theory[8] are not obtained (assuming $\nu^{(0)}$ contains no impulse function).

5.5 Some Boundary Control Problems

As was remarked earlier, some care must be exercised in the formulation and solution of boundary control problems by semi-discrete techniques. Examples are presented in this section for which the optimal boundary control function either differs or remains unchanged when the discretization is performed on the original system equation or on the necessary conditions for optimality.

Consider the problem of minimizing the functional

$$J(m) = \int_o^1 \left[u_d(x) - u(x, t_1)\right]^2 dx + c \int_o^{t_1} f^2(t)\, dt \tag{5.37}$$

subject to the side constraints

$$\frac{\partial u}{\partial t} = \alpha \frac{\partial^2 u}{\partial x^2} \tag{5.38}$$

with boundary conditions

$$u(0, t) = 0 \qquad u(1, t) = f(t) \tag{5.39}$$

and initial conditions

$$u(x, 0) = u_o(x) \tag{5.40}$$

In (5.37), $u_d(x)$ is a given function. Reformulating in semidiscrete form $\left((3.24)\text{-}(3.25)\right)$ and decomposing yields

$$\overset{*}{J}(M) = \frac{J(M)}{h} = \sum_{j=1}^{N} J_j(M_j) \tag{5.41}$$

where

$$J_j = \left[U_{dj} - U_j(t_1)\right]^T \; U_{dj} - U_j(t_1) \qquad (5.42)$$

$$j = 1, \ldots, N-1$$

and

$$J_N = \left[U_{dN} - U_N(t_1)\right]^T \left[U_{dN} - U_N(t_1)\right] + \frac{c}{h}\int_o^{t_1} f^2(t)dt \qquad (5.43)$$

The side constraints become

$$\dot{U}_j = A_j U_j + D_j S_j \qquad (5.44)$$

$$j = 1, \ldots, N-1$$

$$U_j(0) = U_{oj}$$

and

$$j = 1, \ldots, N-1$$

$$\dot{U}_N = A_N U_N + D_N S_N + B_N f \qquad (5.45)$$

$$U_N(0) = U_{oN}$$

where $B_N = [0\ 0\ 0\ k]^T$

and $k = \alpha/h^2$

Note that (5.42) and the first terms in (5.43) depend on values of $U(t_1)$ only. The necessary conditions are obtained in this case from the Mayer formulation[9] of the optimization problem. Equation (5.43) can be written as

$$J_N(t_1) = \left[U_{dN} - U_N(t_1)\right]^T \left[U_{dN} - U_N(t_1)\right] + \phi(t_1) \qquad (5.46)$$

where

$$\dot{\phi} = \frac{c}{h} f^2(t) \qquad (5.47)$$

$$\phi(0) = 0$$

The Hamiltonian then becomes

$$H = \sum_{j=1}^{N-1} \left\{ \lambda_j^T (A_j U_j + D_j S_j) + \rho_j^T \left(\theta_j U_j - S^j \right) \right\} + \lambda_\phi \frac{c}{h} f^2 \qquad (5.48)$$

$$+ \lambda_N^T (A_N U_N + D_N S_N + B_N f) + \rho_N^T \left(\theta_N U_N - S^N \right)$$

where all matrices are as defined in (5.7) and (5.10). Application of the Mayer theory then yields the first-level necessary conditions (5.44), (5.45), (5.47) and

$$\dot{\lambda}_j = - A_j^T \lambda_j - \theta_j^T \rho_j \qquad j = 1, \ldots, N \qquad \dot{\lambda}_\phi = 0 \qquad (5.49)$$

$$f = \frac{-h}{2c} B_N \lambda_N = \frac{-\alpha}{2ch} \lambda_N^4 \qquad (5.50)$$

with terminal conditions

$$\lambda_j(t_1) = \frac{\partial J_j(t_1)}{\partial U_j} = - 2 \left(U_{dj} - U_j(t_1) \right) \qquad \lambda_\phi(t_1) = 1 \qquad (5.51)$$

$$j = 1, \ldots, N$$

The second-level necessary conditions are (using the Gauss-Seidel controller) the same as (5.16).

For comparison the necessary conditions were also determined by discretizing the necessary conditions determined from an analytical approach. This approach uses the extended definition of an operator as discussed by Friedman[23] and applied by Brogan.[11] The criterion functional is given by

$$J(m) = \int_0^1 \left\{ \left[u_d - u(x, t_1) \right]^2 + \delta(x - 1)\, c \int_0^t f^2(t)\, dt \right\} dx \qquad (5.52)$$

and the side constraint is written with homogeneous boundary conditions as

$$\frac{\partial u}{\partial t} = \alpha \frac{\partial^2 u}{\partial x^2} + \alpha\, \delta'(x-1)\, f(t) \tag{5.53}$$

$$u(x, 0) = u_o(x)$$

$$u(0, t) = u(1, t) = 0$$

Then the Hamiltonian

$$H = \int_o^1 \left\{ \delta(x-1)\, c\, f^2(t) + \lambda^T \left(\alpha \frac{\partial^2 u}{\partial x^2} + \alpha\, \delta'(x-1)\, f(t) \right) \right\} dx \tag{5.54}$$

is minimized over f by requiring

$$\int_o^1 \left\{ 2\, \delta(x-1)\, c\, f + \alpha\, \delta'(x-1)\, \lambda(x, t) \right\} dx = 0 \tag{5.55}$$

or using the appropriate identity[23]

$$f(t) = \frac{\alpha}{2c} \frac{d\lambda}{dx}(1, t) \tag{5.56}$$

The adjoint system is

$$\frac{\partial \lambda}{\partial t} = -\alpha \frac{\partial^2 \lambda}{\partial x^2} \tag{5.57}$$

also having homogeneous boundary conditions and a final (time) value of

$$\lambda(x, t_1) = -2\left(u_d(x) - u(x, t_1) \right) \tag{5.58}$$

Discretizing (5.56) yields

$$f(t) = \frac{-\alpha}{2ch} \lambda(1-h, t) \tag{5.59}$$

a result identical to (5.50). To avoid approximating the doublet function in (5.53), the nonhomogeneous form ((5.38)-(5.39)) can be used. The semidiscrete form of the analytical solution is easily seen to be identical to that obtained previously ((5.44)-(5.51)).

In most cases, however, slightly different expressions for the optimal control law are obtained from these two approaches.

Consider for example the problem discussed above $\left((5.37)\text{-}(5.38)\right)$ but with boundary conditions given by

$$u(0, t) = 0 \qquad \frac{\partial u(1, t)}{\partial x} = f(t) - u(1, t) \tag{5.60}$$

Discretizing (5.60) along with (5.38) yields in one place of (5.45)

$$\dot{U}_N = A_N U_N + D_N S_N + B_N \left(\frac{1}{1+h}\right)\left(hf + u_N^4\right) \tag{5.61}$$

$$U_N(0) = U_{No}$$

The corresponding control law is given by

$$f = \frac{-\alpha}{2c(1+h)}\,\lambda_N^4 \tag{5.62}$$

and the adjoint equation is

$$\dot{\lambda}_N^4 = k\left[-\lambda_N^3 + \left(\frac{1+2h}{1+h}\right)\lambda_N^4\right] \tag{5.63}$$

$$\lambda_N^4(t_1) = -2\left(u_{dN}^4 - u_N^4(t_1)\right)$$

The analytical formulation of this problem requires the minimization of (5.52) subject to the side constraint

$$\frac{\partial u}{\partial t} = \alpha \frac{\partial^2 u}{\partial x^2} + \alpha\,\delta(x, 1)\left[f(t) - u(1, t)\right] \tag{5.64}$$

having homogeneous boundary conditions

$$u(0, t) = 0 \qquad \frac{\partial u(1, t)}{\partial x} = 0 \tag{5.65}$$

Writing the Hamiltonian as before requires for a minimum that

$$\int_0^1 \left[2\,\delta(x-1)\,c\,f + \alpha\,\lambda(x, t)\,\delta(x-1)\right] dx = 0 \tag{5.66}$$

or

$$f(t) = -\frac{\alpha}{2c}\,\lambda(1, t) \tag{5.67}$$

The adjoint system is now the same as (5.57) but with boundary conditions given by

$$\lambda(0,t) = 0 \qquad \frac{\partial\lambda(1,t)}{\partial x} = 0 \tag{5.68}$$

Discretizing (5.67) and (5.68) yields

$$f(t) = -\frac{\alpha}{2c}\lambda(1-h,t) \tag{5.69}$$

a result different from (5.62). The adjoint equation using this approach, also differs from (5.63) and is given by

$$\dot{\lambda}_N^4 = k\left[-\lambda_N^3 + \lambda_N^4\right] \tag{5.70}$$

$$\lambda_N^4(t_1) = -2\left(u_{dN}^4 - u_N^4(t_1)\right)$$

All other equations resulting from the different approaches are identical (except (5.61) which differs by the amount that f differs).

Consider finally the boundary control problem of minimizing

$$J(m) = \int_o^{t_1}\left\{\int_o^1 \left[u_d(x,t) - u(x,t)\right]^2 dx + c\,f^2(t)\right\}dt \tag{5.71}$$

subject to side constraints given by ((5.38)-(5.40)). The first-level necessary conditions for this problem are identical to ((5.44)-(5.50)) except that the adjoint system is given by

$$\dot{\lambda}_j = -\left[2\left(U_j - U_{dj}\right) + A_j^T\lambda_j + \theta_j^T\rho_j\right] \tag{5.72}$$

$$\lambda_j(t_1) = 0 \qquad j = 1,\ldots,N$$

The second-level necessary conditions are again given by (5.16).

Other problems could be considered such as the minimum terminal error problem ((5.37) with c = 0) with inequality constrained control. The semidiscrete solution to this problem yields the

expected bang-bang result which agrees with the analytical formulation. Alternatively the problem of (5.37) could be treated but with inequality constrained control. The solution to these problems is similar to the result obtained with distributed control and these problems are formulated in the next section. Obviously many other types of problems are possible.

5.6 Some Problems Involving Control Inequality Constraints

Consider the distributed control problem of minimizing

$$J(m) = \int_o^l \left\{ \left(u_d(x) - u(x, t_1) \right)^2 + c \int_o^{t_1} m^2(x, t)\, dt \right\} dx \tag{5.73}$$

subject to the side constraints $\left((5.2)\text{-}(5.4)\right)$ and

$$m_o(x) \leq m(x, t) \leq m^o(x) \tag{5.74}$$

Decomposing in the standard way, the criterion functional (5.73) becomes

$$\frac{J(M)}{h} = \sum_{j=1}^{N} \left\{ \left[U_{dj} - U_j(t_1) \right]^T \left[U_{dj} - U_j(t_1) \right] + c \int_o^{t_1} M_j^T M_j\, dt \right\} \tag{5.75}$$

and the inequality constraints (5.74) yield

$$\begin{aligned} (M_j - M_{oj}) &\geq 0 \\ (M_j^o - M_j) &\geq 0 \end{aligned} \tag{5.76}$$

Using Valentines[60] technique, the inequalities (5.76) can be changed to equality constraints and appended to the Hamiltonian which is then given by

$$\begin{aligned} H = \sum_{j=1}^{N} \Big\{ & c\, M_j^T M_j + \lambda_j^T (A_j U_j + D_j S_j + M_j) + \rho_j^T \left(\theta_j^T M_j - S^j \right) \\ & - \left(M_j - M_{oj} \right)^T \bar{\nu}_j \left(M_j^o - M_j \right) + \eta_j^T \bar{\nu}_j\, \eta_j \Big\} \end{aligned} \tag{5.77}$$

where

$\bar{\nu}_j$ = n_j dimensional diagonal matrix of Lagrange multipliers

Minimizing (5.77) with respect to M_j gives

$$M_j = \frac{1}{2}\left(c\,I + \bar{\nu}_j\right)^{-1} \left\{ \bar{\nu}_j \left(M_j^o + M_{oj}\right) - \lambda_j \right\} \tag{5.78}$$

and the Clebsch condition requires that each element (ν_j) of ν_j be greater than or equal to zero. Two conditions can therefore arise:

(1) $\nu_j = 0$ and

$$m_j = -\frac{1}{2c}\lambda_j \tag{5.79}$$

or (2) $\nu_j > 0$ and

$$\begin{aligned} m_j &= m_{oj} \quad \text{if} \quad \lambda_j > -2c\,m_{oj} \\ m_j &= m_j^o \quad \text{if} \quad \lambda_j < -2c\,m_j^o \\ j &= 1, \ldots, N \; ; \; i = 1, \ldots, n_j \end{aligned} \tag{5.80}$$

where the element subscript i has been suppressed for notational convenience. Equations (5.79) and (5.80) follow immediately from (5.78). The remaining necessary conditions follow from the earlier results of this chapter.

An alternative problem would require the minimization of (5.73) with $c = 0$ and side constraints given by $\left((5.2)\text{-}(5.4)\right)$ and

$$|m(x,t)| \le m^o(x) \tag{5.81}$$

The Hamiltonian for this problem becomes

$$H = \sum_{j=1}^{N} \left\{ \lambda_j^T \left(A_j U_j + D_j S_j + M_j\right) + \rho_j^T \left(\theta_j U_j - S^j\right) \right\} \tag{5.82}$$

where the inequality constraint

$$|M_j| \le M_j^o \tag{5.83}$$

must be satisfied elementwise. The minimization of H to M_j then yields

$$m_j = -m_j^o \operatorname{sgn} \lambda_j \tag{5.84}$$

$$j = 1, \ldots, N \; ; \quad i = 1, \ldots, n_j$$

where the element subscript i has again been suppressed for notational convenience.

