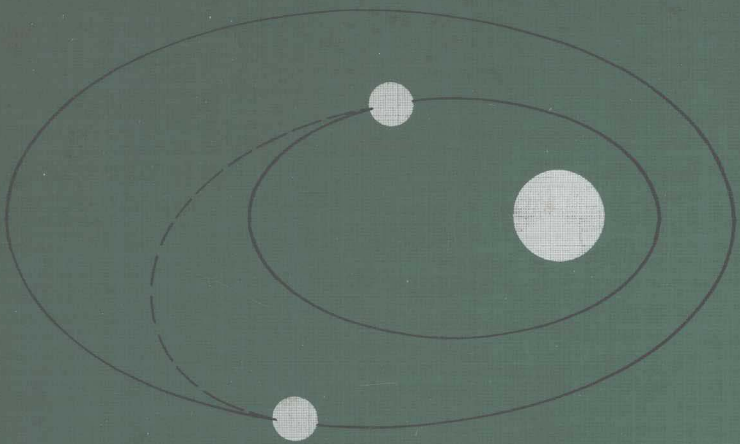

numerical optimisation of dynamic systems

edited by

l. c. w. dixon and g. p. szegö

north-holland

NUMERICAL OPTIMISATION OF DYNAMIC SYSTEMS

Edited by

L. C. W. DIXON
The Numerical Optimisation Centre
School of Information Science
The Hatfield Polytechnic
Hatfield, U.K.

and

G. P. SZEGÖ
University of Bergamo
Bergamo, Italy

1980

NORTH-HOLLAND PUBLISHING COMPANY
AMSTERDAM · NEW YORK · OXFORD

ISBN: 0 444 85494 0

Publishers:

NORTH-HOLLAND PUBLISHING COMPANY
AMSTERDAM • NEW YORK • OXFORD

Sole distributors for the U.S.A. and Canada:
ELSEVIER NORTH-HOLLAND, INC.
52 VANDERBILT AVENUE,
NEW YORK, N.Y. 10017

Library of Congress Cataloging in Publication Data

Main entry under title:

Numerical optimisation of dynamic systems.

1. Mathematical optimization. 2. System analysis.
3. Nonlinear programming. I. Dixon, Laurence Charles
Ward. II. Szegö, G. P.
QA402.5.N86 519.7 80-14255
ISBN 0-444-85494-0

PRINTED IN THE NETHERLANDS

NUMERICAL OPTIMISATION OF DYNAMIC SYSTEMS

PREFACE

It has become evident during the last ten years that more and more of the vital problems facing mankind can be posed in a common mathematical format - namely that of dynamic optimisation. This is as true in the areas of energy usage, industrial investment, home insulation, ecologic systems, econometric and sociological models, as it is in the areas of interplanetary travel. Each of these presents first the problem of model identification and second the problem of optimal decision taking. As resources of all type become scarse it is essential that both problems can be solved efficiently. Their solution must be undertaken numerically on a digital computer as they are far too complex for any other approach. The task challenging us in the next few years can be broadly described as "mastering the complexity" of these problems. The papers in this volume are presented as a contribution to the second of these two problems, a companion volume dealing with the former problem is to be edited by Professors Archetti and Cugiani of the University of Milan.

The papers in this volume are divided into three parts preceded by an introductory paper by the editors. The first part consists of nine papers presenting new ideas and applications in the dynamic optimisation field. As most algorithms for dynamic optimisation are developments of ideas first proposed for solving nonlinear programming problems, the second part consists of seven papers detailing new results in that area which are significant for possible future developments in dynamic optimisation. The third section contains four papers devoted to the theory and application of sensitivity analysis. Little work has been published on the sensitivity analysis of dynamic optimisation problems but it is acknowledged that no solution to any optimisation problem is technically feasible until the sensitivity of the solution to the data and of the objective function value to errors in implementation have been investigated. It was therefore felt to be desirable to include a section in this book devoted to that problem.

L.C.W. Dixon & G.P. Szegö

CONTENTS

PART ONE

DYNAMIC OPTIMISATION

Numerical Optimisation of Dynamic Systems
L.C.W. Dixon and G.P. Szegö (eds.)

