THE UNIVERSITY OF CALGARY

OPTIMAL CONTROL OF DISTRIBUTED PARAMETER SYSTEMS WITH DISCRETE CONSTRAINED INPUTS

by

Mohamed Emad Mousa Rasmy

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF SCIENCE

DEPARTMENT OF MATHEMATICS, STATISTICS AND COMPUTING SCIENCE

CALGARY, ALBERTA

SEPTEMBER 1971

THE UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "Optimal Control of Distributed Parameter Systems with Discrete Constrained Inputs", submitted by Mohamed Emad Mousa Rasmy in partial fulfillment of the requirements for the degree of Master of Science.

M. H. Hamza

Supervisor - Dr. M. H. Hamza
(Mathematics Department)

P. Lancaster.

Dr. P. Lancaster (Mathematics Department)

P A Binding

Dr. P. Binding (Mathematics Department)

Wah Chun Chan

Dr. W. C. Chan (Electrical Engineering)

Date Oct, 1, 1971

ABSTRACT

The problem of minimizing a quadratic performance index subject to an energy-type constraint on the control vector is considered in this thesis. The control vector is restricted to be discrete in time. The problem is formulated as a minimization problem in a Hilbert space and the existence and uniqueness of the optimal control vector is shown.

A necessary and sufficient condition for optimality is derived which is used to yield an equation whose solution gives the optimal control vector. It is shown that if the optimal control vector lies on the boundary of the constrained region, then the Lagrange multiplier must be determined in order to solve the corresponding optimality equation. An algorithm based on Newton's method is presented for the calculation of the Lagrange multiplier, and the convergence of the algorithm is proved.

The convergence of the objective function, corresponding to the optimal discrete control vector, to the same objective function when the control vector is not restricted to be discrete in time, is shown. Furthermore, a condition is derived which, when satisfied, assures that the optimal discrete control vector will converge to the nondiscrete optimal control vector as the number of the sampling periods tends to infinity.

Two examples are presented to show the application of the theoretical results, and numerical solutions are given.

ACKNOWLEDGEMENTS

The author wishes to express his appreciation to Dr. M.H. Hamza for his guidance during the course of this work and for the advice given during preparation of this manuscript.

Thanks are due to the National Research Council of Canada and to the Department of Mathematics, University of Calgary, for their financial assistance during the performance of this research.

The author also wishes to thank Mrs. A. Nordquist for typing this manuscript.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONT'D)

LIST OF TABLES

LIST OF SYMBOLS

$\underline{q}(\underline{x},t)$ = n-dimensional vector representing the state of the system

$\underline{q}_o(x)$ = n-dimensional initial state vector

$\underline{K}(\underline{x},t,\tau),\underline{H}(\underline{x},t)$ = matrix linear operator (nxr) and (nxn) respectively

T = final time and is fixed

$\underline{u}(t)$ = r-dimensional control vector

$I(\underline{u})$ = performance index

Ω = a simply connected open region in an m-dimensional Euclidean space, the spatial domain

$\partial\Omega$ = the boundary of the domain Ω

E = number representing the actual constraint on the input functions

N = number of sampling periods

T_N = the sampling period

$\underline{B}(\underline{x})$ = (nxs) spatial matrix

$\underline{K}_\ell(\underline{x})$ = (nxr) spatial matrix operator

$\underline{u}$ = s-dimensional vector representing the control vector in discrete form

$\underline{q}_d(\underline{x})$ = n-dimensional vector representing the desired state of the system

$\underline{B}^*$ = adjoint of the mapping $\underline{B}$

inf = greatest lower bound

λ_i = eigenvalues as defined for the matrix operator $\underline{B}^*\underline{B}$

$\underline{x}_i$ = eigenfunctions of the matrix operator $\underline{B}^*\underline{B}$

$\rho(A)$ = resolvent set of A

λ = Lagrange multiplier

LIST OF SYMBOLS (CONT'D)

$\underline{I}$ = identity matrix

$\oplus$ = direct sum decomposition

$R(\underline{B}^*\underline{B})$ = range of the transformation $\underline{B}^*\underline{B}$

$N(\underline{B}^*\underline{B})$ = null space of the mapping $\underline{B}^*\underline{B}$

$N^{\perp}(\underline{B}^*\underline{B})$ = orthogonal complement of $N(\underline{B}^*\underline{B})$

$L(X,X)$ = bounded, linear transformation mapping normed space X into itself

$A \geqslant 0$ = the operator A is positive semidefinite

$A > 0$ = the operator A is positive definite

$\nabla g(\underline{x})$ = the gradient of a functional $g(\underline{x})$

$-$ = symbol identifying vectors or matrices

TO

SOUHAIR, MY WIFE

1. INTRODUCTION

1.1 GENERAL

In physical situations, one often encounters systems whose parameters are distributed in both space and time. The dynamic behavior of these systems is governed by partial differential equations, integral equations, integrodifferential equations and sometimes by more general functional equations. The name distributed parameter is used for these systems so as to differentiate them from others whose behavior can be described by ordinary differential equations. Usually, the name lumped parameter is used for these latter systems.

In general, the problem of optimal control arises from attempting to minimize (maximize) a certain functional of the state and of the controlling action. Constraints usually exist due to practical limitations and this leads to restrictions on the state as well as on the controlling functions.

In attempting to formulate these problems one has to make the formulation broad enough so as to include many physical systems. On the other hand, one has to narrow the investigations, since a general formulation leads to results which are usually difficult to apply.

1.2 THESIS OBJECTIVES AND OUTLINE

The main purpose of this thesis is to investigate the problem of minimizing a quadratic performance index under an energy type constraint on the control vector. A problem similar to that of Weigand[1] is treated subject, however, to the additional constraint that the control vector is discrete in time. This restriction is introduced due to the trend of using on-line digital computers to control industrial processes.

We also study the convergence of the optimal discrete problem to the nondiscrete optimal one as the number of the sampling periods tends to infinity. Apart from its theoretical benefit which establishes a link between the discrete and the nondiscrete problem, this study provides us with a tool by which we can approximate a distributed parameter system with measurable inputs by a corresponding discrete one whose solution is much easier to obtain.

The main outline of the thesis is:

Chapter 2 contains a review of the work done in the field of optimal control of distributed parameter systems which is significant to the work reported in this thesis.

In Chapter 3 the performance index is introduced and the problem is formulated as a minimization problem in Hilbert space. The necessary and sufficient condition for optimality is derived and the existence and uniqueness of the optimal control vector are shown.

Chapter 4 is concerned with solving the optimality equation using Newton's method. The convergence properties of the optimal discrete problem are also investigated.

In Chapter 5 two examples are given to demonstrate the application of the theory. Computer results are also presented.

Chapter 6 draws conclusions concerning the results obtained throughout the thesis and gives suggestions for further research.

2. REVIEW

2.1 OPTIMAL CONTROL OF DISTRIBUTED PARAMETER SYSTEMS

Butkovskii and Lerner[2] were among the first to consider the optimal control of distributed parameter systems. They posed a general problem for first and second order partial differential equations which have both functional and amplitude type constraints on the controlling functions and a performance index in the form of a general functional.

In a later paper Butkovskii[3] developed a maximum principle for a distributed parameter system whose motion is described by a nonlinear integral relationship. He used an objective function in the form of a general functional. Using Pontryagin's maximum principle, he obtained an implicit form for the optimal control function which involves the solution of a nonlinear integral equation. The solution of such an equation is not easy in general.

In subsequent papers,[4,5,6] Butkovskii developed a generalized maximum principle for distributed parameter systems which can be put in the form of a set of integral equations. Again the solution is in an implicit form which is difficult to use.

Butkovskii[7] considered a linear system with distributed parameters of the form

$$Q(x,t) = \int_0^t K(x,t-\tau)u(\tau)d\tau, \qquad x \in [0,s] \text{ and } t \in [0,T]$$

where Q represents the system state and u represents the control function. He posed for this system the following problem:

Find a control function u(t), $|u(t)| \leqslant L$ for which the relation

$$Q^*(x) = \int_0^T K(x,t-\tau)u(\tau)d\tau$$

is satisfied, where $Q^*(x)$ is given and the final time T is to be minimized. He solved this problem by using results concerning the L-problem of moments due to Krein.[8] In subsequent papers,[9,10,11] Butkovskii applied his results[6] to solve the optimal control problems of a wave equation and also of a vibrating string.

Egorov[12] examined certain questions in the theory of controlled thermal processes which are connected with the choice of a control that is in some sense optimal. He treated the minimal time problem and a problem with a quadratic cost with an amplitude constraint on the controlling function in both cases. Moreover, he studied the existence and uniqueness of the optimal control. Later on, Egorov[13] treated the problem of optimal control of systems described by a second order parabolic equation and he deduced a maximum principle as a necessary condition for optimality. Butkovskii and Egorov[14] gave a survey paper containing most of the Soviet work in the field of optimal control of distributed parameter systems.

A general discussion on the properties as well as the optimal control problem of distributed parameter systems was presented by Wang and Tung.[15] They discussed (a) the mathematical description of distributed parameter systems, (b) the controllability and observability of these systems, (c) the formulation of optimum control problems and the derivation of a maximum principle for a particular class of systems, and (d) the problem associated with approximating distributed parameter systems by discretization. A more detailed and complete discussion of distributed parameter systems was presented by Wang.[16]

Wang[17] considered the optimal control problem associated with a diffusion system with a free boundary. This arises physically from attempting to control the rate of solidification of a liquid. He formulated the control problem as an infinite dimensional mathematical programming problem (linear), and then approximated it by a corresponding finite dimensional one. He also established the convergence of the approximation.

Sakawa[18] treated the problem of optimal control of a distributed parameter system governed by a heat conduction equation. His objective was to minimize the deviation of the temperature distribution from an assigned distribution at a given time subject to an amplitude constraint on the controlling function. Using calculus of variations he obtained a Fredholm integral equation of the first kind as a necessary condition for the optimal control.

Sakawa[19] treated the more general problem of the optimal control of a one-dimensional linear stationary distributed parameter system controlled by boundary functions which act at both ends of the one-dimensional space. He minimized an objective function of the form

$$I(u) = \int_0^1 [q^*(x) - q(x,T)]^2 dx + \sum_{i=1}^{2} c_i \int_0^T u_i^2(t)dt$$

where $q(x,t)$ is given by

$$q(x,t) = q_o(x,t) + \sum_{i=1}^{2} \int_0^t g_i(x,t-\tau)u_i(\tau)d\tau$$

where

$q^*(x)$ = the desired final temperature distribution

$q(x,T)$ = the actual final temperature distribution

T = the final time of the process

$u_1(t)$ and $u_2(t)$ = the boundary control functions

c_1 and c_2 = positive constants which act as weighting factors.

Using functional analysis he reduced the problem to the minimization of a quadratic functional in a Hilbert space. By using the variational method for the unconstrained case, he obtained a Fredholm integral equation of the second kind as a necessary and sufficient condition. In the case of constraints of the form,

$$a_i \leqslant u_i(t) \leqslant b_i \qquad i = 1,2$$

and by using the Kuhn-Tucker theorem for nonlinear programming, he obtained a system of nonlinear integral equations of a form similar to the integral equation of the Hammerstein type. Under suitable assumptions, he solved this equation using successive approximation.

Axelband[20] presented a solution for the unconstrained optimization problem of distributed systems wherein the control action and control are related by a bounded linear operator. He used function space techniques to prove the existence and uniqueness and to derive necessary and sufficient conditions for the optimal control for a quadratic performance index. Axelband[21] developed an approximate technique for the optimal control of linear distributed parameter systems with an amplitude constraint on the control function. He considered a performance index of the integral squared error type and derived an algorithm for the computation of the optimal control function by a nonlinear programming procedure.