CHAPTER 6

NUMERICAL PROCEDURES AND RESULTS

6.1 Subsystem Optimization

One of the advantages of the multilevel approach to optimization is that the various subsystem optimizations can proceed using different techniques. For example, linear subsystems could be solved in closed form,[18] while nonlinear subsystems would require iterative methods such as second-variational techniques,[10, 33, 58] gradient methods[34] or quasilinearization.[30, 47, 51] However, in order to provide the generality required to solve a number of linear and nonlinear examples, the method of quasilinearization† was used for each subsystem in all computational examples considered here.

Quasilinearization is an iterative technique which satisfies the boundary conditions and the maximum principle (along a given trajectory) exactly and iterates until the system differential (state and adjoint) equations are satisfied. Convergence of this method depends upon the initial guesses of the state and adjoint solution trajectories and, when obtained, the convergence is quadratic.[51] Consider the two-point boundary value problem given by

$$\dot{y} = f(y, t) \tag{6.1a}$$

$$y_i(0) = y_{oi} \ , \quad y_i(t_1) = y_{1i} \ , \quad i = 1, \ldots, \frac{n}{2} \tag{6.1b}$$

where y is an n vector of state and adjoint equations with boundary conditions on (say) the state variables. The quasilinearization solution proceeds for the k^{th} iteration by solving the linearized systems

†Several subroutines were already available from the work of Paine.[51]

$$\dot{z}^{k+1} = f(y^k, t) + \left(\frac{\partial f}{\partial y}\right)^k \left(z^{k+1} - y^k\right) \tag{6.2}$$

$$z^{k+1}(0) = y^k(0)$$

and

$$\dot{Y}^{k+1} = \left(\frac{\partial f}{\partial y}\right)^k Y^{k+1} \tag{6.3}$$

$$Y^{k+1}(0) = I$$

where

$$y^{k+1} = z^{k+1} + Y^{k+1}\alpha^{k+1} \tag{6.4}$$

y^o = initial (guessed) solution which satisfies (6.1b)

$\frac{\partial f}{\partial y}$ = n x n matrix

Y = n x n matrix

α = n dimensional constant vector

The terminal boundary conditions are always satisfied for (6.2) by choosing α^{k+1} such that

$$y_{1i} = z_i^{k+1}(t_1) + \sum_{j=\frac{n}{2}+1}^{n} Y_{ij}^{k+1}\,\alpha_j^{k+1} \qquad i = 1, \ldots, \frac{n}{2} \tag{6.5}$$

As the iteration proceeds the solution to (6.2) converges to the solution of (6.1). As a check on the numerical accuracy and to conserve rapid-access storage, it is convenient to determine $y^{k+1}(t)$ by re-integrating (6.2) with initial conditions $y^{k+1}(0)$. If the final values $y_i^{k+1}(t_1)$ compare favorably with the desired final values y_{1i} $(i = 1, \ldots, \frac{n}{2})$, then numerical accuracy has been preserved. Note that quasilinearization requires considerable storage along the entire trajectory $\left(\text{e.g., } y^k,\ f(y^k, t),\ f_y(y^k, t)\right)$ which may become

inconvenient for large problems. In these cases, other techniques should be investigated.

For the numerical work considered here, the state vector was of fourth order. The number of equations integrated in the first pass of quasilinearization was 40 (8 in (6. 2) and 32 in (6. 3)) and in the second pass 9 (8 in (6. 2) plus the criterion functional). The method of integration used in all cases was the modified predictor-corrector scheme due to Hamming.[26] This method is known to be computationally stable† in the scalar case if[55]

$$\Delta t < \frac{.65}{|f_y|} \tag{6.6}$$

Hamming states that this method may become relatively unstable as f_y becomes very large. No analogous results for systems of differential equations are known to the author. However, from the numerical work done here it is clear that some relation of the general form of (6. 6) must exist. In particular since the factor $\alpha/(\Delta x)^2$ is present in all tridiagonal terms of the matrix f_y, it was necessary to reduce Δt as the spatial increment Δx was refined in order to preserve the integration accuracy. Another drawback of this method is that it is not self-starting. In this work the iterative technique given by Ralston[55] was used as a starting procedure. On the advantageous side, the Hamming method is a fifth order method (truncation error proportional to the fifth derivative of the solution) and is relatively fast, requiring only two evaluations of the derivative for each step in the integration. In addition, the truncation error is easily determined.

---

†Stability implies $\frac{\partial f}{\partial y} < 0$. If $\frac{\partial f}{\partial y} < 0$, the analogous concept is relative stability.[55]

6.2 The Computer Program

The computer program consists of the ten subroutines listed below:

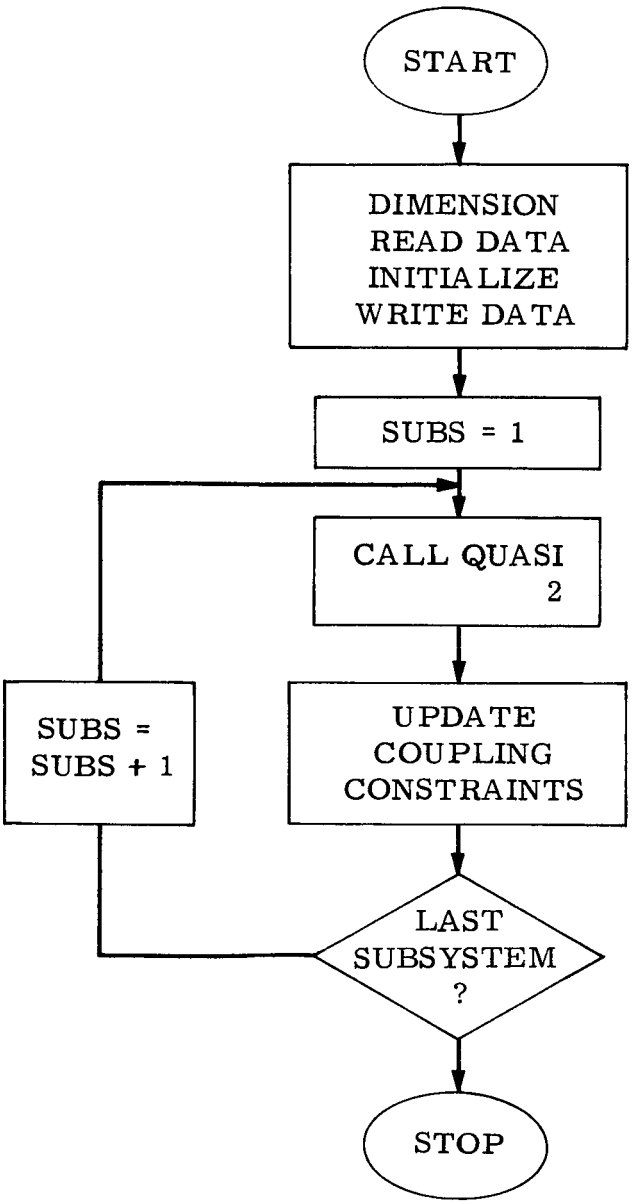
1. Second-level controller
2. Quasilinearization
3. Derivative evaluation, $f(y^k, t)$, $\left(\frac{\partial f}{\partial y}\right)^k$
4. Hamming forward integration
5. Derivative evaluation, $\dot{z}$
6. Control function calculation
7. Matrix inversion
8. Calculate subsystem Hamiltonian and write results
9. Calculate total Hamiltonian and spatial truncation errors
10. Plot results

The general interconnection of these subroutines is indicated in Figures 6.1 and 6.2. The numbers inside certain boxes indicate the presence of the corresponding subroutines numbered above.

Two additional subroutines corresponding to 4 and 5 for a Hamming forward-backward integration were required for the boundary control problems. In this case no two-point boundary value problem existed for subsystems having no control variable. For these subsystems, the state variables were integrated forward and the adjoints variables backward. The flowchart in Figure 6.2 must be suitably altered for such problems.

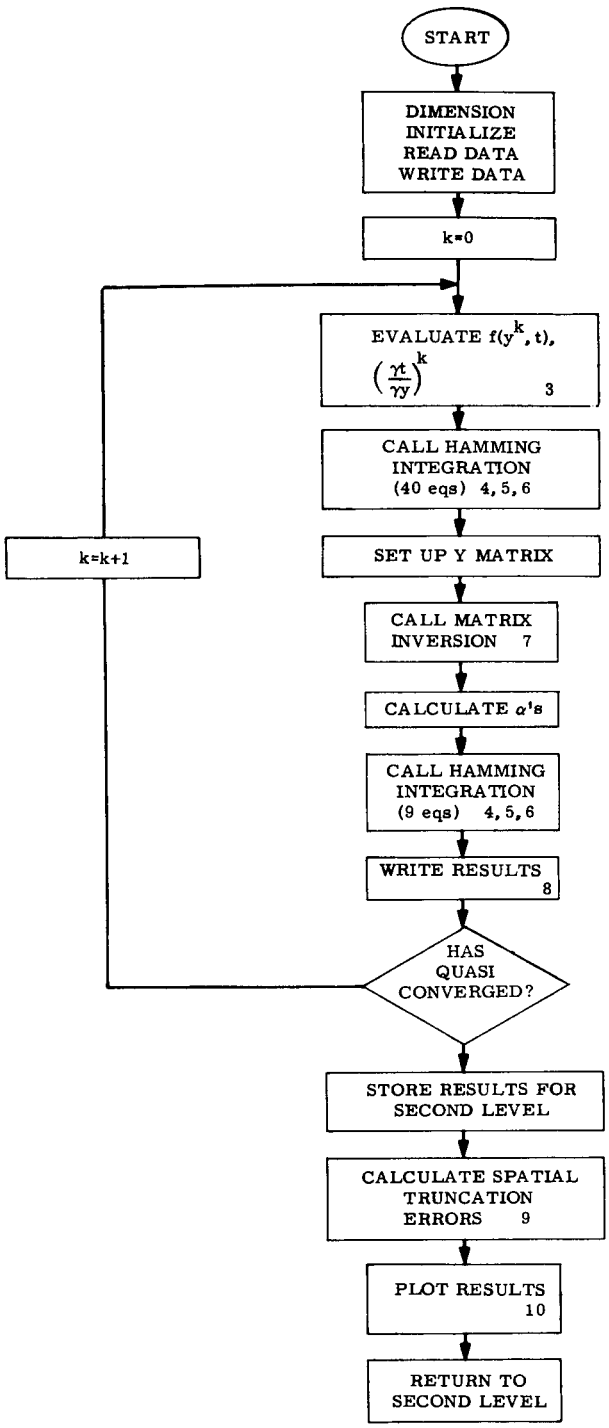
6.3 Minimum Effort, Fixed End, Linear Examples with Distributed Control

The example to be discussed in this section was formulated in Section 5.2 using both a nonfeasible and a Gauss-Seidel type second-level controller. In this example the following numerical values were used



SECOND-LEVEL CONTROLLER PROGRAM

FIGURE 6.1



SUBROUTINE FOR SUBSYSTEM OPTIMIZATION
USING QUASILINEARIZATION

FIGURE 6. 2

$$
\begin{aligned}
l &= 1 \\
\alpha &= .0033 \\
t_1 &= 5 \\
\Delta t &= 0.1
\end{aligned}
\tag{6.7}
$$

This value of α corresponds to the thermal diffusity of steel in units of $ft.^2$ per minute in the case where the problem is thought to be one of heat conduction. This problem is similar to one treated by Brogan[11] and was so chosen in order to check the reasonableness of the results.

As mentioned previously, no acceptable results were obtained using the nonfeasible controller. The primary reason for this is discussed below. Two fourth-order subsystems were considered and the initial guesses of $\rho_1^2(t)$ and $\rho_2^1(t)$ were taken as -0.1. Suceeding values were determined using (5.13). In order to assure that ρ_1^2 and ρ_2^1 remained negative, their values were compared with zero and the step size $\left(a_1, a_2 \text{ in (5.13)}\right)$ was halved each time ρ_j^i became positive. An integration problem arose as portions of ρ_j^i approached zero since then the coefficients of B_j in (5.11) increased significantly. In order to maintain stability of the Hamming integration method, Δt would have to be decreased prohibitively as indicated in Section 6.1. Since the maximum magnitudes encountered in the time varying B_j matrix are not known apriori, it is difficult to prespecify the time increment for any integration technique using a fixed step size. Perhaps variable step size integration methods could be used advantageously here, although it is felt that the time required would still be prohibitive.