THE NUMERICAL OPTIMISATION OF DYNAMIC SYSTEMS: A SURVEY

L.C.W. Dixon
Numerical Optimisation Centre
The Hatfield Polytechnic
England

and

G.P. Szegö
The University of Bergamo
Italy

In this paper we introduce the subject of dynamic optimisation by first indicating some of the problems that have been solved in this area, second describing the historic development of the subject, then the numerical methods of solution and finally introducing briefly the papers of this volume.

INTRODUCTION

Over the last twenty years it has become increasingly recognised that many very important practical problems in different application areas can be posed in a standard mathematical form termed the dynamic optimisation or optimal control problem. Dynamic optimisation problems consist of three elements:- the objective function, state or dynamic relationships, and constraints.

The problem is that of finding the values of decision (or control) variables, that optimise the objective function and satisfy the constraints. The state relationships which provide the structure of the problem, can be viewed as the main constraints of the problem, and provide the connection between the decision and state variables. The constraints consist of additional equations or inequalities which must be satisfied by the decision and state variables. The objective function may depend on both sets of variables, though this double dependence will often be disguised in the formulation of the problem.

The static version of the problem is normally termed the linear or nonlinear programming problem, depending on the nature of the objective function and constraints. In such problems the objective is usually a differential function of the optimisation variables. In contrast, in dynamic optimisation the objective function is usually a functional and frequently a time integral along the trajectory. The problem is still not complete until the initial conditions on the state variables are given and the final desired state specified.

The mathematical structure of this problem is the following:-

Problem 1.1; The optimal control problem

Identify the vector function u(t) belonging to a given class U which minimises the functional

$$X_o = \int_{t_o}^{t_f} f_o\ (x(t),\ u(t),\ t)dt \tag{1.1.1}$$

if such a nector exists. In 1.1.1 the state vector x(t) and the control vector u(t) satisfy the differential constraint

$$\dot{X} = f\ (x(t),\ u(t),\ t) \tag{1.1.2}$$

The given initial condition can be written as

$$X(t_o) = X_o \tag{1.1.3}$$

and the desired final constraints as

$$\psi(X_{t_F}) = 0. \tag{1.1.4}$$

Additional types of constraints will be introduced in Section 3.

In the above $X \in R^n$, $u \in R^m$, $f_o : R^{n+m+1} \rightarrow R^1$

and $f : R^{n+m+1} \quad R^{n+1}$.

The system 1.1.1 → 1.1.3 is one of the simplest types of dynamic optimisation problem. If the simple differential constraint 1.1.2 is replaced by an integro-differential equation, a delayed differential equation or even a partial differential equation then the mathematical treatment becomes more complex. The cases when 1.1.2 is replaced by difference equations or stochastic differential equations have received considerable attention. In all cases however, the basic structure of the problem remains unchanged.

One of the oldest examples of a well-posed optimal control problem will be used to clarify the structure of such problems and, in particular the terminal conditions that can occur. Consider the "soft landing" situation, i.e. the task of sending a satellite from the earth to the moon in such a way as to minimise the energy required (the objective function). The soft landing condition implies that at the instant of landing the lander must not only be in the desired position on the moons surface, but also that its relative velocity with respect to the moon is zero at that instant. In this problem the control function consists of the times at which the various rockets are fired and possibly of the orientation of the engines; the dynamic equations relate the state variables (possibly the position of the satellite and its higher derivatives) to the control function and the constraints impose bounds on the thrust level of the engines and on their maximum displacement.

In most cases the solution of the optimal control problem (1.1) can only be obtained numerically with the aid of a digital computer. Digital computers have become more widely available and their cost has reduced compared with the size and importance of the practical problems which can be solved. Thus more and more practical tasks have been formulated as dynamic optimisation problems and then numerically solved.

The solution of the optimal control problem, the vector u, can either be expressed as a function of time u(t), which is termed open-loop control, or alternatively as a function of the state variables u(X), which is termed closed-loop control. From a mathematical point of view there is no difference between the two types of solutions, the connection being given by the solution of the state equations X(t). In our example of the soft landing problem the open-loop solution will be a set of instructions giving the exact time of firing of the various engines and of orienting their thrust in different ways. The closed-loop control will, in contrast, give the firing and thrust instructions in terms of the state, i.e. position and velocity of the lander. The difference in the two types of solution becomes important when the actual behaviour of the system is not exactly that predicted by the state equations, due to possible disturbances of various types not included in the model. In this case there is often a considerable practical difference in the performance of the open-loop and closed-loop optimal controls.

