UNIVERSITY OF LONDON

u.c

IMPERIAL COLLEGE OF SCIENCE & TECHNOLOGY

A Geometric Approach to the Theory of Optimal Control

THESIS

submitted for the degree of Ph.D
in the Faculty of Engineering.

Dept. of Electrical Engineering
City & Guilds College
Exhibition Road
London S.W.7

Stanley Shapiro, B.Sc(Eng) A.C.G.I.

August 1965

ABSTRACT.

The theory of optimal control is becoming a branch of mathematics, the interests of engineers being left very much in the background. The geometric basis of muchof the theory is only faintly reflected in many mathematical presentations, but here it is kept strictly in the forefront, providing a framework which is easy to grasp and which allows the intuitive motivation to keep pace with the mathematics - an important consideration in engineering mathematics. The techniques used are essentially transformations in linear spaces, and differentiability theorems for differential equations, introduced in a suitable form before application to the control problem. A basic assumption, given good justification, is that optimal trajectories successively occupy regions of different dimension in state space, in each of which the feedback control is differentiable (in a modified sense) with recpect to the state. An analysis of fields of optimal trajectories, based upon the concept of an 'isotim' - a surface of constant cost - leads to a constructive theory for optimal control, requiring no modification for the treatment of inequality constraints. The insight this gives into the behaviour of systems is different from that gained via other techniques, and, together with a second geometric approach, based upon Huygen's construction, suggests useful techniques for dealing with the two-point boundary value problem.

PREFACE

When I started work in 1961 the only readily available source of information on optimal control that was suitable for beginners was Bellman's 'Guided Tour'; local library facilities were such that elementary Russian publications such as Rozonoer's were not widely known. Drs. J. Florentin and J. Pearson, now at Brown University and Case Institute respectively, were my early informants on the subject, and my attempts to understand it from an intuitive and 'physical' point of view led to the present work. Initially it was intended merely to formulate the theory for a specific application - control of a proton accelerator - but this project eventually proved far too difficult and the theory itself became increasingly more interesting.

An important impetus was the opportunity of presenting lecture courses on the subject, for which I am very grateful to my supervisor Professor J. H. Westcott, without whose constant encouragement and help in this and many other ways this work could not have been done. In addition to those already mentioned, the sources of many of my ideas are to be traced to lengthy discussions with colleagues, notably Mr. M. Levine and Mr. S. Mitter.

For financial assistance I am indebted to the now defunct D.S.I.R., and also to the National Institute for Research in Nuclear Science, which, through the good offices of Mr. T. Walsh, sponsored this work for a six - month period.

Typing the manuscript was an indefatigable labour of love on the part of my parents, Mr. & Mrs. A. Shapiro, whose criticisms of style, grammar, etc., contributed much to whatever standard of literacy has been achieved.

תם ונשלם שבח לה׳ בורא עולם

CONTENTS

NOTATION

Contravariant vector components are superscripted : $x=(x^1,x^2,\dots x^n)$

Covariant " " " subscripted : $p = (p_1,p_2,\dots,p_n)$

and the repeated index summation convention is generally used :

$$p_i x^i = \sum_i p_i x^i$$

Subscripts on contravariant vectors and scalars indicate specific values : e.g. $x_o = x(0)$

Partial derivatives are indicated by subscripts : $J_{x^i} = \partial J/\partial x^i$

$$J_x = (J_{x^1}, J_{x^2}, \dots J_{x^n})$$

Total derivatives with respect to t are dotted: $\dfrac{dx}{dt} = \dot{x}$

and with respect to other variables are primed : $\dfrac{dx}{ds} = x'$

(occasionally t' will indicate a particular value of t).

" n-dim." means "n - dimensional".

Scalar products are written a.b or a.(b+c), etc.

The simplified matrix notation does not explicitly show transposition.

If a is avector $(a_1, \dots a_n)$, and A a matrix a_{ij}

$$aA = (a_1, \dots a_n)\begin{bmatrix} a_{11} & a_{1n} \\ & \\ a_{n1} & a_{nn} \end{bmatrix}$$

If b is a vector col $(b_1, \dots b_n)$

$$Ab = \begin{bmatrix} a_{11} & a_{1n} \\ & \\ a_{n1} & a_{nn} \end{bmatrix}\begin{bmatrix} b_1 \\ \\ b_n \end{bmatrix}$$

Chapter 1 INTRODUCTION

1.1 Engineering Mathematics.

This thesis must be regarded as a didactic piece of work, rather than as breaking new ground in engineering techniques. Over the last few years optimal control has become the province of the mathematicians, and although the essential techniques can be reduced to a set of rules, the principles are enshrouded in a wrap of mathematics so obscure that many engineers are neither sufficiently equipped nor interested enough to penetrate it. This is a regrettable but not unavoidable state of affairs, and it arises partly because the engineer is trained to know his place and not to dabble in mysteries beyond his scope and partly because that which interests the mathematician does not necessarily interest the engineer, and vice versa, so that the mathematical discussion is often presented in a form which does not appear immediately relevant to the practical problem.

There is a legitimate divergence of interests between engineers and mathematicians, but it has created a gap which must be bridged, and which is being bridged and even filled by a comparatively new genre---engineering mathematics. This has always been a shabby relation of 'real' mathematics: it makes no pretence to rigour, or firm foundations, or elegance, or to any of the classic virtues associated with mathematics----but it must 'work'. thus it is possible for an engineer to study a course in, say, differential equations, without ever hearing tell of existence or uniqueness theorems!

All this is changing. Modern engineering mathematics is as sophisticated and precise as pure mathematics, but is characterized by its direct relevance to practical problems and its natural evolution from them. The pure mathematician is content to derive theorems from axioms, leaving the axioms themselves in doubt (Russell 13 p.373). He is not compelled to justify his axioms in any way, and if he wishes to draw new quantities 'out of the hat', or make esoteric definitions or manipulations apparently without purpose, he is entitled, by the rules of the game, to do so. The engineer, on the other hand, is very much concerned with the axioms; he will refuse to 'consider the equation....' unless he can be shown good reason for doing so, and if an assumption is made which does not arise naturally or from immediate necessity, he will rightly demur.

This difference in attitudes and in the logical foundations of the two disciplines (for 'mathematics and logic are identical'---Russell 13 intro. ---while engineering is not at all the same as logic) leads to quite different treatment of the same material. Certain mathematical techniques emerge as a response to the requirements of natural or engineering science, and arouse no interest among mathematicians until they can be shown to have a rigorous logical foundation. Probability theory, for example, met with little pure mathematical development until it was found to be similar to the respectable theory of measure. For pure mathematics it is important to free the soul of an idea from its earthy origins, and give it an independent and more general existence; hence the modern trend towards axiomatic mathematics in which even frankly mechanical sciences such as dynamics are given the form of pure mathematics by seeking a set of axioms from which the whole discipline can be deduced without further physical considerations. (Synge 68 p.5, McKinsey, Sugar, Suppes 69, Hamel 70, Truesdell and Toupin 71 p.228, Landsberg 72, Kilminster 73).

While this lack of specificity of mathematics is valuable also to the engineer, for it enables him to apply techniques to situations far from those implied in the origins of the subject, he will in fact be dealing with specific situations, and he must take the mathematical ideas,

"Turn them into shapes, and give to airy nothing
A local habitation and a name."

In doing this, the mathematical techniques must not be treated as something external to the problem, borrowed as tools for a special purpose, but should be derived from the given background of the problem, as a natural consequence of it.

This is desirable both aesthetically and scientifically; aesthetically, because there will be an uncomfortable tension involved in putting together quite different disciplines without smoothing the seams, and scientifically, because a mathematical model is like any analogy, it holds only up to a point, and unless the mathematics is made to relate to the fundamentals of the situation, it is impossible to know just how far it is applicable.

Let us consider how far these differences affect the application of a mathematical theorem to a physical situation. A mathematician will present axioms and assumptions, prove a series of lemmas, and finally the theorem, in as great a generality as possible ——all this without any apparent motivation or connection with the physical system. Then he shows that a mathematical model of the system accords with the assumptions, therefore the theorem holds; again, there need be no physical interpretation of the theorem, it is simply a valid rule.

The engineer, on the other hand, will discuss a basic mathematical model of the system, and how far certain simplifying assumptions can be made, attempting to restrict, not generalize the mathematics, since he is interested only in the one system. Then he develops the theory in such a way that every mathematical step relates to some physical or geometric or other easily conceived property. The result will be the same, equally valid, equally rigorous, but one scheme will have the advantage of generality, and the other the advantage that it really discloses properties of the system, and is easy to grasp in the given context.

It is a valid subject for research to develop an engineering approach to mathematical techniques, and conversely, to find an axiomatic basis for engineering methods. The difference proves to be more than merely formal, and not simply a question of finding a posteriori interpretations for particular variables or equations. The entire development may have to be changed, and results may be trivial in one technique which are difficult in another. In this thesis the first approach will be taken, and an effort made to study the properties of optimal control systems from a point of view which never loses sight of a straightforward geometric model of a dynamic system.

It is interesting to contrast this viewpoint with a recent remark by Halkin (7 p.7).'Any mathematical venture is made up of two parts: geometrical intuition and analytical machinery.......the geometrical intuition always precedes the analytical manipulation in the formation of a theory, and the first is of great help to understand the second. Unfortunately, this duality has a marked tendency to disappear, and the role of geometrical intuition is barely noticeable in the final form of a theory.........The geometric motivation is virtually absent."

His evaluation of the role of geometry may be a little too sweeping (Hadamard 74), but for those with a practical bent it is probably the major element in their mathematical thinking. This thesis represents an attempt to use geometrical ideas in order to present the theory of optimal control of first order differential systems from a simple conceptual basis, giving a direct motivation both for the theory and for techniques of application.

The scope of this work is clear from the list of contents: a brief discussion of the formulation of the problem deals with the implications of certain aspects of the mathematical model, followed by a consideration of evident properties of engineering systems which allow further simplifying assumptions, laying the ground for the geometric construction from which the necessary conditions for optimality are derived. Chapter 6 introduces a new construction, alternative to the first, and amenable to more general treatment, but not permitting such a natural derivation of the theory. Finally, well-known applications of optimal control are considered in the light of this approach, making them easier to comprehend and implement.

Since this work is designed to demonstrate an approach and an interpretation rather than new mechanical techniques, experimental results are less in evidence than is usual in engineering reports. In the present state of the art, the computation of even small problems presents extensive technical difficulties, usually peculiar to that problem, and of little general interest except in the context of work on computing methods as such, while larger problems with a significance of their own would constitute research projects in their own right. On balance, it was considered that time would be more usefully spent on ideas than on extensive computation, in view of the purpose of this thesis. An appendix treats several problems, mostly elementary in themselves, but for that very reason more useful in demonstrating essential features of the geometric approach.

The remainder of ~~this introductory chapter~~ discusses the role of mathematics in control theory, and the current mathematical attitudes to it.

1.2. Mathematics in Control Theory.

Mathematics enters wholeheartedly into engineering when a general mathematical framework can be found which accords with the physical situation. Sometimes the engineering needs come first, developing powerful tools, but without the rigorous foundations that a pure mathematical approach would provide; to some extent this is the present situation in linear feedback theory based upon transform methods and root-locus concepts. The natural process of development in such cases is to find a rigorous basis for the method, and give it a broader and firmer foundation, opening doors to wider fields of application. In other cases, the mathematical theory is well-established, and it is afterwards found that the physical system can be described in a similar way so that it is amenable to the same techniques of solution; this has occurred in modern optimal control theory.

The mathematical treatment of such systems involves two stages--- construction of a model, and solution of the problem. The first is always the more difficult, requiring real originality; the second usually reduces to the extension of well-known techniques or devising computational schemes. The modern state-space model of control systems stems from the classic block-diagrammatic representation whereby a complex system can be broken into distinct parts interacting in a specific way. Each block (Fig 1) has inputs and outputs, the output being some function or operation on the inputs. The common systems contain piecewise smooth functions and integral operators.

By labelling the outputs of the n integrators x^i $(i=1,\ldots,n)$, we obtain a

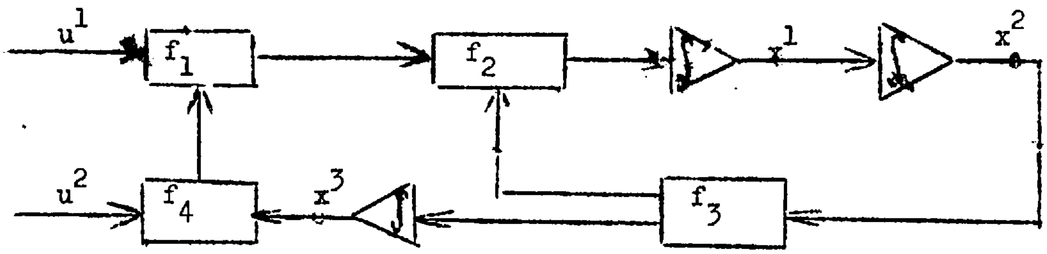


Fig. 1

vector $x = (x^1 \ldots x^n)$ called the 'state vector'. The out put of the system is usually a certain collection of components of the state vector which are chosen to be observed. We shall assume throughout that the state-output relation is sufficiently trivial to be ignored in a theoretical study.

The system of Fig. 1 can be described completely by

$$\dot{x}^1 = f_2(f_3(x^2), f_1(u^1, f_4(u^2, x^3)))$$

$$\dot{x}^2 = x^1$$

$$\dot{x}^3 = f_3(x^2)$$

or generally

$$\dot{x} = f(x,u),$$

a very neat model concerning which a vast literature exists. The classical differential equation is somewhat modified by the inclusion of the indefinite function $u(t)$ representing the manipulable inputs of the physical system. This does not absolutely necessitate a new treatment of the theory, but it does contribute to pure mathematics concepts such as 'reachable zones', which had not yet achieved a place in the natural heritage of differential equation theory, and is still not to be found in standard treatises on the subject.

There are two major defects of this representation of a system. First, its extreme complexity for large systems; it is really a microscopic description, in which the contribution of each part is scrupulously accounted for. A system need not be very complex by usual engineering standards before it runs to hundreds of variables, while its overall response appears comparatively simple. What is required is a model to represent the whole rather than the sum of its parts; this is not yet available, and the result is that useful results can be obtained only for systems of low dimension.

A second fault is the lack of a firm logical foundation, and the rather arbitrary use of differential equation forms, which are obtained, as indicated in Fig.1, by electing to treat the outputs of integrators with special favour. This is done only because other points can be connected by explicit algebraic relations, and seems an arbitrary choice of state variables, for they do not necessarily have any real physical significance. It raises difficulties not only for the purist, but also for the technician, for implementation often depends upon the possibility of measuring all the state variables. The differential form for physical processes is always suspect, for a derivative, or a velocity, is a purely mathematical concept with no empirical basis at all (Russell 13 p.473, Truesdell 71 p.233), though this is easily avoided by using the integral form. There is, however, no absolute necessity to suppose that even that is generally satisfactory, and it could scarcely be used as a basis for an axiomatic theory.

A more satisfactory approach would be via the fundamental systems theory touched on by Zadeh and Desoer (76), and the related theory of automata, which has close connections with control theory (Arbib 77). In general terms, a system may be expressed as a sextet (I,O,S,f,g,t) representing respectively the class of i) inputs, ii) outputs, iii) states,

iv) the input state relation, v) (input-state)-output relation, vi) time. In addition, there may be some measure of performance, and other factors external to the system itself. In our case, a comparatively simple structure is imposed by requiring I,O,S to be vectors or vector functions (e.g., u(t)) in finite-dimensional vector spaces, f a first order differential function, g an algebraic function, and the performance criterion scalar valued.

The performance criterion is not part of the system, but it plays a central role in specifying the problem. Informally, this consists of choosing the input function in such a way that the system behaves satisfactorily. 'Satisfactoriness' or 'acceptability' are difficult concepts to define, especially since we may be dealing with systems involving human interests, or systems in which human operators play a role, and such crucial matters as convenience, security, or psychological considerations do not lend themselves readily to exact evaluation. At present we must restrict serious consideration to purely technical aspects, such as can be ascribed a precise measure, but even here there are difficulties. 'Acceptability' is too general a criterion to provide precise results ---a more restrictive requirement is 'optimality', i.e., that the system shall behave in the best possible way, within a region of admissibility described by inequalities, beyond which the solution is definitely unacceptable. Presumably this implies a unique behaviour in many cases, but in practice there are many conflicting considerations----demands of efficiency, economy, security, do not generally pull in the same direction----and the optimal solution must be a compromise, the precise degree of which must be predetermined by the designer.

In principle one would like to see something along these lines: a factor of optimality for each relevant property of the system, all contributing to an overall measure of optimality according to the desired compromise, leading to a sensitivity analysis to indicate how variations of the control function would affect the various factors. This would provide a satisfactorily flexible programme for practical applications, and would give a good insight into the behaviour of the system. Unfortunately, this is not yet possible, though similar ideas have been mooted (Zadeh 4, Pearson 80); the usual practice is to combine all the relevant factors in a single scalar functional----the performance criterion.

It is clear now that the familiar control problem is only a very special example of a much wider class of as yet unformulated problems in systems analysis, and it is quite evident that the motivation for this particular formulation was its similarity to the well-known problem of Bolza in the calculus of variations. It will not be long before this problem loses its current popularity, and becomes recognized as the correct form only for the low-dimensional ordinary differential system and single objective function such as arises in trajectory problems, but will no longer be regarded as "the optimal control problem". It is, however, the problem with which this thesis is concerned.

The development of this problem is only a chapter in the history of the calculus of variations. The tendency to regard the calculus of variations as outdated (from a control point of view), or incapable of dealing with modern problems, is quite unjustified, and even mean, for there is no modern treatment of control theory which is more than a step away from similar methods used in the older discipline. The fashionable disparagement of that calculus (cf. Pontryagin, 1 p.1, Halkin, 7 p.6) is open to uncharitable interpretation, and is very difficult to understand, in view of, for example, Berkovitz's work (25), merely translating the control problem into a Bolza problem, for which the necessary conditions hold over a wider class of situations than some more popular techniques can handle.

Indeed, despite the fact that the modern problem was only fully stated in 1949 (Hestenes 79) ----and that without state constraints---- it was effectively solved earlier. Bliss, in his textbook in 1946 (5) presents a problem involving differential equality constraints, which, apart from inequality constraints is effectively the modern problem. Finite (state) inequalities had been thoroughly studied (e.g.,Bliss and Underhill 35) and differential (control) inequalities had received some attention (Valentine 81). All that was lacking was the need to bring all these elements together in an engineering context. It was not until 1964 that this was done in a form including state constraints (Guinn 82)but the fault was not that of the calculus of variations.

The classical approach is not, however, a perfect fit to the physical situation; rather the feeling is that the problem has been forced to suit the manner of solution, for a method which treats the minimisation as central and the dynamic system as a mere side constraint is clearly not a natural one to adopt. Developments of the problem for control purposes have been an improvement, treating the system as the basic material of the problem, though this is never given as the primary motivation of the new method. True, the effect of the Lagrange multiplier technique is very similar, ensuring that the constraints are automatically satisfied, but it smacks of the nature of a'device' rather than a basically convincing approach, and the multipliers themselves are difficult to place in the physical scheme. They do have a straightforward interpretation as the 'effort' required to ensure that the constraint is not violated, (Lanczos 28 p84), but this serves only to emphasize the secondary role of the constraint.

The techniques of functional analysis now being applied both to the classical problem (Liusternik & Sobolev 83) and the control form of it (Balakrishnan 84) tend to provide powerful tools, but little modification of the fundamental attitude to the problem, though work such as the little known paper of Dubovitskii & Milyutin (59) take steps in the right direction, for while the approach of linear functionals and the fundamental lemma is the same, only variations admissible with respect to the system and the restricted regions are permitted. The tendency is to allow the system to define permissible operations, rather than regarding it as a constraint - the difference is subtle, but profound.

The purely geometric approach to the classical problem via the geometry of Finsler spaces (Rund 17) goes much further in this direction. The functional to be minimised is supposed to define a metric on the space and the minimising trajectories are geodesics. This leads easily to the canonical equations and Weierstrass's condition, and the refinements of constraints fall naturally into place, though they have received very little attention from the classical practitioners in this context. An approach in a similar spirit is made in Chapter 6, but the powerful tensor calculus, which would seem to be the natural tool to use proves difficult, for the treatment of the classical problem rests heavily on the assumption that the integrand of the cost function is homogeneous of degree one in $\dot{x}$, and this is not true of the control problem.

The shift in emphasis from extremum aspects of the problem to the differential system itself and properties such as controllability, accessibility, stability, etc, has meant that control theory now occupies an established place both in the calculus of variations and in the theory of differential equations. Proper application of geometric and topological techniques, using metrics imposed by the cost functions of control systems, and restricted spaces defined by the reachable zones of a differential system, will probably lead to innovations in differential geometry. It is impossible to foresee what future developements will bring, but it seems likely that in the interplay between mathematics and systems theory the flow of new ideas is likely to run in both directions.

Chapter 2 PRELIMINARIES

2 . 1 The System

We shall be dealing exclusively with systems whose behaviour can be described by a first order ordinary vector differential equation

$$\dot{x} = f(x, u(t)) \qquad 2.1$$

t, the independent variable, is monotonically increasing on the interval $I = [t_o, t_f]$ of the real line.

x, the state vector, is at any instant a point in real, n-dim. Euclidean space, E^n.

u, the control vector, is at any instant a point in real, m - dim. Euclidean space E^m.

As t traverses I, a mapping

$$u : I \rightarrow E^m \qquad 2.2$$

traces a graph $u(t)$, called the control function, which is assumed to be contained entirely within some specified region U of E^m. $u(t)$ will be chosen to be piecewise continuous, which is sufficient to describe physically realizable controls. At points t^1 of discontinuity of $u(t)$ we will accept the convention that $u(t^1) = u(t^1+o)$.

