OPTIMAL TRAJECTORIES FOR SPACE NAVIGATION

LAWDEN

OPTIMAL TRAJECTORIES FOR SPACE NAVIGATION

D. F. LAWDEN, M.A., Sc.D.

Department of Mathematics, University of Canterbury, New Zealand

LONDON
BUTTERWORTHS
1963

PREFACE

Problems of optimizing rocket trajectories have been extensively studied during the last decade and general techniques of both analytical and numerical types have been developed for their solution. This monograph is concerned exclusively with the analytical approach and makes no reference to numerical methods such as those of 'steepest descent' (see Kelley, H. J. (1960) 'Gradient Theory of Optimal Flight Paths.' *Journal of the American Rocket Society* **30**, No. 10, 947–954) and of 'dynamic programming' (see Bellman, R. and Dreyfus, S. (1959) 'An Application of Dynamic Programming to the Determination of Optimal Satellite Trajectories'. *J. Brit. Interplanet. Soc.* **17**, Nos. 3–4, 78–83). It is true that, on account of the complex nature of the majority of practical problems in this field, recourse to numerical techniques is usually inevitable before an acceptable solution is forthcoming, but the analytical approach is none the less valuable for the following reasons: Experience teaches us that the form of an optimal trajectory is rarely, if ever, very critically dependent upon the data of a problem and consequently if, by making suitable simplifying approximations, the actual problem can be transformed into an idealized problem whose solution is analytically tractable, then this latter solution will often provide an excellent substitute for the optimal motor thrust programme in the actual situation. All that then remains to be done is to recompute the trajectory employing this programme and taking account of the real circumstances. Further, it is only by adopting the analytical approach in any field of research, that those general principles, which lead to a real understanding of the nature of the solutions, are discovered. Lacking such an appreciation, our sense of direction for the numerical attack will be defective and, as a consequence, computations will become unnecessarily lengthy or even quite ineffective. As for almost every other application of mathematics to practical affairs, therefore, the analytical and numerical approaches are here complimentary rather than alternative.

The mathematical background assumed of the reader is described in the preliminary section. This is approximately of the standard normally reached by a modern honours mathematics student after two years at the university and the book is accordingly suitable as a text for a third-year course for this class of student. Such a course, supplemented by further material from the calculus of variations, could constitute an optional alternative to some other branch of applied mathematics and might appeal to students who have developed an interest in this old-established branch of analysis which, after a period of comparative neglect, is now attracting considerable attention for its practical applications. But my principal aim has been to provide an introduction to the mathematical theory of optimal trajectories for the rapidly growing body of young space scientists in private and public research establishments both in this country and abroad who are becoming involved with astronautical calculations.

It gives me great pleasure to acknowledge the debt I owe my fellow researchers in this field, the majority of whom are resident in the United States, for the generous way in which they have exchanged ideas with an outsider in a somewhat remote country. I am conscious that the mode of development of the theory I have chosen in this book owes much to the profit I have received from this traffic. The lists of references give some indication of the sources from which many of the subject's fundamental ideas derive and I hope all those who have contributed to my own approach to the subject will find their names at appropriate points in the text. To the Council of the University of Canterbury for granting me a period of absence from normal university duties during which this monograph was written, I offer my sincere thanks. In conclusion, I should also like to express my gratitude to Professor P. J. Hilton (Cornell University) who, during his occupancy of the Chair of Pure Mathematics at Birmingham University, offered me hospitality within his department and provided me with all necessary facilities.