Kim and Erzberger[22] used dynamic programming to obtain the optimum feedback boundary control function for a distributed parameter system which is described by the n-dimensional wave equation. They considered a quadratic performance index for an unconstrained control function. The functional equation for the optimum controller, analogous to the matrix Riccati equation obtained by Kalman for lumped parameter systems, was shown to be a nonlinear partial integrodifferential equation. They showed that, for a certain type of weighting factor in the quadratic error index, the nonlinear functional equations can be solved by using the method of separation of variables.

Weigand[1] in his paper considered the problem of obtaining the optimal control functions, subject to an energy-type constraint, which minimize a performance index of a quadratic type for the control of linear distributed parameter systems. He formulated the problem as a minimization problem in Hilbert space and derived the necessary and sufficient condition for optimality using both functional analysis and variational methods. He obtained the optimal control function by solving the Fredholm integral equation with symmetric kernel and gave an explicit form for the optimal control function in terms of eigenfunction expansions.

Vidyasagar[23] solved the same problem of Weigand for one-dimensional distributed parameter systems. He used the Kuhn-Tucker theorem of nonlinear programming in deriving a necessary and sufficient condition for optimality. In fact, he obtained almost the same results as Weigand, but using a different approach.

Goldwyn[24] et al showed the applicability of the Laplace transformation for the determination of the time optimal control of a linear diffusion process with amplitude constraint on the control. They used a method which can be interpreted as requiring a control whose transform, in combination with the initial condition, places zeros at the poles of the open loop transfer function to derive the optimal control function on the assumption that it is bang bang.

For distributed parameter systems described by parabolic equation, Gal'chuk[25] studied the possibility of translating the system to a stationary regime subject to an amplitude constraint on the control function. He showed that this problem is equivalent to a certain problem of moment and gave conditions for the attainability of stationary states.

Balakrishnan[26] treated the problem of minimizing the distance

$$||Lu - x||$$

where L is a compact linear operator mapping the Hilbert space $L_2^r[o,T]$ into the Euclidean space R^n, x is a given element in R^n, and u is the control vector and is restricted to belong to a closed convex subset of $L_2^r[o,T]$. Without using the finite dimensionability of R^n, he showed the existence and uniqueness of the optimal control. He presented an algorithm based on the steepest descent method to compute the optimal control vector. However, his algorithm is not practical from a computational point of view.

An almost exhaustive and commented bibliography prior to the end of 1969 was given by Robinson.[27] Also, an excellent survey on the optimal control of distributed parameter systems was presented by Lions.[28]

All the aforementioned results for distributed parameter systems gave solutions in terms of control functions which are not discrete. As far as the author knows, only the following papers treated the problem of optimum distributed parameter systems whose control function is discrete in time.

Lorchirachoonkul and Pierre[29] considered the problem of minimizing, at certain discrete points on the spatial domain, the deviation between a desired response and the actual system response of a linear distributed parameter system subject to constraints on both control and state function. Using a discrete control function, they reduced the problem to a linear programming problem whose solution can easily be obtained.

Matsumoto and Kito[30] studied the problem of designing an optimal feedback controller based on a quadratic performance index for a distributed system described by a partial differential equation of the parabolic type with spatially concentrated controls. They assumed the presence of an on-line digital computer and they considered the control function to be discrete with respect to time. Using dynamic programming, they obtained the optimal control as a function of the system state.

Hassan and Solberg[31] treated the unconstrained problem of optimal control of a distributed parameter system with a quadratic cost functional. They restricted their control function to be discrete in time and used the technique of dynamic programming to derive an expression for feedback control in terms of an auxiliary spatial dependent variable. They showed that this variable satisfied a Riccati type functional equation with an unknown final value. Using an orthogonal series expansion, they transformed this equation to a recursive algebraic equation in the coefficients of the expansion. They demonstrated the applicability of the method by an example of an automatic regulator for the flux pattern in a slab nuclear reactor.

3. OPTIMAL CONTROL OF LINEAR DISTRIBUTED PARAMETER SYSTEMS WITH A QUADRATIC CONSTRAINT

3.1 INTRODUCTION

In this chapter we present the optimal control problem which consists of minimizing a quadratic performance index under an energy-type constraint for linear distributed parameter systems. We restrict the control function to be discrete in time. This problem is formulated as a minimization problem in a finite dimensional Euclidean space. Using a necessary and sufficient condition from functional analysis, we arrive at an equation whose solution gives the optimal control function.

3.2 FORMULATION OF THE PROBLEM

Consider a linear distributed parameter system. The system states are assumed to be described by state functions which can be expressed as

$$\underline{q}(\underline{x},t) = \int_0^t \underline{K}(\underline{x},t,\tau)\underline{u}(\tau)d\tau + \underline{H}(\underline{x},t)\underline{q}_o(\underline{x}) \qquad (3.2.1)$$

where $\underline{q}(\underline{x},t)$ is an n-dimensional vector representing the state of the system, $\underline{x}$ is an m-dimensional spatial coordinate, $\underline{x}\varepsilon\Omega$, where Ω is a simply connected open subset of an m-dimensional Euclidean space, $\partial\Omega$ denotes its boundary and t is time $(0 < t \leqslant T)$. $\underline{u}(t)$ is an r-dimensional control vector which could be either a boundary control vector or a spatially concentrated control vector. Moreover, it could be a mixture of both types. $\underline{q}_o(\underline{x})$ is an n-dimensional vector representing the initial state of the system, $\underline{H}(\underline{x},t)$ and $\underline{K}(\underline{x},t,\tau)$ are (nxn) and (nxr) matrix linear operators respectively whose elements are known functions which are determined corresponding to given partial differential equations and initial and boundary conditions.

The following control problem is posed. Find the control vector $\underline{u}(t)$ of minimal norm which minimizes the objective function

$$I(\underline{u}) = \sum_{i=1}^{n} \int_{\Omega} [q_i(\underline{x},T) - q_{d_i}(\underline{x})]^2 d\underline{x} \tag{3.2.2}$$

under the constraint

$$\sum_{i=1}^{r} \int_{0}^{T} u_i^2(t)dt \leqslant E \tag{3.2.3}$$

where T is the final time and $\underline{q}_d(\underline{x})$ is an n-dimensional vector representing a prescribed spatial distribution function of the states.

Let $L_2^n(\Omega)$ denote the real Hilbert space of n-dimensional functions square integrable over Ω and $L_2^r[o,T]$ represents a real Hilbert space of r-dimensional functions square integrable over (o,T). Define the inner product of two vectors $\underline{p}$ and $\underline{q}$ in $L_2^n(\Omega)$ by

$$(\underline{p},\underline{q}) = \int_{\Omega} \underline{p}'(\underline{x})\underline{q}(\underline{x})d\underline{x} \; . \tag{3.2.4}$$

Similarly, denote the inner product of two vectors $\underline{u}$ and $\underline{v}$ in $L_2^r[o,T]$ by

$$(\underline{u},\underline{v}) = \int_{0}^{T} \underline{u}'(t)\underline{v}(t)dt \tag{3.2.5}$$

where ' denotes the transpose of a vector or a matrix.

Define a transformation A from $L_2^r[o,T]$ into $L_2^n(\Omega)$ by

$$A\underline{u} = \int_{0}^{T} \underline{K}(\underline{x},T,\tau)\underline{u}(\tau)d\tau. \tag{3.2.6}$$

It is clear that A is a linear operator. Let us define an (rxr) square matrix $\underline{G}(\underline{x},t,\tau)$ by

$$\underline{G}(\underline{x},t,\tau) = \underline{K}'(\underline{x},T,t)\ \underline{K}(\underline{x},T,\tau). \tag{3.2.7}$$

Furthermore, define another (rxr) square matrix $\Phi(t,\tau)$ as

$$\underline{\Phi}(t,\tau) = \int_{\Omega} \underline{G}(\underline{x},t,\tau)d\underline{x} . \tag{3.2.8}$$

Assuming that

$$\sum_{i,j=1}^{r} \int_{o}^{T}\int_{o}^{T} \Phi_{ij}(t,\tau)^2 dt d\tau < \infty, \tag{3.2.9}$$

It can be shown that A is a completely continuous operator (see appendix). Let us also assume that $\underline{H}$ is a linear operator with range in $L_2^n(\Omega)$.

We will assume that the control vector $\underline{u}(t)$ is a discrete function of time. Assuming that the number of sampling periods is N, the sampling period T_N is given by

$$T_N = T/N \tag{3.2.10}$$

and, hence, $\underline{u}(t)$ will be defined as

$$\underline{u}(t) = \underline{u}_{\ell} \quad , \quad \ell T_N \leq t < (\ell+1)T_N \quad , \quad \ell = 0,1,\ldots,N-1 \tag{3.2.11}$$

Substituting (3.2.11) into (3.2.1) yields

$$\begin{aligned} \underline{q}(\underline{x},T) &= \int_{o}^{T} \underline{K}(\underline{x},T,\tau)\underline{u}(\tau)d\tau + \underline{H}(\underline{x},T)\underline{q}_o(\underline{x}) \\ &= \sum_{\ell=o}^{N-1} \int_{\ell T_N}^{(\ell+1)T_N} \underline{K}(\underline{x},T,\tau)\underline{u}_{\ell}d\tau + \underline{H}(\underline{x},T)\underline{q}_o(\underline{x}) \\ &= \sum_{\ell=o}^{N-1} \underline{K}_{\ell}(\underline{x})\underline{u}_{\ell} + \underline{H}(\underline{x},T)\underline{q}_o(\underline{x}) \end{aligned} \tag{3.2.12}$$

where $\underline{K}_{\ell}(\underline{x})$, $\ell = 0,1,\ldots,N-1$ is an (nxr) spatial matrix given by

$$\underline{K}_\ell(\underline{x}) = \int_{\ell T_N}^{(\ell+1)T_N} \underline{K}(\underline{x},T,\tau)d\tau \qquad \ell = 0,1,\ldots,N-1 \qquad (3.2.13)$$

The energy constraint (3.2.3) will reduce to

$$\sum_{i=1}^{r} \sum_{\ell=o}^{N-1} u_{i_\ell}^2 \leqslant NE/T \qquad (3.2.14)$$

where u_{i_ℓ} denotes the ith component of the vector $\underline{u}(t)$ during the interval $[\ell T_N,(\ell+1)T_N]$. Let us define an (nxs) spatial matrix $\underline{B}$, where $s = rN$, as

$$\underline{B}(\underline{x}) = [\underline{K}_o \quad \underline{K}_1 \quad \ldots\ldots \quad \underline{K}_{N-1}] \qquad (3.2.15)$$

Also define an s-dimensional column vector $\underline{u}$ by

$$\underline{u} = \begin{bmatrix} \underline{u}_o \\ \underline{u}_1 \\ \cdot \\ \cdot \\ \cdot \\ \underline{u}_{N-1} \end{bmatrix} \qquad (3.2.16)$$

where $\underline{u}_\ell$ is defined in (3.2.11), so the $(\ell r+i)$th component of $\underline{u}$ equals u_{i_ℓ}. Taking (3.2.16) into consideration, (3.2.12) and (3.2.14) reduce to

$$\underline{q}(\underline{x},T) = \underline{B}(\underline{x})\underline{u} + \underline{H}(\underline{x},T)\underline{q}_o(\underline{x}) \qquad (3.2.17)$$

$$\sum_{i=1}^{s} u_i^2 \leqslant NE/T. \qquad (3.2.18)$$

Let us define the norm of an element $\underline{h}$ in the s-dimensional Euclidean space by

$$||\underline{h}|| = \left(\sum_{i=1}^{s} h_i^2\right)^{\frac{1}{2}}. \tag{3.2.19}$$

Then $\underline{B}$ defines a compact linear operator from R^s into $L_2^n(\Omega)$ whose range is finite dimensional. Define $\underline{q}^*(\underline{x})$ as

$$\underline{q}^*(\underline{x}) = \underline{q}_d(\underline{x}) - \underline{H}(\underline{x},T)\underline{q}_o(\underline{x}). \tag{3.2.20}$$

It is clear that $\underline{q}^*(\underline{x}) \in L_2^n(\Omega)$, and the problem reduces to finding a control vector $\underline{u} \in R^s$ with minimum norm which minimizes

$$I(\underline{u}) = ||\underline{B}(\underline{x})\underline{u} - \underline{q}^*(\underline{x})||^2 \tag{3.2.21}$$

subject to the constraint

$$||\underline{u}||^2 \leqslant NE/T. \tag{3.2.22}$$

3.3 EXISTENCE AND UNIQUENESS OF THE OPTIMAL CONTROL VECTOR

Theorem 1. A closed convex subset of a Hilbert space contains a unique element of minimal norm.