Corresponding to a particular control function $u(t)$, the solution (if it exists; see section 2.4) of 2. 1 which passes through the point x_1 at $t = t_1$ will be described by the function

$$y (x_1, t_1, u(t); t) \qquad 2.3$$

Such a solution traces a 'trajectory', denoted $x(t)$, in E^n, which will be required to remain within a specified region X of E^n. For an autonomous system, the independent variable can be shifted by an arbitrary constant c so that the trajectory $x(t) = y(x_1, t_1, u(t); t)$ is the same as $x(t^1) = y(x_1, t_1+c, u(t^1); t^1)$ where $t^1 = t + c$.

This useful property of 2. 1 will often be used to allow different solutions to start with a common value of t by adjusting the t – origin suitably for each.

If the solution of 2. 1 is continuous for continuous $u(t)$, and we assume that it is, then an absolutely continuous solution can be constructed for piecewise continuous $u(t)$ by taking the endpoint of a continuous sub-arc for the initial point of the next continous sub-arc. (Pontryagin 1 p12). Physical systems can certainly be constructed whose state variables are not absolutely continuous, but no fundamental principles are overlooked by excluding them.

The regions U, X will be defined by inequalities

$$C_i(x) \leq 0 \qquad 2.\ 4a$$

$$B_j(u) \leq 0 \qquad 2.\ 4b$$

whose left sides are continuous and differentiable. There may be any number of these constraints, which, when equality holds, define the boundaries of X and U, which are piecewise smooth manifolds of at most, $(n-1)$, $(m-1)$-dim. respectively; they might entirely enclose a region, or leave it open on some sides. Where a region is not explicitly bounded in this way it is assumed to extend to infinity.

A question of some delicacy arises regarding the nature of these constraints. Do they designate a region of interest within the domain of definition of the system, or do they themselves specify the domain of definition so that the system cannot be properly described without them. The difference is between supposing $f(x,u)$ to be defined everywhere on $E^n \times E^m$ but x, u permitted to take values only in $X \times U$, and f to be defined only on $X \times U$. The distinction is a real one, for differentiability properties at the boundary will be affected, and certain techniques which permit small excursions beyond the boundaries (e.g Chang 2) will only be possible under the first construction; it is not only a mathematical distinction, for physical systems exhibit both imposed constraints (of the first kind) which must not be violated for reasons of safety, stability, economy, etc, and natural constraints (of the second kind) which cannot possibly be violated. Examples of the latter are mass variables, which cannot be negative, height above ground for an aircraft system, temperatures, which have a natural lower bound, etc, etc. A completely satisfactory theory would reflect both types in the formulation of the mathematical model, but to insist upon this would be pedantic. We may follow the easier practice of adjusting the domain of definition of f to an open neighbourhood of $X \times U$ in $E^n \times E^m$, and let f be continuous with its partial derivatives in all its arguments.

Such a model is really quite restrictive, excluding as it does systems with distributed parameters, delays, and random variables. However, a large class of engineering system do fall into this category, including all ordinary differential systems of any finite order, and non-autonomous systems. The latter occur whenever t appears as an argument of f or of the constraints 2 .4 , and in this case we simply introduce an additional state variable defined by $\dot{x}^0 = 1$, and replace t wherever it occurs by x^0 (except where t is merely the independent variable). This manoeuvre inposes a greater degree of symmetry on the variables and allows us to use the autonomous formulation throughout.

In addition to 2.4 , there may be 'mixed constraints ' of the form

$$R (x,u) \leqq 0 , \qquad 2 . 5$$

restricting (x,u) to a region R of $X \times U$. In the absence of such inequalities, $R = X \times U$. A control function $u(t)$ for which $(x(t),u(t))$ remains in R for all $t \in I$ is an 'admissible control'. The corresponding trajectory is an admissible trajectory.

In a given situation it will be required that the solution of 2.1 shall pass through certain points or subsets of X at various stages along the trajectory. The most important of these are the initial set S and the terminal set T, and in this work T receives special prominence. It will be a smooth $(n-s)$ - dim. manifold in X, defined by a set of s equalities

$$T_i\,(x) = 0 \qquad i = 1, \ldots, s - n \qquad 2.6$$

T_i being continuous and differentiable.

The terminal time t_f of a process starting from any point $x_o \in X$ and any t_o, is defined as the first instant at which the trajectory reaches T; i.e.

$$t_f = \inf\ (\,t^1 : y\,(\,x_o, t_o, u\,(t);\ t\,) \in T\,) \qquad 2.7$$

There are some cases of practical interest which are not covered by this description, such as the problem involving the miss - distance from a given set, (Bridgland 3) but this will serve for the present.

2. 2 The Cost Function

The usual performance criterion takes the form of a scalar functional, measured either at t_f, the termination of the process, or as an integral over the entire interval I. If the former, it will be a function of the terminal values of the state variables; the control will not be relevant, for at t_f it has no effect on performance, seeing that the process ends at that point. If it is an integral some measure of the control may well be involved. Thus we have the alternative scalars

$$g\,(x\,(t_f)\,) \qquad\qquad 2.8^a$$

$$\int_{t_o}^{t_f} L\,(x\,(t),\,u(t)\,)\,dt \qquad\qquad b$$

The term 'performance criterion' is something of a misnomer, for evaluation of the function without knowledge of upper or lower bounds gives no indication of the quality of the performance. A more apt term, if only for minimization problems, is ' cost function '.

The cost function is entirely at our disposal, since it is not part of the system, but reflects the intentions of the engineer concerning it. For convenience let us choose g and L to be continuous and differentiable in all their arguments. If we choose a function of type 2.8a we have, using the terminology of the analogous situation in the calculus of variations, a Mayer problem; if type 2.8b , a Lagrange problem.

Mathematically, the two forms are completely equivalent and can be transformed from one to the other with no mathematical embarrassment. Thus, defining a new variable x^{n+1} by

$$\dot{x}^{n+1} = L\,(x,u) \qquad x^{n+1}\,(t_o) = 0$$

8b becomes simply $x^{n+1}\,(tf)$, or, writing $\frac{dg^o(x)}{dt} = g_x.f$,

we have

$$g(x(t_f)) = \int_{t_o}^{t_f} g_x.f\;dt + g(x(t_o)).$$

Since $\dot{g}(t_o)$ is not involved in the minimization, the terminal point expression or the integral may be used indifferently. (Bliss 5 p189)

Mathematical equivalence is one thing: practical equivalence quite another. In practice a cost function will almost invariably suggest itself in one of the forms 2.8a or b, and to transform it into the other requires either the introduction of a new variable, or a rather strained interpretation of the function. For example , suppose we wish to minimise the magnitude of one variable, say x^1, at a given time. The natural cost function would be $x^1(t_f)$, and putting it into the Lagrange form $\int_{t_o}^{t_f} f^1(x;u)dt$ completely obscures the meaning of the function. Again, the familiar regulator cost function $\int (x^2 + u^2)\, dt$ which measures the integrated error (from zero) and the control effort, could be expressed as a Mayer function $x^{n+1}(t_f)$, $\dot{x}^{n+1} = x^2 + u^2$, which not only obscures the character of the problem, but also unnecessarily extends the state space.

It is common in presenting the theory of optimal systems to reduce all problems to Mayer form (Pontryagin 1, Halkin 6) which, from a purely mathematical viewpoint, is quite legitimate, but the engineer will feel uneasy at this, for if the problem formulates itself it obviously knows what it is doing and should not be forced into a preconceived pattern. Like a difficult child, a problem can be very cooperative if given an opportunity for self-reliance, but obstructive when restrained by artificial regulations. In any case it would be impolitic to submit to the whims of mathematics at this early stage - she will make stronger demands soon enough. Let us be satisfied then, to leave Lagrange as Lagrange, and Mayer as Mayer. We shall find that this independence is rewarded, for the different formulations lead to quite dissimilar insights into the nature of the system of optimal trajectories.

Pontryagin's formulation we may reject for a further reason. His cost function is not permitted to be one of the original state variables of the system, which in many cases means introducing a new variable which is algebraically dependent upon the others, or even (surely a reductio ad absurdum) identical with one of them, (see comments by Halkin 7, Roxin 8). This artificial situation is in stark contrast to the 'natural' approach we have agreed to adopt.

Whichever form the cost function takes, we shall use the following notation:

$$P\ (x_1, t_1, \ u(t)) \tag{2.9}$$

is the value of the cost function evaluated for the trajectory which starts from (x_1, t_1) and terminates on T, corresponding to the control $u(t)$ defined for all t in $[t_1, t_f]$.

2 . 3 The Problem

We are now in a position to formulate precisely the problem of optimal control :

Given a dynamic system $\dot{x} = f(x,u)$;

permissible regions X, U, R defined by 2.4,2.5;

sets $S, T \subset X$;

a cost function 2.8a or b ;

determine the admissible control function for which the corresponding trajectory defined by

$$x(t) = y\ (x_o, t_o, \ u(t)\ ;\ t\ )$$

satisfies

$$\begin{aligned} x\ (t_o) &\in S && \text{a} \\ x\ (t_f) &\in T && \text{b} \quad 2.10 \\ (x(t), u(t)) &\subset R && \text{c} \\ P\ (x_o, t_o, \ u(t)) &\leq P\ (x_o, t_o, v\ (t)) && \text{d} \end{aligned}$$

where $v(t)$ is any admissible control for which 2.10 a -c are satisfied. Such a function $u(t)$ is an 'optimal control function '; the corresponding trajectory is an 'optimal trajectory'.

2 .4 Existence and Uniqueness.

The question of existence is always popular with mathematicians, but usually neglected by engineers, for, if a solution can be found 'existence' is proven, and if not, the knowledge that one exists is not very helpful. Unfortunately this is not always a realistic attitude, for we are dealing not only with real systems, where it is usually obvious whether the objective is attainable or not, but also with their formulation into mathematical problems, where the very concept of a 'solution' is different. Thus there may exist an infinite sequence of control functions, and a corresponding sequence of costs with a lower bound but no minimum. For engineering purposes this is good enough, but, mathematically speaking, no solution exists, and the machinery will break down. The point of raising the question of existence in engineering mathematics is not simply to find out whether there is a solution, but to confirm that the mathematical model is adequate. The technique remains the same: the philosophy is significantly different.

The subject is intimately connected with the existence of the less restricted class of admissible controls for which 10a - c are satisfied, but not necessarily 10d. It will depend upon the constraints, the initial and terminal sets, and the dynamic system, as well as the cost function, and is obviously exceedingly difficult to treat in general, though results have been obtained in particular cases. (Kalman 9, Kalman, Ho, and Novendra 10, Markus & Lee 90, Roxin 91) The best approach to a specific problem is to attempt to construct a solution in the hope that it can be done. If the attempt is successful, well and good; if not, it is advisable to reframe the problem, either by relaxing certain restrictions or reconstructing the system or cost function.

For theoretical purposes it is convenient to overcome this difficulty by the assumption that

from every point in X there exists
an admissible trajectory terminating on 2. 11
T which is optimal in the sense of 10d.

It may turn out that the assumption holds only in a closed subset $\bar{X} \subset X$ bounded by T. If $\bar{X}$ is n-dim. we may construct constraints of the type 4a, circumscribing $\bar{X}$, then we restrict our attention to $\bar{X}$, and this is possible even if $\bar{X}$ is in fact the union of disjoint subsets. In that case 2 .11 can be regarded as equivalent to a collection of state constraints. If $\bar{X}$ is p-dim, then the trajectories of 2.1 occupy only a p-dim. subspace of X, and the n-dim. representation of the system must be redundant. This may be remedied by suitable coordinate transformations. (see Chap.3). If the assumption does not hold at all there is no more to be said.

The related topic of uniqueness is rather different, and less of a hurdle. It should be considered on two levels: the possibility of a finite number of solutions, and of an infinite number. In the former case just one of the possibilities will be chosen, and this choice makes that solution effectively unique. The criterion guiding the choice has the same effect as a more stringent cost function. Thus, in a practical sense we always have a unique solution, but mathematical conditions are difficult to lay down. We shall assume that the assumption 2. 11 is restricted to a unique trajectory. The second possibility cannot be so easily dealt with, but we must avoid it by assuming that it does not occur . (cf. Thau. 49)

There is a further question. Does a given control function give rise to a unique trajectory. This is a comparatively simple problem which depends only upon $f(x,u)$. If f is continuous in x and t (through u) and satisfies a Lipshitz condition in some open region of X then, for a given $u(t)$, there will be a unique trajectory within that region. If the partial derivatives f_x are continuous the Lipshitz condition certainly holds (Lefschetz 12 p.34), and since we have assumed this to be the case, and also $u(t)$ to be piecewise continuous, a unique trajectory is assured.

Chapter 3. THE SOLUTION SPACE.

3.1 Some Physical Considerations

In this chapter we shall develop a picture of the optimal trajectories covering X . (The adjective 'optimal' will generally be omitted in this chapter, but must be understood to apply.) It has already been made clear that in the order of priorities guiding this exposition, simplicity of the geometric concepts takes first place. In order to establish the principles we need have no hesitation in sacrificing generality to simplicity, as long as the restrictions leave us a reasonably large class of situations of engineering interest. It is encouraging to observe that physically realizable functions are usually simple in structure - continuous, many times differentiable, etc., - so that a considerable degree of mathematically stringent restriction can be accepted without unduly affecting the practicability of the results. We may, however, be forced to take as assumptions, or hypotheses, properties that, via more rigorous but less straightforward routes, could actually be proven.

At this stage we are seeking a physically reasonable picture, using heuristic arguments on any material that comes to hand. In succeeding sections a suitable mathematical framework for the resulting ideas will be described, and we shall have to cover some of the same ground again, but with a less cavalier disregard of details.

An important concept arises from the assumption that from every point there is only one trajectory to the terminal set. It implies uniqueness from the left, but not necessarily from the right, and, although only one path can emerge from a point there may be many distinct trajectories converging onto a common point or onto a common trajectory - much like a confluence of tributaries into a main stream, except that in the latter case conservation of flow holds, but with our trajectories this analogy breaks down. We must examine this situation further.

Since the trajectory is completely defined, given the initial point (by virtue of the uniqueness assumption) we may write, for x, at any time t_1,

$$x = y(x_o;t_1).$$

If trajectories meet, so that from distinct points x_o, x'_o they reach x_1, this equation cannot be solved uniquely for x_o. According to the implicit function theorem (Bliss 5 p.270) it could be so solved if

$$\det \left[y_{x_o} \right] \neq 0.$$

In this case, then, either y_{x_o} ceases to exist or the determinant becomes zero. The former implies a discontinuity in $u(t)$, for standard theorems (Bliss 5 p.270) assure the existence of derivatives of solutions of 2.1 if $u(t)$ is continuous. The latter, while admitting the existence of the derivatives, implies that the rank of the transformation $x_o \to x(t)$ is less than maximum, and therefore does not preserve dimension: Points in an n-dim. neighbourhood of x_o are sent into a region of dimension equal to the rank of the Jacobian determinant, less than n. This is an acceptable result - if trajectories meet, thereafter remaining coincident, they must occupy ' less space'.

Just as a given initial point defines the unique trajectory from that point, so it must define the control action from that point, and we have a unique function $u(x_o;t)$, and in particular a unique vector $u(x_o)$, representing the action to be taken at x_o. This argument applies to every point, and we have a unique vector field $u(x)$ defined over the state space X.

From our assumptions so far it is not clear what the properties of $u(x)$ will be. Trajectories may be unique and continuous, and, as a family, cover every point in X, but still be pathologically twisted and knotted, and allow $u(x)$ to be discontinuous except in certain directions. It is easy to see that since $u(t)$ is piecewise continuous (by assumption) and $x(t)$ is continuous, $u(x)$ must be piecewise continuous for x taken along a trajectory, but not necessarily so for x in an arbitrary set. More precisely, if K is a trajectory passing through $x_1(t_1)$, and $u(t)$ is continuous in an interval I_1 containing t_1, then for all $e > 0$ there exists some d such that $\| x(t_i) - x_1 \| < d$ implies $\| u(x(t_1)) - u(x_1) \| < e$; where $x(t_i) \in K$, $t_i \notin I_1$.

If $u(x)$ is not continuous in an arbitrary small neighbourhood of x_1, then the controls u_1, u_2 corresponding to x_1 and x_2 are not necessarily close, however small the distance $\| x_2 - x_1 \|$. In practice this means that if the measurement of x and the implementation of u are not absolutely accurate, the applied control might be hopelessly wrong. There may well be places in the state space where this occurs for a physical system----sharp dividing lines are possible where a decision is either right or wrong, but if it is true everywhere we would be wasting our time even to attempt a scheme of physical control. We may, then, permit the restriction that $u(x)$ is piecewise continuous: there is a partition of X consisting of subspaces X_i $(i = 1, 2, \ldots\ )$ of various dimensions, such that every point in X is contained in one and only one of the X_i, and $u(x)$ is continuous over each X_i.

What of differentiability? It cannot be considered to be as essential as continuity, but it is worthwhile investigating the consequences of such a property, for if they correspond to what would be expected of a real system without disallowing any reasonable possibilities, there is every reason to accept it as a working hypothesis.

Suppose there is a family of sets $\bar{X}_j$ forming a subpartition of the $\bar{X}_i$, and in each $\bar{X}_j$, $u(x)$ is differentiable. (There is no need to be too pedantic at this stage - the conceptswill be made more precise in the next section) The variation equation for 2.1 is

$$\delta \dot{x} = f_x(x,u)\, \delta x + f_u(x,u)\, \delta u \qquad 3.1$$

Confining our attention to an n dim. region this may be written

$$\delta \dot{x} = (f_x + f_u u_x)\, \delta x \qquad 3.2$$

$$= A(t)\, \delta x$$

where the derivatives are evaluated along a particular trajectory. The solution is of the form

$$\delta x(t) = w(t,t_o)\, \delta x(t_o) \qquad 3.3$$

If n vectors $\delta x(t_o)$ are linearly independent the same is true of $\delta x(t)$ if w is non -singular. We have

$$|w(t,t_o)| = |w(t_o,t_o)| \exp\left(\int_{t_o}^{t} \text{trace } A(s)ds\right) \qquad 3.4$$

(Lefschetz 12 p.60) where trace A is the sum of the diagonal elements of A, so that as long as A is defined and w is somewhere non - singular the vectors $\delta x(t)$ span an n-dim. space and remain distinct. If, for finite t, certain elements of A become infinite, then as $A \to +\infty$, $\delta x(t) \to +\infty$ and the system is unstable; as $A \to -\infty$ $|w| \to 0$ and the δx are no longer independent but span a space of lower dimension, a result we have anticipated, and which occurs, for example, at the terminal set, for although it is of dimension less than n, trajectories in n -space must converge to it ; the former result, though undesirable, is a practical possibility.

In a region of dimension less than n the description 2.1 is redundant. This can be remedied by a suitable choice of state variables, then the same arguments hold as before but for a state of reduced dimension. We may conclude that the assumption of piecewise differentiability of $u(x)$ does not violate the laws governing real systems, and we may proceed to base our discussion upon it.

The picture we now have is of trajectories smoothly covering separate regions of various dimensionality (Fig 2). They may go from an n-space into an adjacent n-space ($A(t)$ piecewise continuous), or into an r-space ($A(t)$, $nxn \rightarrow rxr$), or, remaining in the same region, converge into its lower - dimensional boundary ($A(t) \rightarrow -\infty$).

It must be emphasized that the differentiability condition is not an assumption in the sense of a specification of the system, for the properties of the optimal control are entirely determined by the dynamics and the cost function. This property, if true, should be derivable, and indeed can be derived by techniques, which, though more rigorous, lack the intuitive basis that we have chosen to adopt.

In the next section the ideas introduced here will be made more precise.

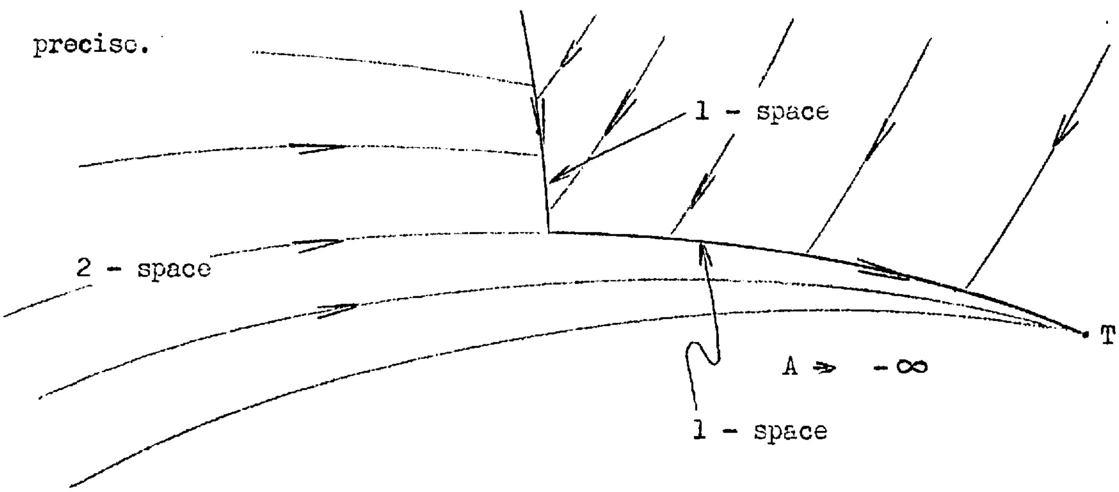


Fig. 2.

3.2 Arithmetic Spaces

3.2.1. Basic Concepts

We are using the terminology of geometry - speaking of spaces, trajectories, etc., - in the context of an analysis of physical systems. Having established, or taken for granted, the field in which we are working, it is as well to pause at this point and examine how far these two branches of science are compatible. This is not a philosophical luxury, for we shall be utilising some very basic concepts, and it is important to know whether the tools are right for the job, or whether we are simply using a baseless analogy. Regretfully, we must leave aside the really fundamental issues which have occupied natural philosophers from time immemorial, and merely touch lightly upon the principles that must be understood in order to use the machinery properly.