DEREK F. LAWDEN

DEPARTMENT OF PURE MATHEMATICS,
UNIVERSITY OF BIRMINGHAM

ENGLAND: BUTTERWORTH & CO. (PUBLISHERS) LTD.
LONDON: 88 Kingsway, W.C.2

AFRICA: BUTTERWORTH & CO. (AFRICA) LTD.
DURBAN: 33–35 Beach Grove

AUSTRALIA: BUTTERWORTH & CO. (AUSTRALIA) LTD.
SYDNEY: 6–8 O'Connell Street
MELBOURNE: 473 Bourke Street
BRISBANE: 240 Queen Street

CANADA: BUTTERWORTH & CO. (CANADA) LTD.
TORONTO: 1367 Danforth Avenue, 6

NEW ZEALAND: BUTTERWORTH & CO. (NEW ZEALAND) LTD.
WELLINGTON: 49/51 Ballance Street
AUCKLAND: 35 High Street

U.S.A.: BUTTERWORTH INC.
WASHINGTON, D.C.: 7235 Wisconsin Avenue, 14

Suggested U. D. C. number: 629.7.086

©
Butterworth & Co. (Publishers) Ltd.
1963

Printed in Great Britain by
The Camelot Press Ltd., London and Southampton

试读结束：需要全本请在线购买： www.ertongbo

CONTENTS

PRELIMINARY REMARKS

The reader of this book is assumed to be familiar with the material normally included in a first honours mathematics course concerned with the analysis of functions of real variables and with the fundamentals of the theory of differential equations. Extensive use is made of techniques and results taken from the calculus of variations, but these are all developed *ab initio* in the first chapter, so that no preliminary acquaintance with this field of mathematics is necessary. However, some background knowledge of the type of problem to which this calculus may be applied and of the results to be expected will no doubt prove helpful in following the argument of the first chapter, and the reader who wishes to obtain some familiarity with this field before proceeding further is advised to consult one of the many texts dealing with the subject, two of the best known of which are Courant and Hilbert (1953) and Fox (1950).

In addition to standard techniques for the differentiation of functions of functions and of functions defined implicitly by means of sets of equations, a theorem relating to the conditions under which a set of implicit functions is properly defined by a set of equations, plays an important role in the argument of Chapter 1 (p. 13). For a proof of this theorem, the reader is referred to Goursat (1904).

The existence of certain sets of functions has been assumed at some points in the first chapter (p. 8, p. 10, p. 13, p. 22). For a rigorous demonstration of their existence, the reader should consult the standard work in the English language on the calculus of variations, *viz.* Bliss (1946). At appropriate points in the argument, this book is referred to by the author's name followed by a page number.

A knowledge of the Newtonian mechanics of a particle, including the basic formulae relating to the motion of a particle in an inverse square law field of attractive force, is also assumed to be possessed by the reader. All such formulae will be found derived in the book by Lawden (1961). Wherever necessary, this book is referred to as 'Lawden', and a page number is stated.

The reader who requires further information bearing on the subject of this monograph should consult the comprehensive bibliography to be found at the close of the survey article by Leitmann, (1962).

REFERENCES

Bliss, G. A. (1946). *Lectures on the Calculus of Variations*. Chicago, University of Chicago Press.

Courant, R. and Hilbert, D. (1953). *Methods of Mathematical Physics*, Volume 1, Chap. 4. New York, Interscience.

Fox, C. (1950). *An Introduction to the Calculus of Variations*. London, Oxford University Press.

Goursat, E. (1904). *A Course in Mathematical Analysis*, Vol. 1, p. 45. Boston, Ginn and Co.

Lawden, D. F. (1961). *A Course in Applied Mathematics*, Vol. 1, p. 215. London, English Universities Press.

Leitmann, G. (1962). 'The Optimization of Rocket Trajectories—A Survey,' *Progress in the Astronautical Sciences 1*, Chap. 4. Amsterdam, North-Holland Publishing Co.

1

THE PROBLEM OF MAYER

1.1. Introduction

This book concerns itself with the problem of finding the trajectory a rocket must follow if it is to accomplish some specified mission in an optimal manner as judged against some criterion of a quantitative nature. The mission may be a military one as, for example, when an intercontinental missile is launched with the object of transporting an atomic bomb payload to a given target area; or the mission may have a scientific object, as when a rocket probe, carrying a payload of observing instruments, is to be guided into an orbit about the Moon or a planet. In the not too distant future, it is to be expected that the mission will frequently have for its object the transport of a human cargo to another body in the solar system or beyond. In most cases, optimization of the trajectory with respect to propellent expenditure will be desired, since this will permit the largest payload to be delivered for a given size of vehicle and we shall accordingly pay particular attention to this case.