Proof. For proof see reference 40, p 243.

Let C denote the closed sphere of radius NE/T in R^s; then we can state the following theorem:

Theorem 2. There exists a unique element $\underline{u}^*$ of minimal norm in C such that

$$\inf_{\underline{u}\in C} ||\underline{B}\,\underline{u} - \underline{q}^*|| = ||\underline{B}\,\underline{u}^* - \underline{q}^*||.$$

Proof. Let $\{\underline{u}_n\}$ be a sequence of elements in C such that

$$\lim_{n\to\infty} ||\underline{B}\,\underline{u}_n - \underline{q}^*|| = \inf_{\underline{u}\varepsilon C} ||\underline{B}\,\underline{u} - \underline{q}^*||.$$

Since C is a bounded closed subset of R^s, i.e., compact, this implies that there exists a convergent subsequence $\{\underline{u}_{n_k}\}$ whose limit $\underline{v} \varepsilon C$. Therefore, we have

$$||\underline{B}\,\underline{v} - \underline{q}^*|| = \inf_{\underline{u}\varepsilon C} ||\underline{B}\,\underline{u} - \underline{q}^*||.$$

Now, consider the set $D = \{\underline{u}\varepsilon C | \underline{B}\,\underline{u} = \underline{B}\,\underline{v}\}$. This is a closed convex set and therefore, by theorem 1, it must contain a unique element $\underline{u}^*$ with minimum norm. This completes the proof.

Let $\underline{B}^*$ be the adjoint operator of $\underline{B}$, and let $\underline{v}$ be any element in R^s and $\underline{q}$ be any element in $L_2^n(\Omega)$. Then $\underline{B}^*$ is defined as

$$(\underline{B}\,\underline{v},\underline{q}) = (\underline{v},\underline{B}^*\underline{q})$$

$$(\underline{v},\underline{B}^*\underline{q}) = \int_\Omega \underline{v}'\underline{B}'(\underline{x})\underline{q}(\underline{x})d\underline{x}$$

$$= \underline{v}' \int_\Omega \underline{B}'(\underline{x})\underline{q}(\underline{x})d\underline{x}$$

$$\Rightarrow \qquad \underline{B}^*\underline{q} = \int_\Omega \underline{B}'(\underline{x})\underline{q}(\underline{x})d\underline{x} \tag{3.3.1}$$

Since $\underline{B}$ is bounded, it follows that $\underline{B}^*\underline{B}$ is a bounded linear transformation defined on R^s, mapping R^s into R^s and, similarly, $\underline{B}\,\underline{B}^*$ is a bounded linear transformation mapping $L_2^n(\Omega)$ into itself. It is worth noting that both $\underline{B}^*\underline{B}$ and $\underline{B}\,\underline{B}^*$ are compact operators. This follows directly from the fact that $\underline{B}$ is compact and $\underline{B}^*$ is continuous (ref. 32, p. 290). Furthermore, $\underline{B}^*\underline{B}$ is an (sxs) positive semidefinite Hermitian matrix which is given by

$$\underline{B}^*\underline{B} = \int_\Omega \underline{B}'(\underline{x})\underline{B}(\underline{x})d\underline{x}. \qquad (3.3.2)$$

Since the matrix $\underline{B}^*\underline{B}$ is positive semidefinite Hermitian, therefore the eigenvalues of $\underline{B}^*\underline{B}$ are real and non-negative. Let $\{\lambda_i\}$, $i = 1,2,\ldots\ldots,p \leqslant s$; $\lambda_1 \geqslant \lambda_2 \geqslant \ldots\ldots \geqslant \lambda_p$, be the nonzero eigenvalues of $\underline{B}^*\underline{B}$ and let the $\underline{z}_i$ be the corresponding orthonormalized eigenvectors. Hence, if $\underline{z} \in R^s$, then $\underline{B}^*\underline{B}\,\underline{z}$ can be written as

$$\underline{B}^*\underline{B}\,\underline{z} = \sum_{i=1}^{p} \lambda_i(\underline{z},\underline{z}_i)\underline{z}_i. \qquad (3.3.3)$$

3.4 NECESSARY AND SUFFICIENT CONDITIONS FOR OPTIMAL CONTROL[33]

Theorem 1. Let $\underline{z}$ be a vector in a Hilbert space H and let W be a convex subset of H. If there exists a $\underline{y}_o \in W$ such that

$$||\underline{z} - \underline{y}_o|| \leqslant ||\underline{z} - \underline{y}|| \qquad \forall\, \underline{y} \in W \qquad (3.4.1)$$

A necessary and sufficient condition for $\underline{y}_o$ to satisfy (3.4.1) is that

$$(\underline{z} - \underline{y}_o, \underline{y} - \underline{y}_o) \leqslant 0 \qquad \forall\, \underline{y} \in W \qquad (3.4.2)$$

Proof.

Necessity: Suppose $\underline{y}_o$ is a minimizing vector, i.e., it satisfies condition (3.4.1) and we want to show that (3.4.2) holds $\forall y \in W$. Suppose that (3.4.2) is false, i.e., there exists a vector $\underline{y}_1 \in W$ such that

$$(\underline{z} - \underline{y}_o, \underline{y}_1 - \underline{y}_o) = \varepsilon > 0.$$

Consider the vectors $\underline{y}_\alpha = (1-\alpha)\underline{y}_o + \alpha\underline{y}_1$, $0 \leqslant \alpha \leqslant 1$. Since W is a convex set, it follows that each $\underline{y}_\alpha \in W$. Also,

$$\|\underline{z} - \underline{y}_\alpha\|^2 = \|(1-\alpha)(\underline{z}-\underline{y}_o) + \alpha(\underline{z}-\underline{y}_1)\|^2$$

$$= (1-\alpha)^2\|\underline{z}-\underline{y}_o\|^2 + 2\alpha(1-\alpha)(\underline{z}-\underline{y}_o,\underline{z}-\underline{y}_1) + \alpha^2\|\underline{z}-\underline{y}_1\|^2$$

$$= (1-\alpha)^2\|\underline{z}-\underline{y}_o\|^2 + 2\alpha(1-\alpha)\{(\underline{z}-\underline{y}_o,\underline{z}-\underline{y}_o) + (\underline{z}-\underline{y}_o,\underline{y}_o-\underline{y}_1)\}$$

$$+ \alpha^2\|\underline{z}-\underline{y}_1\|^2$$

$$= \|\underline{z}-\underline{y}_o\|^2 + \alpha^2\{\|\underline{z}-\underline{y}_1\|^2 - \|\underline{z}-\underline{y}_o\|^2\} - 2\alpha(1-\alpha)(\underline{z}-\underline{y}_o,\underline{y}_1-\underline{y}_o).$$

Choose α sufficiently small such that

$$\alpha^2\{\|\underline{z}-\underline{y}_1\|^2 - \|\underline{z}-\underline{y}_o\|^2\} - 2\alpha(1-\alpha)(\underline{z}-\underline{y}_o,\underline{y}_1-\underline{y}_o) < 0$$

$$\Rightarrow \qquad \|\underline{z}-\underline{y}_\alpha\| < \|\underline{z}-\underline{y}_o\|.$$

This contradicts the minimizing property of $\underline{y}_o$. Hence, no such $\underline{y}_1$ can exist.

Sufficiency: Suppose that $\underline{y}_o \varepsilon W$ and $\underline{y}_o$ satisfies (3.4.2). Hence, for any $\underline{y} \varepsilon W$ such that $\underline{y} \neq \underline{y}_o$, we have

$$\|\underline{z}-\underline{y}\|^2 = \|\underline{z}-\underline{y}_o+\underline{y}_o-\underline{y}\|^2$$

$$= \|\underline{z}-\underline{y}_o\|^2 + 2(\underline{z}-\underline{y}_o,\underline{y}_o-\underline{y}) + \|\underline{y}_o-\underline{y}\|^2 > \|\underline{z}-\underline{y}_o\|^2$$

which implies that $\underline{y}_o$ is a minimizing vector Q.E.D.

The set $\underline{B}(C)$ consisting of all elements of the form $\underline{B}\,\underline{u}$, $\underline{u}\varepsilon C$, is convex. Therefore, we can apply the necessary and sufficient condition of the previous theorem to our problem, i.e., for $\underline{u}^*$ to be an optimal control vector, it is necessary and sufficient that $\underline{u}^*$ satisfies

$$(\underline{q}^* - \underline{B}\,\underline{u}^*, \underline{B}(\underline{u}-\underline{u}^*)) \leq 0 \qquad \underline{u}\varepsilon C \qquad (3.4.3)$$

We have two cases:

a) $\underline{u}^*$ belongs to the interior of the set C, i.e.

$$\sum_{i=1}^{s} u_i^{*2} < ET/N. \tag{3.4.4}$$

Let $\{\underline{e}_i\}$, i = 1,....,s, be the set of orthogonal unit vectors in R^s. Choose ε sufficiently small such that the set of vectors $\{\underline{v}_i\}$, i = 1,....,s, where $\underline{v}_i = \underline{u}^* + \varepsilon\underline{e}_i$, belong to C. Putting $\underline{u} = \underline{v}_i$ in (3.4.3), we get

$$(\underline{q}^* - \underline{B}\,\underline{u}^*, \varepsilon\,\underline{B}\,\underline{e}_i) \leqslant 0 \qquad i = 1,2,\ldots,s$$

$$\Rightarrow \qquad \varepsilon(\underline{q}^* - \underline{B}\,\underline{u}^*, \underline{B}\,\underline{e}_i) \leqslant 0 \qquad i = 1,2,\ldots,s$$

$$\Rightarrow \qquad \varepsilon(\underline{B}^*\underline{q}^* - \underline{B}^*\underline{B}\,\underline{u}^*, \underline{e}_i) \leqslant 0 \qquad i = 1,2,\ldots,s \tag{3.4.5}$$

Since ε can take positive as well as negative values, it follows from (3.4.5) that

$$(\underline{B}^*\underline{q}^* - \underline{B}^*\underline{B}\,\underline{u}^*, \underline{e}_i) = 0 \qquad i = 1,2,\ldots,s \tag{3.4.6}$$

The set $\{\underline{e}_i\}_{i=1}^{s}$ is complete in R^s and hence it follows from (3.4.6) that

$$\underline{B}^*\underline{q}^* - \underline{B}^*\underline{B}\,\underline{u}^* = \underline{0}$$

or

$$\underline{B}^*\underline{B}\,\underline{u}^* = \underline{B}^*\underline{q}^*. \tag{3.4.7}$$

b) $\underline{u}^*$ belongs to the boundary of the set C, i.e.,

$$\sum_{i=1}^{s} u_i^{*2} = EN/T. \tag{3.4.8}$$

In this case $\underline{u}^*$ is the solution of a finite dimensional optimization problem under equality constraint. Hence, we can use the

Lagrange multiplier method to find the optimal control vector $\underline{u}^*$. Our problem, in this case, is to minimize

$$I(\underline{u}) = (\underline{B}\,\underline{u} - \underline{q}^*, \underline{B}\,\underline{u} - \underline{q}^*)$$

$$= (\underline{u},\underline{B}^*\underline{B}\,\underline{u}) - 2(\underline{u},\underline{B}^*\underline{q}^*) + (\underline{q}^*,\underline{q}^*) \qquad (3.4.9)$$

under the equality constraint

$$(\underline{u},\underline{u}) = EN/T. \qquad (3.4.10)$$

Let λ be the Lagrange multiplier, where $\lambda > 0$, then

$$I(\underline{u},\lambda) = (\underline{u},\underline{B}^*\underline{B}\,\underline{u}) - 2(\underline{u},\underline{B}^*\underline{q}^*) + (\underline{q}^*,\underline{q}^*) + \lambda(\underline{u},\underline{u}). \qquad (3.4.11)$$