The above problem is not encountered with the Gauss-Seidel second-level controller since the B_j matrix in (5.15) is not time

varying and Δt , once determined, remains constant. The problem of Section 5.2 was solved by this method considering both two and three fourth-order subsystems. The spatial increment Δx is given by

$$\Delta x = \frac{1}{D+1} \tag{6.8}$$

where

$$D = \sum_{j=1}^{N} n_j$$

In (6.8) D is the total dimension of the state vector. Hence for two fourth-order subsystems, $D = 8$, $\Delta x = 0.111$, and for three such subsystems, $D = 12$ and $\Delta x = 0.077$. The initial guesses required by the Gauss-Seidel controller were taken as

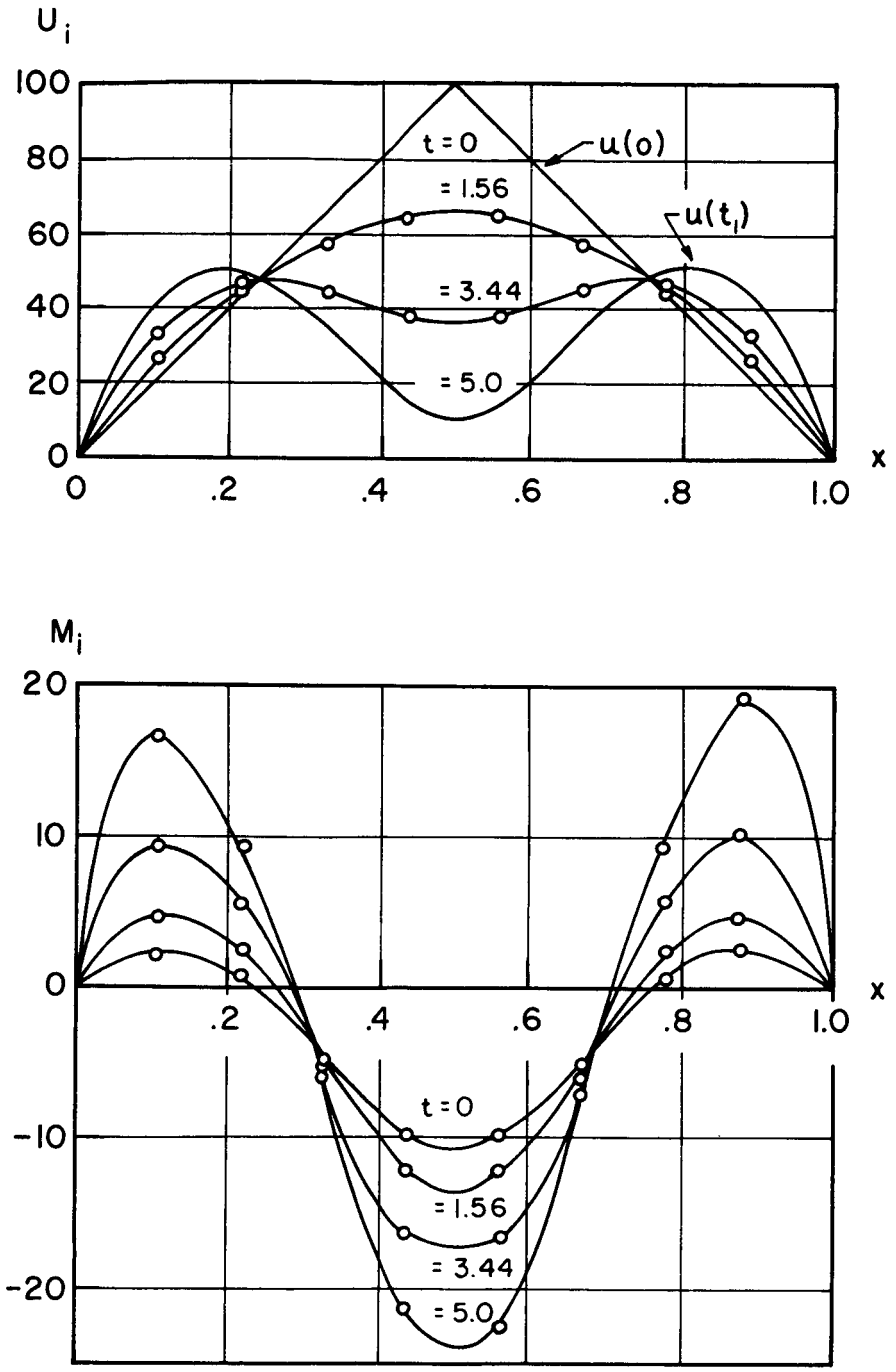
$$s_1^2 = 50$$
$$\rho_1^2 = -0.1$$

for the two subsystem case and

$$s_1^2 = s_2^2 = 50$$
$$\rho_1^2 = \rho_2^2 = -0.1$$

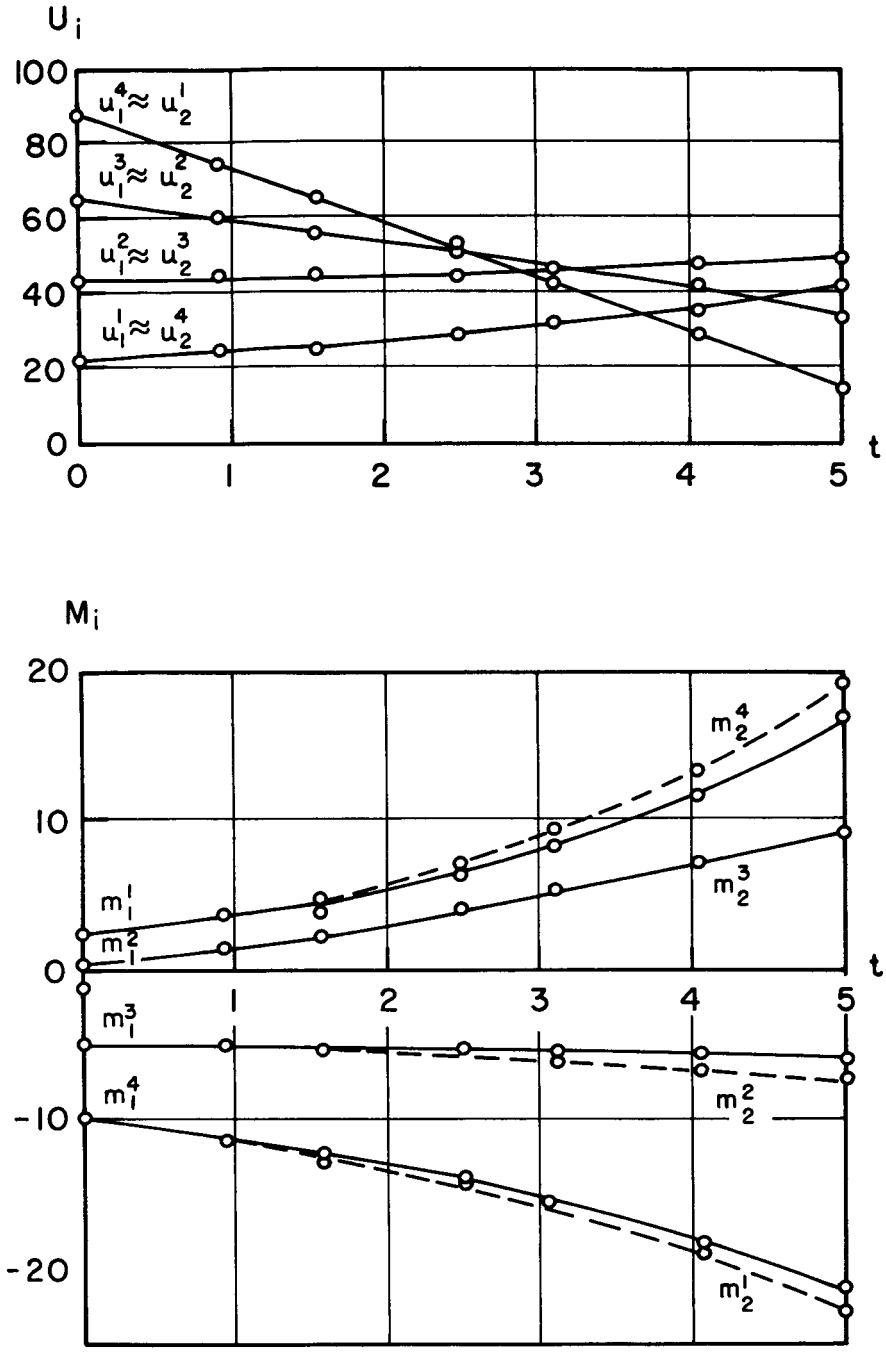
for the three subsystem case.

The initial and final space distribution for the state variables were respectively the triangle and double humped curves shown in Figure 6.3. Several intermediate state trajectories and the corresponding controls for the two subsystem case are also shown on Figure 6.3. These same trajectories are shown as functions of time in Figure 6.4. The initial guesses and final values of the coupling constraints are given in Figure 6.5. The behavior of the Hamiltonian function as the iteration proceeded is shown in Figure 6.6. For this problem the constancy of the Hamiltonian served as a good



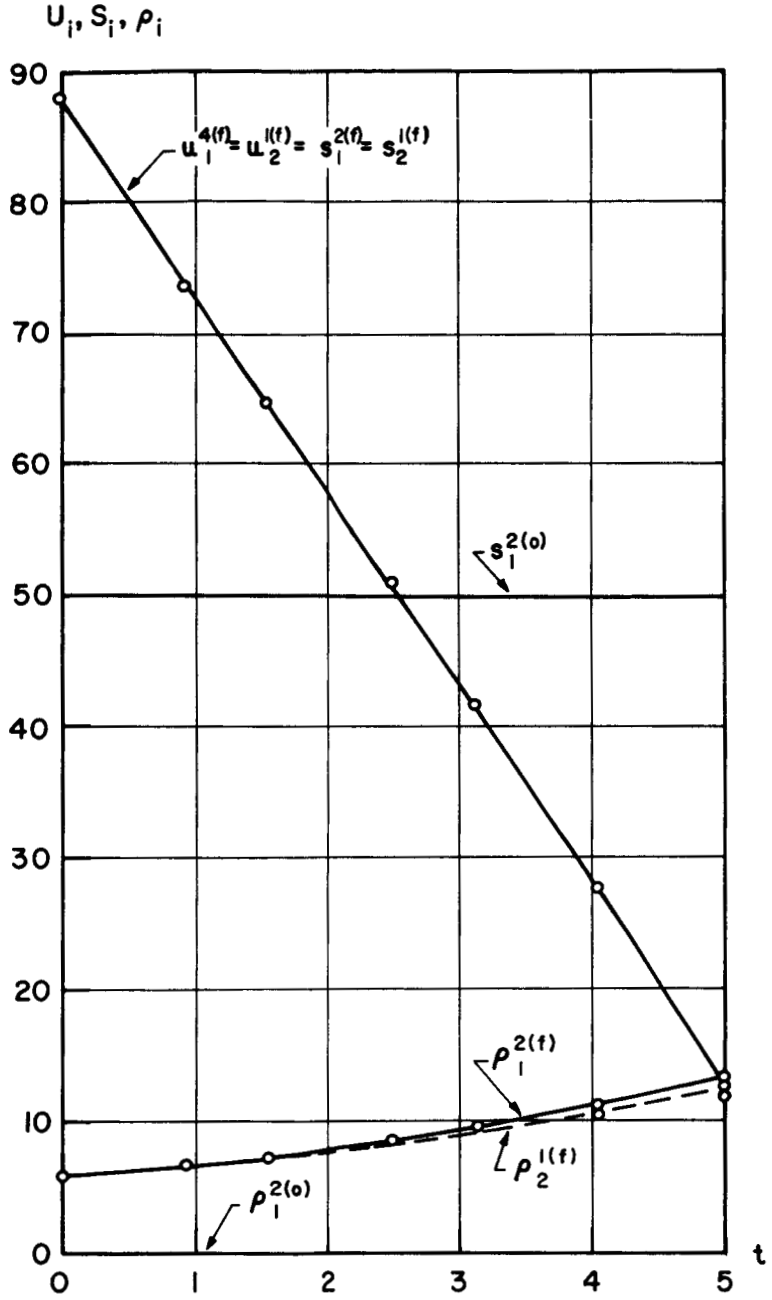
CONTROL AND RESPONSE FOR MINIMUM EFFORT
LINEAR EXAMPLE USING TWO SUBSYSTEMS

FIGURE 6.3



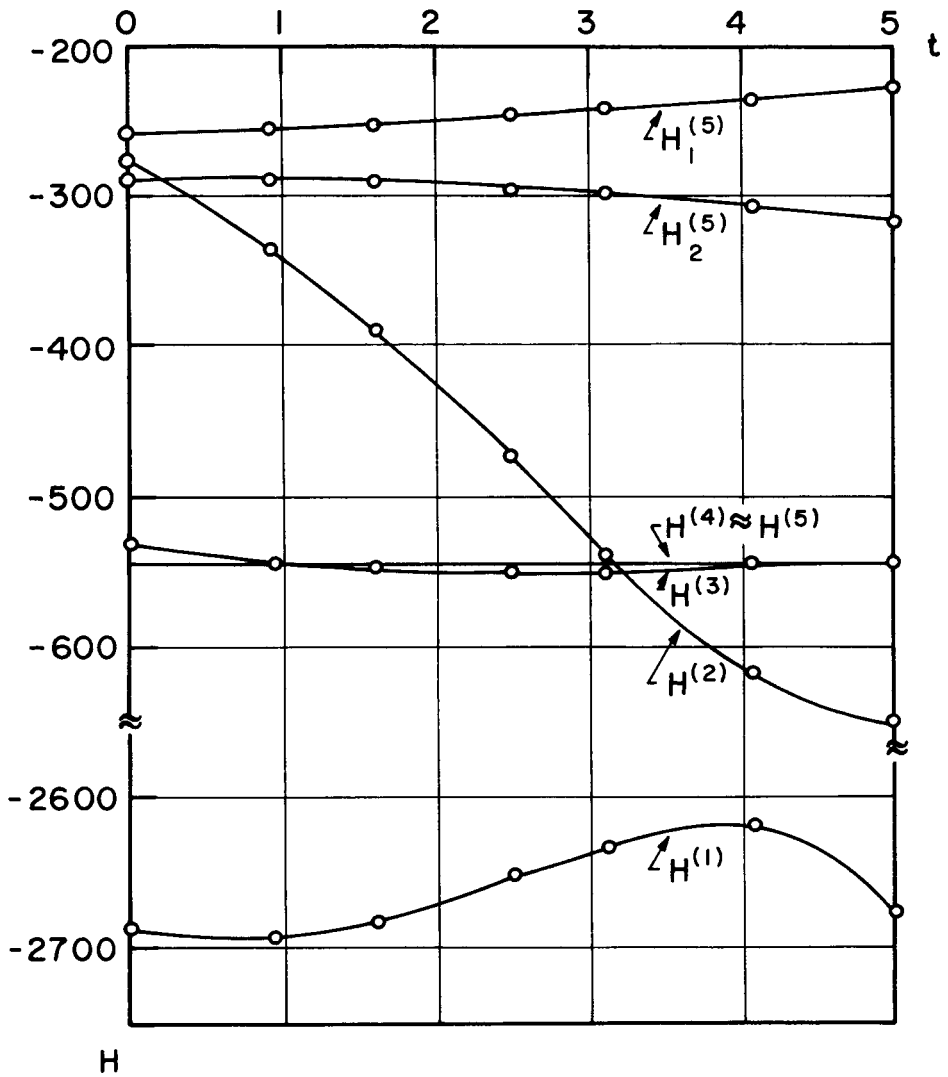
CONTROL AND RESPONSE AS FUNCTIONS OF TIME

FIGURE 6.4



INITIAL AND FINAL VALUES OF COUPLING CONSTRAINTS FOR TWO SUBSYSTEMS

FIGURE 6.5



HAMILTONIAN CONVERGENCE FOR TWO SUBSYSTEMS

FIGURE 6.6

check on the programming accuracy. Note that the total Hamiltonian is constant at -543.33 after 5 iterations, but that the subsystem Hamiltonians, H_1 and H_2, are not constant. This is to be expected since the subsystem solutions are not optimal when taken separately.