It should be recognized at the outset that the spaces we are dealing with are arithmetic point spaces, not geometric spaces. In the latter it is a matter of doubt whether points can be said to exist at all (Russell 13 p.445), but in the former there is no question: ' an ordered set of n (real) numbers will be called an arithmetic point ' (Veblen and Whitehead 14). Two points in the same space are similarly ordered sets, but with different numbers, and it is the relations between points which define the nature of the space. One isolated point can provide no information whatsoever, but just how many points and what type of relations are required to completely specify a space is a question of axiomatics which need not detain us here.

Clearly, the machinery developed for abstract arithmetic spaces will serve to analyse any phenomenon whose properties can be described by a suitable array of numbers. Thus, in a table consisting of columns of numbers each row is an arithmetic point. The class of phenomena that can be described in such a way is vast, and includes engineering systems of the type we are concerned with, where each position in the array corresponds to the value of some measurable property - temperature, velocity, voltage, etc., - and the relations between points are the physical laws of the system. Our familiar Newtonian space can be described in similar fashion, when the points may be measurements of length from a fixed origin in some chosen directions. This type of model can only treat those properties of a system which can be put in this form, and throws no light on the underlying nature of geometric space or of a physical system.

If the language we use is geometric it is because geometry has pre-empted the terminology, not because the nature of the technique is essentially geometric, and to use this approach for engineering systems is by no means using an analogy with geometry, but applying a technique which is even better suited to physical systems with their clearcut physical coordinates, than to geometric spaces. Nevertheless, results obtained from geometric thinking in this context are applicable directly to our systems via the framework of arithmetic spaces, however strange they might appear. For example, in geometric spaces all directions are equivalent and one coordinate system is no more fundamental than another. It is therefore possible to transform points from one coordinate system to another without affecting the properties of the space. In the state space of physical systems the coordinates have a definite meaning, and it would not be obvious that the same licence is valid; nevertheless the techniques are those of arithmetic spaces whatever their apparent interpretation, and we may indeed transform the coordinate system of state space at our convenience, regardless of whether the new coordinate system is physically meaningful or not.

The concept of dimension is important; it is the number of numbers required to define a point. If it is found, by comparing a sufficient number of points, that fewer numbers are required than are actually given, then the description is redundant. Thus. if n coordinates are given, but by suitable manipulation and application of the laws of the space, n - r of them could be deduced from the remaining r, then the dimension of the space is r. This may be quite straightforward; for example, for an electrical network it is not necessary to be informed of all the voltages and currents, because some are derivable from the others by the physical laws - Kirchhoff's, etc. If some law or dependence between the variables applies in some region but not others, then the space is of variable dimension. For example, there is some relation between intelligence and size of feet in human beings up to a certain age, but not beyond that time; the biological space is not of constant dimension. In a geometric space the points in, for example, a room, may have three degrees of freedom in the interior, but only two on the walls. This last example demonstrates the type of situation that led to the construction of constraints such as 2.4, but this picture of 'hard' constraints is too crude for our needs, and the other point of view, of regions of variable dimension, or variable dependence between the components of the points, is more suitable, though they might be expressed in the same algebraic form. This is the idea behind the partition of state space introduced in the previous section.

To deal with variable dimensions we might treat each region on its merits, as a p-, q-, r- or whatever - dim. space it happened to be, regardless of neighbouring regions. In physical state space this is not the best approach, for the coordinates are, after all, all there - we have their values, redundant though they be. It is best to regard the reduced space as embedded in a higher dimensional space, and retain the redundant variables, so that when the process moves from one region to another there is no change in the specification of the points, as in fact there is no change in the nature of the real system.

The essential techniques: transformation of coordinates, and embedding into higher - order spaces, will be discussed in the next section.

3.2.2 Linear and tangent spaces.

A point $x = (x^1,\ldots,x^r)$ defined in a linear space X can be transformed into a point $y = (y^1,\ldots,y^r)$ in another linear space Y, by multiplication with an $r \times r$ matrix $A = \{a_{ij}\}$. Thus

$$y = A\,x. \qquad 3.5$$

If A is non-singular Y has the same dimension as X and the correspondence is one -to -one; also, given points x,y it is possible to find an A to satisfy 3. 5. This type of operation is sometimes regarded as expressing the same point in terms of a different coordinate system, as was hinted in the previous section. This is very dubious, for if a point is defined to be a certain array of numbers, it is difficult to interpret a quite different array to be the ' same point'. Alternatively 3. 5 may be said to transform x into another point in X : A is a transformation of X into itself. This is a little better, but still difficult to support in terms of a state space of a real system, for there is no physical process corresponding to an arbitrary transformation, and the only way in which a point can move in state space is according to the dynamic equation $\dot{x} = f(x,u)$. Such an interpretation is perfectly satisfactory if A is in fact the transition matrix for a linear dynamic system, otherwise it is best to accept the original characterisation as a transformation to an abstract space Y, and there is no call to interpret Y in any physical sense.

Suppose that X, though r-dim., is part of a larger space containing regions of up to n-dim. In order to treat the whole problem in the same way it is desirable to express x in the form of an n-vector. (The terms 'vector', 'point', are equivalent. (Veblen and Whitehead 14 p.2)). This can easily be done by introducing an additional $(n-r)$ numbers such that they can be expressed as linear combinations of the original r independent components. Thus, given x^i, $i = 1,\ldots.r$, introduce numbers x^j $j = r+1,\ldots,n$ such that for each j,

$$x^j = c_{ji}x^i$$

where the c_{ji} are arbitrary numbers, and the repeated index summation convention applies. The n- coordinate vector x is now

$$x = (x^1,\ldots,x^r, c_{r+1\ i}x^i,\ldots\ c_{ni}x^i)$$

Let us show that an arbitrary non-singular $n \times n$ matrix A associates with this x a vector y in Y with n coordinates, of which only r are independent.

Let A be the array $\left\{ a_{km} \right\}$ $k,m = 1,\ldots\ldots,n$. The k'th coordinate of y is

$$\begin{aligned} y^k &= a_{km}x^m \\ &= a_{ki}x^i + a_{kj}x^j \\ &= (a_{ki} + a_{kj}c_{ji})\, x^i \end{aligned}$$

i,j ranging over $1,\ldots,r;\ r+1,\ldots,n$ as before.

The coordinates are dependent if there exist n numbers d_k such that

$$d_k y^k = 0,$$

and y is r-dim. if there are $n-r$ independent sets of such d_k.

$$d_k y^k = d_k (a_{ki} + a_{kj} c_{ji}) x^i$$
$$= d_k a^1_{ki} x^i$$

where a^1_{ki} is an arbitrary nxr array of rank r. Choose any r of the d_k arbitrarily, say d_p, p=1, ...,r. Then

$$d_k y^k = d_p a^1_{pi} x^i + d_{n-p} a^1_{n-p,i} \; x^i = 0$$

which is a non - homogenous set of n-r equations in the n-r variables d_{n-p}, and has a unique solution. Since the d_p can be chosen in r independent ways, y is r-dim.

Furthermore, it is always possible to find some A which will give $y = (y^1, \ldots, y^r, o, \ldots, o)$, for,

$$y^k = (a_{ki} + a_{kj} c_{ji}) x^i$$

and it is only necessary to choose the a_{km} so that for each k=r+1,...,n and each i =1,...,r, there holds

$$a_{ki} + a_{kj} c_{ji} = 0 \qquad 3.6$$

This is a set of (n-r)r equations for the n^2 elements of A, so there is no difficulty in satisfying it. In that case any r-vector can be given n coordinates and remain r-dim. under any non-singular transformation, by the simple expedient of adding n-r zero components.

Let us apply these ideas to an r-dim. region R in the state space of an n-dim. system. Unfortunately R might not be linear, but we will suppose the n-r degrees of dependence to be expressed by some set of non-linear differentiable functions $M^p(x)$ p = r+1,...,n. Since they are differentiable

$$M^p(x+\Delta x) - M^p(x) = m^p_i \Delta x^i + m^p(x, \Delta x) \, \| \Delta x \|$$

for all x, $x+\Delta x$, and each p. m^p is a function which tends to zero as $\Delta x \to 0$, and the m^p_i are finite functions of x, in fact the partial derivatives M_{x^i}. (Berge 15 p.195) If x, $x + \Delta x$ both satisfy $M^p(x) = 0$, and if $\|\Delta x\|$ is small,

$$m^p_i \Delta x^i = 0$$

so that there is a linear dependence between all vectors in directions tangential to the r-dim. smooth manifold represented by the intersection of the $M^p(x) = 0$. In short, there is a linear r-dim. tangent space at x in which the appropriate vectors are differentials dx or derivatives $\dot{x}$.

Using the notation of tensor calculus, a transformation from a linear tangent space X_t to a similar space Y_t is :

$$dy^i = \frac{\partial y^i}{\partial x^j} dx^j$$

for the contravariant vectors dx, $\dot{x}$, and

$$G_{y^i} = \frac{\partial x^j}{\partial y^i} G_{x^j}$$

for co variant vectors G_x, $G(\dot{x})$ being a differentiable scalar function.

Any r-dim. vector $\dot{x}$ can be embedded in an n-space by the addition of n-r zero components, and arbitrary non-singular n x n transformations $\frac{\partial y}{\partial x}$ will preserve the dimension.

It will always be assumed in future that any dependence between the components of x is expressed by some set of <u>differentiable</u> scalar functions $M(x) = 0$. When trajectories reach/a such region the velocity vector $\dot{x}$ can be resolved into a set of components normal to the differentiable manifold,

$$\dot{y} = M_x . \dot{x} = 0$$

or, if there are n-r such functions $M(x)$,

$$\dot{y}^j = M^j_{x^i} \dot{x}^i = 0 \qquad i = 1,\ldots,n; j = r+1,\ldots,n.$$

and a set in the tangent space. Since every tangent vector is normal to every one of the zero vectors which point out of the manifold, we have

$$\dot{y}^k . \dot{y}^j = 0$$

$$k = 1,\ldots,r \qquad j = r+1,\ldots,n$$

i.e. $$\sum_{i=1}^{n} \frac{\partial y^k}{\partial x^i} M^j_{x^i} (\dot{x}^i)^2 = 0,$$

a set of $r(n-r)$ equations.

In addition, it is convenient to choose a Cartesian coordinate system for the r- vectors in the tangent space, so that they are all mutually orthogonal:

$$\frac{\partial y^k}{\partial x^i} \cdot \frac{\partial y^m}{\partial x^i} (\dot{x}^i)^2 = 0$$

$k , m = 1 , \ldots , r$. If the M_x are explicitly known, there remain nr elements of the transformation to be chosen subject to $r (n - r) + \frac{1}{2} r (r - 1)$ equations. They can always be so chosen if

$$nr \geq nr - \tfrac{1}{2} r (r + 1) ,$$

which is always true.

If the $M (x)$ are not given explicitly, we must use the $r (n - r)$ equations 3.6 and we have a total of $2r (n - r) + \frac{1}{2}r (r - 1)$ conditions to be satisfied by the n^2 elements, which again can always be done, together with the further natural requirement that the vector shall not be changed in magnitude under the transformation. This is equivalent to requiring $$\frac{\partial y}{\partial x} = 1 .$$

The equations $M (x) = 0$ are given only when certain constraints are imposed in the specification of the problem, for example defining the boundary of state space. When the optimal control is itself of such a nature that it forces the trajectories to lie in restricted regions , the form of the subspace is not known a priori, and indeed it may be difficult to determine it even with full knowledge of the optimal control function. The most familiar situation of this nature occurs in linear bang-bang control where trajectories may lie on a switching surface. In such a case we can only assume that tangent spaces can be constructed , i.e., that the switching surfaces are differentiable, and the transformation will be performed implicitly.

3.2.3 Solutions of differential equations.

We have airily proposed that the differentiability properties of $u(x)$ may vary in different regions of state space, and that solutions of the set of n differential equations of the dynamic system may be confined in some regions to subspaces of dimension less than n. Such a situation renders inapplicable some standard theorems concerned with differential equations, and these must be modified to some extent before we can proceed to the analysis of optimal systems.

Definition 3.1 : Let $f(x) = f(x^1, \ldots, x^n)$ be defined on some region G of n - dim. space R^n, and let x_0 be some point in G. $f(x)$ is continuous at x_0 in G if for all $e > 0$ there is some $d(x_0) > 0$ such that

$$\| x_1 - x_0 \| < d(x_0) \text{ implies } | f(x_1) - f(x_0) | < e \qquad 3.7$$

whenever x_1 is in G. $\left(\|x\| \overset{\text{def}}{=} \sup_i |x^i| \right)$.

Note that this definition allows x_0 to be a boundary point of G.

Definition 3.2 : $f(x)$ is uniformly continuous in G if for all $e > 0$ there is some $d > 0$ such that for all $x_0 \in G$

$$\| x_1 - x_0 \| < d \text{ implies } | f(x_1) - f(x_0) | < e$$

whenever x_1 is in G.

Straightforward extensions of these definitions are

Definition 3.3 : Let $H \subset R^r$, $r \leq n$, be a subset of G, and x_0 be in H : $f(x)$ is continuous at x_0 in H if 3.7 holds for x_1 in H ; Uniform continuity in H is defined by an analogous modification of definition 3.2.

If $f(x)$ is continuous in G it is continuous in H , for x_1 is certainly in G if it is in H ; the converse, however, is not necessarily true.

Other possibilities suggest themselves, for example, $f(x)$ may be continuous at x_0 in H if x_0 is a point of closure of H not contained in H , but we shall use them only if the need arises.

Definition 3.4 : $f(x)$ is differentiable in H if for all x , $x + \Delta x$ in H ,

$$f(x + \Delta x) - f(x) = a_i(x)\Delta x^i + a(x, \Delta x)\|\Delta x\| \qquad 3.8$$

where the product $a . \Delta x$ is finite for finite Δx (the repeated index summation convention is implied), and $a(x, \Delta x)$ tends to zero when Δx tends to zero.

If H is an open n-dim. set (i.e. $r = n$) this definition is equivalent to the usual one for differentiability (Berge 15 p.195) and under those conditions we merely say 'f(x) is differentiable' without qualification. Definition 3.4 allows x to be a boundary point of H , obviating the need for concepts such as right-or -left derivatives. The real strength of the definition is that the Δx are not arbitrary, but are restricted to a particular set; this has important consequences, as we shall see.

Definition 3.5 : $f(x)$ admits a partial derivative with respect to x^i if
$$\left[f(x^1, \ldots, x^i + h, \ldots x^n) - f(x)\right]\frac{1}{h}$$
tends to a limit when $h \to 0$.

It is an easy consequence of definition 3.4 that if $f(x)$ is differentiable (i.e. in an open region of R^n) then it admits continuous partial derivatives with respect to all x^i , but if $f(x)$ is differentiable only in H, then it may be that no partial derivatives exist, for $(x^1, \ldots, x^i + h, \ldots x^n)$ might not be in H for any x^i, for h however small. For example, $f(x)$ might admit a directional derivative along a curve, when H would correspond to the tangent, but a variation of any one component of x would take x out of H.

Now we can say that there is a partition X_i $i = 1,2\ldots$ of X such that $u(x)$ is differentiable in each X_i. That is to say, the state space X is defined by 2.4 b and 2.11 et. seq., and is divided into regions such that every point in X is contained in one and only one X_i. The dimension r of each region is constant throughout that region, but r may vary from region to region. The differentiability properties of $u(x)$ are clearly the same for $f(x, u(x))$, so that in practice f might admit no partial derivatives. This is inconvenient, for in practice the operation of differentiation can only be carried out coordinate by coordinate, which is disallowed here. However, we can show that by a suitable transformation of the tangent space of X_i certain partial derivatives can be guaranteed to exist.

Let $f(x)$ be differentiable at a point x_0 in H, H being r-dim. ('Differentiable at x_0' means that in definition 3.4 'for all x, $x + \Delta x$' is replaced by 'for x_0 and all Δx such that $x_0 + \Delta x$ is ... and in 3.8 x is replaced by x_0). A suitable transformation $\partial y/\partial x$ takes Δx to

$$\Delta y = (\Delta y^1, \Delta y^2 \ldots\ldots \Delta y^r, 0 \ldots\ldots 0) \qquad 3.9$$

If any component Δx^p is identically zero on H, it is convenient to choose a particular Δy^q to correspond:

$$\Delta y^q = \frac{\partial y^q}{\partial x^p} \Delta x^p$$

with $\partial y^q/\partial x^p = 1$, $\partial y^q/\partial x^m = 0$, $m \neq p$.

At the same time $a_i(x)$ (see 3.8) may be subjected to a covariant transformation

$$b_j = \frac{\partial x^i}{\partial y^j} a_i \quad .$$

Now,

$$b_j \Delta y^j = \frac{\partial x^i}{\partial y^j} a_i \frac{\partial y^j}{\partial x^i} \Delta x^i$$

$$= a_i \Delta x^i \qquad 3.10$$

If certain of the Δx^i in 3.8 are identically zero in H , the corresponding a^i (the partial derivatives) are not defined, although the product 3.10 is. Since it is an invariant of the transformation, the elements of the transformation may be chosen in any way compatible with 3.9, and the product $b_i \Delta y^j$ will remain defined.

Substituting 3.10 into 3.8 we have

$$f(x_o + \Delta x) - f(x_o) = b_j \Delta y^j + a(x_o, \Delta x) \|\Delta x\| \qquad 3.11$$

Strictly speaking, 3.9 and 3.10 are valid only for limiting values of Δx , which are precisely in the tangent space to the differentiable subspace X_i. Non-infinitesimal vectors Δx for which x, $x+\Delta x$ are in X_i are not in the tangent space at x, but have a projection $\Delta x'$ onto it, where

$$\Delta x' = \Delta x + e(\Delta x) \|\Delta x\|$$

where $e(\Delta x)$ tends to zero with Δx. A transformation of vectors in the tangent space gives

$$\Delta y^j = \frac{\partial y^j}{\partial x^i} \Delta x'^i$$

$$= \frac{\partial y^j}{\partial x^i} \left[\Delta x^i + e^i \|\Delta x\| \right]$$

$$b_j \Delta y^j = a_i \Delta x^i + a_i e^i(\Delta x) \|\Delta x\|$$

and it would be more correct to write 3.11 as

$$\Delta f = b_j \Delta y^j + \left[a(x_o, \Delta x) - a_i e^i \right] \| \Delta x \|$$

but the alteration would add nothing of significance since we are not con cerned with the exact form of the function multiplying $\|\Delta x\|$. In future vectors of the form Δx will be treated as tangent vectors without further comment.

Consider the index sets $p = r+1, \ldots, n$; $q = 1, \ldots, r$. The Δy^q are independent (cf. 3.9) so the corresponding partial derivatives f_{y^q} exist and are equal to $b_q(x_o)$, but the b_p, multiplied by zero components, are undefined. In general we can say that if $f(x)$ is differentiable in an r-dim. set it is possible to find a transformation such that $f(x)$ admits r partial derivatives f_y.

We are now in a position to tackle the differential equation $\dot{x} = f(x)$. An important theorem for the system of full dimension is that the solutions are differentiable with respect to initial conditions. This property is of prime importance to this work, so an analogue of this result must be proved for the conditions of state space. We begin by showing that solutions are continuous with respect to initial conditions.

Theorem: Let X be a region in n-dim. space, and G an r-dim. region contained in X. H is an r-dim. subset of G whose closure is in G. For every point z in H let there be a unique solution $y(z,o;t)$ (cf.2.3) of $\dot{x} = f(x)$ remaining in H for $t < t_z$. Suppose $f(x)$ is bounded and continuous in H. Then for all $e > 0$ there is a $d > 0$ such that

$$\| z_1 - z_2 \| < d \quad \text{implies} \quad \| y(z_1;t) - y(z_2;t) \| < e$$

for $t < \min (t_{z_1}, t_{z_2})$ and for z_1, z_2 in H.

Proof: Choose a point x_o in H , and some n-dim. neighbourhood D of x_o containing H. Let $E = D \cap G$ and $\alpha = \inf_{w \in E-H} \| w-x_o \|$. Consider the r-dim. box B: $\| z-x_o \| \leq \alpha$ $z \in E$. Every point z in B is in H, for the nearest point in the r-dim. subspace that is not in H is at a distance α from x_o. The equality is permitted only when the corresponding boundary point of H is actually in H.
Choose an interval $I = [0, T)$ such that every trajectory starting in B remains in H in the interval, e.g. if $y(z,0;t)$ reaches a point outside of H at $t = t_z$, let $T = \min_{z \in B} (t_z)$.

For $t < T$,

$$y(z,o;t) = z + \int_o^t f(y(z,o;s)) \, ds \qquad 3.12$$

remains in H, and, since f is bounded, y(t) is uniformly bounded and uniformly continuous. That is, for all $e > 0$ there is a $d(z) > 0$ independent of t such that

$$|t_1-t_2| < d(z) \text{ implies } \| y(z;t_1) - y(z;t_2) \| < e \qquad 3.13$$

This is true for any $0 < t_1, t_2 < T$ and each z, so that for e there must be some $d = \min_z \{ d(z) \}$ for which 3.13 holds uniformly in z. Then the family $y(z,o;t)$ issaid to be equicontinuous.

Ascoli's theorem (Coddington and Levinson 16 p.5) states that every infinite family $\{ g(t) \}$ uniformly bounded and equicontinuous on a bounded interval contains a sequence $\{ g_n(t) \}$ which is uniformly convergent on that interval.

As $z \to x_o$ we have such a family, and there is a convergent sequence of which every member, and therefore also the limiting member, satisfies 3.12 and so is a solution. By assumption, this solution is unique, so that all such sequences converge uniformly to $y(x_o,0;t)$ however z tends to x_o. This proves continuity at x_o in B.