A necessary condition for $\underline{u}^*$ to be a minimizing vector is that

$$\left.\frac{dI}{d\underline{u}}\right|_{\underline{u}=\underline{u}^*} = \underline{0}\,. \qquad (3.4.12)$$

Using this necessary condition into equation (3.4.11), we get

$$\underline{B}^*\underline{B}\,\underline{u}^* - B^*\underline{q}^* + \lambda\,\underline{u}^* = \underline{0} \qquad (3.4.13)$$

$$\Rightarrow \qquad (B^*B+\lambda\underline{I})\underline{u}^* = \underline{B}^*\underline{q}^* \qquad (3.4.14)$$

where u^* has to satisfy the equation

$$(\underline{u}^*,\underline{u}^*) = EN/T. \qquad (3.4.15)$$

Let us now show that $\underline{u}^*$, which is the solution of equations (3.4.14) and (3.4.15), does in fact satisfy our necessary and sufficient condition for optimality, (3.4.3). We have

$$(\underline{q}^* - \underline{B}\,\underline{u}^*,\ B(\underline{u}-\underline{u}^*)) = (\underline{B}^*\underline{q}^* - \underline{B}^*\underline{B}\,\underline{u}^*,\ (\underline{u}-\underline{u}^*))$$

$$= \lambda(\underline{u}^*,\ \underline{u} - \underline{u}^*). \qquad (3.4.16)$$

Since $\lambda > 0$, in order to show that (3.4.16) satisfies (3.4.3), it is enough to show that $(\underline{u}^*, \underline{u} - \underline{u}^*) \leq 0 \quad \forall \underline{u} \varepsilon C$. But, we have

$$|(\underline{u}^*,\underline{u})| \leq ||\underline{u}^*||\ ||\underline{u}|| \leq ||\underline{u}^*||^2 = (\underline{u}^*,\underline{u}^*) \qquad \forall \underline{u}\varepsilon C\ . \qquad (3.4.17)$$

The second inequality follows from the assumption that $||\underline{u}^*||^2 = EN/T$. Therefore (3.4.17) implies that

$$(\underline{u}^*,\underline{u}) \leq (\underline{u}^*,\underline{u}^*)$$

$$=> \qquad (\underline{u}^*,\ \underline{u} - \underline{u}^*) \leq 0. \qquad (3.4.18)$$

This completes the proof.

We can summarize the results of this section as follows:

1) Find the solution of equation (3.4.7) with minimal norm.
2) If the norm of the solution of (3.4.7) satisfies (3.4.4), then the optimal control vector has been determined.
3) If the solution of (3.4.7) does not satisfy (3.4.4), then the optimal control vector can be obtained from the solution of (3.4.14) and (3.4.15).

4. SOLUTION OF THE OPTIMALITY EQUATIONS AND TWO CONVERGENCE PROPERTIES OF THE DISCRETE PROBLEM

4.1 INTRODUCTION

The solution of the optimality equations (3.4.7) and (3.4.14) which resulted from the application of the necessary and sufficient condition is presented in this chapter. An algorithm will be given, based on Newton's method, for determining the Lagrange multiplier λ of equation (3.4.14), for the case when the control vector lies on the boundary of the constraint region. The convergence of the algorithm will be proved. Moreover, we are going to show that the objective function of the discrete case will converge to the objective function of the nondiscrete one. Furthermore, we will show that if the linear bounded transformation A^*A is positive definite, then we have the stronger result of the convergence of the optimal discrete control functions to the nondiscrete optimal control functions.

4.2 SOLUTION OF THE EQUATIONS OF OPTIMALITY

We now proceed to find the optimal control vector $\underline{u}^*$. Consider first case a) where

$$\underline{B}^*\underline{B}\,\underline{u} = \underline{B}^*\underline{q}^*. \tag{4.2.1}$$

This equation possesses a unique solution if and only if the homogeneous equation

$$\underline{B}^*\underline{B}\,\underline{u} = \underline{0} \tag{4.2.2}$$

has only the trivial solution. In this case the unique solution of (4.2.1) is given by

$$\underline{u}^* = (\underline{B}^*\underline{B})^{-1}\underline{B}^*\underline{q}^*. \tag{4.2.3}$$

In other words, (4.2.1) possesses a unique solution if and only if $\underline{B}^*\underline{B}$ is positive definite.

On the other hand, suppose (4.2.2) has a nontrivial solution. Then (4.2.1) possesses a solution only for these vectors $\underline{B}^*\underline{q}^*$ which belong to the set $N^{\perp}(\underline{B}^*\underline{B})$. To show that this condition is satisfied, let $\underline{v}$ be any nontrivial solution of (4.2.2). Therefore, $\underline{v}$ satisfies the equation

$$\underline{B}^*\underline{B}\,\underline{v} = \underline{0}$$

$$\Rightarrow \qquad 0 = (\underline{v}, \underline{B}^*\underline{B}\,\underline{v}) = (\underline{B}\,\underline{v}, \underline{B}\,\underline{v})$$

$$\Rightarrow \qquad \underline{B}\,\underline{v} = \underline{0}.$$

Therefore,

$$(\underline{v}, \underline{B}^*\underline{q}^*) = (\underline{B}\,\underline{v}, \underline{q}^*) = 0. \tag{4.2.4}$$

Equation (4.2.4) implies that $\underline{B}^*\underline{q}^* \varepsilon N^{\perp}(\underline{B}^*\underline{B})$, and hence (4.2.1) always possesses a solution.

Since the transformation $\underline{B}^*\underline{B}$ is self adjoint, therefore, we have the following direct sum decomposition of the Euclidean space R^s,

$$R^s = N(\underline{B}^*\underline{B}) \oplus R(\underline{B}^*\underline{B}).$$

This means that if $\underline{v}\varepsilon R^s$, then $\underline{v} = \underline{y}+\underline{z}$, where $\underline{y}\varepsilon N(\underline{B}^*\underline{B})$ and $\underline{z}\varepsilon R(\underline{B}^*\underline{B})$ and $\underline{y}$ and $\underline{z}$ are defined uniquely. Furthermore, since $\underline{y}$ is orthogonal to $\underline{z}$, we have

$$||\underline{v}||^2 = ||\underline{y}||^2 + ||\underline{z}||^2.$$

Suppose that $\underline{u}$ is a solution of equation (4.2.1); therefore, $\underline{u} = \underline{u}_1+\underline{u}_2$, where $\underline{u}_1 \varepsilon N(\underline{B}^*\underline{B})$ and $\underline{u}_2 \varepsilon R(\underline{B}^*\underline{B})$. Substituting for $\underline{u}$ into equation (4.2.1) we get

$$\underline{B}^*\underline{B}\,\underline{u} = \underline{B}^*\underline{B}(\underline{u}_1+\underline{u}_2) = \underline{B}^*\underline{B}\,\underline{u}_1 + \underline{B}^*\underline{B}\,\underline{u}_2 = \underline{B}^*\underline{B}\,\underline{u}_2 = \underline{B}^*\underline{q}^*.$$

Therefore, $\underline{u}_2$ is also a solution of (4.2.1) and since we are looking for a solution with minimum norm, and from the fact that $||\underline{u}||^2 = ||\underline{u}_1||^2 + ||\underline{u}_2||^2$, so the optimal control vector $\underline{u}^*$ has to belong to $R(\underline{B}^*\underline{B})$. This means that the optimal control vector has to be of the form

$$\underline{u}^* = \sum_{i=1}^{p} \alpha_i\, \underline{z}_i. \tag{4.2.5}$$

Let us express $\underline{B}^*\underline{q}^*$ in terms of the eigenvectors of $\underline{B}^*\underline{B}$

$$\begin{aligned}
\underline{B}^*\underline{q}^* &= \sum_{i=1}^{s} (\underline{B}^*\underline{q}^*,\underline{z}_i)\underline{z}_i \\
&= \sum_{i=1}^{s} (\underline{q}^*,\ \underline{B}\ \underline{z}_i)\underline{z}_i \\
&= \sum_{i=1}^{p} (\underline{q}^*,\ \underline{B}\ \underline{z}_i)\underline{z}_i. \qquad (4.2.6)
\end{aligned}$$

Substituting (4.2.6) and (4.2.5) into (4.2.1), we get

$$\sum_{i=1}^{p} (\underline{q}^*,\underline{B}\ \underline{z}_i)\underline{z}_i = \sum_{i=1}^{p} \alpha_i \underline{B}^*\underline{B}\ \underline{z}_i = \sum_{i=1}^{p} \alpha_i \lambda_i \underline{z}_i. \tag{4.2.7}$$

Since the $\underline{z}_i$'s are linearly independent, we deduce from (4.2.7) that

$$\alpha_i = \frac{(\underline{q}^*,\underline{B}\ \underline{z}_i)}{\lambda_i} \qquad i = 1,\ldots,p \tag{4.2.8}$$

Therefore, the optimal control vector $\underline{u}^*$ is given by

$$\underline{u}^* = \sum_{i=1}^{p} \frac{(\underline{q}^*, \underline{B}\,\underline{z}_i)}{\lambda_i} \underline{z}_i. \qquad (4.2.9)$$

Case b)

In this case the optimal control vector is obtained through the solution of the following two equations for $\lambda > 0$

$$(\underline{B}^*\underline{B} + \lambda\underline{I})\underline{u} = \underline{B}^*\underline{q}^* \qquad (4.2.10)$$

$$(\underline{u},\underline{u}) = EN/T. \qquad (4.2.11)$$

Since $\lambda > 0$ and $\underline{B}^*\underline{B}$ is positive semidefinite, therefore $-\lambda$ does not belong to the spectrum of $\underline{B}^*\underline{B}$, i.e., the inverse of $(\underline{B}^*\underline{B}+\lambda\underline{I})$ does exist and hence $\underline{u}^*$ is given by

$$\underline{u}^* = (\underline{B}^*\underline{B} + \lambda\underline{I})^{-1}\underline{B}^*\underline{q}^* \qquad (4.2.12)$$

where λ is to be chosen such that $\underline{u}^*$ satisfies (4.2.11). Substituting (4.2.12) into (4.2.11), we get

$$((\underline{B}^*\underline{B}+\lambda\underline{I})^{-1}\underline{B}^*\underline{q}^*, (\underline{B}^*\underline{B}+\lambda\underline{I})^{-1}\underline{B}^*\underline{q}^*) = EN/T$$

or

$$(\underline{B}^*\underline{q}^*, (\underline{B}^*\underline{B}+\lambda\underline{I})^{-2}\underline{B}^*\underline{q}^*) = EN/T. \qquad (4.2.13)$$

Define a functional f mapping the open set $(0,\infty)$ into R by

$$f(\lambda) = (\underline{B}^*\underline{q}^*, (\underline{B}^*\underline{B}+\lambda\underline{I})^{-2}\underline{B}^*\underline{q}^*) - EN/T. \qquad (4.2.14)$$

The problem of finding the optimal control vector in this case turns out to be the problem of finding λ which solves the equation $f(\lambda) = 0$. To find λ, which satisfies $f(\lambda) = 0$, we are going to use Newton's method.[34] In order to prove the convergence of Newton's method we need to show that $f(\lambda)$ is a strictly convex function and we need the following theorems and lemmas to prove this result.

Let $R(\lambda)$ denote the resolvent operator of $\underline{B}^*\underline{B}$, i.e., $R(\lambda) = (\underline{B}^*\underline{B}+\lambda\underline{I})^{-1}$, where $-\lambda\epsilon\rho(\underline{B}^*\underline{B})$, and $\rho(\underline{B}^*\underline{B})$ denotes the resolvent set of the operator $\underline{B}^*\underline{B}$. It is clear that $R(\lambda)$ is a positive definite operator mapping R^s onto itself.