In order to evaluate the speed of convergence of the Gauss-Seidel second-level controller, the following norms are defined

$$||e_1||_i = \int_o^{t_1} \left\{ |u_1^4 - s_2^1|_i + |k\lambda_1^4 - \rho_2^1| \right\} dt$$
$$||e_2||_i = \int_o^{t_1} \left\{ |u_2^1 - s_1^2|_i + |k\lambda_2^1 - \rho_1^2| \right\} dt \qquad (6.9)$$

where the absolute values are on the coupling constraints (5.16) after the i^{th} iteration. For the two subsystem problem, (6.9) behaved as follows

Iteration (i)	$\|\|e_1\|\|_i$	$\|\|e_2\|\|_i$
1	$.176 \times 10^3$	$.320 \times 10^3$
2	$.213 \times 10^2$	$.253 \times 10^2$
3	.92	$.10 \times 10$
4	$.31 \times 10^{-1}$	$.37 \times 10^{-1}$
5	$.14 \times 10^{-2}$	$.17 \times 10^{-2}$

The subsystem optimization by quasilinearization converged rapidly whenever the time increment Δt was sufficiently small to insure integration accuracy. In each case the initial guesses for the state trajectories were linear curves satisfying the boundary conditions and the initial guesses for the adjoints were zero. The quasilinearization convergence rate was monitored by the norm

$$||E||_j = \max_{i=1,\ldots,8} \max_t |y_{ij}(t) - y_{i,j-1}(t)| \qquad (6.10)$$

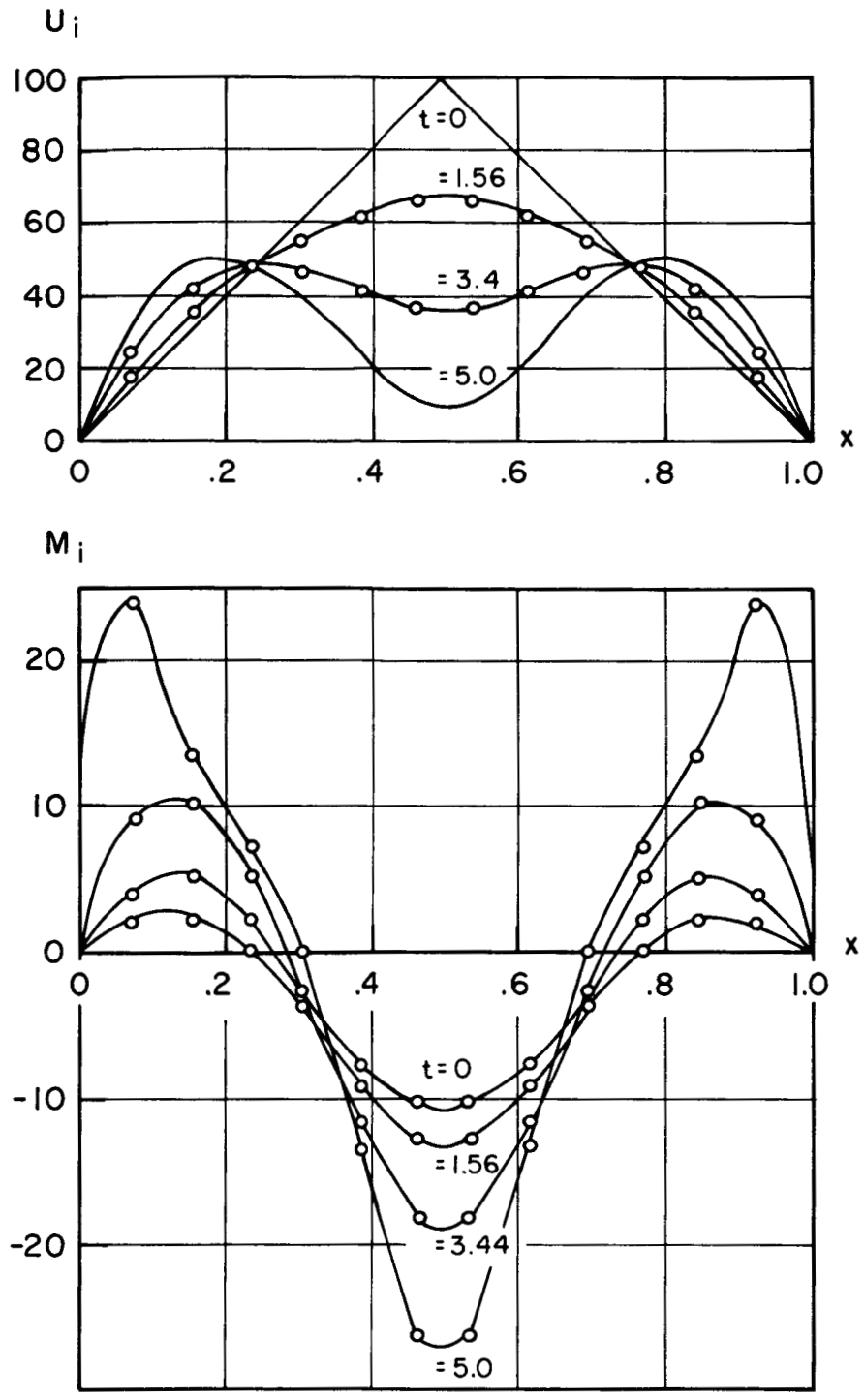
where j is the iteration number and i ranges over the four state and four adjoint solution trajectories. For a typical pass through the first-level control, (6.10) is evaluated as

Iteration (j)	Subsystem 1 $\|\|E\|\|_j$	Subsystem 2 $\|\|E\|\|_j$
1	$.786 \times 10^{2}$	$.42 \times 10^{2}$
2	$.133 \times 10$	$.584 \times 10$
3	$.91 \times 10^{-5}$	$.13 \times 10^{-4}$

The control functions and state variable response from the solution of this symmetric example using three subsystems are shown in Figure 6.7. The corresponding initial guesses and final coupling constraints are given in Figure 6.8. The second-level convergence for this case using (6.9) is given by

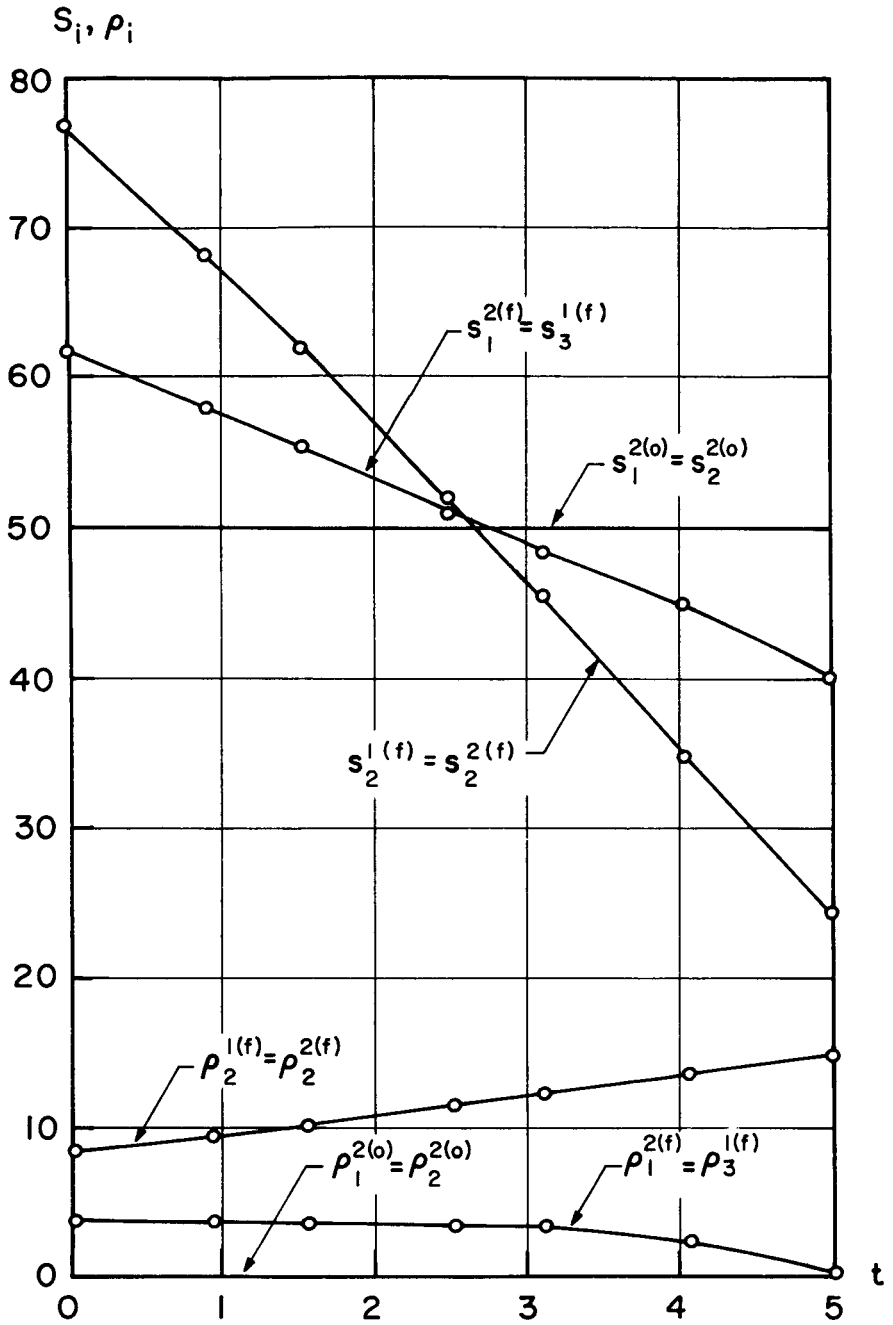
Iteration (i)	Direction	$\|\|e_1\|\|_i$	$\|\|e_2\|\|_i$	$\|\|e_3\|\|_i$
1	forward	$.161 \times 10^{3}$	$.291 \times 10^{3}$	$.337 \times 10^{3}$
2	backward	$.601 \times 10^{2}$	$.160 \times 10^{3}$	
3	forward		$.704 \times 10^{2}$	$.752 \times 10^{2}$
4	backward	$.622 \times 10$	$.187 \times 10$	
5	forward		$.755 \times 10$	$.808 \times 10$
6	backward	.582	.239	
7	forward		.711	.857
8	backward	.114	$.676 \times 10^{-1}$	
9	forward		.132	

It is interesting to note that the first and third subsystems converge monotonically while the second does not. However, the second subsystem does converge monotonically for all forward passes and all