However, although economy of propellent expenditure will often be the most pressing requirement, circumstances in which allowance must also be made in the optimization criterion for other factors, are easily envisaged. For example, in the case of a spaceship transporting a human cargo, the quantities of food and supplies which have to be carried will be proportional to the time of transit between the two terminals for the journey and, for this reason, it may prove desirable to reduce this time at the cost of some additional expenditure of propellent, with the ultimate object of minimizing the overall weight of the vehicle. The criterion for optimization will then involve a combination of the two quantities, mass of propellent and time of transit. The techniques we shall develop will be applicable to optimization problems of this more complex nature, but we have thought it advisable, in an introductory work such as this monograph is intended to be, to confine our attention to problems whose mathematical statement is simply expressed, in order that the principal features of our methods should not be obscured by irrelevant details. It should, however, be mentioned here that our analysis will reveal that the equations governing the arcs from which the optimal trajectory must be constructed are quite independent of the optimization criterion. Much of the argument, therefore, is perfectly relevant to the issue, no matter what criterion is ultimately adopted.

Although the mathematical theory of this chapter has been constructed with the express object of applying it to rocket trajectory problems, it can also be employed in many other problem situations. In particular, it has received application to the problem of optimizing the response of control systems and servomechanisms and, in this connection, reference may be made to the work of the Russian mathematician Pontryagin who has based similar techniques upon an alternative mode of development of the mathematical theory. An account in English of his approach will be found in a paper due to R. E. Kopp (1962).

1.2. Synopsis of Results

As was first noticed by Cicala and Miele (1956) and Miele (1958), the problem of optimizing a rocket trajectory is a particular case of a general mathematical problem from the calculus of variations associated with the name of Mayer. In this chapter, we state the Mayer problem in a form which differs from that given by Bliss (p. 189) in that certain parameters, termed the control functions, are permitted to enter into the constraining eqs (1.1), (1.3), whilst their derivatives are absent from these conditions. We then obtain a number of sets of conditions which are necessarily satisfied by a solution to the problem, the line of argument being that due to Bliss, modified where necessary to make allowances for the intrusion of the control functions. The boundary conditions (eqs (1.2), (1.4)) are not in the very general form given by Bliss, but are sufficiently so for our present purpose. The sets of necessary conditions will be found stated on pages 16, 19, 25 and 26.

In Chapter 2, we make our first application of the general

试读结束：需要全本请在线购买： www.ertongbo

theory to a number of problems relating to terrestrial rocket trajectories, *viz.* maximization of the range of a rocket missile, optimal launching of a satellite into orbit and optimization of the performance of an atmospheric sounding rocket.

The theory of optimal rocket trajectories in a general gravitational field, when it is assumed that there is no atmospheric resistance, is developed in Chapter 3. It is shown (p. 59) that the conditions to be satisfied by the required trajectory are conveniently expressed in terms of a *switching function* κ and a *primer vector* $\mathbf{p}$, κ determining the instants of transition from one motor thrust phase to the next and $\mathbf{p}$ the direction of the motor thrust. In the special, but practically important, case when the possible motor thrust is assumed to be of unlimited magnitude, it is demonstrated that the conditions can all be expressed in terms of the primer vector alone (p. 63). The cases when optimization is to be carried through with respect to propellent expenditure and the net final vehicle energy (the escape problem), are given particular consideration.

The simplification which occurs when the gravitational field is uniform is indicated in Chapter 4, and it is demonstrated that a fairly complete theory can be offered to cover this case. It is shown (p. 72) that, in general, there can be at most three distinct thrust phases and that these occur in the sequence, maximum thrust, null thrust, maximum thrust. In special circumstances, when the boundary conditions are favourable, a phase during which the motor thrust is not a maximum can occur but, in this case, the solution is not unique and a solution involving null and maximum thrust phases alone is always available (p. 71).