Lemma 1. $$R(\lambda) - R(\mu) = (\mu-\lambda)R(\mu)R(\lambda). \tag{4.2.15}$$

Proof. From the definition of $R(\lambda)$ and $R(\mu)$, we have

$$(\mu-\lambda)\underline{I} = (\underline{B}^*\underline{B}+\mu\underline{I}) - (\underline{B}^*\underline{B}+\lambda\underline{I})$$

$$= (\underline{B}^*\underline{B}+\mu\underline{I})[R(\lambda)-R(\mu)](\underline{B}^*\underline{B}+\lambda\underline{I})$$

$$\Rightarrow \qquad R(\lambda) - R(\mu) = (\mu-\lambda)R(\mu)R(\lambda).$$

As a consequence of the above lemma, we have:

Corollary: $R(\mu)$ and $R(\lambda)$ commute.

Theorem 1. Suppose $A,B\epsilon L(X,X)$ and that A and B commute. Then $A \geqslant 0$ and $B \geqslant 0$ implies that $AB \geqslant 0$.

Proof. For proof see reference 32, p 415.

Lemma 2. Suppose $A\epsilon L(X,X)$; then $A \geqslant 0$ and A^{-1} exists $\Leftrightarrow A > 0$.

Proof. ($\Rightarrow$) Suppose false, i.e., there exists $z\epsilon X$, $z \neq 0$ such that $(z,Az) = 0$. Let B denote the square root of A, i.e., $B^2 = A$; then B exists since $A \geqslant 0$ and B is self adjoint.* Moreover, $B\epsilon L(X,X)$ and we have

$$0 = (z,Az) = (z,B^2z) = (Bz,Bz)$$

$$\Rightarrow \qquad Bz = 0 \qquad \Rightarrow \qquad B^2z = Az = 0.$$

*Reference 32, p 422.

Since A^{-1} exists, therefore $z = 0$, contradicting our hypothesis and, hence, no such x exists.

(<=) follows trivially.

As a result of the last theorem and lemma, we have the following result.

Corollary: Suppose A and B ε L(X,X) and that A and B commute. Then $A > 0$ and $B > 0$ implies that $AB > 0$.

Theorem 2. Let g be a differentiable functional on an open set $\Gamma \subset R^n$. A necessary and sufficient condition that g be strictly convex* on Γ is that for each $\underline{z}_1$ and $\underline{z}_2 \; \varepsilon \; \Gamma$

$$(\nabla g(\underline{z}_2) - \nabla g(\underline{z}_1), \underline{z}_2 - \underline{z}_1) > 0.$$

Proof. For proof see reference 35, p 87.

Theorem 3. The functional f defined by (4.2.14) which maps the set $(0,\infty)$ into R^1 is one to one and strictly convex.

Proof. f is one to one. To show this, suppose false, i.e., there exists $\lambda,\mu > 0$ and $\lambda \neq \mu$ such that $f(\lambda) = f(\mu)$ =>

$$0 = f(\lambda)-f(\mu) = (\underline{B}^*\underline{q}^*, [R^2(\lambda)-R^2(\mu)]\underline{B}^*\underline{q}^*)$$

$$= (\mu-\lambda)(\underline{B}^*\underline{q}^*, R(\lambda)R(\mu)[R(\lambda)+R(\mu)]\underline{B}^*\underline{q}^*). \quad (4.2.16)$$

But, since $R(\lambda) > 0$, $R(\mu) > 0$, and both $R(\lambda)$ and $R(\mu)$ commute, it follows from the corollary of lemma 2 that $R(\lambda)R(\mu)[R(\lambda)+R(\mu)] > 0$ =>

$$(\underline{B}^*\underline{q}^*, R(\lambda)R(\mu)[R(\lambda)+R(\mu)]\underline{B}^*\underline{q}^*) > 0. \quad (4.2.17)$$

*For definition, see reference 35, p 56.

Using the result of (4.2.17) in (4.2.16), we deduce that $\mu - \lambda = 0$. This contradicts our hypothesis and thus establishes that f is one to one.

To show that f is strictly convex, consider

$$\nabla f(\mu) = -2(\underline{B}^*\underline{q}^*, R^3(\mu)\underline{B}^*\underline{q}^*).$$

Therefore,

$$\begin{aligned}
[\nabla f(\mu)-\nabla f(\lambda)](\mu-\lambda) &= -2(\mu-\lambda)(\underline{B}^*\underline{q}^*, [R^3(\mu)-R^3(\lambda)]\underline{B}^*\underline{q}^*) \\
&= -2(\mu-\lambda)(\underline{B}^*\underline{q}^*, [R(\mu)-R(\lambda)][R^2(\mu)+R(\mu)R(\lambda)+R^2(\lambda)]\underline{B}^*\underline{q}^*) \\
&= 2(\mu-\lambda)^2(\underline{B}^*\underline{q}^*, R(\mu)R(\lambda)[R^2(\mu)+R(\mu)R(\lambda)+R^2(\lambda)]\underline{B}^*\underline{q}^*) \\
&> 0
\end{aligned}$$

which implies that f is strictly convex.

If we denote the solution of equations (4.2.10) or (4.2.12) by $\underline{u}^\lambda$ and the solution of equation (4.2.1) by $\underline{u}^o$, we can have the following result.

Lemma 3. $\underline{u}^\lambda \to \underline{u}^o$ as $\lambda \to 0$. Furthermore, $\underline{u}^\lambda \to \underline{0}$ as $\lambda \to \infty$.

Proof. Let us express the solution of (4.2.10) in terms of the eigen-vectors of $\underline{B}^*\underline{B}$; therefore, the expression of $\underline{u}^\lambda$ will take the form

$$\underline{u}^\lambda = \sum_{i=1}^{s} \alpha_i \underline{z}_i .$$

Substituting into (4.2.10) we get

$$(\underline{B}^*\underline{B} + \lambda \underline{I}) \sum_{i=1}^{s} \alpha_i \underline{z}_i = \sum_{i=1}^{p} (\underline{q}^*, \underline{B}\,\underline{z}_i)\underline{z}_i$$

$$=> \quad \sum_{i=1}^{p} \alpha_i(\lambda_i+\lambda)\underline{z}_i + \sum_{i=p+1}^{s} \alpha_i \lambda \underline{z}_i = \sum_{i=1}^{p} (\underline{q}^*, \underline{B}\,\underline{z}_i)\underline{z}_i$$

$$=> \quad \alpha_i = \begin{cases} \dfrac{(\underline{q}^*, \underline{B}\,\underline{z}_i)}{\lambda_i + \lambda} & i = 1,2,\ldots,p \\ 0 & i = p+1,\ldots,s \end{cases} \tag{4.2.18}$$

Therefore,

$$\underline{u}^\lambda = \sum_{i=1}^{p} \frac{(\underline{q}^*, \underline{B}\,\underline{z}_i)}{\lambda_i + \lambda} \underline{z}_i. \tag{4.2.19}$$

Since the expression of $\underline{u}^o$ is given by

$$\underline{u}^o = \sum_{i=1}^{p} \frac{(\underline{q}^*, \underline{B}\,\underline{z}_i)}{\lambda_i} \underline{z}_i,$$

it is clear that $\underline{u}^\lambda \to \underline{u}^o$ as $\lambda \to 0$. Moreover, it follows from (4.2.19) that $\underline{u}^\lambda \to \underline{0}$ as $\lambda \to \infty$.

Theorem 4. The equation $f(\lambda) = 0$ has a solution and this solution is unique.

Proof. Let us write the expression of $f(\lambda)$ as

$$f(\lambda) = (\underline{u}^\lambda, \underline{u}^\lambda) - EN/T.$$

From lemma 3 we know that $\underline{u}^\lambda \to \underline{u}^o$ as $\lambda \to 0$, and since $||\underline{u}^o||^2 > EN/T$, therefore, if λ is sufficiently small => $||\underline{u}^\lambda||^2 > EN/T$, i.e., $f(\lambda) > 0$ for sufficiently small λ.

Also, we have

$$\nabla f(\lambda) = -2(\underline{B}^*\underline{q}^*, R^3(\lambda)\underline{B}^*\underline{q}^*) < 0.$$

Therefore, the function $f(\lambda)$ is monotonically decreasing. Moreover, since $\underline{u}^{\lambda} \to 0$ as $\lambda \to \infty$ => $f(\lambda) \to -EN/T$ as $\lambda \to \infty$. From the above results and since f is one to one, it follows that there exists a unique solution λ, $0 < \lambda < \infty$, for $f(\lambda) = 0$.

4.3 NEWTON'S METHOD

In this section we will give an iterative scheme based on Newton's method for computing λ which is the solution of the equation $f(\lambda) = 0$.

Let λ_o be chosen such that $f(\lambda_o) > 0$. This can easily be done if we choose λ_o sufficiently small as seen from theorem 4. Choose λ_n inductively such that

$$\lambda_{n+1} = \lambda_n - f(\lambda_n)/\nabla f(\lambda_n). \qquad (4.3.1)$$

Before proving the convergence of the sequence $\{\lambda_n\}$ we need the following theorem.

Theorem 5. Let g be a differentiable functional on an open convex set $\Gamma \subset R^n$. g is strictly convex on Γ if and only if for each $\underline{z}_1$, $\underline{z}_2 \in \Gamma$

$$g(\underline{z}_2) - g(\underline{z}_1) > \nabla g'(\underline{z}_1)(\underline{z}_2-\underline{z}_1). \qquad (4.3.2)$$

Proof. For proof see reference 35.

We have all the necessary information we need to prove the convergence of Newton's method.

Theorem 6. The sequence $\{\lambda_n\}$ defined by equation (4.3.1) is a monotonically increasing convergent sequence. Furthermore, $\lambda_n \to \lambda$ as $n \to \infty$, where λ satisfies the equation $f(\lambda) = 0$.

Proof. The proof that $\{\lambda_n\}$ is monotonically increasing will be by induction. Assume that $f(\lambda_n) > 0$ and since $\nabla f(\lambda_n) < 0$, it follows from (4.3.1) that $\lambda_{n+1} > \lambda_n$. In order for this to hold for all n we have to show that $f(\lambda_{n+1}) > 0$. From (4.3.2), we have

$$f(\lambda_{n+1}) > f(\lambda_n) + \nabla f(\lambda_n)(\lambda_{n+1}-\lambda_n). \tag{4.3.3}$$

Since from (4.3.1) the right hand side of (4.3.3) is equal to zero, it follows that $f(\lambda_{n+1}) > 0 \Rightarrow \{\lambda_n\}$ is a monotonically increasing sequence. Since $f(\lambda_n) > 0$ and f decreases monotonically with λ, this implies that the sequence $\{\lambda_n\}$ is bounded from above by λ which satisfies $f(\lambda) = 0 \Rightarrow$ the sequence $\{\lambda_n\}$ converges.

Furthermore, from the convergence of $\{\lambda_n\}$ it follows that $\frac{f(\lambda_n)}{\nabla f(\lambda_n)} \to 0$ as $n \to \infty$. But since $\nabla f(\lambda_n)$ is bounded, $f(\lambda_n) \to 0$ as $n \to \infty$ and this completes the proof.

4.4 SOME CONVERGENCE PROPERTIES OF THE DISCRETE PROBLEM

The approximation of a continuous optimal control problem by a discrete one for lumped parameter systems was not considered until recently. Both Budak et al[36] and Cullum[37] treated this problem and showed, under certain reasonable assumptions, that the solution of the discrete optimization problem converged to the solution of the continuous optimal problem. Moreover, Cullum in her paper showed that the two point boundary value problem, encountered usually in the solution of continuous optimal control problems, can be overcome by discretization. This is an important result, since the solution of discrete problems is usually easier than that of continuous ones.

In this section we are going to show that the objective function of the discrete optimization problem will converge to the objective function of the nondiscrete problem. By the nondiscrete problem we mean the minimizing of the objective function (3.2.2) under the constraint (3.2.3), where the control vector $\underline{u}(t) \in L_2^r[0,T]$ and $\underline{u}(t)$ is not subject to any further constraints.