Without attempting a comprehensive survey we will mention a number of major achievements made possible by dynamic optimisation techniques. Of these the most spectacular must surely be the successful Space programmes of the U.S.A. and U.S.S.R. governments. These included so many major achievements that it is difficult to know which to mention - the first artificial satellite - the landing on the moon and Mars - the probing of Venus' atmosphere - the photographs of Jupiter and Saturn during the Pioneer missions - the meansurements of the sun's electromagnetic field and the solar wind - the network of weather stations and the telephone and T.V. satellites. Each in its own way has revolutionized our understanding of the universe.

It is worth pointing out that each of these tasks entailed solving many optimal control problems. In the above space missions we can usually identify at least the following three dynamic optimisation problems:-

(1) the computation of the optimal trajectory and the firing thrust and orientation plan which enables that trajectory to be followed
(2) the attitude control problem, i.e. the firing and thrust orientation schedule to ensure the spacecraft maintains a desired aspect in space, which is difficult in the absence of air damping and
(3) the launch and landing strategies where both attitude and trajectory control

must be handled at the same time and many other safety factors must be considered.

On the industrial front many of the early achievements of dynamics optimisation were made by Chemical Engineers; who were interested in first modelling complex chemical processes and then modifying them to achieve the desired result. In this era of fuel crisis we should not forget that our oil refineries are very sophisticated dynamic systems that have been tuned over the years to produce highly efficient low cost fuels. One of the major difficulties in introducing a new fuel system would lie in catching up with all the experience that has gone into optimising the present system.

The oil industry is, of course, not the only example of the use of optimisation in the energy field. Many of the early optimisation codes were developed to solve the problem of distributing electricity safely and cheaply - the cost being the objective function and the safety requirements introducing additional constraints. Similarly the design of hydroelectric systems on the rivers in France was an early successful application of dynamic optimisation techniques.

More recently governments faced with the shortage of energy have commenced to model the complete energy systems of nations, in an effort to predict the effects of sudden changes in fuel prices and to optimise the investment in new equipment. Early results with these models enphasize once again the fact that predictions obtained by such models are highly dependent on accuracy of the model. If important relations are left out or modeled inaccurately, or certain possibilities not included, then very misleading forecasts can occur.

For this reason in any such excersise the complementary problem of identifying the structure of the system and then the parameter values, must be investigated thoroughly, before any dynamic optimisation can be attempted. Mathematically the functions $f_i(X, u)$ must be known before the problem can be solved. In many problems this identification stage is by far the more difficult, though frequently this can itself be posed as a dynamic optimisation problem in terms of likelihood functions.

In the econometric and ecologic areas the problem is still usually to understand the system. Whilst models of the economy of many countries have been built, most econometricists are more concerned with investigating their accuracy and introducing perturbations into the models and studying their effects, rather than attempting to optimise among them. Indeed it would be difficult, if not impossible, to obtain agreement on an objective function for a national economy and many economists claim, with good reason, that it is impossible to construct a utility function (i.e. objective function in an uncertain world) for any group of individuals.

Smaller scale economic problems, such as investment, allocation and portfolio selection under uncertainty lead to better defined optimisation problems. These problems can be formulated as stochastic optimal control problems and in many cases an analytic solution can be obtained. From this analytic solution the whole theoretic framework of these problems can be better understood and the numerical solutions obtained for more difficult problems of the same nature better appreciated.

A similar situation arises in the theory of growth, the branch of economic theory which investigates the optimal policy for economic growth given limited resources. In this area too the basic result has been an analytic solution around which the whole field of Hamiltonian economic systems has been developed, Shell (1967), Kuhn & Szegö (1968) and Cass & Shell (1977).