The extension of this result to hold for all points in H is strightforward, for it applies, as it stands, to any point in B, and a similar box can be constructed centred on any point in B up to points for which $\alpha = 0$. Similarly, if $x_1 = y(z,0;t_1)$ $t_1 < T$, the proof applies for points in a box centred on x_1, with another suitably chosen T. The process is repeated for T arbitrarily small, thereby covering all points in H. We conclude that the solution of a differential equation is continuous uniformly with respect to the coordinates of its initial point as long as it remains in the same region of continuity of $f(x)$.

If $f(x)$ is differentiable in H, the solution will also be differentiable with respect to the initial conditions.

For, consider the solutions from two points z_o, z_1.

$$\begin{aligned} x_o(t) &= y(z_o, 0;t) = z_o + \int_0^t f(y(z_o;s))\,ds \\ x_1(t) &= y(z_1, 0;t) = z_1 + \int_0^t f(y(z_1;s))\,ds \end{aligned} \qquad 3.14$$

and restrict t to an interval within which the solution remains in H.

Let $x_1(t) - x_o(t) = \Delta x(t)$; $z_1 - z_o = \Delta z$.

Applying 3.8 to each element of x we have

$$\Delta x(t) = \Delta z + \int_0^t A(x(s))\Delta x(s) + a(x(s),\Delta x(s))\|\Delta x(s)\|\,ds \qquad 3.15$$

where the matrix A comprises the elements a_i of 3.8 for each component of f, and $a(x,\Delta x)$ is a vector each of whose components tends to zero with $\Delta x(s)$.

The solution of 3.15 is of the form

$$\Delta x(t) = g(\Delta z, 0;t),$$

a continuous function of Δz, so that 3.15 can be written in the form

$$\Delta x(t) = \Delta z + \int_0^t A(s)\Delta x(s) + \underline{a}(s,\Delta z)\|\Delta z\|\,ds \qquad 3.16$$

$\underline{a}$ being a function that tends to zero with Δz. Differentiating,

$$\Delta \dot{x} = A(t)\Delta x + \underline{a}(t, \Delta z)\|\Delta z\| \qquad 3.17$$

Consider the n-dim. vector equation

$$\dot{u} = P(t)\, u + p(t, u_0) \qquad 3.18$$

where u_0 is the initial value $u(o)$, and p tends to zero with u_0. 3.18 is of similar form to 3.17, and if n linearly independent solutions can be found, has the solution

$$u(t) = W(t,0)\left[u_0 + \int_0^t W(o,s)p(s,u_0)\, ds\right] \qquad 3.19$$

where W is the solution to the homogeneous matrix equation

$$\dot{U} = P(t)U$$

with initial condition $U(o)$ = unit matrix. In this form it is not quite comparable to 3.17, which does not admit n linearly indepent solutions. However, suppose 3.18 to be in fact r-dim. so that it includes $n-r$ degrees of redundancy. It can be transformed into

$$\dot{v} = Q(t)v + q(t, v_0) \qquad 3.20$$

where $v = (v^1, \ldots, v^n)$ and $v^{r+1} = \ldots = v^n = 0$,

by choosing a matrix $R(t)$ and a vector $r(t)$ such that

$$u = Rv + r$$

$$\therefore \quad \dot{u} = \dot{R}v + R\dot{v} + \dot{r}$$

$$= PRv + Pr + p(u_0)$$

$$\dot{v} = R^{-1}[PR - \dot{R}]\, v + R^{-1}\left[Pr - \dot{r} + p(u_0)\right] \qquad 3.21$$

$$\therefore \quad \dot{v} = Qv + q(v_0)$$

Using the indices ~~k = 1,...,r~~ $m = r+1, \ldots, n$ we see that if R, r are chosen to satisfy the differential equations

$$Q_{mi} = 0 \qquad i = 1, \ldots, n$$

$$q^m = 0$$

with initial values

$$v_o^m = \left\{ R^{-1} \left[u_o - r_o \right] \right\}^m = 0 \qquad 3.22$$

then the last $n - r$ equations in 3.20 become zero identities and the remaining equations form a normal r - dim. set with a solution of the form 3.19. There is no need to actually carry out such a transformation, but the knowledge that it is feasible enables us to write the solution of 3.17, supposedly transformed into normal r - dim. form, but without altering the notation; as

$$\Delta x = W(t, 0)\Delta z + \int_0^t \underline{a}(s, \Delta z)ds \|\Delta z\| \qquad 3.23$$

The integral tends to zero with Δz, uniformly in t, so that 3.23 conforms with the condition that x should be differentiable in H with respect to z, the initial condition.

We must now consider the situation when trajectories enter a neighbouring subspace. The two trajectories defined in 3.14 have initial points in a region H_1 and reach H_2, a space of possibly different dimension, at t_o, t_1 respectively. Let $t > t_1 > t_o$, and $x_1(t) - x_o(t) = \Delta x(t)$;

$$\begin{aligned}\Delta x(t) = \Delta z &+ \int_0^{t_o} A(x(s))\Delta x(s) + a(x, \Delta x)\|\Delta x(s)\| \, ds + \\ &+ \int_{t_o}^{t_1} f(y(z_1; s)) - f(y(z_o; s)) \, ds + \\ &+ \int_{t_1}^{t} B(x(s))\Delta x(s) + b(x, \Delta x)\|\Delta x(s)\| \, ds\end{aligned}$$

A, a, B, b, correspond to the elements involved in the definition of differentiability (3.8), the first pair applying in H_1, the second in H_2. The integral from t_o to t_1 involves two functions whose arguments are taken from different regions of space, for in that interval the first trajectory has already crossed the border. $f(x)$ may be discontinuous at such a point, and its values at t_o in H_1, H_2 respectively will be indicated by

$-$, $+$. Then

$$\int_{t_0}^{t_1} f(y(z_1 ; s)) - f(y(z_0 ; s))ds = \int_{t_0}^{t_1} f(y(z_1 ; s)) -$$

$$-f(y(z_0 , t_0^+)) - \dot{f}(y(z_0 , t_0^+))(s - t_0) + F(t_0 , s)\left|s - t_0\right| +$$

$$+f(y(z_0 , t_0^-)) - f(y(z_0 , t_0^-)) \ ds$$

(where $F \rightarrow 0$ as $s \rightarrow t_0$),

$$= \left[f(y(z_0 , t_0^-)) - f(y(z_0 , t_0^+))\right](t_1 - t_0) - \tfrac{1}{2}\dot{f}(y(z_0 , t_0^+))(t_1 - t_0)^2 +$$

$$+ \int_{t_0}^{t_1} F(t_0 , s)\left|s - t_0\right| + f(y(z_1 ; s)) - f(y(z_0 ; t_0^-)) \ ds \qquad 3.25$$

The last two functions in the intergrand take their values in the same region H_1 in which $f(x)$ is differentiable. Since y is differentiable with respect to both z and t , the integrand involves only terms of the order of magnitude of Δz and $(s - t_0)$. If the time of reaching the boundary is a continuous function $t(z)$, then as $\Delta z \rightarrow 0$ the only significant contribution to the discontinuity of $\Delta x(t)$ is

$$\left[f(z_0 , t_0^-) - f(z_0 , t_0^+)\right](t_1 - t_0) \qquad 3.26$$

and if f proves to be continuous at t_0 then the partial derivative $\partial x(t)/\partial z$ is continuous: otherwise the derivatives are continuous only within single regions, and after transition points $x(z)$ is not necessarily differentiable, since 3.24 does not have the form of a linear equation. If, however, $t(z)$ is differentiable, $(t_1 - t_0)$ can be written as a linear function of Δz together with terms of higher order, and $\partial x/\partial z$, though not continuous in time, does exist, and is the solution of

$$\frac{\partial x(t)}{\partial z} = E + \int_0^{t_0} A(s)\frac{\partial x(s)}{\partial z} ds + \left[f(t_0^-) - f(t_0^+)\right]\frac{\partial t_0}{\partial z} + \qquad 3.27$$

$$+ \int_{t_0}^{t} B(s)\frac{\partial x(s)}{\partial z} ds$$ etc. (E being the unit matrix).

If H_1 , H_2 are p-dim., q-dim., respectively, then the proper transformations make A a p-dim. row vector of partial derivatives, $\partial x(s)/\partial z$ a $p \times p$ matrix (in the first integral), $\partial t_0/\partial z$ a p-dim. vector, B a q-dim. vector, and $\partial x/\partial z$ in the second integral a $q \times p$ matrix.

3.3 Isotims.

3.3.1. Hypersurfaces of constant cost.

With these basic and crucial results established, we may return to the problem of optimal control. The equation $\dot{x} = f(x)$ forming the basis of the analysis above is in fact the equation satisfied by a dynamic plant under optimal control, where the control function can be expressed $u(x)$. In establishing this 'feedback' form in section 3.1 we agreed that a point in X is sufficient data (given the specification of the problem) to determine a control function, a trajectory, and, since the cost function depends only on these items, also a unique value of cost, which we shall write $J(x)$. Referring to 2.9 and 2.10

$$\begin{aligned} J(x) &= \min_{v(t)} P(x , 0 , v(t)) \qquad 3.31 \\ &= P(x , 0 , u(t)) \end{aligned}$$

Consider the points which satisfy the equation

$$J(x) = c$$

They form a set of points each of which has the same value of cost, and this set may appropriately be called an 'isotim' (Greek $\tau\iota\mu\eta$ = cost). The equation of the isotim expresses one degree of interdependence between the arguments of $J(x)$, the coordinates of the point x, so that the points in an r-dim. region which are on an isotim constitute an (r-1) -dim. subspace. As we shall see, $J(x)$ is piecewise differentiable and the isotims are hypersurfaces with a normal (not necessarily unique) at each point.

It is evident that every point in the state space lies on some isotim, and that isotims cannot meet, for this would imply that the single point has two values of optimal cost, which is ruled out by the assumption of uniqueness.

Whether the cost function is of the form 2.8a or 2.8b , it is clear that it can be expressed as the solution of a differential equation. This point was emphasized in section 2.2 together with the possibility of adjoining this differential equation to the dynamic equations of the plant. That particular step was rejected, but the differential-equation character of the cost function is not to be overlooked, and was in fact a major (though unmentioned) motive in deriving the results of section 3.2.

For both the Lagrange and Mayer cost function we may write an equation of the form $\dot{J}(x) = w(x)$

$w(x)$ having the same differentiability properties as $u(x)$, for it is really $w(x, u(x))$. Then

$$J(x_o) = \int_0^{t_f(x_o)} w(x(t))\, dt \qquad 3.32$$

with $x(t)$ taken along an optimal trajectory, i.e., $x(t) = y(x_o , 0 ; t)$. x_o is independent of t , and $J(x_o)$ will be differentiable if both the integrand and $t_f(x_o)$ are differentiable for x_o in some given region. $w(x)$ certainly admits certain partials, for $w(x , u)$ is designed by the specification of the cost function to be differentiable for x and u , and $u(x)$ is differentiable to a degree. The solutions $y(x_o ; t)$ have been shown to be differentiable so that under suitable transformations certain derivatives of the form $w_{x_o} = \left(w_x + w_u u_x\right) y_{x_o}$ will exist.

It remains to investigate $t_f(x_o)$. t_f is defined (see 2.7) as the first instant at which the solution reaches the terminal set T which is itself defined by a set of differentiable functions

$$T(x) = 0 . \tag{3.33}$$

Each such function represents an $(n - 1)$-dim. manifold, and if T is the intersection of s such manifolds it is $(n - s)$-dim. Immediately before reaching T the trajectories are in a region of dimension $p > n-s$, of which T is a boundary.

Consider trajectories starting at $t = 0$ from x_a, x_b, and let $x_b - x_a = \Delta x_a$, and let the region X_o containing both x_a and x_b be r-dim. They reach T at t_a, t_b at points $x_1 = y(x_a, 0 ; t_a)$, $x_2 = y(x_b, 0 ; t_b)$ respectively. Define $x_2 - x_1 = \Delta x_1$. For each component $T(x)$ of 3.33 there holds

$$T(x_1) = T(x_2) = 0 \tag{3.34}$$

and since $T(x)$ is differentiable,

$$T(x_2) - T(x_1) = T_z \cdot \Delta z + \tau(x_1, \Delta x_1) \|\Delta x_1\| \tag{3.35}$$

where a transformation $\Delta z = Z \Delta x_1$ expresses Δx_1 as an r-dim. vector so that $\Delta z^{r+1} = \ldots = \Delta z^n = 0$, and the $T_{z^1}, \ldots, T_{z^r}$ are partial derivatives.

$$\begin{aligned} \Delta x_1 &= \int_o^{t_b} f(y(x_b ; t))\, dt - \int_o^{t_a} f(y(x_a ; t))\, dt \\ &= \int_o^{t_a} f(y(x_b ; t)) - f(y(x_a ; t))\, dt + \int_{t_a}^{t_b} f(y(x_b ; t))\, dt \end{aligned} \tag{3.36}$$

3.34, 3.35, 3.36 give

$$T_z \cdot Z \left[\int_o^{t_a} f(y_b) - f(y_a)\, dt + \int_{t_a}^{t_b} f(y_b)\, dt\right] + \tau \|\Delta x_1\| = 0 \tag{3.37}$$

the notation being abbreviated in a self-explanatory fashion.

$y_b = y(x_b, 0; t)$ is a continuous function of t, and the mean value theorem applies in this case stating that for each component f^j there is some $t_a \leq t^{j'} \leq t_b$ such that

$$\int_{t_a}^{t_b} f^j(y(x_b; t))\, dt = f^j(y(x_b; t^{j'}))\left[t_b - t_a\right] \qquad 3.38$$

$\therefore$ 3.37 gives

$$t_b - t_a = \frac{-T_z . Z\left[\int_0^{t_a} f(y_b) - f(y_a)dt\right] + \tau\|\Delta x_1\|}{T_z . Z\, f(y(x_b; t'))} \qquad 3.39$$

The integrand can be expressed in terms of $\Delta y(t)$, hence in terms of Δz and t, and it is easy to manipulate 3.39 into a form corresponding to 3.8, showing that $t_f(x_o)$ is differentiable, admitting r partial derivatives. Assuming the proper transformations to have been made, we will have

$$J_{x_o} = \int_0^{t_f(x_o)} w_{x_o}\, dt + w(t_f)(t_f)_{x_o} \qquad 3.40$$

Since $J(x)$ is differentiable its partial derivatives will be continuous as long as x remains in one region of differentiability, but as the trajectories move from one region to another, the partials at boundary points might suffer discontinuities. We shall investigate this possibility. A similar problem arose in studying the properties of solutions of the equations at such boundaries, although the situation is not quite the same, for there the initial point was fixed and the solution moved, while here it is the initial point itself which moves.

Consider two sequences $\{x_i\}$, $\{x_j\}$ in X_o, converging to x_o, x_1 respectively in X_1. X_o, X_1 are r-dim., p-dim. respectively. 3.40 holds for each point x_i, x_j, x_o, x_1, but r partial derivatives exist for the points in X_o, and p for those in X_1. For the two points x_s, x_t

from $\{x_i\}$ $\{x_j\}$,

$$J(x_s) - J(x_t) = h(x_s).\left[x_s - x_t\right] + j(x_s, x_s - x_t) \|x_s - x_t\|$$
$$= J_z(x_s).\Delta z + j \|x_s - x_t\|$$

under a suitable transformation. Also,

$$J(x_o) - J(x_1) = J_w.\Delta w + j'(x_o, x_o - x_1) \|x_o - x_1\|$$

Δw being a vector in a p-dim. tangent space at x_o.

The difference

$$\left[J(x_s) - J(x_t)\right] - \left[J(x_o) - J(x_1)\right]$$

can be made as small as desired, as $\{x_i\}$ $\{x_j\}$ converge to x_o, x_1, therefore the difference between the right hand sides approaches

$$J_z(x_o).\Delta z - J_w(x_o).\Delta w + \left[j(x_o, x_o - x_1) - j'(x_o, x_o - x_1)\right] \|x_o - x_1\| = 0.$$

Since there are terms ofboth first and second order in Δx, it must be that $j = j'$, and

$$J_z.\Delta z - J_w.\Delta w = 0$$

Δz and Δw are transformations of the same vector $x_o - x_1$ but the former is the limit of a sequence of vectors in r-dim. tangent spaces, and the latter is p-dim. Suppose $r > p$ (we choose this for definiteness, but it could equally be $r \leq p$), then vectors in the p-dim. space corresponding to Δw can be embedded into an r-dim.space at the same point in such a way that the components $p + 1, \ldots, r$ are all zero. In particular Δw can be given r coordinates, and be made identical with Δz, when correspondingly $J_w \rightarrow J_z'$, and

$$(J_z - J_z').\Delta z = 0$$

demonstrating that when the components of Δz are not zero the expressions J_z, J_z' are equal and are the partial derivatives.

In many practical situations a 'switching surface' is the boundary between one n-dim. region and another, in which case $r = p = n$, and all derivatives J_x are continuous across the surface, indicating that the isotims suffer no discontinuities of slope, and a diagram of the isotims alone would not reveal the existence of control discontinuities of this kind. However, it may be that the trajectories remain on this surface, when we have $r > p$, and only p derivatives are continuous, and the isotims degenerate to $(p - 1)$-dim. hypersurfaces. Another way of expressing the continuity of certain partial derivatives is to say that that component of the isotim gradient which is tangential to the 'switching surface' (or whatever surface it happens to be) is continuous. The boundary of state space presents a similar situation, though most likely without any discontinuity of the control, and again the component of grad J tangential to the boundary is continuous, and other components are undefined on the boundary. This topic will be taken up again later on, but in the next section we shall find that the components of grad J play a central role in the determination of the necessary conditions for optimality.

At this stage we must distinguish between the Lagrange problem and the Mayer problem, for the cost function is differently defined, and the term 'points of same cost' leads to quite dissimilar constructions of isotims.

3.3.2 The Lagrange isotim.

2.8.6 defines the cost function in this case, and for an optimal control policy,

$$J(x) = \int_t^{t_f} L(x(s), u(x(s)))\, ds$$

where t corresponds to the point for which $J(x)$ is evaluated. The value of the cost changes from point to point along a trajectory and we have

$$\dot{J} = -L(x, u(x)). \qquad 3.41$$

Thus trajectories cross the isotims at a rate depending upon the control at the point, and if L is non-negative the isotims will be met in a monotonically decreasing sequence, reaching a value of zero at T, where $t = t_f$. The isotim $J(x) = 0$ contains the entire terminal set, though the reverse is not necessarily true, for there may be points not in T from which T can be reached with zero cost.

It usually turns out that L. is non-negative throughout the interval $[t_o, t_f]$, though exceptions are conceivable, especially when the admissible state space is peculiarly shaped. In such cases it might be possible to construct an equivalent cost function which is always positive, but which has the same optimisation properties as the original one. This might be achieved by adding to L a total differential,

$$L^1 = L + S_x . f(x,u)$$

$$\therefore \int L^1 dt = \int L dt + S(x(t_f)) - S(x(t_o)).$$

If such a function can be found which has a constant value over T and a constant value over the initial set of points so that the choice of optimal initial and final points is not affected, and if $S_x.f$ is sufficiently positive, L^1 will be positive but equivalent to L.

In the analogous situation of the classical calculus of variations for a cost integrand $L(x,\dot{x})$ a function $S(x)$ can be found to satisfy

$$L^1(x,\dot{x}) = L(x,\dot{x}) + S_x(x).\dot{x} > 0$$

if there exists some line element $(x_o,\dot{x}_o)$ at a point x_o where

$$L(x_o,\dot{x}) - L_{\dot{x}}(x_o,\dot{x}_o).\ \dot{x} > 0 \qquad 3.42$$

for all $\dot{x} \neq k\dot{x}_o$. * A condition of this nature forthe control problem is lacking, though the form of 3.42 suggests analogies. At the

* (Rund 17 p.5 gives no proof, but refers to Caratheodory 18 p.243.)

moment it is helpful to note that it is a condition which will usually be satisfied.

In an n-dim. region the isotims always have unique normals at each point but in a region of lower dimension this is no longer so. When a normal is not uniquely defined on a continuous manifold there is a 'ridge' at that point, and this is in fact the situation here. In an r-dim. region trajectories are always on a ridge of the isotim, though if a narrow view is taken, restricted only to the interior of that subspace the isotims are quite smooth- just as the corner of a box is a ridge in 3-dimensions, but merely a straight line if viewed from the background of, say, one wall. Fig 3 shows 3.dim. regions A,B separated by a 2-dim. surface C. An isotim has an edge in C, but an observer whose panorama is restricted only to C will see no 'edge' but a perfectly smooth curve. The components of $J_x(A)$, $J_x(B)$ in the tangent plane to C are equal to $J_x(c)$ (J_x is used here as a symbol for the normal vector.)