Let $\underline{u}_N^*$ denote the optimal control vector of the discrete optimization problem and $\underline{u}^{**}(t)$ denote the optimal control vector of the nondiscrete problem. It has been shown that $\underline{u}^{**}$ exists. We will show that $\underline{u}_N^*$ will tend to $\underline{u}^{**}$ as N tends to infinity provided that the linear bounded transformation A^*A is positive definite.

Let U denote the set of all elements in $L_2^r[0,T]$ which satisfies (3.2.3). Let U_N denote the set of $\underline{u}$ (3.2.11) satisfying (3.2.18).

Theorem 7. If $I_N^* = \inf_{\underline{u}_N \in U_N} I(\underline{u}_N)$, then

$$\lim_{N\to\infty} I_N^* = I^* = I(\underline{u}^{**}).$$

Proof. Since the set of continuous functions is dense in $L_2^r[0,T]$, we can choose a vector of continuous functions $\underline{v}(t) \in U$ such that $||\underline{u}^{**}-\underline{v}|| < \delta$ and hence

$$\begin{aligned}
\left|I(\underline{u}^{**}) - I(\underline{v})\right| &= \left| ||A\underline{u}^{**}-\underline{q}^*||^2 - ||A\underline{v}-\underline{q}^*||^2 \right| \\
&= \left| ||A\underline{u}^{**}-\underline{q}^*|| - ||A\underline{v}-\underline{q}^*|| \right| \left| ||A\underline{u}^{**}-\underline{q}^*|| + ||A\underline{v}-\underline{q}^*|| \right| \\
&< k||A(\underline{u}^{**}-\underline{v})|| \\
&< k||A|| \; ||\underline{u}^{**}-\underline{v}|| \\
&< \varepsilon/2 \qquad (4.4.1)
\end{aligned}$$

where $\varepsilon = 2k\delta||A||$.

Since each component $v_i(t)$, $i = 1,\ldots,r$, of the vector $\underline{v}(t)$ is continuous on $[0,T]$, therefore each $v_i(t)$ is uniformly continuous on $[0,T]$. Hence given $\varepsilon_1 > 0$, there exists $\delta_1 > 0$ such that $|v_i(t_1) - v_i(t_2)| < \varepsilon_1$, $\forall\ t_1, t_2 \in [0,T]$ such that $|t_1 - t_2| < \delta_1$, $i = 1,2,\ldots,r$.

If we choose N sufficiently large such that $T_N < \delta_1$, then we can be sure that the oscillation of any component $v_i(t)$ in any subinterval $[\ell T_N,(\ell+1)T_N]$, $\ell = 0,\ldots,N-1$, is less than ε_1. Therefore, the piecewise constant vector $\underline{u}_N(t,\varepsilon_1)$, coinciding on each subinterval $[\ell T_N,(\ell+1)T_N)$, $\ell = 0,\ldots,N-1$, with the value of $\underline{v}(t)$ on $[\ell T_N,(\ell+1)T_N)$ which is closest to zero, will approximate $\underline{v}(t)$ uniformly on $[0,T]$ with accuracy ε_1. Moreover, $\underline{u}_N(t,\varepsilon_1)$ belongs to U_N by our choice. Also, if we choose $\varepsilon_1 < \delta/\,Tr$, then $||\underline{v}(t) - \underline{u}_N(t,\varepsilon_1)|| < \varepsilon_1\,Tr < \delta$. Following a similar procedure to that used in deriving (4.4.1), we can show that

$$|I(\underline{v}) - I(\underline{u}_N(t,\varepsilon_1))| < \varepsilon/2. \tag{4.4.2}$$

Combining equations (4.4.1) and (4.4.2), we get

$$|I(\underline{u}^{**}) - I(\underline{u}_N(t,\varepsilon_1))| < \varepsilon.$$

From this inequality and the fact that

$$I(\underline{u}^{**}) \leq I(\underline{u}_N(t,\varepsilon_1)),$$

it follows that

$$I(\underline{u}^{**}) \leq I(\underline{u}_N(t,\varepsilon_1)) \leq I(\underline{u}^{**}) + \varepsilon \tag{4.4.3}$$

Since I_N^* is the greatest lower bound of $I(\underline{u})$ taken over all $\underline{u}_N \varepsilon U_N$, we have the obvious inequality

$$I_N^* \leq I(\underline{u}_N(t,\varepsilon_1)). \tag{4.4.4}$$

Combining equations (4.3.3) and (4.3.4), we get

$$I(\underline{u}^{**}) \leq I_N^* \leq I(\underline{u}_N(t,\varepsilon_1)) \leq I(\underline{u}^{**}) + \varepsilon \tag{4.4.5}$$

Since ε can be chosen arbitrarily small, it follows from (4.4.5) that $I_N^* \to I(\underline{u}^{**})$ as $N \to \infty$.

In the next theorem we are going to show that if the linear bounded transformation A^*A is positive definite, then we will have the stronger result of the convergence of the optimal control vector of the discrete problem to the optimal control vector of the nondiscrete problem. Hence, given any nondiscrete optimization problem of the quadratic form considered in this thesis with $A^*A > 0$, we can obtain an approximate solution to this optimization problem using discrete inputs. Moreover, the approximate solution can be made as close as we wish to the exact solution.

Theorem 8. If $A^*A > 0$, then $\underline{u}_N^* \to \underline{u}^{**}$ as $N \to \infty$.

Proof. Consider the objective function corresponding to an optimal discrete vector $I(\underline{u}_N^*)$. We have

$$\begin{aligned} I(\underline{u}_N^*) &= (A\underline{u}_N^*-\underline{q}^*, A\underline{u}_N^*-\underline{q}^*) \\ &= (A\underline{u}^{**}-\underline{q}^*+A[\underline{u}_N^*-\underline{u}^{**}], A\underline{u}^{**}-\underline{q}^*+A[\underline{u}_N^*-\underline{u}^{**}]) \\ &= I(\underline{u}^{**}) + 2(A[\underline{u}_N^*-\underline{u}^{**}], A\underline{u}^{**}-\underline{q}^*) + ([\underline{u}_N^*-\underline{u}^{**}], A^*A[\underline{u}_N^*-\underline{u}^{**}]) \end{aligned} \tag{4.4.6}$$

$$\Rightarrow \quad ([\underline{u}_N^*-\underline{u}^{**}], A^*A[\underline{u}_N^*-\underline{u}^{**}]) = I(\underline{u}_N^*)-I(\underline{u}^{**})+2(\underline{q}^*-A\underline{u}^{**}, A[\underline{u}_N^*-\underline{u}^{**}]). \tag{4.4.7}$$

Using the necessary and sufficient condition for optimality (3.4.2), it can be seen that $(\underline{q}^*-A\underline{u}^{**}, A[\underline{u}_N^*-\underline{u}^{**}]) \leq 0$. Substituting this last inequality into (4.4.7), we deduce that

$$([\underline{u}_N^* - \underline{u}^{**}], A^*A[\underline{u}_N^* - \underline{u}^{**}]) < I(\underline{u}_N^*) - I(\underline{u}^{**})$$

$$< \varepsilon. \qquad (4.4.8)$$

If N is chosen large enough, it is seen that (4.4.8) follows from theorem 7. Also, since A^*A is linear bounded positive definite operator, then there exists two constants, k_1 and k_2, such that

$$k_1||\underline{u}||^2 < (\underline{u}, A^*A\underline{u}) < k_2||\underline{u}||^2. \qquad (4.4.9)$$

Substituting the left inequality of (4.4.9) into (4.4.8), we get

$$k_1||\underline{u}_N^* - \underline{u}^{**}||^2 < \varepsilon$$

$$\Rightarrow \qquad ||\underline{u}_N^* - \underline{u}^{**}||^2 < \varepsilon/k_1. \qquad (4.4.10)$$

Hence, the theorem follows directly from the inequality (4.4.10)

5. APPLICATION OF THE THEORETICAL RESULTS TO TWO SPECIFIC EXAMPLES

5.1 INTRODUCTION

In the last two chapters we were concerned with obtaining a necessary and sufficient condition for optimality from which we derive an optimality equation whose solution yields the optimal control vector.

In this chapter we are going to apply the results obtained in chapters two and three to solve two examples. The first example consists of minimizing the total energy of a vibrating string in a given time subject to an energy constraint on the control function. The second example is concerned with a system described by a diffusion equation and our objective is to attain a temperature distribution along the slab which is as close as possible to a specified temperature distribution. Also, an energy type constraint is imposed on the control function.

5.2 TRANSVERSE VIBRATION OF A STRING

Consider the transverse vibration of a string whose displacement $w(x,t)$ from the equilibrium position is given by

$$\rho \frac{\partial^2 w(x,t)}{\partial t^2} = \tau \frac{\partial^2 w(x,t)}{\partial x^2} \qquad 0 < x < \ell,\ t \in [0,T] \tag{5.2.1}$$

with boundary conditions

$$w(0,t) = 0 \quad ; \quad w(\ell,t) = f(t) \tag{5.2.2}$$

and initial conditions

$$w_t(x,0) = 0 \quad , \quad w_x(x,0) = \frac{f(0)}{\ell} \tag{5.2.3}$$

where ρ denotes the density of the string material, τ denotes the tension and ℓ denotes the length of the string. If we assume for simplicity that $\ell = \pi$ and $\rho = \tau = 1$, then (5.2.1) will take the form

$$w_{tt} = w_{xx} \qquad 0 < x < \ell,\ 0 < t \leq T. \tag{5.2.4}$$

Using the transformation of variable

$$v(x,t) = w(x,t) - \frac{x}{\ell} f(t) \tag{5.2.5}$$

and substituting the expression of $v(x,t)$ from (5.2.5) into equation (5.2.4), (5.2.2) and (5.2.3), we get

$$v_{tt} = v_{xx} - \frac{x}{\pi} u(t) \tag{5.2.6}$$

where

$$u(t) = \frac{d^2 f(t)}{dt^2} \tag{5.2.7}$$

with boundary conditions

$$v(0,t) = 0 \quad , \quad v(\ell,t) = 0 \tag{5.2.8}$$

and initial conditions

$$\begin{aligned} v_t(x,0) &= w_t(x,0) - \frac{x}{\ell} f'(0) \\ v_x(x,0) &= w_x(x,0) - \frac{f(0)}{\pi} \quad . \end{aligned} \tag{5.2.9}$$

Assuming that $f'(0) = 0$, the initial conditions with respect to the new variable $v(x,t)$ will be

$$v_t(x,0) = 0 \quad , \quad v_x(x,0) = 0. \tag{5.2.10}$$

We are going to take u(t) as our control function,[38] i.e., we are controlling the vibration of the string by applying a control function in the form of an acceleration on the free end of the string. Our aim is to minimize the total energy of the string which is given by

$$I(u) = \frac{1}{2}\int_0^{\ell} (\rho w_t^2 + \tau w_x^2)dx$$

$$= \frac{1}{2}\int_0^{\pi} (w_t^2 + w_x^2)dx \tag{5.2.11}$$

subject to the energy constraint

$$\int_0^T u^2(t)dt \leqslant E \tag{5.2.12}$$

where E is a given constant.