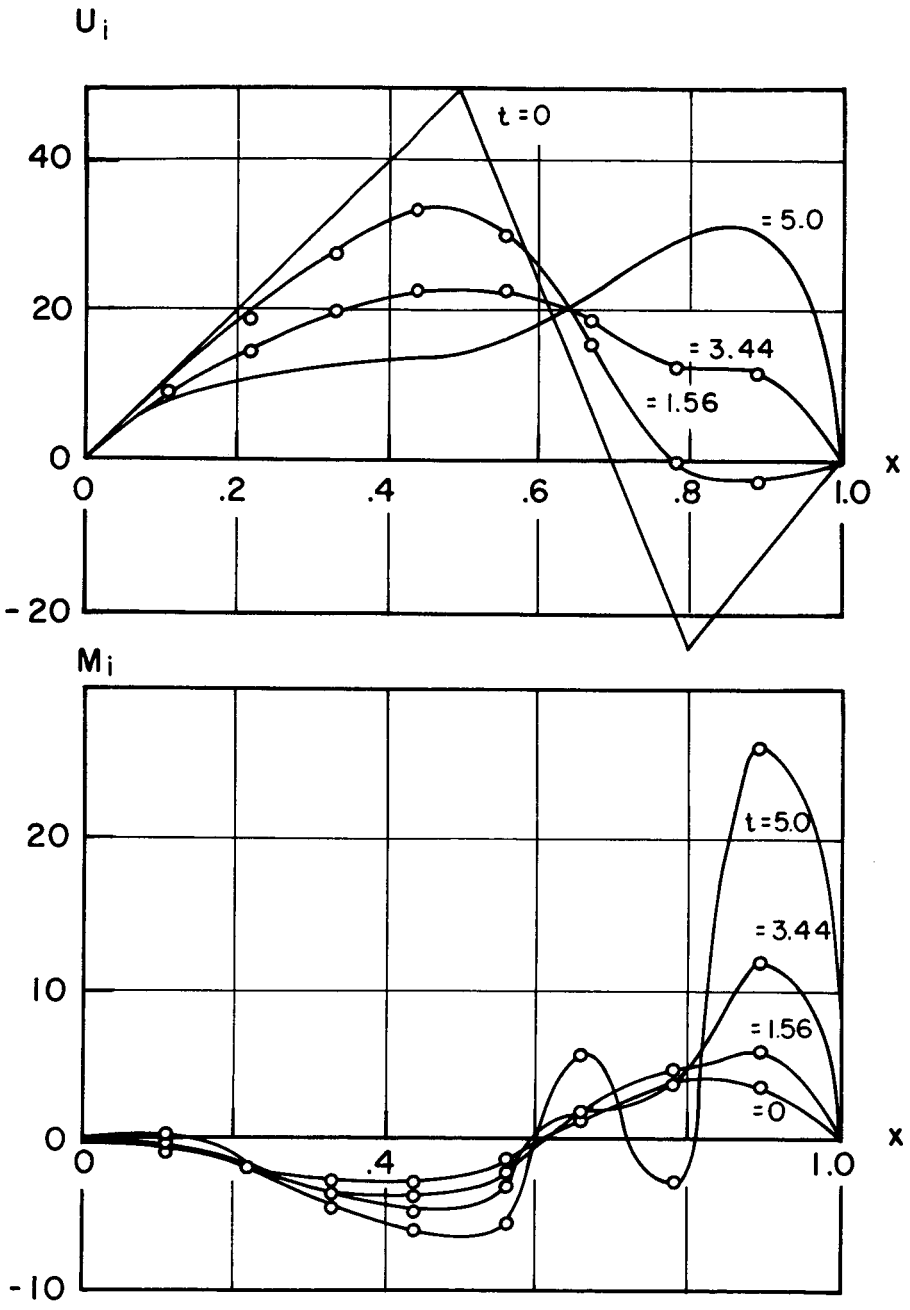
CONTROL AND RESPONSE FOR THREE SUBSYSTEMS

FIGURE 6. 7



COUPLING CONSTRAINTS FOR THREE SUBSYSTEMS

FIGURE 6.8



CONTROL AND RESPONSE FOR NONSYMMETRIC MINIMUM EFFORT EXAMPLE

FIGURE 6. 9

backward passes taken separately. The convergence rate is seen to be slower for the three subsystem problem than for the equivalent two subsystem problem due to the decrease in Δx. This point was mentioned in Chapter 4.

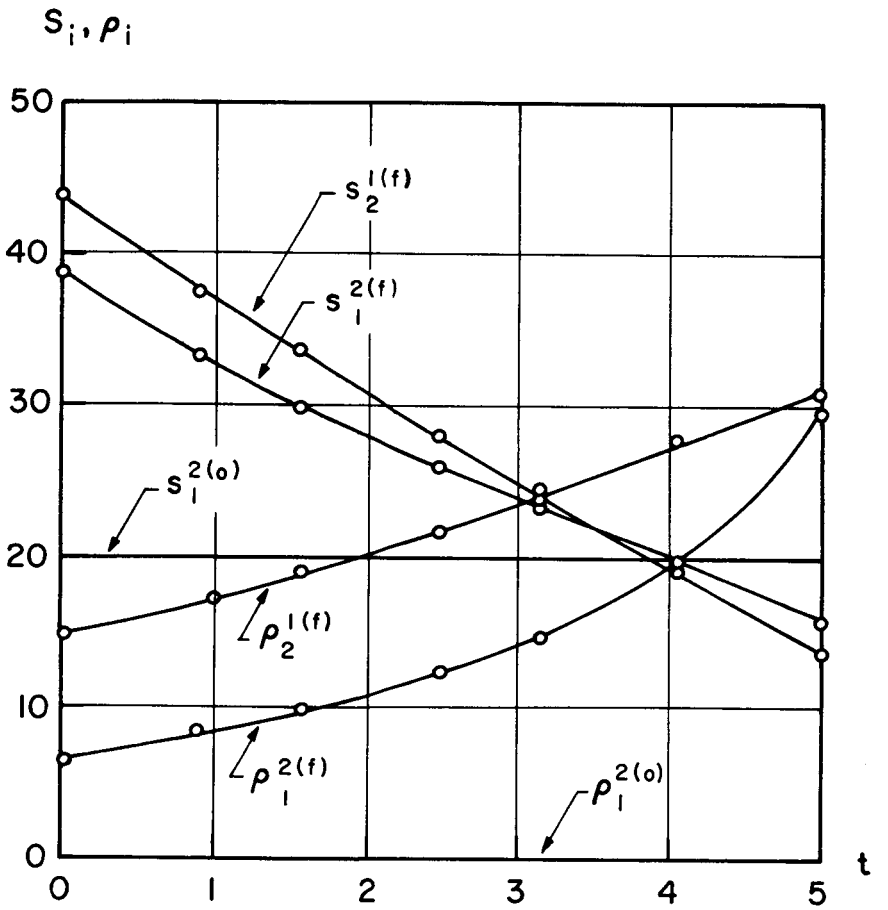
An example of a minimum effort, fixed end, linear problem was also solved using nonsymmetric initial and final spatial distributions for two subsystems. These distributions, along with intermediate trajectories and corresponding controls, are given in Figure 6.9. The initial guesses and final coupling constraints are shown in Figure 6.10. The second-level convergence for this example was similar to the symmetric two subsystem results given above.

6.4 Minimum Effort, Fixed End, Nonlinear Example with Distributed Control

The nonlinear problem formulated in Section 5.3 was solved using the numerical values in (6.7). Because of the nonlinearity, the elements of the matrix f_y are now proportional to the state variables which are of such a magnitude as to require a considerable reduction in Δt to preserve the integration accuracy. However, in this example, computation time can be saved by rescaling the state variables with a substitution

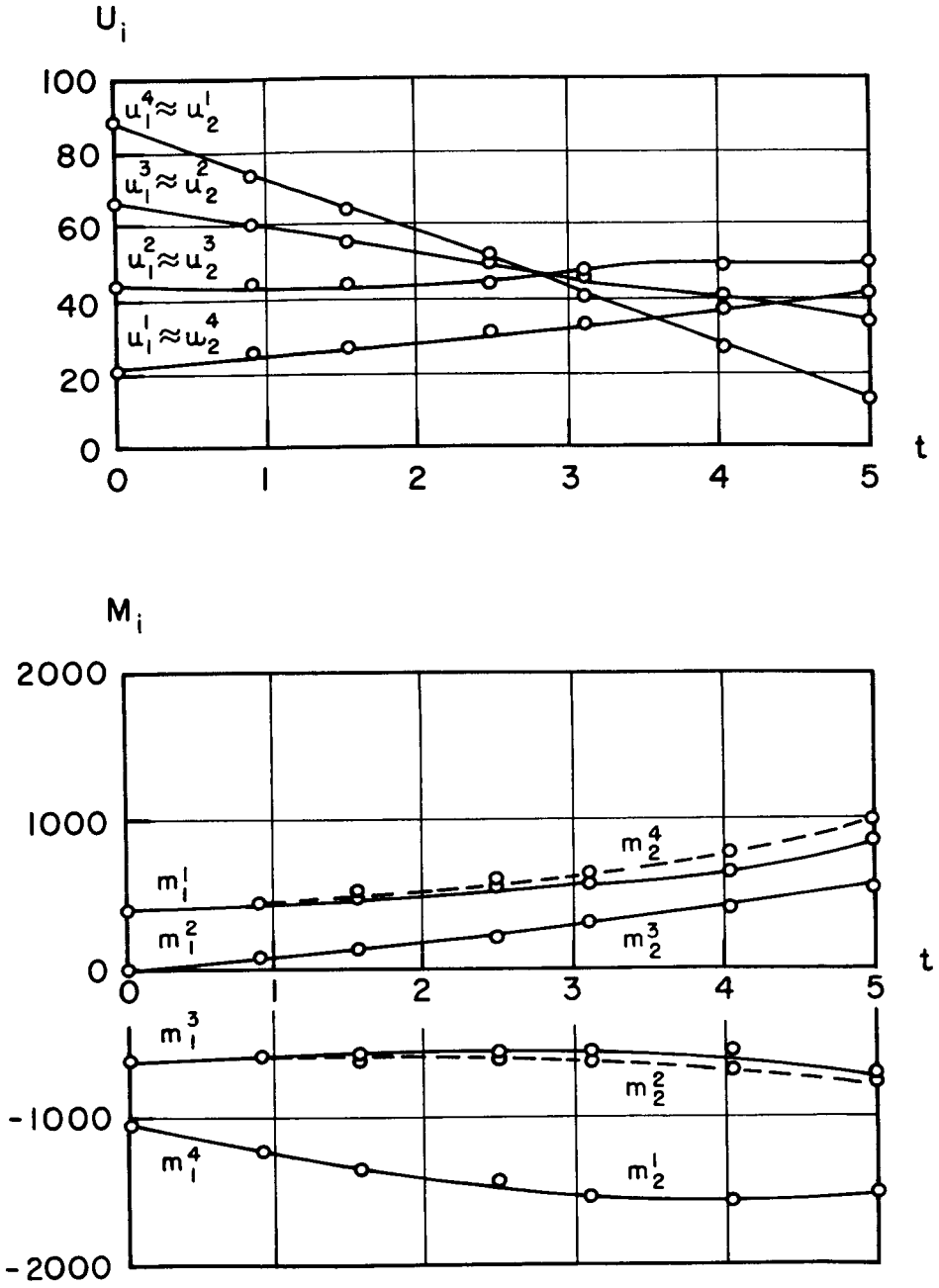
$$w = a u \qquad 0 < a < 1 \tag{6.11}$$

By choosing $a = .01$, the value of Δt used for the previous examples was sufficiently small to maintain integration accuracy. The controls and state variable response for this example are shown in Figure 6.11, where the initial and final space distributions are the symmetric ones used in Section 6.3. In this figure, the results have



COUPLING CONSTRAINTS FOR NONSYMMETRIC EXAMPLE

FIGURE 6.10



CONTROL AND RESPONSE FOR NONLINEAR EXAMPLE

FIGURE 6.11

been scaled back up for comparison with previous examples. Most noteworthy is the fact that the control magnitude increases appreciably.

The convergence of the second-level Gauss-Seidel controller in this example was comparable to the linear example of Section 6.3. This is not surprising since the mode of operation of the second-level controller is not affected by the nonlinearities even though the coupling constraint (5.22) is slightly more complex than (5.16). The convergence of the subsystems by quasilinearization as measured by (6.10) for a typical pass through the first level is given below:

Iteration (j)	Subsystem 1 $\|\|E\|\|_j$	Subsystem 2 $\|\|E\|\|_j$
1	.55	.39
2	$.25 \times 10^{-1}$	$.53 \times 10^{-1}$
3	$.86 \times 10^{-2}$	$.94 \times 10^{-2}$
4	$.19 \times 10^{-2}$	$.31 \times 10^{-2}$
5	$.53 \times 10^{-3}$	$.10 \times 10^{-2}$
6		$.37 \times 10^{-3}$

As expected, the subsystems converged somewhat slower than in the corresponding linear case.