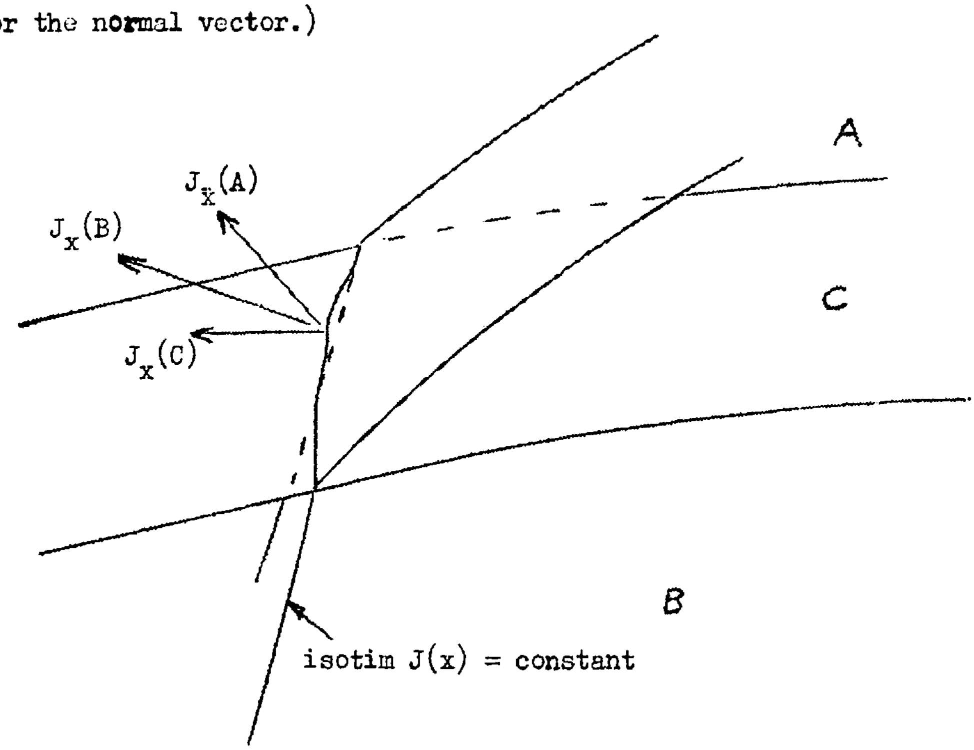


Fig.3

3.3.3 The Mayer isotim

The cost function, defined by 2.8a, is evaluated only at the terminal point, and is therefore constant with respect to all points on one trajectory. Along a trajectory, then, we have $\dot{J} = 0$, showing that the trajectory remains on the same isotim throughout its entire range. An isotim is in fact not merely a collection of points but a set of trajectories, forming a 'sheat' or perhaps a 'tube' in state space, which meets the terminal set. Fig 4 shows the reduction of a surface isotim to a single trajectory at a 2-dim. subspace. The tangential components of J_x are again continuous.

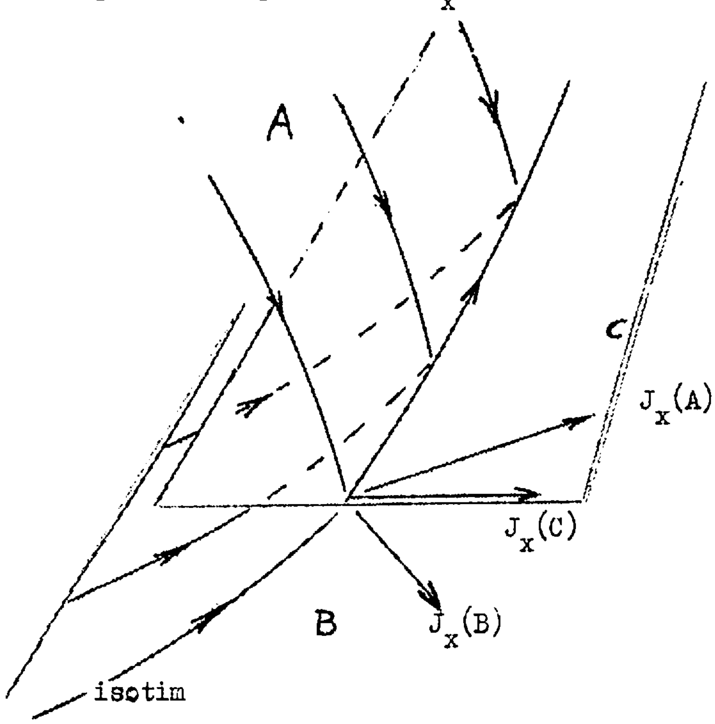


Fig 4

Each isotim divides the state space into two regions in which the cost is greater than and less than the value of the isotim. This apparently trivial observation points out a property which will prove to be quite profound and far-reaching.

Chapter 4. NECESSARY CONDITIONS FOR OPTIMALITY.

4.1. The Problem of Lagrange.

4.1.1. The minimum principle.

The cost function

$$P(x_1, t_1, v(t)) = \int_{t_1}^{t_f} L(x(t),\ v(t))dt \qquad 4.1$$

is evaluated for one point x_1, along the trajectory proceeding from x_1, t_1 with control $v(t)$ to the terminal set. * It is therefore a path integral. Along an optimal path the integrand $L(x,u)$ is a measure of the 'rate of descent' of the state x down the 'hill of cost', the isotims being the contours. A non-optimal trajectory also crosses isotims, but not at the same rate at which its optimal cost $J(x)$ decreases. We must find some expression for the rate at which arbitrary trajectories cross isotims.

Consider two points x_1, x_2 on the same trajectory.

$$x_2 - x_1 = \int_{t_1}^{t_2} f(x(t), v(t))dt.$$

According to the mean value theorem there is some t' such that

$$x_2 - x_1 = f\,(x(t'), v(t'))\,\left(t_2 - t_1\right)$$

$$t_1 \le t' \le t_2 .$$

Strictly speaking t' is not necessarily the same for each component of f, but that is no matter here.

$$J(x_2) - J(x_1) = p(x_1).(x_2 - x_1) + j(x_1, x_2 - x_1)\ \| x_2 - x_1 \|$$

since $J(x)$ is differentiable.

$$\frac{J(x_2) - J(x_1)}{t_2 - t_1} = p(x_1).\ f(x(t'), v(t')) + j(x_1, x_2 - x_1)\ \frac{\| x_2 - x_1 \|}{t_2 - t_1}$$

*(In this section trajectories are not optimal unless specifically designated such.)

$$= p(x).f(t') + j . \sup \left| f^i(t') \right|$$

therefore, letting $t_2 - t_1 \to 0$,

$$\dot{J} = p(x).\ f(x,v) \qquad 4.2$$

It must be emphasized that f is a contravariant vector in the tangent space at x, and therefore transforms precisely according to the usual rules. $p(x)$ under these conditions transforms to give the partial derivatives in restricted regions, and in n-dim. regions $p(x)$ is identically $J_{\dot{x}}$.

Any trajectory from x_1 on $J(x_1)=c_1$ reaches, after an arbitrary interval $[t_1, t_2]$, some point x_2, $J(x_2) = c_2$.

The cost over this interval is

$$\int_{t_1}^{t_2} L\ (x,v)\ dt,$$

and since it is not optimal,

$$c_1 \leq \int_{t_1}^{t_2} L(x,v)dt + c_2 \qquad 4.3$$

That is to say, the optimal cost from x_1 cannot be greater than the cost of a trajectory which is non-optimal for an arbitrary interval and thereafter optimal. Now

$$\begin{aligned} c_2 - c_1 &= J(x_2) - J(x_1) \\ &= \int_{t_1}^{t_2} \frac{dJ}{dt}\ dt \end{aligned} \qquad 4.4$$

evaluated along th said trajectory. Using 4.2, 4.3,

$$\int_{t_1}^{t_2} L\ (x,v) + p(x).\ f(x,v) \quad dt \geq 0$$

but the interval $[t_1, t_2]$ is arbitrary, so that

$$L(x,v) + p(x).\ f(x,v) \geq 0. \qquad 4.5$$

Along an optimal trajectory the equality will hold, and $v = u$. If we define

$$H(x,v) = L(x,v) + p(x).f(x,v) \qquad 4.6$$

then the optimal control satisfies

$$H(x,u) = \min_v H(x,v) = 0 \qquad 4.7$$

In words, the minimisation operation 4.7 means 'for fixed x find that value of v for which the function H takes a value less than that for any other admissible value of v'. 'Admissible' means a value which does not violate the constraints specified in formulating the problem, for example 2.4b, 2.5.

It is also possible to describe 4.7 as a minimization with respect to x : ' for fixed u find that point x for which the function H takes a value less than that for any other admissible point'. We confirm easily that this is so, for at a point $x + \Delta x$ the corresponding optimal control is $u(x+\Delta x)$. If we evaluate H at $x+\Delta x$, but retaining the value $u(x)$, H will not take its minimum value, for $u(x)$ is not optimal at $x+\Delta x$, so that

$$H(x+\Delta x, \ u(x)) \geq H(x+\Delta x, \ u(x+\Delta x)) = 0 \qquad 4.8$$

We can now extend 4.7 to the elegant expression

$$\min_{v,x} H = 0 \qquad 4.9$$

which contains the most essential necessary conditions and the tools with which to construct optimal control functions. 4.9 may be called the 'minimum principle'.

The most significant steps in the analysis are 1) the removal of the integration sign at 4.5, and 2) the recognition that $H(x,u)$ is a minimum for x, with u fixed. The former has the effect of reducing the problem from that of the minimization of a functional with an associated differential system to the minimization of a function - a much simpler problem which can be solved either by ordinary calculus or, often, by inspection. In terms of the behaviour of trajectories, we find at each point the control which takes the trajectory in the optimal direction with respect to the isotim.

$p.f$ is after all only a measure of the angle between the isotim normal and the velocity vector if the actual magnitudes $\|p\|$, $\|f\|$ are disregarded; if the magnitudes are considered significant 4.5 implies optimization of the descent rate of the trajectory, ensuring that $\dot{J} = -L$, for $J_x = p$ indicates not merely a normal direction but also an 'isotim density' or gradient.

The other crucial step, minimization with respect to x, we shall now develop further. 4.8 implies

$$H(x+\Delta x, u(x)) - H(x, u(x)) \geq 0 \qquad 4.10$$

We have, for all x, $x+\Delta x$ in X,

$$\begin{aligned} L(x+\Delta x,u) - L(x,u) &= L_x.\Delta x + K(x,x+\Delta x)\|\Delta x\| \\ f(x+\Delta x,u) - f(x,u) &= f_x.\Delta x + F(x,x+\Delta x)\|\Delta x\| \end{aligned} \qquad 4.11$$

for L, f are differentiable, by hypothesis, in anopen region containing X. Although 4.11 holds throughout X wewould prefer, as usual, to transform the velocity vector to separate its components in the tangent space from the zero components directed out of that space. Unfortunately that cannot be done at this stage, for the vector $f(x+\Delta x, u(x))$ is not optimal, and does not necessarily lie in the local manifold of optimal trajectories.

4.10 and 4.11 give

$$\begin{aligned} L_x(x,u).\Delta x + p(x+\Delta x).\left[f_x(x,u)\Delta x + F\|\Delta x\|\right] + \\ + \left[p(x+\Delta x) - p(x)\right].\ f(x_1u) + K\|\Delta x\| \geq 0 \end{aligned} \qquad 4.12$$

for any x, $x+\Delta x$ in the state space.

If we consider the effect of an arbitrary variation of one component Δx^i we will have to deal with the expression

$$\frac{p(x+\Delta x^i) - p(x)}{\Delta x^i} . f(x,u) \qquad 4.13$$

which does tend to a definite limit. This does not imply that the vector $\left[p(x+\Delta x^i) - p(x)\right] / \Delta x^i$ tends to a limit, and indeed it cannot do so in general, for that would mean that every component of p admits continuous partial derivatives, and a fortiori is itself continuous throughout the interior of X. Since p is only continuous in local regions that conclusion must be false, and 4.13 in fact seems to offer little information.

But let Δx be a special variation, in the direction of the optimal trajectory, $\Delta x = hf(x,u)$, and suppose x to be a point of continuity of $u(x)$. Then

$$(L_x + pf_x).f(x,u) + \frac{1}{h}\left[p(x+\Delta x) - p(x)\right].f(x,u) = 0$$

ignoring small quantities and noting that the inequality in 4.12 is inadmissible since the scalar h may be positive or negative. Then

$$\left[L_x + pf_x + \frac{\Delta p}{h}\right] \cdot f(x,u) = 0 \qquad 4.14$$

The quantity in brackets is a vector parallel to f. It must therefore be zero, and $\Delta p/h$ tends to a limit, which can be identified with $\dot{p}$. Thus for each component,

$$L_{x^i} + pf_{x^i} + \dot{p}_i = 0 \qquad 4.15$$

At this point the expected transformation is possible. The equations 4.15 are analogous to the Lagrangian equations of classical dynamics which are known to be invariant for local point transformations. The transformation is performed explicitly in Appendix B, with the result that

$$L_{z^i} + q_j(g^j)_{z^i} + \dot{q}_i = 0 \qquad 4.16$$

$i,j=1,\dots,r$, the local subspace being r-dim. The only variables involved in 4.16 are those relevant to behaviour as observed from within the local manifold, the q_i, q_j being uniquely defined partial derivatives. The remaining n-r variables q_k are undefined.

At the boundary with an adjacent region the partial derivatives are continuous if they are equivalently defined in both regions, that is to say if the dimension of the regions is the same and the transformations correspond component by component. If the second region has a lower dimension certain components of J_z must be dropped, the others remaining continuous: if a higher dimension new components must be added from that point.

4.1.2 Boundary conditions

A differential equation without boundary values is a feeble thing, as far as applications are concerned, unless a general analytic solution is available (an event, in control, most conspicuous by its almost invariable non-occurrence), for numerical solution is required, and this is a process which cannot begin without initial values. Fortunately it is always possible to produce enough boundary values to solve the dynamic equations and the auxiliary equations 4.16. Suppose the initial set S is defined by $n-r$ relations $S^j(x) = 0 \quad j = r+1,\dots,n$. S is r-dim. and $J(x)$ will have at least r-partial derivatives defined there; ('at least r' because S is a boundary to a region of dimension probably greater than r, and we are concerned with behaviour as x approaches S rather than at S itself). The optimal initial point x_o has the property that for all $x_o+\Delta x$ in S,

$$J(x_o+\Delta x) - J(x_o) \geq 0.$$

$$\therefore \quad J_z . \Delta z \geq 0$$

and since all vectors Δz are admissible,

$$J_{z^i} = 0, \quad i = 1,\dots,r. \qquad 4.17$$

At the terminal set a similar result holds, for T coincides with the zero isotim, and for all $x_f + \Delta x$ in T there holds

$$J(x_f+\Delta x) - J(x_f) = 0, \text{and } J_{z^k} = 0 \quad K=1,\dots,S \text{ , where T is S-dim.}$$

There remains the question of the new components of J_z introduced when a trajectory moves from a region X_1 of higher to X_2 of lower dimension. These have to be given values when they first appear. Suppose the transition point is x_1 in X_2. In the limit as x approaches x_1 from X_1 let the optimal control be u_1, and let it be u_2 at x_1 in X_2. Suppose the dimensions of X_1 X_2 are r,s respectively, $s > r$. Equating $H(x,u)$ in X_1 and X_2 we have

$$L(x_1 u_1) + J_{z^i} \cdot g^i(x_1,u_1) = L(x_1,u_2) + J_{z^i} \cdot g^i(x_1,u_2) + \\ + J_{z^j} \; g^j(x_1,u_2) \qquad 4.18$$

$i = 1,\ldots,r \quad j = r+1,\ldots,S$; g is the transformed velocity vector.

$$\therefore J_{z^j} \, g^j(x_1,u_2) = L(u_1) - L(u_2) + J_{z^i} \cdot \left[g^i(x_1,u_1) - g^i(x_1,u_2) \right]$$

for corresponding components of J_z are continuous. If $s-r=1$, this equation can be solved for J_{z^s}, otherwise there is no means of providing for the new components. This difficulty will be discussed further when we come to deal with computational methods.

4.2 The Problem of Mayer

4.2.1 Reachable sets.

The cost function $g(x(t_f))$ is evaluated only at the terminal point; in words, a Mayer problem requires the trajectory to move to that point in T for which the function $g(x)$ is least. The cost does not, apparently, depend upon the path taken to that point. Obviously this is quite a different requirement from that of the Lagrange problem, and indeed it is rather more subtle in its implications, and some interesting properties of optimal systems can be deduced from the mere statement of the problem.

The terminal set T is known a priori, and it is possible to evaluate

$g(x)$ over the whole of T without reference to the dynamic system. Suppose the point $x_f \in T$ gives

$$g(x_f) = \min_{x \in T} g(x)$$

At first sight it seems possible to restate the problem 'find a control for which the solution $y(x_o,o;u(t))$ reaches the point x_f'. This involves no optimization and is simply a boundary value problem, albeit difficult to solve. A moment's consideration of any typical Mayer problem — time optimality to the origin, maximum orbital velocity, etc. — shows that in general this naive interpretation overlooks a crucial fact, viz., that the apparently optimal point cannot be reached by the system. Such an interpretation will usually propose a point at infinity if no constraints prevent it. This makes it clear that there are certain points that cannot be reached, however $u(t)$ is chosen, and however great the time interval. There are some exceptional systems for which this is not so - completely controllable systems (Kalman 9) have the property that any point can be transferred to any other in finite time. For these systems the superficial interpretation is correct, but the system must be hedged about by control or state constraints to make it sensible. In general, however, we may define, for any point x, a set of points $R(x)$ which are reachable from it. The correct restatement of the Mayer problem would then read: ' find a control for which the solution of $\dot{x} = f(x,u)$ from x_o reaches the point in $R(x_o) \cap T$ at which $g(x)$ takes its least value '. There is, if our assumption of existence and uniqueness holds good, a unique terminal point corresponding to everypoint in X and therefore a unique value of optimal cost for every point.

We have, then,

$$J(x) = g(x_f); \qquad x_f = y(x\ ,\ 0\ ,\ u(t)\ ;\ t_f)$$

$u(t)$ being the optimal control function. We have seen that all trajectories cross isotims at a rate

$$\dot{J} = p(x).\ f(x, v)$$

(4.2). Suppose that at some (x_1, t_1) $\dot{J}$ is negative.

Let $x_2 = x_1 + f(x_1, v(t_1))\delta t$. Then

$$J(x_2) = J(x_1) + \dot{J}\delta t$$
$$< J(x_1)$$

implying that from x_2 it is possible to reach a point of T for which the cost is less than for points reachable from x_1. But if x_2 is reachable from x_1, so is that terminal point, in which case $J(x_1)$ was not optimal. We can only conclude that x_2 is not reachable from x_1, and there is no admissible control vector which can make $\dot{J}$ negative. We know, however, that on an optimal trajectory $\dot{J} = 0$ (section 3.3.3). Each isotim, then, is a boundary between the reachable and unreachable zones, and the optimal trajectory is always at the very limit of what is attainable. This interpretation gives a sharper edge to the term 'optimal'.

4.2.2. <u>The minimum principle.</u>

It is more useful to compare the optimal trajectory with other possibilities than with impossibilities, and it is the properties of the optimum as a member of the attainable set of trajectories that enables an immediate 'minimum principle' to be derived for the Mayer problem.

At t_1 let $\dot{J}$ be positive. Then

$$J(t_1 + \delta t) > J(t_1)$$

and since $\dot{J}$ can never be negative, the optimal terminal point previously reachable is now beyond our scope. The moral is that $\dot{J}$ must not be permitted to take positive values.

Defining $H(x\ ,\ v)\ =\ p(x).f(x\ ,\ v)$ 4.19

(cf. 4.6), the optimal control u must satisfy

$$H(x,\ u)\ =\ \min_{v}\ H(x,\ v)$$
$$=\ 0$$

The discussion following the analogous result for the Lagrange problem applies without modification except the removal of $L(x\ ,\ u)$. 4.16 holds here in the form

$$\dot{J}_{z^i}\ +\ J_{z^m}\ g^m_{z^i}\ =\ 0 \qquad 4.20.$$

$i\ ,\ m\ =\ 1\ ,\ldots\ldots\ ,\ r.$

4.2.3. Boundary conditions.

Boundary values for 4.20 are found in a very similar fashion as for the Lagrange problem. For the initial set the argument is identical; for the terminal set we have identically

$$J(x_f)\ =\ g(x_f). \qquad 4.21$$

Since this is an identity, it follows immediately that $J_z\ =\ g_z$, when the derivatives exist.

4.3 Construction of the Solution.

4.3.1 The interior of state space.

The equations 4.16 , 4.20 are sufficient to describe the behaviour of the local gradients as the trajectory traverses the 'hill of cost', but are not in a satisfactory form to solve and determine the optimal control. Their development depended upon particular transformations which cannot be chosen in advance except in special cases where the local subspace is known, such as on the boundary of state space, and there too it cannot be assumed that other surfaces will not unexpectedly appear and have to be dealt with. To overcome this difficulty it is necessary to be able to write the equations in such a form that the solution is unaffected whether the transformation is done or not, though where a suitable transformation is known, it is better to apply it.

In fact, we already have the equations in a suitable form. 4.15 is a set of n differential equations from which the local auxiliary equations 4.16 were derived. By inverting the transformation we regain 4.15. The process is entirely analogous to that whereby an r-dim. contravariant vector was expressed, in chapter 3, in a form involving n coordinates. In Appendix B it is shown that the auxiliary equation transforms as a covariant vector, and, seeing that 4.14 is an invariant, we need only augment the r-dim. velocity vector to an n-form in the usual way, adding n-r zero components, and similarly augment the r-dim. 'gradient' vector 4.16 with n-r <u>arbitrary</u> components, and a point transformation will give the n equations 4.15.

It may be objected that we have merely returned to the starting point, wasting considerable ingenuity and effort, but in fact we have gained an enormous insight into the structure of the solution. Furthermore, when the local subspace is known there is no need to retain all n equations, and we may work more conveniently in a reduced space.

To confirm that we have sufficient information to construct the solution to the optimal control problem, let us follow a trajectory along its entire range, assuring ourselves that every predicament met with can be satisfactorily handled. It is convenient to use the more economic notation of the Mayer problem which in fact covers both types of problem, for one of the components of x can be regarded as equal to cost, with an associated p component equal to either zero, for the Mayer problem, or to unity, for the Lagrange problem.

The initial point x_o is chosen to satisfy the condition 4.17. Since this must be applied without a transformation we must repeat the argument of that section for the untransformed variables. x_o has the property that

$$dJ(x_o) = p(x_o) \cdot dx = 0 \qquad 4.22$$

for all $x_o + dx$ in S. That is, we have S defined by the r equations

$$S(x_o) = 0 \qquad 4.23$$

and dx must satisfy

$$S_x(x_o) \cdot dx = 0 \qquad 4.24.$$

This allows r components of dx to be expressed in terms of the remaining $n-r$, which are then arbitrary, and whose coefficients in 4.22 are zero. This, together with the r equations 4.23 give a total of n conditions for the required $2n$ initial values x_o, p_o.