Using the method of separation of variables,[39] the solution of equation (5.2.6), with zero initial and boundary conditions, is given by

$$v(x,t) = \frac{2}{\pi} \sum_{n=1}^{\infty} \{\frac{(-1)^n}{n^2} \int_0^t \sin[n(t-\tau)]u(\tau)d\tau\}\sin(nx). \tag{5.2.13}$$

Therefore, the expression of $w_t(x,t)$ and $w_x(x,t)$ will take the form

$$\begin{aligned} w_t(x,t) &= v_t(x,t) + \frac{x}{\pi} f'(t) \\ &= v_t(x,t) + \frac{x}{\ell}\int_0^t u(\tau)d\tau \\ &= \int_0^t \{\frac{x}{\pi} + \frac{2}{\pi}\sum_{n=1}^{\infty} \frac{(-1)^n}{n} \sin(nx)\cos[n(t-\tau)]\}u(\tau)d\tau \end{aligned} \tag{5.2.14}$$

and

$$w_x(x,t) = v_x(x,t) + \frac{f(t)}{\pi}$$

$$= v_x(x,t) + \frac{1}{\pi}\left[f(0) + \int_0^t (t-\tau)u(\tau)d\tau\right]$$

$$= \frac{f(0)}{\pi} + \int_0^t \left\{\frac{(t-\tau)}{\pi} + \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^n}{n}\cos(nx)\sin[n(t-\tau)]\right\}u(\tau)d\tau. \tag{5.2.15}$$

Hence, at the final time T, $w_t(x,T)$ and $w_x(x,T)$ will be

$$w_t(x,T) = \int_0^T \left\{\frac{x}{\pi} + \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^n}{n}\sin(nx)\cos[n(T-\tau)]\right\}u(\tau)d\tau \tag{5.2.16}$$

$$w_x(x,T) = \frac{f(0)}{\pi} + \int_0^T \left\{\frac{(T-\tau)}{\pi} + \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^n}{n}\cos(nx)\sin[n(T-\tau)]\right\}u(\tau)d\tau. \tag{5.2.17}$$

Comparing this problem with our original problem represented by equations (3.2.1) and (3.2.2), we can easily deduce that

$$\underline{q}(x,T) = \begin{bmatrix} q_1(x,T) \\ \\ q_2(x,T) \end{bmatrix} = \int_0^T \underline{K}(x,T,\tau)u(\tau)d\tau + \begin{bmatrix} 0 \\ \\ \frac{f(0)}{\pi} \end{bmatrix} \tag{5.2.18}$$

where

$$\underline{K}(x,T,\tau) = \begin{bmatrix} K_{11}(x,T,\tau) \\ \\ K_{21}(x,T,\tau) \end{bmatrix}$$

and

$$K_{11}(x,T,\tau) = \frac{1}{\pi}\left\{x + 2\sum_{n=1}^{\infty}\frac{(-1)^n}{n}\sin(nx)\cos[n(T-\tau)]\right\} \tag{5.2.19}$$

$$K_{21}(x,T,\tau) = \frac{1}{\pi}\left\{(T-\tau) + 2\sum_{n=1}^{\infty}\frac{(-1)^n}{n}\cos(nx)\sin[n(T-\tau)]\right\}. \tag{5.2.20}$$

The objective function I(u) can be expressed as

$$I(u) = \frac{1}{2} \sum_{i=1}^{2} \int_{0}^{\pi} [q_i(x,T) - q_{d_i}(x)]^2 dx \tag{5.2.21}$$

where

$$q_{d_i}(x) = 0, \qquad i = 1,2. \tag{5.2.22}$$

Using equation (3.2.20), the expression for $\underline{q}^*(x)$ is

$$\underline{q}^*(x) = \begin{bmatrix} q_1^*(x) \\ \\ q_2^*(x) \end{bmatrix} = \begin{bmatrix} 0 \\ \\ -\frac{f(0)}{\pi} \end{bmatrix}. \tag{5.2.23}$$

From (3.2.13), the expression of $\underline{K}_\ell(\underline{x})$ is given by

$$\underline{K}_\ell(x) = \int_{\ell T_N}^{(\ell+1)T_N} \underline{K}(x,T,\tau)d\tau \qquad \ell = 0,1,\ldots,N-1. \tag{5.2.24}$$

Hence, performing the above integration yields

$$\underline{K}_\ell(x) = \begin{bmatrix} \frac{1}{\pi}\{xT_N + 4 \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} \sin(nx)\sin(\frac{nT_N}{2})\cos[n(T - \frac{2\ell+1}{2} T_N)]\} \\ \\ \frac{1}{\pi}\{T_N(T - \frac{2\ell+1}{2} T_N) + 4 \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} \cos(nx)\sin(\frac{nT_N}{2})\sin[n(T-\frac{2\ell+1}{2}T_N)]\} \end{bmatrix}$$

$$\ell = 0,1,\ldots,N-1. \tag{5.2.25}$$

By denoting the square matrix $\underline{B}^*\underline{B}$ by $\underline{C}$ and taking into account the definition of the matrix $\underline{B}^*\underline{B}$, as given by (3.3.2), and performing the necessary multiplication and integration needed, it follows that an element C_{ij}, $i,j = 1,\ldots,N$, of the matrix $\underline{C}$ can be expressed as

$$C_{ij} = \frac{\pi T_N^2}{3} + \frac{T_N^2}{\pi}\left(T - \frac{2i-1}{2}T_N\right)\left(T - \frac{2j-1}{2}T_N\right) + \frac{8}{\pi}\sum_{n=1}^{\infty}\left\{\frac{1}{n^4}\sin^2\left(\frac{nT_N}{2}\right)\right.$$

$$\left.\cos[n(i-j)T_N]\right\} - \frac{4T_N}{\pi}\sum_{n=1}^{\infty}\frac{1}{n^3}\sin\frac{nT_N}{2}\left\{\cos\left[n\left(T - \frac{2i-1}{2}T_N\right)\right] + \right.$$

$$\left.\cos\left[n\left(T - \frac{2j-1}{2}T_N\right)\right]\right\}. \tag{5.2.26}$$

Furthermore, it can be shown from the definition of $\underline{B}^*$ and $\underline{q}^*$, equations (3.3.1) and (5.2.23) respectively, that an element $(\underline{B}^*\underline{q}^*)_i$ of $\underline{B}^*\underline{q}^*$ can be expressed as

$$(\underline{B}^*\underline{q}^*)_i = -\frac{f(0)}{\pi}T_N\left(T - \frac{2i-1}{2}T_N\right) \qquad i = 1,\ldots,N \tag{5.2.27}$$

Using (5.2.26) for calculating $\underline{B}^*\underline{B}$, formula (5.2.27) for calculating $\underline{B}^*\underline{q}^*$ and (5.2.23) for calculating $\underline{q}^*$, and taking $f(0) = 0.2$, $E = 0.2\text{x}10^{-3}$ and $T = 12$, the following results were obtained for 10 sampling periods:

Lagrange multiplier $\lambda = 0.948$

Objective function $= 0.432\text{x}10^{-4}$

The optimal discrete control vector is given in Table 5.1.

TABLE 5.1

OPTIMAL CONTROL VECTOR $\underline{u}$

Sampling Period	Control Function $\underline{u}$
1	-0.541×10^{-2}
2	-0.396×10^{-2}
3	-0.399×10^{-2}
4	-0.401×10^{-2}
5	-0.568×10^{-2}
6	0.926×10^{-3}
7	0.405×10^{-2}
8	0.399×10^{-2}
9	0.400×10^{-2}
10	0.280×10^{-2}

The Lagrange multiplier λ was calculated using Newton's method and it took six iterations to compute it with an accuracy of the order of 10^{-4}.

The final displacement and velocity of the string corresponding to this optimal control vector were calculated and are tabulated in Table 5.2.

TABLE 5.2

FINAL VELOCITY AND DISPLACEMENT OF THE STRING

x/ℓ	Displacement	Velocity
0	0	0
0.1	-0.279×10^{-5}	0.247×10^{-4}
0.2	-0.144×10^{-4}	0.543×10^{-4}
0.3	-0.352×10^{-4}	0.818×10^{-4}
0.4	-0.381×10^{-4}	-0.172×10^{-3}
0.5	0.161×10^{-3}	-0.131×10^{-2}
0.6	0.752×10^{-3}	-0.302×10^{-2}
0.7	0.159×10^{-2}	-0.530×10^{-2}
0.8	0.243×10^{-2}	-0.625×10^{-2}
0.9	0.316×10^{-2}	-0.874×10^{-2}
1.0	0.316×10^{-2}	-0.874×10^{-2}

To show the convergence of the objective function of the discrete case to the objective function of the nondiscrete one, the same example has been solved for different sampling periods and the results are given in Table 5.3.

TABLE 5.3

VALUE OF OBJECTIVE FUNCTION

FOR DIFFERENT NUMBER OF SAMPLING PERIODS

Number of Sampling Periods	Objective Function
5	0.786×10^{-4}
10	0.432×10^{-4}
15	0.408×10^{-4}
20	0.371×10^{-4}
30	0.356×10^{-4}
40	0.350×10^{-4}
50	0.345×10^{-4}

5.3 HEATING A SLAB OF METAL

In this section we consider the heating of a slab of metal whose thickness is unity. We are going to control the temperature distribution along the slab by controlling the temperature of the gas medium adjacent to the surface of the slab. The equation which describes the temperature distribution at any instant of time is given by the diffusion equation

$$\frac{\partial^2 q(x,t)}{\partial x^2} = \frac{\partial q(x,t)}{\partial t} \qquad t \in [0,T],\ x \in [0,1] \tag{5.3.1}$$

with initial condition

$$q(x,0) = 0 \tag{5.3.2}$$

and boundary conditions

$$\left.\frac{\partial q(x,t)}{\partial x}\right|_{x=0} = \alpha\{q(0,t) - u(t)\} \tag{5.3.3}$$

$$\left.\frac{\partial q(x,t)}{\partial x}\right|_{x=1} = 0 \tag{5.3.4}$$

where

α = heat transfer coefficient, $u(t)$ = temperature of gas medium and is the control function and $q(x,t)$ = temperature distribution of the metal slab.

The objective is to minimize a quadratic performance index of the form

$$I(u) = \int_o^1 \{q(x,T) - q_d(x)\}^2 dx \tag{5.3.5}$$

subject to the constraint

$$\int_o^T u^2(t)dt \leq E. \tag{5.3.6}$$

Solving the diffusion equation (5.3.1) under the initial and boundary conditions (5.3.2), (5.3.3) and (5.3.4) we get

$$q(x,T) = 2\int_o^T \sum_{i=1}^{\infty} \frac{\cos((1-x)\beta_i)}{(\frac{1}{\alpha} + \frac{1+\alpha}{\beta_i^2})\cos\beta_i} e^{-\beta_i^2(T-\tau)} u(\tau)d\tau \tag{5.3.7}$$

where β_i's are the roots of the transcendental equation $\beta\tan\beta = \alpha$. Therefore, $K(x,T,\tau)$ and $q_o(x)$ will be given by

$$K(x,T,\tau) = 2\sum_{i=1}^{\infty} \frac{\cos((1-x)\beta_i)}{(\frac{1}{\alpha} + \frac{1+\alpha}{\beta_i^2})\cos\beta_i} e^{-\beta_i^2(T-\tau)} \tag{5.3.8}$$

and

$$q_o(x) = 0. \tag{5.3.9}$$

If we take $q_d(x) = q_d$ = constant, it follows from (3.2.20) that

$$q^*(x) = -q_d. \tag{5.3.10}$$

Also, it can be shown using (3.2.13) that

$$K_\ell(x) = 2 \sum_{i=1}^{\infty} \frac{e^{-\beta_i^2 T}\cos((1-x)\beta_i)}{(\frac{\beta_i^2}{\alpha} + 1 + \alpha)\cos\beta_i} \{e^{(\ell+1)\beta_i^2 T_N}(1-e^{-\beta_i^2 T_N})\}. \tag{5.3.11}$$

Again, by denoting the matrix $\underline{B}^*\underline{B}$ by $\underline{C}$, the element C_{ij} of the matrix C will be given by

$$C_{ij} = 2 \sum_{\ell=1}^{\infty} \frac{e^{-2\beta_\ell^2 T}(1 + \alpha \{\frac{\cos\beta_\ell}{\beta_\ell}\}^2)}{\cos^2(\beta_\ell)(1 + \alpha + \frac{\beta_\ell^2}{\alpha})^2} (1 - e^{-\beta_\ell^2 T_N})e^{(i+j)\beta_\ell^2 T_N}. \tag{5.3.12}$$

Using equation (5.3.10) and the definition of $\underline{B}^*$ as given by (3.3.1), it can be shown that an element $(\underline{B}^* q^*)_j$ of the vector $\underline{B}^* q^*$ will have the form

$$(\underline{B}^* q^*)_j = 2q_d\alpha \sum_{i=1}^{\infty} \frac{e^{-\beta_i^2 T}}{\beta_i^2(1 + \alpha + \frac{\beta_i^2}{\alpha})^2} \{e^{j\beta_i^2 T_N}(1 - e^{-\beta_i^2 T_N})\}$$

$$j = 1,\ldots,N \tag{5.3.13}$$

Taking T = 0.4, α = 10, N = 10, q_d = 0.2 and E = 0.06, the following results were obtained: Lagrange multiplier = 0.456×10^{-3}. The optimal control vector corresponding to this Lagrange multiplier is given in Table 5.4.