Quite recently the techniques of stochastic optimisation have been applied to a wide range of financial problems, including the portfolio selection problem, the theory of rational option policy, the theory of multicurrency investment selection, the analysis of demand of indexed bonds, the analysis of optimal long term investment under uncertainty and finally the basic problem of the optimal financial structure of a firm. A complete account of these developments can be found in the recent paper by R.C. Merton (1977).

The dynamic models of how our industrial society interacts with the ecological system are even less well understood than the large econometric systems and must be considered a major area for future research. The great achievements that can be expected to result will be easily appreciated by Londoners who have witnessed the end of winter smogs and by people in Northern England who have seen the dramatic improvement in the quality of water in some of the rivers.

The current energy crisis will imply the solution of many more dynamic optimisation problems in order to conserve energy. A typical important problem of this type recently solved has been the optimisation of the water recourses for hydroelectric plants. This problem will become of increasing importance with the advent of nuclear and/or solar power plants that produce power at predeterminal rate (or time) and must be integrated into the electric power network, thus requiring an efficient technique for power storage. This can be provided by hydroelectric plants as the surplus power of nonconventionally controlled resources can be used to pump water into the upper basins, where it can then be used to generate power in peak demand periods.

On a smaller scale one of the revolutions in our lifestyle in the next twenty years will stem from the introduction of microprocessors to control everyday activities in house and industry. Most of the problems to be solved will be expressible as dynamic optimisation problems and their solution should ease the load of many humdrum tasks. As a case in point we may mention the possibility of installing a microprocessor into a car to select the type of fuel mixture to be used at each given designed speed and acceleration.

Having summarised so many interesting and important applications of dynamic optimisation we hope we have motivated the reader to wish to read more about the solution of the problem numerically on a digital computer. Having emphasized that so many problems have been successfully investigated it may seem strange to state that the algorithms used in their solution are mainly not those discussed in this volume. The reason for this lies in the fact that methodology for solving the dynamic optimisation problem, is itself dynamic and new more efficient algorithms continue to be suggested. The main purpose of this volume is to outline some of these new approaches.

2. THE EARLY HISTORICAL DEVELOPMENT

It is very difficult to trace the first attempts to solve the dynamic optimisation problem or to assess the preliminary formulations of more recent times. From one point of view the mathematical problem could be viewed as having been solved in the nineteen thirties by the contributions of the Chicago school of the calculus of variations.

One of the early statements of an applied problem in the dynamic optimisation framework is given in Flugge-Lotz (1953), in which many earlier results published by that author in German from 1943 to 1948 are also presented. In this monograph the state equations are assumed to be linear and hence the control discontinuous and for this problem the standard terms of optimal control theory such as switching points, starting points and end points for discontinuous solutions are introduced and the phenomenon of "chattering" described, in particular the case of an increasing switching frequency leading as $t \to \infty$ to a point in two-dimensional phase space which is not the desired equilibrium point. Although no real optimal solution is given, according to the author "the theory is sufficiently developed to allow the design of discontinuous control systems with optimum efficiency".

Further similar publications appeared in the same year (1953) by Bushaw [6], [7] and La Salle [8], and further mathematical refinements and results appeared in Bellman, Glicksberg and Gross [9], where the problem is formulated exactly as a minimum time control problem and a complete solution provided for the linear case. At approximately the same date the works by R.W. Gamkrelidze (1957-58) and N.N. Krasovskii (1957) appeared mostly devoted to the time optimal control problem.

A complete formulation of the optimal control problem and a critical discussion of the previous results is contained in the monograph by H. Tsien [13], where for the case n = 2 the optimal control problem is formulated in its full generality (see section 14.7 page 212) and correctly regarded as a calculus of variations problem with a differential equation (the state equation) as an auxilliary condition. The problem is treated there as a "problem of Bolza" solvable through the Euler-Lagrange technique, but it is recognised by Professor Tsien that no analytical solution can be achieved and a numerical approach is proposed.

A complete presentation of the optimal control theory via the classical results of the calculus of variations was presented in the papers of Berkowitz (1961), who referred in particular to the results of Valentine (1937). Valentine had investigated a problem very similar to the modern optimal control problem that even contained hard inequality constraints on the control functions. A presentation of the classical calculus of variation approach to optimal control is given in the book by L. Markus & E.B. Lee (1967) and a more extensive investigation in the volume by L.S. Pontryagin et al (1962 ch. V).