6.5 Minimum Effort, State Inequality Constrained Example

The minimum effort problem described in Section 6.3 was also solved for two subsystems with the state inequality constraint

$$u_1^4 \leq c(t) \tag{6.12}$$

This problem was formulated in Section 5.4. The boundary $c(t)$ was taken as an ellipse

$$\frac{(c-a)^2}{d^2} + \frac{(t-b)^2}{e^2} = 1 \qquad (6.13)$$

where

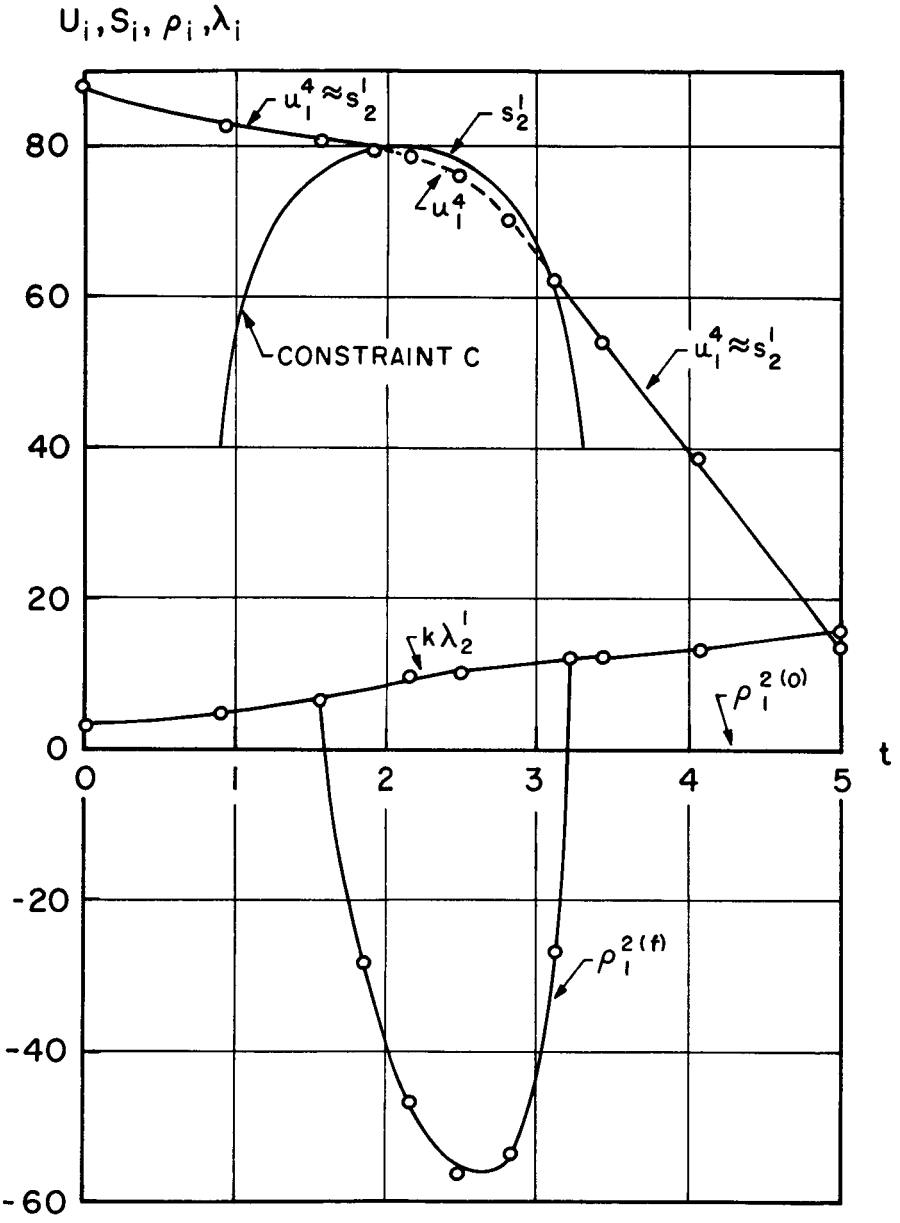
$$a = d = 40$$

$$b = 21\ \Delta t$$

$$e = 12\ \Delta t$$

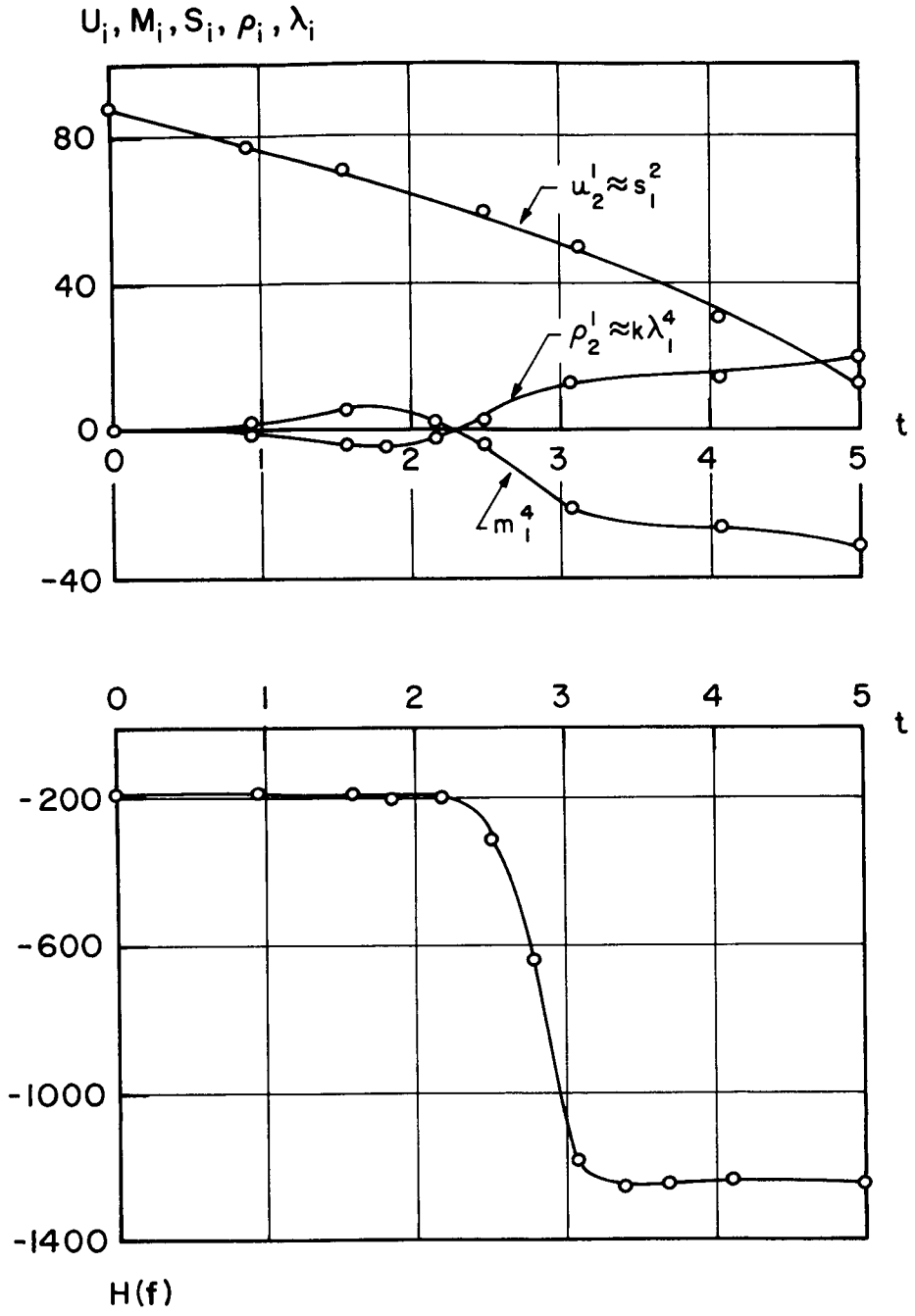
This boundary is plotted in Figure 6.12 along with the associated state and adjoint coupling constraints. As expected, the adjoint coupling constraint is not satisfied when the state variable is on (or near) the boundary. The Gauss-Seidel second-level controller was used to solve this problem except that along the boundary, a gradient technique was used to determine $\rho_1^2(t)$ as discussed in Section 5.4. Improvements in ρ_1^2 and the coupling constraints were made simultaneously at each iteration. Eight iterations were required to obtain the results shown in Figure 6.12. The step size (a) in (5.32) was taken as 0.95. The convergence of the first-level controllers by quasilinearization was similar to the corresponding linear example in Section 6.3.

The remaining two coupling constraints are shown in Figure 6.13 along with the control associated with the constrained state variable. It is this control alone as a function of λ_1^4 which acts in subsystem one to drive u_1^4 toward the boundary. Of course λ_1^4 in turn depends upon ρ_1^2 which is determined iteratively at the second level by (5.32). The remaining control variables and state responses were similar to those in Figure 6.3 for the unconstrained case. Figure 6.13 also shows the total Hamiltonian which is constant over both subarcs on which the state variable u_1^4 is unconstrained, but changes drastically when the state is forced (nearly) onto the constraint boundary.



STATE RESPONSE AND COUPLING CONSTRAINT FOR STATE INEQUALITY CONSTRAINED EXAMPLE

FIGURE 6.12



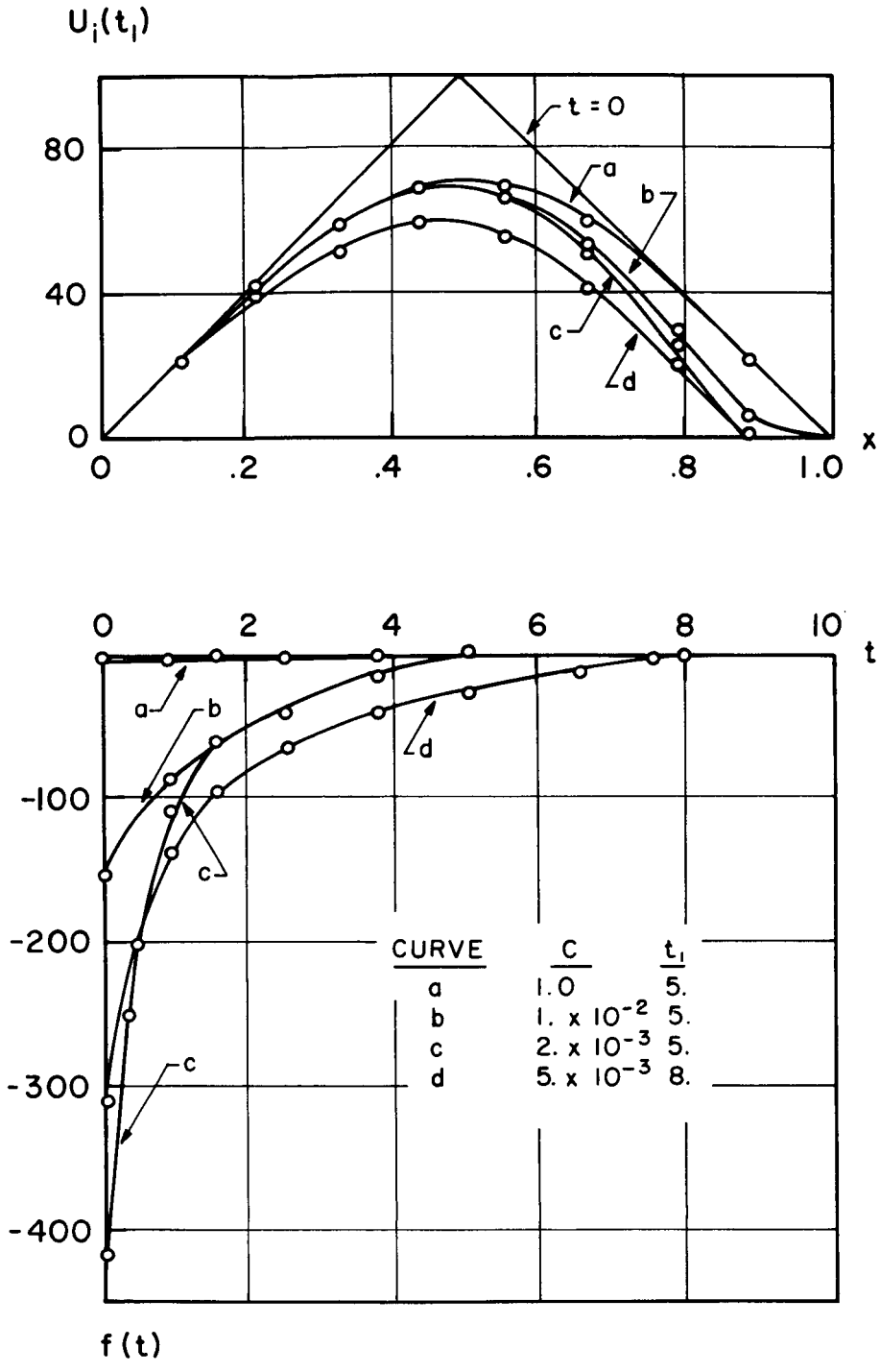
COUPLING CONSTRAINTS, CONTROL, AND HAMILTONIAN FOR THE STATE CONSTRAINED EXAMPLE

FIGURE 6.13

6.6 Minimum Error Plus Effort Example Using Boundary Control

The boundary control problem posed in Section 5.5 was solved using the criterion functional given by (5.71) for minimizing the error from some desired trajectory along the entire (time) path plus the control effort. The desired trajectory $u_d(x,t)$ was taken to be identically zero and the triangular initial distribution used previously was employed. Solutions were determined for a variety of values of the weighting factor c and several terminal times. The terminal states and the corresponding boundary control functions for these cases are given in Figure 6.14. The terminal error is seen to be reduced by decreasing the weighting on the control effort (and thereby increasing the absolute magnitude of the control), or by increasing the final time. The latter effect seems to be the more prominent in this example. Consider, for instance, a long thin rod being heated (or cooled) at one end only. The only way for a desired temperature profile to be attained is by heat conduction along the length of the rod and this rate of conduction is limited by the diffusivity α. Thus the desired trajectory can only be reached by increasing either the final time, the absolute magnitude of the control, or both. The relative effect of these measures depends on the magnitude of α.

Since the final states are free in this example, the final values of the boundary control are always zero. Furthermore, the large negative values of the control drive the state variable nearest the controlled boundary negative over a portion of the time interval.



BOUNDARY CONTROLS AND TERMINAL STATES FOR MINIMUM ERROR PLUS EFFORT EXAMPLE

FIGURE 6.14

CHAPTER 7

CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER STUDY

The (approximate) solution of optimal control problems involving partial differential equations is studied by discretizing the space domain and considering the resultant set of ordinary differential equations. This approach is certainly not new. However, earlier efforts along this path have been hampered by the difficulties involved in solving the optimal control problem for the very large sets of interacting ordinary differential equations which arise as the discretization interval is decreased. This dissertation suggests the use of multilevel control techniques to overcome this computational difficulty.