The extremal arcs, from which any optimal trajectory must be constructed (irrespective of the optimization criterion) when the gravitational field obeys an inverse square law of attraction to a point, are discussed in Chapter 5, and the form taken by the primer vector on such arcs is calculated. The results obtained are employed in the final chapter to analyse a number of orbital transfer problems in a plane. It is proved (p. 106) that, if the optimal trajectory contains a circular arc and if the motor thrust is unlimited, then the trajectory is formed from conic arcs with their axes all aligned and tangential to one another at their apses; the thrusts are then all impulsive and are applied at these apses in a tangential direction. In particular, the case of transfer between two circular orbits is discussed in detail and the Hohmann mode of transfer via a single cotangential ellipse is proved to be optimal provided the ratio of their radii is not too great (p. 110). Optimal escape from any orbit is proved (p. 111) to be effected by the application of impulsive thrusts alone and the escape trajectory to consist of conic arcs with their axes aligned and tangential to one another at their apses. Finally, the optimal two-impulse transfer manœuvre between any two coplanar orbits is studied in some detail and a set of equations derived (p. 118) from which the elements of the transfer orbit can be calculated in any particular case.

1.3. Statement of the Mayer Problem

The calculus of variations is concerned with the problem of minimizing or maximizing functionals, a *functional* being a quantity whose value depends upon the sets of values taken by certain associated functions over domains of their variables for which they are defined. Thus, the quantity I defined by the equation

$$I = \int_0^1 f(x)\,\mathrm{d}x$$

is a functional, for its value depends upon the values assumed by the function $f(x)$ over the interval $0 \leqslant x \leqslant 1$. Clearly, a functional is a mathematical entity which has a more complex nature than a function, since it depends for its value, not on the values taken by a finite set of variables, but on the, in general, infinite set of values assumed by a function over its domain of definition.

The problem of Mayer which we are about to formulate relates to the minimization of a quantity which depends upon a number of functionals, the values of which are related to the forms taken by certain unknown functions occurring as parameters in a given set of differential equations. Specifically, given m functions $a_j(t)$ $(j=1, 2, \ldots, m)$, n further functions $x_i(t)$ $(i=1, 2, \ldots, n)$ are to satisfy differential equations of the form

$$\dot{x}_i = f_i(x_1, \ldots, x_n, a_1, \ldots, a_m, t), \tag{1.1}$$

where, as usual, $\dot{x}_i$ denotes $\mathrm{d}x_i/\mathrm{d}t$. These equations are to be valid for $t_0 \leqslant t \leqslant t_1$ and the α_j are defined throughout this interval as continuous functions apart from a finite number of finite discontinuities. The functions f_i are continuous in all their arguments and possess continuous partial derivatives of an order sufficient to validate all our subsequent arguments and these derivatives are defined over a region sufficiently extensive to include all values of the x_i, α_j and t to be encountered, as interior points. The initial values of the x_i at $t=t_0$ are specified by the equations

$$x_i = x_{i0} \tag{1.2}$$

and it follows that the eqs (1.1) determine the functions $x_i(t)$ uniquely over the interval (t_0, t_1) when the $\alpha_j(t)$ are given. The x_i so determined will be continuous functions of t, but their derivatives may be discontinuous at points of discontinuity of the α_j.

The functions $\alpha_j(t)$ will be termed the *control functions* and the $x_i(t)$ will be referred to as the *state functions*; they are required to satisfy certain constraining equations, *viz.*

$$g_k(x_1, \ldots, x_n, \alpha_1, \ldots, \alpha_m, t) = 0, \tag{1.3}$$

where $k=1, 2, \ldots, p<m$ and the g_k are continuous and possess continuous partial derivatives of sufficiently high order in all their arguments. The control functions are also required to be such that the functions $x_1, x_2, \ldots, x_q$ take prescribed values at $t=t_1$. Thus, for $t=t_1$,

$$x_l = x_{l1}, \quad l=1, 2, \ldots, q. \tag{1.4}$$

and $q \leqslant n$. The control functions can otherwise be chosen arbitrarily. The existence of control functions satisfying the constraints (1.3), (1.4) will be assumed.