The remaining n conditions will be found later, but proceeding with the trajectory into the first subspace the set of auxiliary equations 4.15 together with the dynamic equations $\dot{x} = f(x,u)$ are supplied with values of control obtained from the principle $\min_v H(x,v)$, which is a separate operation for each component of control.

At the boundary between two regions the question arises of the continuity of the variables p. We have seen in section 3.3.1 that if corresponding components of the partial derivatives of $J(x)$ are defined in neighbouring regions, these components will be continuous. If the trajectory moves into a region of the same dimension then that part of p which, when suitably transformed, corresponds to the partial derivatives, is continuous. The remaining part, since it is undefined in both regions, can be made continuous. It is not possible in practice to distinguish between the defined and arbitrary components of p, so they must all remain continuous.

If the second region is of lower dimension than the first (say $h < g$), then $(g-h)$ derivatives of $J(x)$ cease to exist, and again may arbitrarily be set to be continuous. When, on the other hand, $h > g$, $(h-g)$ derivatives cease to be arbitrary, and take uniquely defined values. There is one relation applicable here, viz., 4.18, which in terms of the untransformed variables simply expresses the continuity of $H = p \cdot f(x, u)$.
In terms of the transformed variables 4.18 reads

$$J_{x^i} \cdot (f_1^i - f_2^i) - J_{x^j} \cdot f_2^j = 0 \qquad 4.25.$$

where $_1$, $_2$ indicate values measured in the first and second regions respectively at the transition point, and J_{x^j} are the new variables introduced in the second region but arbitrary in the first.
($i = 1, \ldots, g$; $j = g + 1, \ldots, h$).

If $u = u(x, J_x)$, determined from the operation $\min\limits_{v} H$, is substituted into 4.25, there results a relation between J_{x^i} and J_{x^j}. It is impossible to say in general what information this gives, for it depends upon the function $u(x, J_x)$, but if $h - g = 1$, 4.25 could be solved for the single variable $J_{x^{g+1}}$. It is doubtful, however, whether even this result is available in practice, for there will be no indication whether such transition points occur at all, since we are dealing with subspaces which emerge from the structure of the solution and are not known in advance.

Regretfully we must conclude that the local partial derivatives cannot be determined in such a case. This does not prevent the auxiliary equations from being solved, for p is susceptible of another interpretation which is more amenable to treatment----that of a directional derivative.

A directional derivative is defined in exactly the same way as derivatives in a restricted set (definition 3.4), except that the set of points x, $x + \triangle x$ must lie on a curve --- a one-dim. manifold. In this case it is the set

$$x = x_o + \int_o^t f(x, u)\, dt. \qquad t \in [0, t_f]$$

Since partial derivatives are continuous from region to region as long as they are defined, the single derivative in the direction of the optimal trajectory is certainly defined everywhere, always being tangential to the local manifold, and is therefore absolutely continuous.

p has been regarded as comprising two components, one tangential and one normal to the local manifold. If we now consider the more restricted interpretation of a component in the direction of the trajectory and one normal to it, the former is always continuous, and in the n-coordinate form the solutions of 4.15 will remain continuous even in the unusual case of a transition from lower to higher dimensional regions. The equations can now be solved, but we have lost the uniqueness of the h partial derivative in an h-dim. region. In practice it can always be confirmed whether or not this situation has arisen, by examining the trajectories after solving the equations.

Whenever the trajectory is in an unrestricted region in the interior of state space, p will be continuous, and all that is needed in practice are the $2n$ differential equations, the $2n$ boundary values, and some specification of the interval $[0, t_f]$. n of the boundary values have already been found, the others arise by using a similar argument at the terminal set, where, if T is described by q equations

$$T(x) = 0$$

there holds also the q equations

$$dT = T_x . dx = 0 \qquad 4.26$$

enabling q components of dx to be eliminated. For the Lagrange problem the coefficients of the remaining $n-q$ components of dx are zero, in the equation $dJ = p . dx = 0$,

giving a total of n relations as required. For the Mayer problem the coefficients of those components are zero in

$$(p - g_x) . dx = 0 , \qquad 4.27$$

using 4.21.

There remains to be determined the terminal time t_f . Fortunately there is also a relationship not yet used, viz., $H(x , u) = 0$. This need be applied only at one point, for $H(x , u)$ is constant if u is optimal. In order to prove this, we must show that a) H is absolutely continuous, b) $\dot{H} = 0$ almost everywhere.

Using the definition 4.6 or 4.19 , H may be regarded as a function of the three variables x , p , u . This does not negate the dependence of p on x , though not expressing it explicitly, and is an enormous simplification.

At a point of continuity of $u(t)$, H is evidently continuous, for $p(t)$, $x(t)$ are. At a point t' of discontinuity, suppose the right and left limits of u are u^+ , u^- . x , p are continuous, therefore $x^- = x^+$, $p^- = p^+$. u^- is chosen to minimize H , therefore

$$H(x , p , u^-) \leq H(x , p , u^+) .$$

Similarly u^+ is chosen to minimize H, giving the reverse inequality, hence $H^- = H^+$, and H is absolutely continuous.

Let the minimum value of $H(x , p , u)$ be $\underline{H}(x , p)$. Let $u(t)$, the optimal control, be continuous at t'. At $t \neq t'$,

$$\underline{H}(t) = H(x(t) , p(t) , u(t)) \leq H(x(t) , p(t) , u(t'))$$

$$\therefore \underline{H}(t)-\underline{H}(t') \leq H(x(t), p(t), u(t')) - H(x(t'), p(t'), u(t')) .$$

Letting t approach t' from the right, so that $t - t' > 0$, and passing to the limit,

$$\therefore \dot{\underline{H}} \leq H_x \cdot f + H_p \cdot \dot{p} \Big|_{t = t'} = 0$$

Repeating for $t - t' < 0$, the sense of the inequality is reversed, and we conclude that $\dot{\underline{H}} = 0$ at points of continuity of u, i.e., almost everywhere. $\underline{H}$ is therefore constant, and equal to zero throughout.

When this condition is used, the total of unknowns is equal to the number of conditions to be satisfied, and the problem can be solved without redundancy.

4.3.2. Boundaries of state space.

If the trajectory is at any time on a state space boundary, the situation is no more complicated in principle, though it is in practice. Suppose the boundary is described by the single equation

$$C(x) = 0 \qquad 4.28$$

If the trajectory remains on it for finite time, there holds

$$\dot{C}(x) = C_x \cdot f(x , u) = 0 . \qquad 4.29$$

4.29 is a constraint on the control variables and must be satisfied when $H(x , u)$ is minimised. If it occurs that 4.29 does not involve u then we have $\ddot{C}(x) = 0$

and similarly for higher derivatives. Writing $\frac{d^k}{dt^k} C(x) = C^{(k)}(x)$, a q'th order boundary gives

$$C(x) = C^{(1)}(x) = \dots\dots = C^{(q)}(x, u) = 0 \qquad 4.30$$

the q'th derivative being the first involving u.

The q conditions

$$C(x) = \dots\dots = C^{(q-1)}(x) = 0 \qquad 4.31$$

describe an $(n-q)$-dim. manifold, for which an explicit transformation gives

$$\dot{z}^k = C^{(k)}_{x^i} \cdot f^i$$

$k = 0, \dots, q-1$; $i = 1, \dots\dots, n$, the remaining velocity variables being chosen, as usual, mutually orthogonal, orthogonal to the $\dot{z}^k$, and such that the determinant of the total transformation should be unity. On the boundary only the $(n - q)$-dim. system need be considered, for which J_{z^j}, $j = 1, \dots, n - q$ is continuous at the transition point.

The time at which the boundary is reached must also be found, and as before there is a single relation to fill the gap, not, this time, $H = 0$, but 4.25, expressing the continuity of H. In 4.25, suffixes $_1$, $_2$ represent values at the terminal point measured on the boundary and in the interior of X, respectively, for the direction of motion is from a region of higher to one of lower dimension. Along the boundary the original system equations must still be retained, for the transformed system does not give the actual value of the state variables. Recalling that f, not x, has been transformed, we see that only operations or constraints involving f and J_x, e.g., min H, can be dealt with in the transformed system. This implies that the differential equations for z need not actually be integrated.

At some point the trajectory returns to the interior of X , say an n-dim. region, (this will be the most common occurrence). The q components of J_x excluded from consideration on the boundary, have to be reintroduced, and it is best to apply the inverse transformation at the transition point to restore the system to its natural form. In this case the difficulty of finding the correct value of p cannot be avoided by giving it a spurious continuity, for these components had been discarded completely along the boundary. There are q extra unknowns, introduced because of the boundary, but there are also q extra conditions 4.31 to be satisfied. The remedy, in fact, appeared before the ailment, and this time there is no difficulty. Again, 4.25 supplies the extra piece of information needed to determine the instant of exit from the boundary.

It is not suggested, in using expressions involving $H(x , u)$, that they really have direct reference to the time t_f or the time of reaching a boundary ---they do not. The important fact is that the total number of unknowns must equal the total number of constraining conditions, and it is simpler to correlate them in this way, as a matter of convenience, without implying a real connection. Similarly the boundary conditions 4.31 do not involve an explicit reference to the unknown components of p with which they were associated above. The real meaning of the 'corner condition' 4.25 is that it acts as a link between the two parts of the trajectory before and after that point. Ideally it would be nice to separate entirely the different sections of the trajectory, and deal with each in isolation, were it not that the behaviour of one affects other arcs. This corner condition expresses in simple form the complex interaction between the two parts of the trajectory. To digress for a moment on this point; it would be legitimate to divide the trajectory into a number of isolated subarcs only if

$$\min \sum_i \int_{t_i}^{t_{i+1}} L(x,u)dt = \sum_i \min \int_{t_i}^{t_{i+1}} L(x,u)\,dt \qquad 4.32$$

where successive arcs are taken over $[t_o, t_1)$ $[t_1, t_2)\ldots$, and the Lagrange notation is used. The effect of the corner condition is that 4.32 is always valid if the right hand side is subjected to the restriction that $H=p.f$ is continuous at each junction t_i.

This process is entirely analogous to regarding the cost function integral as an infinite sum, and minimising each summand separately :

$$\min \int_{t_o}^{t_f} L(x\ ,\ u)dt = \int_{t_o}^{t_f} \min L(x\ ,\ u)dt \qquad 4.33$$

which again is valid as long as the right hand side is restricted by $\dot{x} = f(x\ ,\ u)$, which relates successive values of x. In Appendix C a crude but suggestive derivation of the auxiliary equations is carried out on this basis.

State constrained problems raise no particular difficulties; in fact, as should be expected, they simplify the situation by reducing the state space. (cf.Bellman and Dreyfus 19 p.20).

4.3.3. Singular Trajectories.

A possibility not yet treated is that of surfaces of explicitly-given form arising in the interior of state space. Evidently they can be treated in the same way as boundary surfaces, by performing explicit transformations and discarding redundant variables. It may be that they are specified in the formulation of the problem, as constraints, and then their effect is not significantly different from that of the usual constraints, but this situation is rare; a more familiar phenomenon is that associated with singular surfaces.

This is one of a number of cases in which the minimum principle and similar techniques break down, either because there is no unique solution (or no solution at all) for the problem in that form, or through some shortcoming in the technique. Singularity is an example of the latter, and occurs when the minimisation of H does not provide a unique value of u. Typically, this occurs when H is linear in u :

$$H = h(x, p) + g(x, p)u = 0 \tag{4.34}$$

If at some point $g(x,p) = 0$, H is insensitive to u, and the minimisation is worthless. If this happens at an isolated point it causes no trouble, for $u(t)$ may well be undefined on a set of zero measure, but when a finite interval is involved difficulties arise.

This problem has only recently received serious attention. Kreindler (94) remarked on its relation to the flatness of the reachable set boundary; Johnson & Gibson (20) noted that if $g(x) = 0$ then, from 4.34, also $h = 0$ and furthermore $\dot{g} = \ddot{g} = \ldots = 0$, $\dot{h} = \ddot{h} = \ldots = 0$ until u appears as an argument, just as in the case of state boundaries. Each of these relations defines a surface in the 2n-dim. phase space of (x,p), but only a relation involving x alone defines a surface in X. Techniques introduced by Faulkner (93) and Kelley (21) use special coordinate transformations, in a spirit akin to that of transformation theory in analytical dynamics, seeking more easily integrated forms of the equations, rather than a reduction in state space, though this, too, was suggested by Kelley.

These arguments provide a method for constructing singular controls, but give no indication as to their optimality. This property was investigated for a particularly simple problem by Wonham and Johnson (24) and a pertinent test, based on the Clebsch condition, is given by Kelley (58), and there are other techniques (Snow 23, Than 49) designed to obtain reasonable solutions in these and similar degenerate cases rather than to understand the fundamentals of the situation. The true nature of singularity is just beginning to emerge (Hermes 22) in work based (as we have come to expect) on ideas of Carathéodory's, and appears to be connected with optimal accessibility and controllability.

4.3.4. Summary.

In brief, the solution of the optimal control problem involves the following principal stages.

If the trajectories remain entirely in an unrestricted region of stage space the n system equations

$$\dot{x} = f(x,u) \tag{4.35}$$

together with n auxiliary equations

$$\dot{p} = -L_x - pf_x \tag{4.36}$$

are solved to satisfy the $2n+1$ conditions imposed by the initial and terminal sets, the boundary conditions derived therefrom, and $H(x,u) = 0$. The solutions of 4.36 are continuous throughout and, in an n-dim. region represent the partial derivatives of cost J_x. In regions of reduced dimension they may be transformed in such a way as to give the restricted partial derivatives of a local subspace, except at points after a transition from a region of lower to one of higher dimension, when p represents a directional derivative of cost along the optimal trajectory. On a constraining manifold the control is chosen to maintain the trajectory on the manifold, either retaining 4.35 in that form, or transforming it to a reduced set, but in any case retaining all n equations 4.35.

4.35 A Pictorial view of boundaries

A further insight into the behaviour of the solution space at boundaries can be gleaned from a purely pictorial view of the effect of imposing an equality constraint onto a field of trajectories.

Fig 5 represents part of a field of unrestricted optimal trajectories for a Lagrange problem, and a boundary is to be imposed at C so that the trajectories for the new problem must all lie below C. Consider the trajectory B; it and all points below will be quite unaffected by the imposition of a constraint at C. The same applies to all points to the right of A which are on or below C. In this region the isotims and trajectories are unchanged, but points to the left of aA and above B will lie on isotims of greater value than before, because trajectories from these points must be lie, for part of their range, on C, and therefore 'cost more '; the isotims will be distorted to the right.

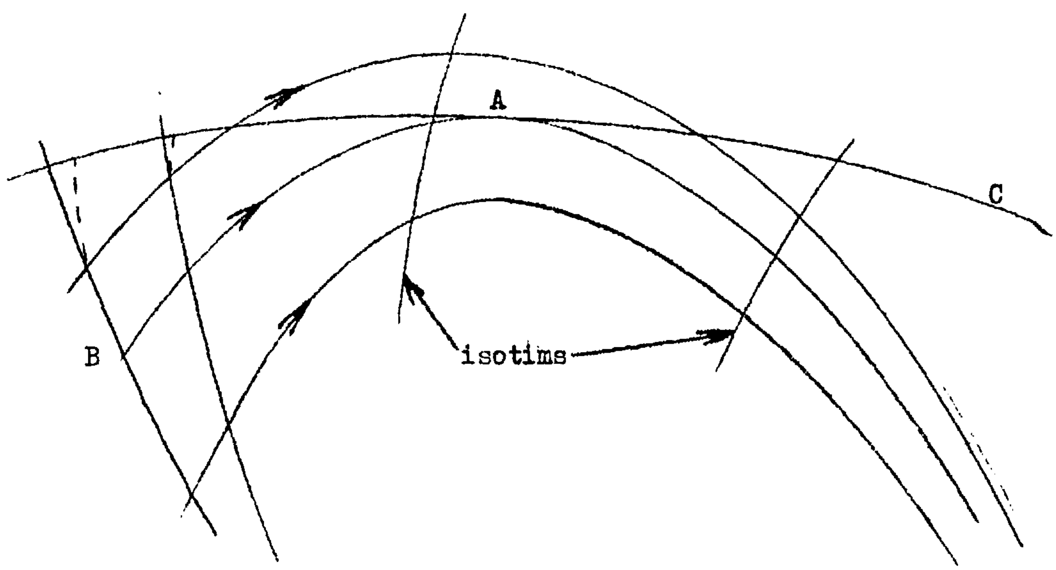


Fig 5

Clearly the trajectories will leave the boundary at A, and follow B thereafter. They cannot leave before, for the optimal direction at such points is away from the interior, and if they remain on the trajectory after A there must be more than one optimal trajectory from A. On the boundary trajectories all coincide, expressing the reduction from 2- to 1- space, and must thereafter remain coincident, occupying the 1- space B. Since B is not known in advance, the trajectory must be treated as a member of a 2.-dim. field.

The behaviour of the isotims at the boundary is interesting, and we will need, to investigate it, the concept of a'penalty function'. This is a function which, added to the cost function, has the effect of relaxing a 'hard constraint ' to an ' elastic constraint'. For a constraint $C(x) = 0$ a suitable penalty function is $k/C(x)$, which tends to a hard constraint as $k \to 0$ Fig 6

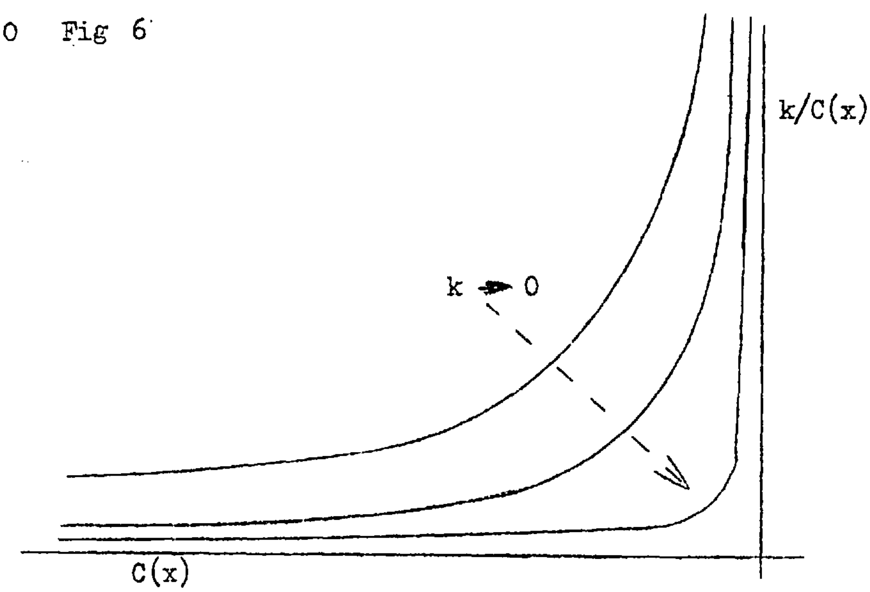


Fig 6.

The cost of being at points close to C is very high, and it will be impossible to actually cross C. The constraint has the effect of transforming the isotims covering the whole of infinite space into the region enclosed by C, and their density will be very high (Fig 7)

As $K \to 0$ the isotims collapse towards C, for the cost at each point not on C decreases (Fig 6), and in the limit they will actually lie on C for part of their range, though points in the interior will not be at all affected by the addition of the penalty function. There will be a sharp discontinuity manifesting itself as a corner in the isotim as it meets the boundary (Fig. 8).

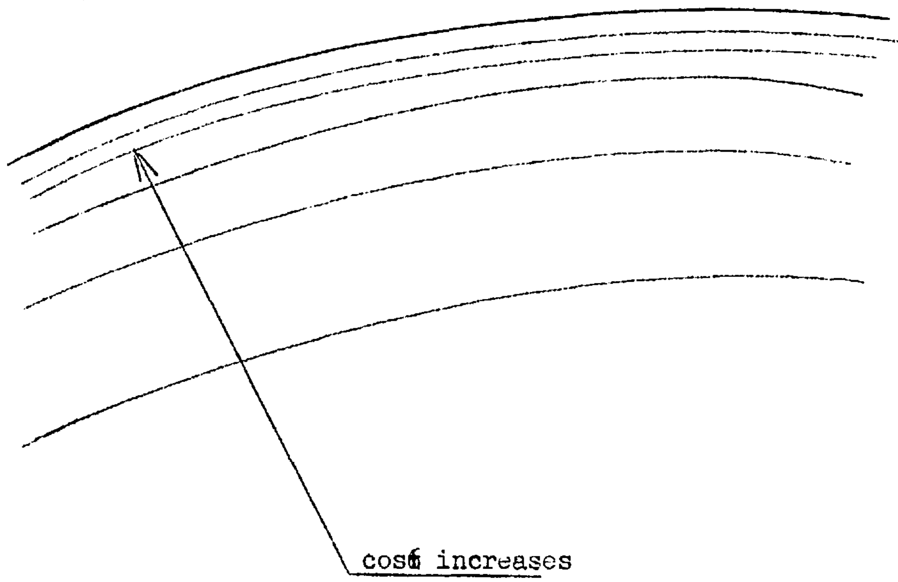


Fig 7

Since the isotims actually coincide on the boundary, J_x, expressing the change in cost for a variation in x must be undefined for δx normal to C, though it will be defined for δx tangential to C, and as a consequence

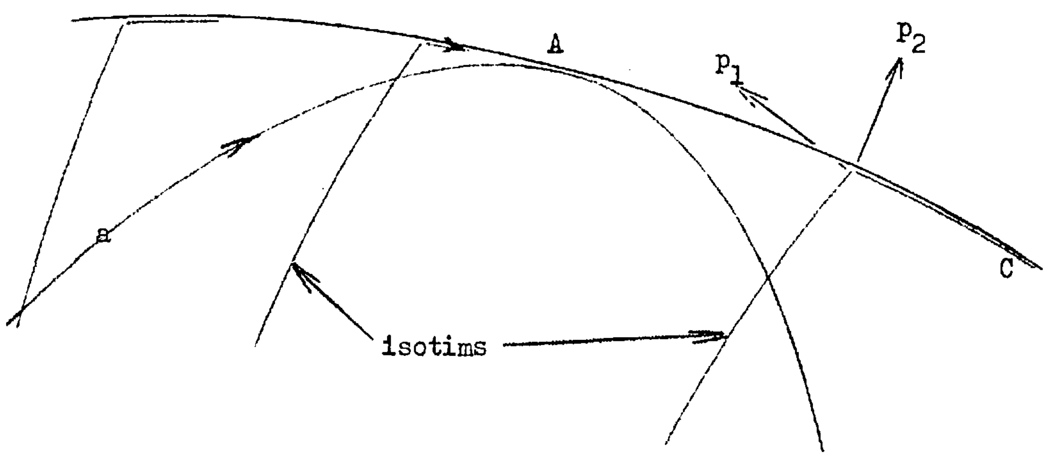


Fig 8

any multiple of C_x will satisfy the equation describing the normal to the isotim. The jump at the point of intersection with the boundary is caused by the sharp corner, and the actual component that we use is the projection of p_1 onto C (Fig 8).