The major conclusion stemming from this research is that the multilevel approach does appear feasible in solving the optimal control problem for certain classes of distributed parameter systems; namely, linear and nonlinear parabolic or elliptic equations. The convergence properties of the second-level controller are of paramount importance in accomplishing this task. Of the three types of second-level controllers discussed here (feasible, nonfeasible, and Gauss-Seidel), the only one considered suitable in this application is the Gauss-Seidel controller. It is extremely simple and was found to have good convergence properties for this type of problem. In particular, the systems of semidiscrete equations may become very large and the number of subsystems likewise. By its very nature, the performance of the Gauss-Seidel controller does not seem to be degraded by increasing the number of subsystems as long as the discretization interval (Δx) is not "too small." The reason for the

good convergence properties observed for this controller is largely the tridiagonal (Jacobi[24]) form of the A_j matrix.

This method of solution appears particularly attractive for obtaining a "rough cut" for problems which will require extensive further study. However, the multilevel form of solution can also be used to obtain more accurate results at the expense of considerable use of computer time. One restriction on the improved accuracy which can be attained is the amount of fast-access computer memory available. One significant advantage of this approach is that it may permit the solution of problems not otherwise possible. In particular, analytical results are very difficult to achieve for problems which are (1) nonlinear, (2) time and/or space varying, or (3) of space dimension greater than one. However, all of these complications can be handled in the framework of the multilevel solution described here.

It should be noted that these optimistic results are not universal for multilevel control techniques. In particular, Bauman[6] states that decomposition and multilevel control should be used only on systems having "one or two coupling equations between subsystems." This recommendation is based on computational experience with a number of fairly simple and low-order systems. It is felt that the satisfactory results reported in this dissertation for higher-order systems are due mainly to the type of problem and its formulation as well as the type of second-level controller employed.

Several areas stand out as fruitful for further research. In particular more work should be done towards obtaining necessary as well as sufficient conditions for the convergence of the Gauss-Seidel (and other) controllers. Other types of second-level controllers with

improved convergence properties would also be worthwhile. One possibility stems from various n-step methods such as the conjugate gradient method.[27, 29]

As pointed out earlier, the number of optimal control problems which can be formulated for distributed parameter systems is very large. As always, more computational experience is warranted. Of particular interest would be (1) larger problems having more subsystems, (2) highly nonlinear problems, (3) problems having higher space dimension than one, (4) boundary control problems. Many interesting questions in the last area remain to be resolved.

The steady state optimization problem resulting from discretizing elliptic partial differential equations was briefly treated here but no computational experience was obtained. Some effort could seemingly be directed toward both the theoretical and computational aspects of this problem.

REFERENCES

1. Ahiezer, N.I., and M. Krein, Some Questions in the Theory of Moments, Providence, R.I., American Mathematical Society, 1962 (translated from the 1938 Russian edition).

2. Antosiewicz, H.A., and W.C. Rheinboldt, "Numerical Analysis and Functional Analysis," Survey of Numerical Analysis, John Todd, Editor, New York, McGraw-Hill, 1962.

3. Athans, Michael, and Peter L. Falb, Optimal Control, New York, McGraw-Hill, 1966.

4. Axelband, Elliot I., The Optimal Control of Certain Classes of Linear Distributed Parameter Systems, Ph.D., Dept. of Engineering, University of California, Los Angeles, 1966.

5. Balakrishnan, A.V., and H.C. Hsieh, "Proceedings of the Optimum System Synthesis Conference," Flight Dynamics Laboratory, Wright-Patterson Air Force Base, Ohio Technical Report, ASD-TDR-63-119, February 1963.

6. Bauman, Edward J., Multilevel Optimization Techniques with Application to Trajectory Decomposition, Ph.D., Dept. of Engineering, University of California, Los Angeles, 1966.

7. Bellman, Richard E., Dynamic Programming, Princeton, Princeton University Press, 1957.

8. Berkovitz, L.D., "On Control Problems with Bounded State Variables," Jour. Math. Anal. and App. Vol. 5, pp. 488-498, 1962.

9. Bliss, Gilbert A., Lectures on the Calculus of Variations, Chicago, The University of Chicago Press, 1946.

10. Breakwell, J.V., et.al., "Optimization and Control of Non-linear Systems Using the Second Variation," Jour. SIAM on Control, Vol. 1, No. 2, pp. 193-223, 1963.

11. Brogan, W.L., Optimal Control Theory Applied to Systems Described by Partial Differential Equations, Ph.D., Dept. of Engineering, University of California, Los Angeles, 1965.

12. Brosilow, C.B., et.al., "Feasible Optimization Methods for Interconnected Systems," Preprints of the 1965 Joint Automatic Control Conference, June 22-25, 1965, Rensselear Polytechnic Institute, Troy, N.Y.

REFERENCES (Continued)

13. Bryson, A.E., Jr., et. al., "Optimal Programming Problems with Inequality Constraints, I: Necessary Conditions for Extremal Solutions," AIAA Journal, Vol. 1, No. 11, pp. 2544-2550, November 1963.

14. Bryson, A.E., Jr., et. al., "Lift of Drag Programs That Minimize Re-Entry Heating," Journal of the Aerospace Sciences, Vol. 29, pp. 420-430, April 1962.

15. Butkovskii, A.G., and A.Y. Lerner, "The Optimum Control of Systems with Distributed Parameters," Automation and Remote Control, Vol. 21, No. 6, pp. 472-477, 1960.

16. Butkovskii, A.G., "Optimum Processes in Systems with Distributed Parameters," Automation and Remote Control, Vol. 2, No. 1, pp. 13-21, 1961.

17. Butkovskii, A.G., "The Broadened Principle of the Maximum for Optimal Control Problems," Automation and Remote Control, Vol. 24, No. 3, pp. 292-304, 1963.

18. Collina, G.L., and P. Dorato, "Application of Pontryagin's Maximum Principle: Linear Control Systems," Polytechnic Institute of Brooklyn, Microwave Research Institute, Brooklyn, N.Y., Report No. PIBMRI-1015-62, June 1962.

19. Dantzig, G.B., "Linear Control Processes and Mathematical Programming," Jour. SIAM on Control, Vol. 4, No. 1, pp. 56-60, February 1966.

20. Dantzig, G.B., and P. Wolfe, "Decomposition Principle for Linear Programs," Operations Research, Vol. 8, No. 1, pp. 101-111, 1960.

21. Dreyfus, S.E., "Variational Problems with State Variable Inequality Constraints," The RAND Corporation, Technical Report, No. P-2605-1, July 1962; Revised, August 1963.

22. Egorov, A.I., "On Optimal Control of Processes in Distributed Objects," Jour. App. Math. and Mech. (PMM), Vol. 27, No. 4, pp. 1045-1058, 1963.

23. Friedman, Bernard, Principles and Techniques of Applied Mathematics, New York, Wiley, 1956.

24. Gelfand, Izrail M., and Sergel V. Fomin, Calculus of Variations, Englewood Cliffs, N.J., Prentice-Hall, 1963, (translated by R.A. Silverman.

REFERENCES (Continued)

25. Gilbert, E. G., "Controllability and Observability of Multivariable Control Systems," Jour. SIAM on Control, Vol. 2, No. 1, pp. 128-151, 1963.

26. Hamming, R. W., "Stable Predictor Corrector Methods for Ordinary Differential Equations, Jour. ACM, Vol. 6, pp. 37-47, January 1959.

27. Hays, R. M., Iterative Methods of Solving Linear Problems on Hilbert Space, Ph. D., Dept. of Mathematics, University of California, Los Angeles, 1952.

28. Hestenes, Magnus R., Calculus of Variations and Control Theory, New York, Wiley (to be published).

29. Hestenes, Magnus R., and E. Stiefel, "Methods of Conjugate Gradients for Solving Linear Systems," Jour. Res. Natl. Bur. Stds., Vol. 49, No. 6, pp. 409-436, December 1952.

30. Kalaba, R., "On Nonlinear Differential Equations, the Maximum Operation, and Monotone Convergence," Jour. Math. Mech., Vol. 8, No. 4, pp. 519-574, July 1959.

31. Kalman, R. E., "On the General Theory of Control Systems," (First International Congress on Automatic Control, Moscow, 1960), Automatic and Remote Control, Vol. 1, J. F. Coales, Editor, London, Butterworths, pp. 481-492, 1961.

32. Keller, H. B., "The Numerical Solution of Parabolic Partial Differential Equations," Mathematical Methods for Digital Computers, Anthony Ralston and Herbert S. Wilf, Editors, New York, Wiley, 1960.

33. Kelley, H. J., "Guidance Theory and Extremal Fields," Transactions IRE on Automatic Control, Vol. AC-7, October 1962.

34. Kelley, H. J., "Method of Gradients," Optimization Techniques, George Leitman, Editor, New York, Academic Press, 1962 (Mathematics in Science and Engineering, 5).

35. Kelley, H. J., et. al., "Air Vehicle Trajectory Optimization," presented at Symposium on Multivariable System Theory, Fall meeting of Soc. Ind. Appl. Math., Cambridge, Mass., 1962.

REFERENCES (Continued)

36. Kolmogorov, A.N. and S.V. Fomin, Elements of the Theory of Functions and Functional Analysis, Vol. 1, Rochester, N.Y., Graylock Press, 1957 (translated from 1954 Russian edition).

37. Kranc, G.M., and P.E. Sarachik, "An Application of Functional Analysis to the Optimal Control Problem," Air Force Office of Scientific Research, Washington, D.C., Technical Report, No. 2102, January 22, 1962.

38. Kron, Gabriel, Tensor Analysis of Networks, New York, Wiley, 1949.

39. Kuhn, H.W., and A.W. Tucker, "Nonlinear Programming," Proceedings of the Second Berkeley Symposium of Mathematical Statistics and Probability, Berkeley, University of California Press, pp. 481-492, 1950.

40. Kulikowski, T., "Optimum Control of Multidimensional and Multilevel Systems," Advances in Control Systems, Vol. 3, Cornelius T. Leondes, Editor, New York, Academic Press, 1966.

41. Lasdon, L.S., "A Multilevel Technique for Optimization," Systems Research Center Case Institute of Technology, Cleveland, Report No. SRC-50-C-64-19, April, 1964.

42. Lasdon, L.S. and J.D. Schoeffler, "A Multilevel Technique for Optimization," Preprints of the 1965 Joint Automatic Control Conference, Rensselaer Polytichnic Institute, Troy, N.Y., 1965.

43. Lefkowitz, I., "Multilevel Approach Applied to Control System Design," Preprints of the 1965 Joint Automatic Control Conference, Reusselear Polytechnic Institute, Troy, N.Y., 1965.

44. Liusternik, L.A., and V.J. Sobolev, Elements of Functional Analysis, New York, Ungar Publishing Co., 1961.

45. Macko, D., and J.D. Pearson, 'A Multilevel Formulation of Nonlinear Dynamic Optimization Problems,' "Papers on Multilevel Control Systems," Systems Research Center Case Institute of Technology, Report No. SRC 70-A-65-25, Cleveland, 1965.

REFERENCES (Continued)

46. McGill, R., "Optimal Control Inequality State Constraints and the Generalized Newton-Raphson Algorithm," Jour. SIAM on Control, Vol. 3, No. 2, pp. 291-298, 1965.

47. McGill, R. and P. Kenneth, "Solution of Variational Problems by Means of a Generalized Newton-Raphson Operator," AIAA Journal, Vol. 2, No. 10, pp. 1761-1766, October 1964.

48. Mesarovic, M. D., "A General Systems Approach to Organization Theory," Systems Research Center Case Institute of Technology, Cleveland, Report No. SRC 2-A-62-2, 1961.

49. Mesarovic, M. D., et. al., "Advances in Multilevel Control," Paper presented at the International Federation of Automatic Control Symposium, August 1965, Tokyo.

50. Newman, M., "Matrix Computations," Survey of Numerical Analysis, John Todd, Editor, New York, McGraw-Hill, 1962.

51. Paine, Garrett, The Application of Quasilinearization to the Computation of Optimal Control, Ph. D., Dept. of Engineering, University of California, Los Angeles, 1966.

52. Pearson, J. D., "Duality and a Decomposition Technique," Jour. SIAM on Control, Vol. 4, No. 1, pp. 164-172, February 1966.

53. Pearson, J. D., "Reciprocity and Duality in Control Programming Problems," Jour. Math. Anal. and App., Vol. 10, pp. 388-408, 1965.

54. Pontryagin, Lev S., et. al., The Mathematical Theory of Optimal Processes, L. W. Neustadt, Editor, New York, Interscience-Wiley, 1962 (translated by K. N. Trirogoff).

55. Ralston, A., "Numerical Intergration Methods for the Solution of Ordinary Differential Equations," Mathematical Methods for Digital Computers, Anthony Ralston and Herbert S. Wilf, Editors, New York, Wiley, 1960.

56. Sakawa, Y., "Solution of an Optimal Control Problem in a Distributed Parameter System," IEEE Trans. on Automatic Control, Vol. AC-9, pp. 420-426, October 1964.

57. Sakawa, Y., "Optimal Control of a Certain Type of Linear Distributed Parameter Systems," IEEE Trans, on Automatic Control, Vol. AC-11, pp. 35-41, January 1966.

REFERENCES (Continued)

58. Scharmack, D. K., "Proceeding of the Optimum System Synthesis Conference," Flight Dynamics Laboratory, Wright-Patterson Air Force Base, Ohio, Technical Report, ASD-TDR-63-119, February 1963 (pp. 119-158)

59. Takahara, Y., 'On the Synthesis of a Multilevel Structure for a Class of Linear Dynamic Optimization Problems,' "Papers on Multilevel Control Systems," Systems Research Center Case Institute of Technology, Report No. SRC 70-A-65-25, Cleveland, 1965.