Let $x_{q+1,\,1}, x_{q+2,\,1}, \ldots, x_{n1}$ be the values of the x_i at $t=t_1$ not fixed by the constraints (1.4). Then our problem is to find control functions α_j determining the x_i so that the constraints (1.3), (1.4) are satisfied and also so that a given function

$$J(x_{q+1,\,1}, x_{q+2,\,1}, \ldots, x_{n1}) \tag{1.5}$$

is minimized.

The problem can be generalized by permitting t_1 to be variable.

In this case, we shall permit J to depend upon this quantity also and write

$$J = J(x_{q+1,\,1}, \ldots, x_{n1}, t_1). \tag{1.6}$$

J is then to be minimized with respect to t_1 also.

J will be supposed continuous in all its variables and to possess continuous partial derivatives of sufficiently high order.

1.4. Admissible Variations

In general, there will exist infinitely many sets of functions $x_i(t)$, $\alpha_j(t)$ satisfying the eqs (1.1)–(1.4) and to each set there will correspond a value of J. Among these sets, we shall suppose that there is one which generates a minimum value for J. This *minimal set* will henceforward be denoted by $x_i(t)$, $\alpha_j(t)$.

However it will be convenient first to consider a wider class of sets of functions for which J is defined and including the minimal set as a member. This is the class of *admissible sets* comprising those functions which satisfy the constraints (1.1), (1.3), but do not necessarily satisfy the end conditions (1.2), (1.4). It will be assumed that a one-parameter family of such admissible sets can be found including the minimal set as one of its members. This sub-class of admissible sets will be denoted by $x_i(t, \epsilon)$, $\alpha_j(t, \epsilon)$, where ϵ is the parameter and $\epsilon = 0$ corresponds to the minimal set. Thus

$$x_i(t, 0) = x_i(t), \quad \alpha_j(t, 0) = \alpha_j(t). \tag{1.7}$$

The functions $x_i(t, \epsilon)$, $\dot{x}_i(t, \epsilon)$ and $\alpha_j(t, \epsilon)$ will be supposed to possess continuous first derivatives with respect to ϵ, for ϵ satisfying $|\epsilon| < \epsilon_0$ and for t in the interval (t_0, t_1); these functions will also be supposed to possess, for the same values of ϵ and t, the continuity properties relative to t enunciated in section 1.3. In the case when t_1 is variable, we shall also make this quantity dependent upon ϵ and take $t_1(0) = t_1$ to be the end value appropriate to the minimal set.

Substituting

$$x_i = x_i(t, \epsilon), \quad \alpha_j = \alpha_j(t, \epsilon) \tag{1.8}$$

into eqs (1.1), these must be satisfied identically with respect to both t and ϵ. Hence, differentiating both sides with respect to ϵ, it is found that

$$\frac{\partial^2 x_i}{\partial \epsilon\, \partial t} = \frac{\partial f_i}{\partial x_r}\frac{\partial x_r}{\partial \epsilon} + \frac{\partial f_i}{\partial \alpha_j}\frac{\partial \alpha_j}{\partial \epsilon}, \tag{1.9}$$

it being understood that the two terms of the right-hand member of this equation are to be summed with respect to r (1, 2, . . ., n) and j (1, 2, . . ., m) respectively, according to the well-known repeated index summation convention. (This summation convention will be operative throughout the book and it is convenient here to state the integral values over which the various literal subscripts employed in this chapter are supposed to range. These are as follows:

$$\left.\begin{array}{lll} i=1,2,\ldots,n; & j=1,2,\ldots,m; & k=1,2,\ldots,p; \\ l=1,2,\ldots,q; & r=1,2,\ldots,n; & s=q+1,q+2,\ldots,n; \\ & t=p+1,p+2,\ldots,m.) & \end{array}\right\} \tag{1.10}$$