It is easy to conceive this situation for a 3-space. C is a surface and A (Fig 5) a curve. The trajectories form a sheet on the boundary, leaving it in the curve A, which is the intersection of

$$C(x) = 0 \quad ; \quad C_x . f(x,u) = 0$$

u being optimal without consideration of C. The isotims are surfaces meeting C at a sharp edge. When trajectories leave C they do so tangentially, for they follow the unconstrained trajectories B, which are tangential if the control is continuous. This strongly suggests that in all cases where the unconstrained optimum is stationary the trajectories will be tangential to the boundary. It is not legitimate to transfer this argument to the point of entry to the boundary, for optimal trajectories are not, in general, symmetric.

The Mayer problem presents a somewhat different picture, involving the concept of reachable set. In accordance with the principle that constraints should be regarded as part of the background to the problem, rather than as extra conditions, the reachable set must be considered in terms of the restricted state space, for no points outside of it can be reachable in the context of the problem. The state space boundary forms a natural boundary to the reachable set, and also to the isotim which coincides with the extreme points of the reachable set.

The situation is rather like that of a toy balloon blown up close to a fixed surface, the neck of the balloon corresponding to the source point of the reachable set which evolves in time. As it expands, the walls of the balloon reach the surface and will lie upon it with a sharp edge at the meeting point. A nest of the reachable sets will share the same boundary over this region. Since a trajectory must remain upon the same isotim throughout its range, boundary or no, the curve in which the isotim surface meets the state space boundary must itself be a trajectory, and since this trajectory corresponds to an edge of the isotim the normal J_x is not uniquely defined along it. As for the Lagrange problem, the coincidence of isotims on the boundary rules out the possibility of uniquely defining a normal component of J_x, and it is only the projection of J_x onto the tangent plane (tangent space for higher dimensions) that may be considered.

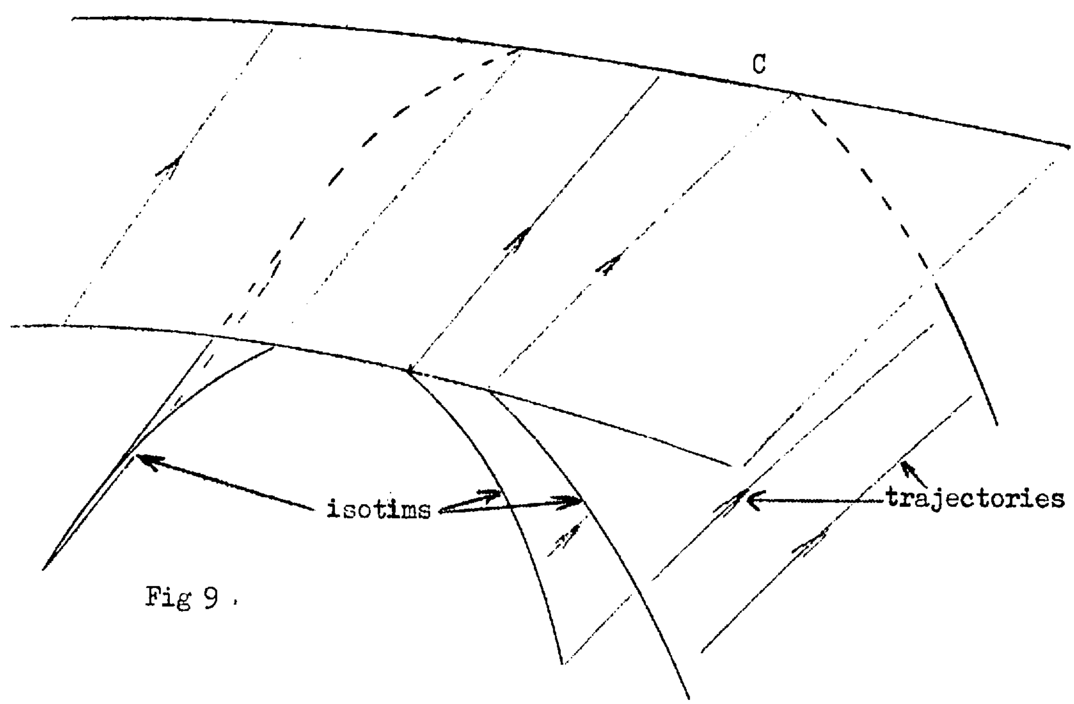


Fig 9 .

Chapter 5 CRITICAL REVIEW OF OPTIMAL CONTROL THEORY

Having established the principles and major properties of optimal systems from a particular viewpoint we must compare these results, and , perhaps even more important, this attitude to the problem, with those familiar from other studies. It will be necessary to discuss the state-constrained problems separately, for although no real distinction has been made hitherto, other methods tend to introduce new techniques to deal with such constraints, treating inequalities with quite unmerited respect.

5.1 Unrestricted State Space

5.1.1. Classical calculus of variations

It was suggested in the introduction that the calculus of variations has developed sufficiently to be able to cope quite satisfactorily with the mathematical difficulties of the problem of optimal control. The details of the particular techniques required may be found in Berkovitz (25) and Hestenes (78), but what is more important to the present work is the general approach of the classical calculus, for which the simpler, formal derivation by Troitskii (26) will suffice.

What is described here as 'classical' is the method based upon the construction of a linear functional representing the first variation of cost, and the application of the 'fundamental lemma' to obtain the necessary conditions for a stationary extremum with respect to weak variations.

Given a cost function

$$g(x(t_f)) + \int_{t_o}^{t_f} L(x,u)dt, \tag{5.1.}$$

a dynamic system

$$\dot{x} = f(x,u) \tag{5.2}$$

5.3

initial and terminal sets defined by

$$S(x) = 0 \qquad 5.3$$
$$T(x) = 0$$

An augmented cost function is formed by adjoining all the vector equality constraints to the minimand with undetermined vector multipliers. Thus,

$$P = g(x(t_f)) + a\,S(x(t_o)) + bT(x(t_f)) + \int_{t_o}^{t_f} L(x,u) + p(\dot{x} - f(x,u))\,dt \qquad 5.4$$

Extending the idea of differentiability of a function, a functional is said to be differentiable if its variation can be expressed as a linear functional of variations of its arguments plus terms of higher order but insignificant magnitude. (Gelfand & Fomin 27 p.11) The first variation of 5.4 is

$$\begin{aligned} \delta P &= g_x.\Delta x(t_f) + \delta a.S + a\,S_x.\Delta x(t_o) + \delta b.T + \\ &+ b\,T_x.\Delta x(t_f) + \left\{ L + p(\dot{x} - f) \right\} \delta t \Big|_{t_o}^{t_f} + \\ &+ \int_{t_o}^{t_f} L_x.\,\delta x + L_u.\,\delta u + \delta p.\,(\dot{x} - f) + \\ &+ p.(\delta\dot{x} - f_x.\,\delta x - f_u.\,\delta u)\,dt \end{aligned} \qquad 5.5$$

All the variations being arbitrary and independent, their coefficients may be set equal to zero, which gives, after integrating $p.\delta\dot{x}$ by parts and setting $\Delta x = \delta x + \dot{x}\,\delta t$,

δp :	$\dot{x} - f(x, u) = 0$		a
$\delta x(t)$:	$\dot{p} + pf_x - L_x = 0$		b
$\delta u(t)$:	$pf_u - L_u = 0$		c
δa :	$S(x(t_o)) = 0$	5.6	d
δb :	$T(x(t_f)) = 0$		e
$\Delta x(t_o)$:	$p(t_o) - aS_x = 0$		f
$\Delta x(t_f)$:	$p(t_f) + g_x + bT_x = 0$		g

$$\delta t_o\,,\ \delta t_f : \qquad \Big[L - p.f\Big]_{t_o}^{t_f} = 0 \qquad \text{h}$$

If two arcs meet at a manifold $M(x(t')) = 0$ it must be adjoined as a constraint in a similar way, giving relations between the right and left limits at t'_+, t'_- :

$$p(t'_-) - p(t'_+) + cM_x = 0 \qquad \text{i}$$

$$\Big[L - p.f\Big]_{t'_+}^{t'_-} = 0 \qquad \text{j}$$

5.6

The use of Lagrange multipliers involves a concept that we rejected in the earlier chapters, namely, that equality relations are 'constraints' preventing the system from behaving in a natural way. Some relations do in fact have this effect——terminal conditions, for example——for if the problem were posed without them the system would behave quite normally, finding a 'natural' unconstrained solution with, as it happens, a lower value of cost. It would, however, be stretching this interpretation to the limit to regard, say, the differential equations in this light; rather, they define the system, and the supposedly unconstrained system that would exist in their absence, has no physical significance at all. The practice of regarding the equations in this light is in the tradition of the calculus of variations, in which the problem is quite sensibly posed without differential side constraints which merely reduce the number of degrees of freedom; in engineering, the equations obviously have a different significance. This distinction, between essential and inessential conditions (which is similar to the different types of inequality constraints discussed in 2.1) does not emerge at all in this formulation, though it is important in practice.

The effect of a constraint is to remove one degree of freedom from the system; the role of the multiplier is to replace the lost freedom by the addition of a new variable, upon which the system may be treated as unconstrained. The result is an increase in the complexity of the system description, now involving more variables, but a simplification in its behaviour, and with finite constraints it is usually a matter of taste whether this technique is used, as a pure device, obscuring the true nature of the situation, or a direct reduction in the system is made, by elimination or transformation. When differential equations behave as constraints, the latter course is not available, and multipliers must be used. The magnitude of the multiplier is a measure of the degree of restriction represented by the constraint——the effort, as it were, that is required to ensure that the system conforms with the constraint. (Lanczos 28 p.84).

5.5 illustrates this interpretation of the multiplier as a force in a different way. If the cost is regarded as a potential, a potential gradient being a force, the multipliers become potential gradients in the constraint space. For example, $a = \partial P/\partial S$, the change in cost due to variations (violations) of S ; similarly p represents the effort of satisfying the dynamic constraint. If any constraint would have been satisfied automatically by the unconstrained system, (for example, a terminal condition which happens to be the same as the free terminal point) the corresponding multiplier is zero. In the same way, since all the variables in 5.4 are on the same footing, 5.6b represents a gradient in the function space of $x(t)$. The corresponding equation 4.15 was derived, it will be recalled, as the gradient of $H(x)$. The distinction between the gradients in the space of $x(t)$ and in the space of x is related to the reduction of the minimisation of an integral to the minimisation of a function in section 4.1.1.

Another, equally fundamental, process involved in 5.5 is the use of the condition for a stationary value, i.e., that the first variation of cost shall vanish. In the control problem, where inequality constraints are the rule, the minimum is often not stationary. This highlights a feature of the classical problem which can be overlooked when the true minimum turns out to be stationary. The classical treatment does not set out to find a true minimum at all, but a stationary value, which, even in the classical context is only a secondary property of the minimum. This is a serious matter, for it implies that the whole approach to the problem is not sufficiently fundamental.

The problem did not arise until late in the development of the calculus, for it did not affect the treatment of state-inequalities, these being handled by the theory of unilateral variations (Hadamard 33), but only differential inequalities—-in this context a very sophisticated refinement. The way in which writers fought shy of this problem is an indication that it raised extraordinary questions; Valentine(81) produced a technique to get round the difficulty, but other reference to the problem is rare. His method uses a slack variable to convert the inequality to an equality, which is then dealt with in the usual way. Thus two additional variables are introduced, a slack variable and an undetermined multiplier, when, on the face of it, the inequality should simplify the problem by reducing the region in which a minimum may be sought. As a device, it is satisfactory, and will solve problems involving control-variable inequalities (Berkovitz 25),
A comparison of the results obtained here and those derived from the classical approach shows little difference as far as practical application is concerned. It is usual, using the latter method, to impose on constraints

of the type

$$R_i(x,u) \leqq 0 \qquad i = 1, \ldots, r \qquad 5.7$$

the condition that, when the first S components of R vanish, the matrix $\partial R_j / \partial u^k$ (k = 1, . . . , m ≥ s) shall have maximum rank.(25, 78). This enables the zero components of R to be adjoined to H with multipliers q, obtaining, together with the constraint

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

the stationarity condition

$$H_u + qR_u = 0 \qquad 5.8$$

Furthermore, it allows the Clebsch condition to be derived from the Weierstrass inequality (cf. Bliss 5 p.224). The condition is

$$e(H_{uu} + qR_{uu})e \geqq 0 \qquad 5.9$$

where $R_u\, e = 0$,

only those components of R being taken which are zero.

If 5.9 is satisfied with a strict inequality, the trajectories will always occupy an n-dim. region, for u will have a unique differentiable solution $u(x,p)$ and the differential equations for the state and auxiliary variables will have differentiable right hand sides. This result is of the greatest importance in applying our earlier technique, for while we assumed that $u(x)$ was differentiable in certain subspaces, it was always doubtful whether $p(x)$ should be interpreted as the n-dim. normal to the isotim or merely a directional derivative of $J(x)$. This test indicates immediately the type of field to expect, and should be applied before attempting any computation. It is, in fact, a test of the singularity of $(H_{uu} + qR_{uu})$.

5.1.2 Paths of steepest descent.

It is quite evident that a method based upon stationary minima does not get to the root of the problem, and a radically different approach is called for. Caratheodory provided such an approach, (29, Bliss 5 p.77), which, though not designed specifically to overcome the difficulties of constraints, nevertheless can lead to a satisfactory treatment.

The differentiability of the extremals and their slope functions $\dot{x} = f(x)$ leads, in the classical calculus, to the possibility of embedding an optimal trajectory in an entire family of such trajectories. In a field of trajectories a one-parameter family of hypersurfaces

$$W(x) = w$$

can be constructed, crossed by trajectories at a rate

$$\dot{w} = W_x . \dot{x}$$

For a problem with cost function $J = \int L(x,\dot{x})dt$ the relation between the cost and the parametric value of the surfaces is expressed by

$$\begin{aligned} \frac{dJ}{dw} &= \frac{dJ}{dt} \cdot \frac{dt}{dw} \\ &= L(x,\dot{x}) \,/\, W_x . \dot{x} \end{aligned} \qquad 5.10$$

The direction of steepest descent is that which minimises dJ/dw, giving, for a stationary minimum,

$$\begin{aligned} L_{\dot{x}} W_x . \dot{x} - L \; W_x &= 0 \\ \therefore \qquad L_{\dot{x}} &= mW_x \\ L &= mW_x . \dot{x} \end{aligned} \qquad 5.11$$

m being a factor of proportionality. If the surfaces are chosen in such a way that the value of the cost along any curve of steepest descent between two fixed surfaces is the same, m becomes unity. Such surfaces were called by Caratheodory 'geodesically equidistant', for which the last equation in 5.11 is the Hamilton - Jacobi equation.

The requirement of steepest descent does not imply that the minimum of 5.10 with respect to $\dot{x}$ should be stationary. We might ask directly for

$$\min_{\dot{x}} \left(L/W_x . \dot{x} \right)$$

and the presence of differential inequality constraints would not affect the formulation at all.

To translate this into control terms requires merely the replacement of $\dot{x}$ by $f(x,u)$, and minimisation with respect to u. This overcomes two difficulties at one stroke; first, the question of the equality relation $\dot{x} = f(x,u)$, which is handled here by substitution rather than by multipliers, secondly, the inequality constraint on u, which merely reduces the range of the minimisation.

We have, then,

$$\min_{u} \; L(x,u) \;/\; W_x . f(x,u) \quad = \quad 1.$$

J and W are equivalent, apart from arbitrary constants, so we may write, if $L > 0$,

$$\min_{u} \; (L - J_x . f) = 0$$

This was the source of the ideas of chapter 4, the supporting arguments and constructions being no more than refinements designed to align these ideas more closely with the needs of engineering systems.

It is at this stage that the translation is made from the minimisation of a functional to that of a function. The requirement of steepest descent for a family of geodesically equidistant surfaces is equivalent to the observation that the increase in optimal cost J represented by a movement from one surface to another cannot be greater than the cost actually accumulated $\int L \, dt$ - cost must be earned, it does not appear by jumps. This fascinating interpretation becomes somewhat dulled in reduction to mathematical form, and appears more prosaically in equations 4.3, 4.4.

5.1.3 Dynamic programming.

This technique was developed as a computational algorithm to treat a large class of multistage decision processes of which the optimal control problem is a special case. It may be summed up in the simple 'principle of optimality': 'every part of an optimal process is optimal ', which , for the Lagrange problem, leads to the following.

Let the optimal cost from x_o to the terminal set be J_o and let the optimal trajectory from x_o pass through x, on J_1 at t_1, Then

$$J_o = \int_o^{t_1} L(x,u)dt + J_1 ,$$

a simplified version of 4.3. A formal analogy between this principle and the results of the classical calculus of variations has been discussed by Bellman and Dreyfus (19) using principles similar to those of chapter 4, but without careful consideration of the differentiability properties of $J(x)$ - they require second partial derivatives in all arguments - and without the geometric background used here. One might say that the work presented here is a geometric interpretation of dynamic programming, though this would give a mistaken impression of the genesis of the 'minimum principle', which was a direct evolution from Caratheodory rather than Bellman. This discussion throws a clear light on the antecedents of dynamic programming, and one is surprised not to find adequate acknowledgement made to Caratheodory in Bellman's works. (cf. Osborn 30)

5.1.4 Pontryagin's maximum principle.

The concept of a field of optimal trajectories leading to the Weierstrass condition, dynamic programming and isotims, marks a watershed in the calculus of variations. On one side a trajectory is seen as a member of a family of similar optimal trajectories; on the other it appears as the unique optimal member of a family of trajectories whose other members are non-optimal. The latter approach is usually made the basis of the derivation of the Euler-Lagrange equations, the optimal trajectory $x(t)$ being regarded as that member of the family $x(t) + ey(t)$ for which e, a scalar parameter, is zero, $y(t)$ being arbitrary. Minimisation is then carried out with respect to the single variable e. (e.g. Bliss 5 p.9; Bolza 31 ; Forsyth 32 ; Hadamard 33). This chain of tradition meets control engineering in the 'maximum principle ' of Pontryagin, which can also (but, amazingly, doesn't) claim roots far back in the classical theory.

The essential step taken towards modern problems was the introduction of differential side constraints. (This is perhaps putting the cart before the horse, for the representation of systems by differential equations in this context probably owes a great deal to this development in the calculus.) These are different from finite constraints, for while a relation $g(x) = 0$ merely reduces the dimension of the admissible space, and can be dealt with by eliminating one variable, a differential relation $g(x,\dot{x}) = 0$ limits the directions in which a trajectory may move from x. If an n-dim. tangent space X_t is centred at x, the vectors $\dot{x}$, if unconstrained, will span the whole of X_t, but if restricted to satisfy $g(x,\dot{x}) = 0$, will only sweep out a cone in X_t, (Mc Shane 34) possibly of reduced dimension.

This tangent space contains not only the direction vectors $\dot{x}$, but also the differentials $dx = \dot{x}\,dt$, representing the points reachable from x in the interval dt. This set can be extended to include points reachable in finite time, and will still be a cone but will only lie in X_t at points infinitely close to the vertex. The trajectories must lie either in the interior or along the boundary of this reachable set, and we have already seen (section 4.2.1) that the latter condition holds for the optimal trajectories of a Mayer problem. Recalling that a Lagrange multiplier is a component of a generalized gradient in the direction of a violation of a constraint (i.e. a variation of $g\,(x,\dot{x})$) it is a straightforward interpretation to describe the multiplier which adjoins the differential constraint to the cost function as a vector directed towards the unreachable zone, and if the boundary of the reachable set has a unique normal it will coincide with this multiplier.

If there are no constraints on $\dot{x}$ the concept of a reachable set is meaningless, for all points are reachable, but in the context of the classical problem a set of r differential equalities reduces the number of degrees of freedom of the directions $\dot{x}$ to $n-r$, (obviously $r \leq n$). It is shown by Bliss (5 p.224) that a set of r constraints

$$g_i(x,\dot{x}) = 0 \qquad\qquad i = 1,\dots,r$$

can be extended by the addition of

$$h_j(x,\dot{x}) = z_j \qquad\qquad j = r+1,\dots,n$$

the Jacobian

$$\left| \begin{matrix} \partial g/\partial \dot{x} \\ \partial h/\partial \dot{x} \end{matrix} \right|$$

being non-zero, while retaining the same freedom for $\dot{x}$. The restriction represented by each additional constraint is relaxed again by the introduction of new variable z.

The dynamic system of control problems is very similar, a set of n differential equations involving the $n+m$ variables $\dot{x},u$ being equivalent, in some sense, to a set of $n-m$ constraints $g(x,\dot{x}) = 0$. Evidently the concepts of admissible cones, etc., introduced by McShane for the classical problem, can be applied without essential modification to the control problem, and this was done by Pontryagin. His technique is to formulate every problem in Mayer form, and introduce the auxiliary variables p defined by

$$\dot{p} = -pf_x$$

which, though not given any important geometric significance, can be recognized as normals to the reachable set. (Pontryagin 1 pp.86, 99; Roxin 8).