The Soviet scientists at this time investigated the mathematical aspect in great depth and rigorously proved that when the control set U is an open set, then the optimal control problem is equivalent to the Lagrange problem of the calculus of variations, but that the calculus of variations was ill-equipped to handle the case when U is a compact set. To handle this case the famed maximum principle was introduced by Acad. L. Pontryagin, the formal proof being given in V.G. Bottyanskii R.V. Gambrelidze & L.S. Pontryagin (1956). We recall that until the publication in English of the famous book by Pontryagin et al [18] in 1962 the only available papers in the West containing the detailed proof were the translations of the series of three papers by L. Rozonoer (1959). The maximum principle provides a set of necessary conditions which the optimal control $u \in U$ must satisfy if it exists. It can be stated either as a maximum or minimum principle depending on a sign conventions and takes the following form:-

2.1 The Minimum Principle

Consider the real valued function

$$H(X, p, u) = f_o(X, u) + \sum_{i=1}^{n} p_i f_i(X, u) \qquad (2.1.1)$$

and the related set of 2n differential equations

$$\dot{X}_i = \frac{\partial H}{\partial p_i} = f_i(X, u)$$

$$\dot{p}_i = -\frac{\partial H}{\partial X_i} = -\frac{\partial f_o}{\partial X_i} - \sum_{j=1}^{n} p_j \frac{\partial f_j}{\partial X_i} \qquad (2.1.2)$$

Let $u^*(t) \in U$ be such that (1.1.3) and (1.1.4) are satisfied and $X^*(t)$ be the corresponding solution of (1.1.2), then in order for u^* to be optimal it is necessary that there exists a function $p^*(t)$ which satisfies the system (2.1.2) and for which

$$H(X^*(t), p^*(t), u^*(t)) \leq H(X(t), p(t), u(t)) \quad (2.1.3)$$

for all $u \in U$ at every time instant and

$$H(X^*(t), p^*(t), u^*(t)) = 0. \quad (2.1.4)$$

In the case when the final state is not fixed in the form

$$X_i(t_F) = X_{F_i}$$

but is for instance a smooth manifold of R^n

$$\psi_i(X_t(t_F)) = 0 \qquad i = 1, \ldots, k \quad (2.1.5)$$

then define

$$M(X^*(t_F)) = \left\{ Z : \sum_j \frac{\partial \psi_i}{\partial X_j} (Z_j - X^*_j(t_F)) = 0 \quad i = 1, \ldots, k \right\} \quad (2.1.6)$$

then additional transversality conditions of the type

$$\sum_j p^*_j(t_F)\ (Z_j - X^*_j(t_F)) = 0 \quad (2.1.7)$$

must be satisfied for all $Z \in M(X^*(t_F))$.

An alternative form of the transversality condition can be obtained by introducing additional Lagrange Multipliers on the constraints (2.1.5) and the necessary condition then becomes: There exist Lagrange Multipliers μ_i such that

$$p^*_j(t_F) + \sum_i \mu_i \frac{\partial \psi_i}{\partial X_j} = 0. \quad (2.1.8)$$

Operationally the identification of an optimal control $u^*(t)$ that satisfy the above conditions requires the integration of the canonical system (2.1.2) which is not trivial.

The existence of an optimal control should of course be proved before attempting its identification and an existence theory was proposed by R.E. Kalman (1960) and termed "controllability". Kalman derived controllability conditions for linear state equations and these are extended to nonlinear systems in Lee & Markus (1967) ch. 6.

Prior to the proof of the maximum principle a totaly different approach to the optimisation of dynamic systems, termed dynamic programming was proposed by the American mathematician R. Bellman. The theory was developed at RAND Corporation in the early fifties and first presented in a systematic form in the mongraph by R. Bellman on Dynamic Programming published in 1957. While the maximum principle may be regarded as an outgrowth of the Hamiltonian approach to variational problems, the method of dynamic programming may be viewed as related to the Hamilton-Jacobi approach to variational problems. Indeed we may quote Bellman (1961,pp 66)