Putting $\epsilon=0$ in eq (1.9), this can be written

$$\dot{y}_i = \frac{\partial f_i}{\partial x_r} y_r + \frac{\partial f_i}{\partial \alpha_j}\beta_j, \tag{1.11}$$

where

$$y_i(t) = \left(\frac{\partial x_i}{\partial \epsilon}\right)_{\epsilon=0}, \quad \beta_j(t) = \left(\frac{\partial \alpha_j}{\partial \epsilon}\right)_{\epsilon=0}, \tag{1.12}$$

and the arguments of $\partial f_i/\partial x_r$, $\partial f_i/\partial \alpha_j$ are the functions $x_i(t)$, $\alpha_j(t)$ of the minimal set. The functions y_i, β_j possess the continuity properties of the functions x_i, α_j respectively and will be termed the *variations* of the minimal set with respect to the family.

Again, substituting from eqs (1.8) into the constraints (1.3) and differentiating with respect to ϵ, it is found that

$$\frac{\partial g_k}{\partial x_i}\frac{\partial x_i}{\partial \epsilon} + \frac{\partial g_k}{\partial \alpha_j}\frac{\partial \alpha_j}{\partial \epsilon} = 0. \tag{1.13}$$

Putting $\epsilon=0$, this reduces to

$$\frac{\partial g_k}{\partial x_i} y_i + \frac{\partial g_k}{\partial \alpha_j}\beta_j = 0, \tag{1.14}$$

where the arguments of $\partial g_k/\partial x_i$, $\partial g_k/\partial \alpha_j$ again refer to the minimal set.

It has been shown, therefore, that the one-parameter family of admissible sets of functions must be such that eqs (1.11) and (1.14) are satisfied by its variations identically with respect to t. We shall assume, conversely, that if any set of functions $y_i(t)$, $\beta_j(t)$ is found to satisfy the eqs (1.11), (1.14), then a one-parameter family of admissible sets, including the minimal set, can be found, whose variations are these functions. A proof of this *imbedding theorem* will be found in Bliss (p. 196). Such a set of functions $y_i(t)$, $\beta_j(t)$ will be termed a set of *admissible variations* for the minimal set $x_i(t)$, $\alpha_j(t)$.

1.5. The First Variation of J

Constructing J (eq (1.6)) for the members of the family of admissible sets defined in the previous section, it becomes a function of the parameter ϵ. The value of $dJ/d\epsilon$ at $\epsilon=0$ will be termed the first variation of J with respect to the family and we shall write

$$\left(\frac{\mathrm{d}J}{\mathrm{d}\epsilon}\right)_{\epsilon=0}=J_1. \tag{1.15}$$

Substituting

$$x_{s1}=x_s[t_1(\epsilon),\epsilon], \quad t_1=t_1(\epsilon), \tag{1.16}$$

in eq (1.6) and differentiating with respect to ϵ, it follows that

$$\frac{\mathrm{d}J}{\mathrm{d}\epsilon}=\frac{\partial J}{\partial x_{s1}}\left(\frac{\partial x_s}{\partial t_1}\,\frac{\mathrm{d}t_1}{\mathrm{d}\epsilon}+\frac{\partial x_s}{\partial \epsilon}\right)+\frac{\partial J}{\partial t_1}\,\frac{\mathrm{d}t_1}{\mathrm{d}\epsilon}. \tag{1.17}$$

Putting $\epsilon=0$, this reduces to

$$J_1=\frac{\partial J}{\partial x_{s1}}\,(\dot{x}_{s1}u_1+y_{s1})+\frac{\partial J}{\partial t_1}u_1, \tag{1.18}$$

where

$$u_1=\left(\frac{\mathrm{d}t_1}{\mathrm{d}\epsilon}\right)_{\epsilon=0}, \tag{1.19}$$

and the arguments of the partial derivatives of J refer to the minimal set. u_1 will be called the *variation of the end point*.

We shall now transform the expression (1.18) for J_1 into a form more convenient for the subsequent argument. This is carried out by introducing certain auxiliary functions $\lambda_i(t)$,

试读结束：需要全本请在线购买： www.ertongbo