60. Valentine, Frederick A., "The Problem of Lagrange with Differential Inequalities as Added Side Conditions," Contributions to the Calculus of Variations, Chicago, University of Chicago Press, 1937.

61. Van Norton, R., "The Solution of Linear Equations by the Gauss-Seidel Method," Mathematical Methods for Digital Computers, Anthony Ralston and Herbert S. Wilf, Editors, New York, Wiley, 1960.

62. Varga, Richard S., Matrix Iterative Analysis, New York, Prentice-Hall, 1962.

63. Varaiya, P. P., "Nonlinear Programming and Optimal Control," Electronics Research Laboratory, University of California, Berkeley, ERL Tech. Memo. M-129, September 1965.

64. Wang, P. K. C., "Control of Distributed Parameter Systems," Advances in Control Systems, Vol. 1, Cornelius T. Leondes, Editor, New York, Academic Press, 1964.

65. Wang, P. K. C., and F. Tung, "Optimum Control of Distributed Parameter Systems," Preprints of the 1963 Joint Automatic Control Conference, University of Minnesota, Minneapolis, June 19-21, 1963 (pp. 16-31).

66. Young, D. M., Jr., "The Numerical Solution of Elliptic and Parabolic Partial Differential Equations," Survey of Numerical Analysis, John Todd, Editor, New York, McGraw-Hill, 1962.

APPENDIX A

AN EXAMPLE OF SUFFICIENT CONVERGENCE CONDITIONS FOR THE GAUSS-SEIDEL CONTROLLER

Consider the homogeneous linear differential equation

$$\dot{Z} = AZ \qquad Z(0) = Z_o \tag{A.1}$$

where

$$Z^T = [U, \lambda]$$

and A is a Jacobi type[24] matrix of constants. The sufficient convergence criterion given by Kolmogorov[36] requires that (A.1) satisfy a Lipschitz condition with respect to Z given by (2.36). The Lipschitz constant L of (2.36) can be determined as follows:

$$\left|f_i(Z^1) - f_i(Z^2)\right| = \sum_{j=1}^{n} a_{ij}\left(z_j^2 - z_j^1\right) \tag{A.2}$$

$$\leq \sum_{j=1}^{n} \left|a_{ij}\left(z_j^2 - z_j^1\right)\right| \tag{A.3}$$

By the Hölder inequality[36] (A.3) is

$$\leq \sum_{j=1}^{n} |a_{ij}|\left|z_j^2 - z_j^1\right| \tag{A.4}$$

$$\leq \max\left\{\left|z_j^2 - z_j^1\right| : 1 \leq j \leq n\right\} \sum_{j=1}^{n} |a_{ij}| \tag{A.5}$$

To find the value of L which is sufficiently large to satisfy (2.36) for all i, choose

$$L = \max_i \sum_{j=1}^{n} |a_{ij}| \qquad i = 1, \ldots, n \tag{A.6}$$

As a specific example, consider the minimum effort problem discussed in Sections 5.2 and 6.3. Using the formulation for a Gauss-Seidel type second-level controller yields

$$L = k\left[|1| + |-2| + |1|\right] + |-0.5| \qquad k = \frac{\alpha}{(\Delta x)^2} \qquad \text{(A.7)}$$

$$= 4k + 0.5$$

For the two subsystem case $k = 0.268$, and for the three subsystem case $k = 0.557$. For these two cases, $L = 1.572$ and $L = 2.728$ respectively. According to the sufficiency condition (2.38) repeated here for convenience

$$L(t_1 - t_o) < 1$$

or for two subsystems (taking $t_o = 0$)

$$t_1 < \frac{1}{1.572} = 0.635 \qquad \text{(A.8)}$$

and for three subsystems

$$t_1 < \frac{1}{2.728} = 0.367 \qquad \text{(A.9)}$$

However, these examples were solved using $t_1 = 5$ and the Gauss-Seidel controller converged very rapidly as shown in Section 6.3. The two subsystem example was also solved using $t_1 = 10$ with equally good results.

Thus (as expected) the sufficient conditions are seen to be quite conservative for the Lipschitz constant determined above. What is really needed are necessary conditions for convergence. The determination of such conditions is a topic for future research.

APPENDIX B

CONSISTENCY AND CONVERGENCE OF THE SEMIDISCRETE APPROXIMATION OF A LINEAR PARABOLIC PARTIAL DIFFERENTIAL EQUATION

Consider the linear one-dimensional parabolic partial differential equation given by

$$L\left[u(x,t)\right] \triangleq u_t - a(x,t)\,u_{xx} - 2b(x,t)\,u_x + c(x,t)\,u = d(x,t) \tag{B.1}$$

where

$$a(x,t) > 0 \tag{B.2}$$

A solution of (B. 1) is uniquely determined over the semi-infinite strip

$$R:\left[0 \le x \le L;\ t \ge 0\right] \tag{B.3}$$

by specifying appropriate initial and boundary conditions; say

$$\begin{aligned} u(x,0) &= f(x) && 0 \le x \le L \\ u(0,t) &= g_o(t) && t > 0 \\ u(L,t) &= g_1(t) && t > 0 \end{aligned} \tag{B.4}$$

Define a grid on the x domain by

$$R_h:\left[x_j = jh,\quad j = 0, \ldots, J+1\right] \tag{B.5}$$

where

$$h = \frac{L}{J+1}$$

and let

$$u_j = u(x_j, t)$$

$$d_j = d(x_j, t)$$

Using the semidiscrete approximation defined by

$$u_x(x_j, t) \approx \frac{1}{2h}\left(v_{j+1} - v_{j-1}\right)$$

$$u_{xx}(x_j, t) \approx \frac{1}{h^2}\left(v_{j+1} - 2v_j + v_{j-1}\right) \qquad \text{(B. 6)}$$

$$u_t(x_j, t) = \frac{du_j}{dt} = \frac{dv_j}{dt}$$

where $v_j = v(x_j, t)$ is the solution to the ordinary differential equations obtained by substituting (B. 6) into (B. 1), yields for (B. 1)

$$L_j\left[v(x_j, t)\right] \triangleq \frac{dv_j}{dt} - \lambda a_j(t)\left[v_{j+1} - 2v_j + v_{j-1}\right] - h\lambda b_j(t)\left[v_{j+1} - v_{j-1}\right]$$
$$+ c_j(t)\, v_j = d_j(t) \qquad \text{(B. 7)}$$

where

$$\lambda = \frac{1}{h^2}$$

The initial and boundary conditions become

$$v_j(0) = f(x_j) \qquad 0 \le j \le J + 1$$

$$v_o(t) = g_o(t) \qquad t > 0 \qquad \text{(B. 8)}$$

$$v_{J+1}(t) = g_1(t) \qquad t > 0$$

Using a natural extension of the definition given by Keller,[32] the semidiscrete approximation (B. 7) is said to be consistent with (B. 1) if

$$\lim_{h \to 0}\left[L[u(x, t)] - L_j[u(x, t)]\right] = 0 \qquad \text{(B. 9)}$$

This condition insures that the equations (B. 7) actually do approximate the partial differential equation (B. 1). By employing Taylor's formula, the derivative terms appearing in the difference in (B. 9) can be written

$$u_x(x_j, t) - \frac{1}{2h}\left(u_{j+1} - u_{j-1}\right) = \frac{h^2}{6}\,\bar{u}_{xxx} = \tau^{(1)}$$

$$u_{xx}(x_j, t) - \frac{1}{h^2}\left(u_{j+1} - 2u_j + u_{j-1}\right) = \frac{h^2}{12}\,\bar{u}_{xxxx} = \tau^{(2)} \qquad \text{(B. 10)}$$

$$u_t(x_j, t) - \frac{du_j}{dt} = 0$$

where the bar indicates that the derivatives are evaluated at appropriate intermediate values, and $\tau^{(1)}$ and $\tau^{(2)}$ are the truncation errors for the respective approximations. Thus

$$L\left[u(x_j, t)\right] - L_j\left[u(x_j, t)\right] = -a_j(t)\,\tau^{(2)} - 2\,b_j(t)\,\tau^{(1)} = O(h^2) \qquad \text{(B.11)}$$

Assuming that the derivatives of u and the coefficients a and b are bounded, the right hand side of (B. 11) goes to zero as h goes to zero and consistency is proved.

Again extending a definition of Keller, the ordinary differential equations (B. 7) are said to be convergent if their solution satisfies

$$\lim_{h \to 0} \left|u(x_j, t) - v_j(t)\right| = 0 \qquad \text{(B.12)}$$

Convergence insures, at least for a sufficiently fine mesh, that the solution of (B. 7) is a "close" approximation to the solution of (B. 1). Before proving that (B. 7) is convergent, it is necessary to prove the following lemma.

Lemma

On every net R_h satisfying

$$2\lambda a(x, t) + c(x, t) \geq 0 \qquad \text{(B.13)}$$
$$a(x, t) - h\left|b(x, t)\right| \geq 0$$

the solution $v_j(t)$ of the ordinary differential equation (B. 7) and (B. 8) is bounded by

$$V(t) \leq \max\left[\frac{D}{C}, B\right] \qquad C > 0 \qquad \text{(B.14a)}$$

$$\leq \max\left[G, \left(F + \frac{D}{|C|}\right) e^{+|C|t}\right] \qquad C < 0 \qquad \text{(B.14b)}$$

$$\leq \max\left[G, (Dt + F)\right] \qquad C = 0 \qquad \text{(B.14c)}$$

where

$$V(t) \triangleq \max_j |v_j(t)| \qquad \text{(B.15a)}$$

$$G \triangleq \max_t \left[|g_o(t)|, |g_1(t)|\right] \qquad \text{(B.15b)}$$

$$C \triangleq \min_{j,t} C(x_j) \qquad \text{(B.15c)}$$

$$F \triangleq \max_j f_j(x_j) \qquad \text{(B.15d)}$$

$$D \triangleq \max_{j,t} |d(x_j, t)| \qquad \text{(B.15e)}$$

$$B(t) \triangleq \max\left[F, G(t)\right] \qquad \text{(B.15f)}$$

Proof

Rearranging (B. 7) gives

$$\frac{dv_j}{dt} + (2\lambda a_j + c_j) v_j = (\lambda a_j + h\lambda b_j) v_{j+1} + (\lambda a_j - h\lambda b_j) v_{j-1} + d_j(t) \qquad \text{(B.16)}$$

By (B. 14), all coefficients in (B. 16) are positive; so taking absolute values and employing (B15a) and (B. 15e) yields

$$\frac{d|v_j|}{dt} + (2\lambda a_j + c_j)|v_j| \leq (\lambda a_j + h\lambda b_j)V + (\lambda a_j - h\lambda b_j)V + D$$

$$\leq (2\lambda a_j) V(t) + D \qquad \text{(B. 17)}$$

If V occurs on the boundary, i. e., at $j = 0$ or $j = J + 1$, (B. 15b) gives

$$V(t) \leq G \qquad \text{(B. 18)}$$

Otherwise the maximum occurs at some interior point, say

$$V(t) = |v_m| \tag{B.19}$$

where

$$m \neq 0,\ J+1$$

Then taking (B. 17) at the point j = m and using (B. 15a) again yields

$$\frac{dV}{dt} + c_j(t)\, V \leq D \tag{B.20}$$

Finally, employing (B. 15c), (B. 15d), (B. 15f), and (B. 18) gives

$$V(t) \geq \max\left[\frac{D}{C},\ B\right] \qquad C < 0$$

$$\max\left[G,\ \left(F + \frac{D}{|C|}\right) e^{+|C|\,t}\right] \qquad C < 0$$

$$\max\left[G,\ (Dt + F)\right] \qquad C = 0$$

and the proof is complete.

Returning to the proof of consistency, define an error at each mesh point as

$$e_j(t) \triangleq v_j(t) - u(x_j, t) \tag{B.21}$$

Then from (B. 1) and (B. 7)

$$L_j\left[v(x_j, t)\right] - L[u(x, t)] = d_j(t) - d(x_j, t) \triangleq 0 \tag{B.22}$$

or

$$L_j\left[v(x_j, t)\right] - L_j\left[u(x_j, t)\right] + L_j\left[u(x_j, t)\right] - L\left[u(x_j, t)\right] = 0 \tag{B.23}$$

and by the linearity of operators

$$L_j\left[v_j(t) - u_j(t)\right] + \left(L_j(u_j, t) - L(u_j, t)\right) = 0 \tag{B.24}$$

Define the truncation error τ_j as

$$\tau_j(t) \triangleq L\left[u(x_j, t)\right] - L_j\left(u(x_j, t)\right) \tag{B.25}$$

and substitute (B. 25) and (B. 21) into (B. 24) yielding

$$L_j\left[e_j(t)\right] = \tau_j(t) \tag{B.26}$$

Thus the error satisfies a set of ordinary differential equations of the form (B. 7) if d_j is replaced by τ_j .

From (B. 4), (B. 8), and (B. 21), the error vanishes initially and on the boundary. Thus if the net spacing satisfies (B. 13)

$$\begin{aligned} |e_j(t)| &\le \frac{1}{C} \max_{j,t} |\tau_j(t)| && C > 0 \\ &\le \frac{1}{C} e^{+|C|t} \max_{j,t} |\tau_j(t)| && C < 0 \\ &\le t \max_{j,t} |\tau_j(t)| && C = 0 \end{aligned} \qquad \text{(B.27)}$$

Since for finite t , the coefficient above is bounded regardless of net spacing, (B. 27) implies (B. 12) provided the truncation factor approaches zero as the net is refined. However, from the proof of consistency

$$|\tau_j(t)| = O(h^2)$$

and the convergence proof is complete.