TABLE 5.4

OPTIMAL CONTROL VECTOR

Sampling Period	Control Function $\underline{u}$
1	0.367
2	0.397
3	0.426
4	0.455
5	0.476
6	0.481
7	0.447
8	0.334
9	0.097
10	0.194

Seven iterations were required to compute the Lagrange multiplier to an accuracy of the order of 10^{-4}. The final temperature distribution along the slab corresponding to the optimal control vector was calculated and is given in Table 5.5.

TABLE 5.5

FINAL TEMPERATURE DISTRIBUTION ALONG THE SLAB

x	Temperature
0	0.188
0.1	0.199
0.2	0.201
0.3	0.202
0.4	0.202
0.5	0.201
0.6	0.197
0.7	0.193
0.8	0.189
0.9	0.186
1.0	0.185

The computations were performed on the IBM 360/50 digital computer of the University of Calgary.

6. CONCLUSIONS AND AREAS FOR FURTHER RESEARCH

6.1 CONCLUSIONS

The optimal control of linear distributed parameter systems, which are representable by a linear vector integral equation, has been discussed. The problem of minimizing the mean squared-error between specified desired final state functions and the actual state functions at a prescribed final time, subject to an energy constraint on the control actions, has been solved using a necessary and sufficient condition from functional analysis. An algorithm, based on Newton's method, has been derived to compute the Lagrange multiplier λ for the case where the control vector lies on the boundary of the constrained region. The computation needed for this algorithm is relatively small.

The convergence of the objective function corresponding to the optimal discrete control vector to its corresponding optimal one when the control vector is not restricted to be discrete in time has been proved. Moreover, the convergence of the optimal discrete control vector to the optimal measurable control vector has been established for the case where the linear bounded transformation A^*A is positive definite. These convergence properties are important, since in solving the nondiscrete optimization problem corresponding to the same objective function considered in this thesis, one faces the difficulty of having to solve a vector Fredholm integral equation (see Weigand[1]). To obtain a numerical solution for the Fredholm integral equation, an approximation technique must be used. The convergence properties of the discrete problem provide us with an alternative way of obtaining an approximate solution for the nondiscrete problem, namely, by solving the corresponding discrete problem. The solution of the later problem is relatively easy, since it only involves the solution of matrix equations.

6.2 SUGGESTIONS FOR FURTHER RESEARCH

The work done in this thesis can be extended by restricting the control vector to belong to a closed convex set. In this case, it will not be possible to obtain an explicit equation whose solution yields the optimal control vector. An approach for solving this problem could be to try to use the necessary and sufficient condition for optimality (4.4.3) to derive an algorithm for computing the optimal control vector.

For partial differential equations, it is not always easy to obtain an exact solution. Thus, ultimately we usually seek to obtain some sort of an approximate solution. It would be interesting to investigate the possibility of convergence of the optimality problem corresponding to the approximate solution to the same problem if we use the exact solution instead. It is also worth determining under what conditions will the optimal control problem, with discrete control vector and using the approximate solution of the partial differential equation, converge to the same problem without any approximation or restriction on the control vector to be discrete in time.

7. REFERENCES

1. W.A. Weigand and A.F. D'Souza, "Optimal control of linear distributed parameter system with constrained inputs," Trans. of ASME, J. of Basic Engineering, June 1969, pp. 161-167.

2. A.G. Butkovskii and A. Ya. Lerner, "The optimal control of systems with distributed parameters," Automation and Remote Control, Vol. 21, No. 6, June 1960, pp. 472-477.

3. A.G. Butkovskii, "Optimum processes in systems with distributed parameters," Automation and Remote Control, Vol. 22, No. 1, January 1961, pp. 13-21.

4. A.G. Butkovskii, "The maximum principle for optimum systems with distributed parameters," Automation and Remote Control, Vol. 22, No. 10, October 1961, pp. 1156-1169.

5. A.G. Butkovskii, "Some approximate methods for solving problems of optimal control of distributed parameter systems," Automation and Remote Control, Vol. 22, No. 12, December 1961, pp. 1429-1438.

6. A.G. Butkovskii, "The broadened principle of the maximum for optimal control problems," Automation and Remote Control, Vol. 24, No. 3, March 1963, pp. 292-304.

7. A.G. Butkovskii, "The method of moment in the theory of optimal control of systems with distributed parameters," Automation and Remote Control, Vol. 24, No. 9, September 1963, pp. 1106-1113.

8. N.I. Aheizer and M. Krein, "Some questions in the theory of moments," Article 4, Amer. Math. Soc. Publication, 1962.

9. A.G. Butkovskii and L.N. Poltavskii, "Optimal control of a distributed oscillatory system," Automation and Remote Control, Vol. 26, No. 11, November 1965, pp. 1835-1848.

10. A.G. Butkovskii and L.N. Poltavskii, "Optimal control of two dimensional distributed oscillatory system," Automation and Remote Control, Vol. 27, No. 4, April 1966, pp. 553-563.

11. A.G. Butkovskii and L.N. Poltavskii, "Optimal control of a wave processes," Automation and Remote Control, Vol. 27, No. 9, September 1966, pp. 1542-1547.

12. Yu. V. Egorov, "Some problems in the theory of optimal control," U.S.S.R. Computational Mathematics and Mathematical Physics, Vol. 3, No. 5, 1963, pp. 1209-1232.

13. A.I. Egorov, "Optimality conditions for systems containing distributed parameter components," Soviet Mat. Dokl., Vol. 7, No. 6, 1966, pp. 1629-1633.

14. A.G. Butkovskii, A.I. Egorov and K.A. Lurie, "Optimal control of distributed systems (a survey of Soviet publications)," SIAM J. Control, Vol. 6, No. 3, 1968, pp. 437-475.

15. P.K.C. Wang and F. Tung, "Optimum control of distributed parameter systems," Trans. of ASME, J. of Basic Engineering, March 1964, pp. 67-79.

16. P.K.C. Wang, "Control of distributed parameter systems," Advances in Control Systems, Edited by C.T. Leondes, Academic Press, Inc., New York, Volume 1, 1964.

17. P.K.C. Wang, "Control of a distributed parameter system with a free boundary," Int. J. Control, Vol. 5, No. 4, 1967, pp. 317-329.

18. Y. Sakawa, "Solution of an optimal control problem in a distributed parameter system," IEEE Trans. on Automatic Control, Vol. AC-9, No. 4, October 1964, pp. 420-426.

19. Y. Sakawa, "Optimal control of a certain type of linear distributed parameter system," IEEE Trans. on Automatic Control, Vol. AC-11, No. 1, January 1966, pp. 35-41.

20. E.I. Axelband, "Function space methods for the optimum control of a class of distributed parameter systems," Proceedings of the Joint Automatic Control Conference, Rensselaer Polytechnic Institute, Troy, New York, 1965, pp. 374-380.

21. E.I. Axelband, "An approximation technique for the optimum control of linear distributed parameter systems with bounded inputs," IEEE Trans. on Automatic Control, Vol. AC-11, No. 1, January 1966, pp. 42-45.

22. M. Kim and H. Erzberger, "On the design of optimum distributed parameter system with boundary control function," IEEE Trans. on Automatic Control, Vol. AC-12, No. 1, February 1967, pp. 22-28.

23. M. Vidyasagar, "Optimal control of a distributed parameter system with quadratic constraint," Proceedings of the Joint Automatic Control Conference, Boulder, Colorado, August 1969, pp. 694-702.

24. R.M. Goldwyn, K.P. Sririams and M. Graham, "Time optimal control of a linear diffusion process," J. SIAM Control, Vol. 5, No. 2, 1967, pp. 295-308.

25. L.I. Gal'chuk, "Optimal control of systems described by parabolic equations," J. SIAM Control, Vol. 7, No. 4, 1969, pp. 546-558.

26. A.V. Balakrishnan, "An operator theoretic formulation of a class of control problems and a steepest descent method of solution," J. SIAM Control, Vol. 1, No. 2, 1963. pp. 109-127.

27. A.C. Robinson, "A survey of optimal control of distributed parameter systems," Automatica, Vol. 7, No. 3, May 1971, pp. 371-388.

28. J.L. Lions, "Optimal control of distributed parameter systems," Survey Paper, Proc. IFAC Symposium on the Control of Distributed Parameter Systems, Banff, June 1971.

29. V. Lorchirachoonkul and D. Pierre, "Optimal control of multivariable distributed parameter systems through linear programming," Proceedings of 1967 Joint Automatic Control Conference, Philadelphia, Penn., pp. 702-710.

30. J. Matsumoto and Kito, "Feedback control of distributed parameter systems with spatially concentrated controls," Int. J. Control, 1970, Vol. 12, No. 3, pp. 401-419.

31. M.A. Hassan and K.O. Solberg, "Discrete time control of linear distributed parameter systems," Automatica, Vol. 6, May 1970, pp. 409-417, Pergoman Press.

32. G. Bachman and L. Narici, <u>Functional Analysis</u>, Academic Press, Inc., New York, 1968.

33. D.G. Luenberger, <u>Optimization By Vector Space Methods</u>, John Wiley & Sons, Inc., 1969.

34. L.V. Kantorovich and G.P. Akilov, <u>Functional Analysis in Normed Spaces</u>, Pergoman Press, Inc., New York, 1964.

35. O.L. Mangasarian, <u>Nonlinear Programming</u>, McGraw-Hill, Inc., New York, 1969.

36. B.M. Budak, E.M. Berkovich and E.N. Solov'eva, "Difference approximation in optimal control problems," J. SIAM Control, Vol. 7, No. 1, February 1969, pp. 18-31.

37. J. Cullum, "Discrete approximations to continuous optimal control problems," J. SIAM Control, Vol. 7, No. 1, February 1969, pp. 32-49.

38. K. Malanowski, "On optimal control of the vibrating string," J. SIAM Control, Vol. 7, No. 2, May 1969, pp. 260-271.

39. A.N. Tychonov and A.A. Samarski, Partial Differential Equations of Mathematical Physics, Volume 1, Holden Day, Inc., San Francisco, 1964.

40. A.E. Taylor, Introduction to Functional Analysis, John Wiley & Sons, Inc., New York, 1967.

APPENDIX

COMPLETE CONTINUITY OF THE OPERATOR A

We are going to show that A is a completely continuous transformation by showing that A^*A is completely continuous (ref. 32, p 373).

From (3.2.6), the linear transformation A is defined by

$$A\,\underline{u} = \int_0^T \underline{K}(\underline{x},T,\tau)\underline{u}(\tau)d\tau. \tag{A.1}$$

Applying the adjoint operator of A on (A.1), we obtain

$$A^*A\,\underline{u} = \int_0^T \{\int_\Omega \underline{K}'(\underline{x},T,t)\underline{K}(\underline{x},T,\tau)d\underline{x}\}u(\tau)d\tau. \tag{A.2}$$

From (3.2.7) and (3.2.8) it follows that

$$A^*A\,\underline{u} = \int_0^T \underline{\phi}(t,\tau)\underline{u}(\tau)d\tau. \tag{A.3}$$

The complete continuity of A^*A follows directly from (A.3) and assumption (3.2.9). For proof see Taylor (ref. 40, p 77).