In section 6.3 a version of Pontryagin's use of p will be applied in a similar context, and need not be reproduced here, but some important remarks should be made about this approach. It is a definite improvement upon the multiplier technique, from the point of view of engineering mathematics, in that it presents optimality as a property of dynamic systems rather than as an abstract mathematical problem, but it does not go far enough in this direction. Insisting upon the Mayer form is admittedly consistent, but it lacks the true generality of the classical technique which can deal with both forms together (see 5.4); ignores the physical meaning of the cost function, and quite overlooks the different geometric interpretation of the Lagrange problem. The further demand that the cost variable x^0 shall not appear in the other differential equations even raises mathematical difficulties

The inclusion of the latter property ensures that the auxiliary variable p_o corresponding to x^0 shall be constant, and since $H = p.f = 0$ is homogeneous in p, p_o can be set equal to one, unless it is zero. This allows us to use the classical concept of 'normality' for a solution is normal if p_o is one, and abnormal if zero. Approaching the problem from the point of view of fields precludes the possibility of a normality analysis, for no field can be constructed for abnormal trajectories. It is a merit of Pontryagin's method that it leaves the door open to such considerations.

The treatment of p, however, is not at all satisfactory. These variables are very difficult to motivate a priori, and their geometric significance can only be grasped in the light of the analysis; They cannot be used in this way for a satisfactory 'engineering' derivation. Mixed constraints $R(x,u) \leq 0$ are not dealt with directly by Pontryagin's principle, and require a complicated construction to handle them (1 Chap.6.) Halkin(7) adopts a somewhat similar approach, using the concepts of reachable sets and their boundaries, but where Pontryagin uses infinitesimal cones comprising vectors $\delta x(t)$ such that for two trajectories,

$$x_1(t) = x_0 + \int_0^t f(x,u)\,ds$$

$$x_2(t) = x_0 + \int_0^t f(x,v)\,ds$$

$\delta x(t) = x_1 - x_2$ is given by

$$x\ (t) = \int_0^t f_x\ \delta x + \left[f(x_1(s),u) - f(x_1(s),v)\right] ds \qquad 5.12$$

only when $\|\delta x\|$ is small, Halkin uses a space of vectors given precisely by 5.12, introducing an'approximate system'

$$\dot{x} = f(x(u,t),v)$$

where

$$x(u,t) = x_0 + \int_0^t f(x,u)\,ds.$$

The basic approach is in the McShane -Pontryagin tradition though the construction is different. The Mayer form is retained, but without requiring x^0 to be an independent variable; mixed constraints are not considered.

It is interesting to note that, as Halkin hints in his introduction (quoted in section 1.2 above), the ideas reflected in the mathematics are not quite those which originally motivated the scheme. The geometric basis of his analysis is clear enough, concerning the trajectories of certain 'approximate systems' and their reachable sets. His introduction discusses a geometric construction which is quite beside the point, relating to a different approach to the problem, which he develops in (6); it is based upon Huygen's construction, a majestic device, which is sufficiently fundamental to demand a separate chapter.

5.2 Restricted State Space

State - variable inequality constraints

$$C(x) \leq 0 \qquad 5.13$$

are not of the same nature as mixed or control constraints, for they impose no immediate restriction on the control - any choice of $u(t_1)$ where $C(x(t_1)) \leq 0$ is apparently satisfactory. When the strict inequality is satisfied the constraint can be ignored, but when equality holds 5.13 implies that u must be chosen such that

$$\dot{C}(x) = C_x.f(x,u) \leq 0 \qquad 5.14$$

If $\partial \dot{C}/ \partial u = 0$, then we turn to higher time-derivatives, and a q'th order constraint is

$$C(x) = C^{(1)}(x) = \dots = C^{(q-1)}(x) = 0 \qquad 5.15$$

together with

$$C^{(q)}(x,u) \leq 0 \qquad 5.16$$

where $C^{(r)} = \dfrac{d^rC}{dt^r}$

The recognition that the problem can be treated in two parts - the interior section, ignoring these constraints, and the boundary section, on which 5.15 and 5.16 (equality) hold - bring this situation into the realm of the classical problem. (Bliss & Underhill 35). A set of terms

$$\pi_k \; C^{(k)}(x(t)) \qquad k= 0,\dots, q \qquad 5.17$$

are added to the cost function 5.4 within the integral, and two similar sets with $k = 0,\dots,q-1$ independent of the integral, for the instants at which the trajectory meets and leaves the boundary. $\pi(t)$ is identically zero for the inequality in 5.13. Variation of $x(t)$, $u(t)$ gives extended versions of 5.6 b,c

$$\dot{p} + pf_x - L_x + \pi_k C_x^{(k)} = 0 \qquad 5.18\ a$$

$$pf_u - L_u + \pi_q C_u^{(q)} = 0 \qquad b$$

while the corner conditions 5.6 i,j appear as

$$p(t'_-) - p(t'_+) + m_k C_x^{(k)} = 0 \qquad 5.18\ c$$

$$\Big[L - p.f\Big]_{t'_+}^{t'_-} = 0 \qquad d$$

5.18.b is equivalent to 5.8 , the mixed inequality being R in one case and $C^{(q)}$ in the other. Similarly an equation of the form 5.18a holds for the case of the mixed constraint, but only involving additional terms of the type qR_x , when q can be eliminated. Now, however, the multipliers π_k , $k = 0 , \dots , q - 1$ cannot be eliminated, and remain unknown. All that can be stated for certain is that they are negative, being gradients of cost, for if any member of 5.15 could increase, the cost would diminish. Similar remarks apply to m in 5.18c .

It is here that the transformation technique proves itself, for all these awkward variables are transformed away. Indeed, far from the boundary involving extra variables, the fact that its form is explicitly given enables us to reduce the complexity of the problem. (Bellman and Dreyfus 19 p.20). A glance at 5.18 shows that all the additional terms are in fact multiples of the components of the outward normal to the boundary; this is why they are indeterminate and vanish from a discussion which is restricted to trajectories in the boundary.

If p, π is a solution to 5.18a , then also $p + q_k C_x^{(k)}$; $(\pi_k - q_{k-1})$ is a solution, $(q_{-1} = 0)$, as may be seen by direct substitution. The numbers q_k are arbitrary, implying that the addition of any vector with the direction of C_x has no effect on the solution, as we predicted. It is evident from 4.14 that such a vector may be added not only to p but even to the equation 4.15 ; being normal to the trajectory's tangent at every point it is immaterial which of the terms $\pi_k C_x^{(k)}$ are added to the equation, if any. This accounts for the divergence between the results obtained by various writers. Gamkrelidze (1 chapter 6) dealing with the case $q = 1$ includes only the last term $\pi_q C_x^{(q)}$. Berkovitz (36) has a similar result, for in the extension of the linear functional he includes only $C^{(q)}(x,u)=0$ as a constraint to be satisfied along the boundary. This is reasonable, for if the remaining constraints $C^{(k)}$, $k < q$, are satisfied at the point of entry to the boundary, they will remain so if the q'th is satisfied. (Again, he deals only with $q = 1$). Dreyfus (39), by reducing the state space, obtains a result for a similar case which has been shown (Berkovitz and Dreyfus 37) to be equivalent to that of Berkovitz and Gamkrelidze , despite a difference of form. Chang (2) does not have the q'th term, but only the first (where $q = 1$), viz., $\pi_0 C_x(x)$.

The so-called 'jump conditions' 5.18c are subject to a similar interpretation, and are given in that form by most writers (Berkovitz and Dreyfus 37). In our consideration of reduced spaces, it was clear that the jump is caused, at entry to the boundary, by the disappearance of certain components of J_x (or rather, their ignoration, for they do not really vanish), and at re-entry to the interior, by their re-emergence as essential variables. The problem of determining the magnitude of the jump does not appear to have been satisfactorily dealt with. Gamkrelidze chooses the jump at entry to reduce to zero the component of p normal to the boundary (1 p.269), but is reticent concerning the exit point. Bryson and Denham (39), forced to deal with the problem in order to obtain solutions rather than principles, give the jump condition at the entry point, but ignore it at the end, presumably on the grounds that the equations of constraint are automatically satisfied there, and to supply a constraint would be superfluous. As a result they obtain continuity of p at the exit. However, they assert, without proof, that a combined jump, at entry and exit, is determined by the problem, but the distribution between the two points is arbitrary——p may be chosen to be continuous at either end, but will turn out to be discontinuous at the other.

This result is not incorrect, though the reasoning is not clear. We showed in section 4.3.2 that the number of extra variables appearing when the trajectory re-emerged into n-dim. space exactly compensated for the number of extra constraints imposed by the boundary. Had it been desired to retain all n equations along the boundary, there would have been required the same number of extra variables, this time in the form of 'jumps' in p (for there is one multiplier to each boundary equation) and it is quite immaterial whether they are introduced at the beginning or end—— or even the middle—— of the boundary arc.

The other corner condition 5.18d expresses continuity of H. In the classical problem it is merely the application of the Weierstrass condition comparing the two directions of a trajectory at a corner, giving both inequalities, hence equality. This result is given by Hadamard (33) who notes that in many cases it precludes the possibility of a discontinuous direction vector. This applies to the control problem too: in many cases corners are ruled out (see 4.25 et seq.), but each example must be investigated separately - there is as yet no general rule.

The test suggested in 5.1.1. for the dimension of the optimal space applies on boundaries too. The q equalities 5.15 constitute an (n-q) -dim. manifold, on which the inequality 5.16 and possibly other permanent inequalities R hold. If the matrix $H_{uu} + \mu_o C^{(q)}_{uu} + \mu R_{uu}$ is nonsingular the trajectories do form an (n-q) -dim. field, and u(x) is differentiable in the boundary.

Chapter 6. HUYGENS' CONSTRUCTION

6.1 Reachable Sets and Waves.

Huygens' construction, one of the most beautiful in the entire literature of dynamics, is a major link between the sciences of particle mechanics, continuum mechanics and geometry. Amazingly, it appeared before the study of dynamics was at all advanced - Galileo, the founder of modern dynamics, was a contemporary of Huygens; Euler, pioneer of field theory, was born in 1707, twelve years after Huygens' death, and it was only with Hamilton's work that optics and particle dynamics were finally mated in a geometric coup which cannot be quite absolved of responsibility for the upheavals of 20th-century physics and the modern fervour for unified theories.

Its relevance to the calculus of variations is often remarked (e.g. Courant & Hilbert 40 vol.2 p.124, Gelfand & Fomin 27 p.209), usually in the spirit of an interesting aside rather than as a fundamental principle, and it has been applied to the optimal control problem by Halkin (6, 7) to derive the necessary conditions. The construction and even the formulation of the problem presented here is quite different from Halkin's, and is intended to emphasize the geometric rather than analytic situation.

Suppose the problem to be of Lagrange form, with cost function $\int L(x,u)dt$; $L(x,u)$ is supposed to be positive for all admissible x,u. Let x_o be a point on the isotim J_o, and let there be constructed all possible admissible trajectories from x_o. For a given number w there will be a point y on every trajectory such that

$$\int_o^{t_y} L(x,u)\,dt = w \qquad 6.1$$

$$y = x_o + \int_o^{t_y} f(x,u)\,dt$$

Designate the set of all such points by

$$Q(x_o, J_o, w).$$

That is, Q is the set of all points reachable from x_o with cost w.

The topological properties of this set may be complicated in general, but the particular properties we require turn out to be quite simple.

Definition 6.1. y is a boundary point of $Q(x_o,J_o,w)$ if i) $y \in Q(x_o,J_o,w)$; ii) there exists some $w' > w$ and a set $Q(x_o,J_o,w')$ such that every neighbourhood of y contains points of $Q(x_o,w') - Q(x_o, w)$

This implies that the slightest extension of a trajectory from a boundary point can attain points that are only reachable from x_o with cost greater than w.

$Q(x_o,w)$ contains no points reachable optimally with cost greater than w, for, suppose z is such a point; if it is in $Q(x_o,w)$ it is reachable with cost equal to w, therefore the optimal cost cannot be greater than w.

$Q(x_o,w)$ might not be bounded, but if there is an optimal trajectory through x_o, and $J_o > w$, there must be at least one boundary point, for, consider the point on the optimal trajectory with cost from x_o equal to w; if it is interior, a small extension of the trajectory will remain interior to $Q(x_o,w)$, but the cost will exceed w. This point must be a boundary point of Q. It is not suggested that there is only one boundary point or that Q is entirely enclosed by a boundary - neither is in general true - but the unique boundary point through which the optimal trajectory passes is the only one of immediate interest.

The optimal trajectory meets the isotim with value (J_o-w) at the point at which the optimal cost from x_o is w, that is, at the boundary of Q.

The set $Q(x_o, J_o, w)$ meets the set $J(x) = J_o - w$ at only one point, all other points in Q lying on isotims with greater value. This gives a characterization of the necessary conditions similar to 4.3, for the optimal control must take the trajectory to that point for which

$$J(y) - (J_o - w) \qquad 6.2$$

is least, under the constraints 6.1.

The set $Q(x_o, w)$ is a wavelet issuing from x_o. Every point on $J(x)=J_o$ is the source of a similar wavelet meeting the isotim $(J_o - w)$ at one point. Since all the wavelets lie on one side of the isotim it may be regarded as a wavefront. In the study of optics for a homogeneous medium the wavelets are spheres. Here they are considerably more complex - not necessarily closed or containing their source - but the essential construction is the same.

To investigate the implications of 6.2 it is necessary to make comparisons between different paths. Unfortunately the interval of integration $[0, t_y]$ depends upon the path taken, which is most inconvenient, and to make the interval uniform we may use a simple parameter transformation which gives the problem a new, but familiar, form.

6.2 <u>The generalized time - optimal problem.</u>

A justification for treating the Lagrange and Mayer problems separately was that practical problems fall naturally into one or other form. But in one outstanding case the distinction fails. The problem of time - optimality can be regarded equally well as a path integral ($\int^{t_f} dt$) or a terminal value $(g(x(t_f)) = t_f)$ problem, without any drastic reinterpretation. The equations resulting from either formulation are precisely the same, for the expressions min $(1 + \dot{J})$ and min $\dot{J}$ are equivalent (cf 4.6, 4.19). The only difference lies in the magnitude of the vector J_x (or p), depending upon whether $J_x.f + 1 = 0$ or $J_x.f = 0$, but it is only the ratios of J_{x^i} which are relevant. The pictorial significance and nature of the isotims are different, but this is not fundamental.

Since the two formulations of the general problem are mathematically equivalent, and the optimal - time problem is actually identical in both forms, it suggests that the latter is the link between the two, and that the general equivalence can be traced to a basic similarity between the optimal - time problem and the general problem.

That all Mayer problems are basically equivalent is apparent, for the auxiliary variables p satisfy the same equations $\dot{p} = -pf_x$, differing only in the boundary values derived from the cost function $g(x)$. In particular they are all equivalent to the time - optimal problem, for which $g(x) = t_f$. (Yashilev 41) In fact the equivalence extends also to all problems of Lagrange form, for we may introduce an arbitrary parameter s, (indicating $\frac{d}{ds}$ by the prime), then the cost function is

$$L\,(x,u)t'ds \tag{6.3a}$$

and the dynamic system is

$$x' = f(x,u)\ t'. \tag{6.3b}$$

Defining $\qquad t' = {}^1/L(x,u)$

6.3 becomes

$$ds \tag{6.4a}$$

$$x' = \frac{f(x,u)}{L(x,u)} = \underline{f}\ (x,u) \tag{6.4b}$$

which has the form of a time optimal problem for the new system.

This allows considerable simplification of many aspects of the problem. for the two monotonically increasing scalars - time and cost - have been reduced to one. The condition min $(L + \dot{J})$ no longer represents a compromise between rapid reduction of optimal cost J and increase in actual cost L, but simply reduces to a condition of steepest descent in J, since L is constant (=1). Similarly 6.2, representing a minimization of $J(y)$ under the rather clumsy constraint 6.1 now becomes a minimization and a trivial constraint,

$$\int_0^{s_y} ds = w \; : \; s_y = w$$

The set $Q(x_o, J_o, w)$ becomes the set of points reachable from x_o by the system 6.4 b in time w, and the isotims are isochrones. The transformation is possible only if $L > 0$, when it is simply the replacement of one monotonically increasing parameter by another, together with a scale factor.

6.3 Properties of the wavelet.

We can now interpret the necessary conditions obtained in chapter 4 in terms of the wavelet issuing from a point.

Suppose a trajectory to be constructed in accordance with the equation

$$\min \; p.\, \underline{f}\,(x,u) \qquad \text{a}$$

$$p' = -p\underline{f}_x(x,u) \qquad 6.5 \;\; \text{b}$$

$$p(0).\underline{f}(0) + 1 = 0 \qquad \text{c}$$

No reference is made to the origin of the expressions, and no assumption concerning optimality. Our purpose is to investigate the trajectory that emerges from this construction. The procedure is essentially the same as Pontryagin's .

The trajectory and control corresponding to 6.5 is $x(s)$, $u(s)$; a neighbouring trajectory $x_1(s)$ corresponds to some admissible control $v(s)$, which is arbitrary except that $\| x_1(s) - x(s) \|$ is small.

We have
$$x(s) = x_o + \int_0^s \underline{f}\,(x,u)ds$$

$$x_1(s) = x_o + \int_0^s \underline{f}\,(x_1,v)\,ds$$

If $\delta x(s) = x_1(s) - x(s)$ is to be a small quantity of first order, the control variation $v(s) - u(s)$ may be either a) first order for finite time, or b) finite over a first order interval. A 'perturbed' control may be constructed in the following way.

In the interval

$$I = [s_o, s_f]$$

choose instants s_i, s_j $(i,j = 1,2,.....)$, as many as desired, non-negative finite numbers r_i, and non-negative first order quantities δs_j such that the intervals

$$I_i = (s_i, s_i + r_i] \qquad I_j = (s_j, s_j + \delta s_j]$$

are disjoint, and if

$$K_i = \bigcup_i I_i \ ; \ K_j = \bigcup_j I_j \ ,$$

then

$$K_i \bigcup K_j \subseteq I .$$

The control function $v(s)$ is defined as

$$v(s) = \begin{cases} u(s) \ , & s \in I - K_i \bigcup K_j = I_k \\ v_j \ , & s \in K_j \\ v_i(s) \ , & s \in K_i \end{cases} \qquad 6.6$$

where v_j is an arbitrary point of the m.dim. control space U; $v_i(s)$ is an arbitrary continuous function to the control space with the restrictions that $\quad \| \delta u(s) \| = \| v_i(s) - u(s) \| \quad s \in K_i$

shall be of first order, and $x_1(s)$, $v(s)$ shall be admissible for all s I.

It is not assumed that the state space is unrestricted. Both $u(s)$ and $v(s)$ are constructed to ensure that the trajectories remain admissible. In the case of $u(s)$ this means that certain constraints are implicitly satisfied in addition to 6.5. Thus, the minimization of H is carried out subject to the requirement (which might completely define u) that the trajectory does not leave X. It will be recalled from section 5.2 that the degree of arbitrariness involved in the boundary equations allows 6.5b to hold even on the boundary of state space.

Using 6.6, $x_1(s)$ becomes

$$x_1(s) = x_o + \int_{s\in I_k} \underline{f}\,(x_1,u)\,ds + \int_{s\in I_i} \underline{f}\,(x_1,v_i)\,ds + $$
$$+ \int_{s\in I_j} f(x_1,v(s))\,ds$$

$$\therefore\ \delta x(s) = x_1(s) - x(s)$$
$$= \int_{s\in I_k} \underline{f}\,(x_1,u) - \underline{f}\,(x,u)\,ds +$$
$$+ \int_{s\in I_i} \underline{f}\,(x_1,v_i) - \underline{f}\,(x,u)\,ds + \qquad 6.7$$
$$+ \int_{s\in I_j} \underline{f}\,(x_1,v(s)) - \underline{f}\,(x,u)\,ds.$$

Applying the mean - value theorem, these three integrals become, respectively

$$\int \underline{f}_x\,(x^*u)\,\delta x\ ds$$
$$\int \underline{f}_x\,(x^*,u)\,\delta x + \underline{f}_u(x,u^*)\,[v_i - u_i]\ ds$$
$$= \int \underline{f}_x\,(x^*,u)\,\delta x + \underline{f}\,(x,v_i) - \underline{f}\,(x,u_i)\ ds \qquad 6.8$$
$$\int \underline{f}_x\,(x^*,u)\,\delta x + \underline{f}_u(x,u^*)\,\delta u\ ds$$
$$= \int \underline{f}_x\,(x^*,u)\,\delta x + \underline{f}\,(x,v) - \underline{f}\,(x,u)\ ds$$

where $x^* = x + \alpha\,(x_1-x) \qquad 0 \le \alpha \le 1$

$u^* = u + \beta\,(v-u) \qquad 0 \le \beta \le 1$

6.9

If the solution of an equation

$$y' = \underline{f}_x\,(x^*(s),\ u(s))\ y$$

is $$y(s_2) = A(s_2,s_1)\ y(s_1), \qquad 6.9$$

we may apply 6.9 to 6.7, which, in view of 6.8, is a linear non-homogeneous equation in x. Treating, for example, the typical intervals

$$(s_1, s_1 + r_1] \qquad (s_1 + r_1, s_2] \qquad (s_2, s_2 + \delta s_2]$$

we obtain, recalling that

$$A(s_3,s_1) = A(s_3,s_2)\ A(s_2,s_1)$$

and $$A(s_1,s_1) = \text{unit matrix},$$

$$\delta x(s_2+\delta s_2) = A(s_2+\delta s_2,\ s_2)\left[\delta x(s_2) + \left\{\underline{f}\,(v(s_2)) - \underline{f}(u(s_2))\right\}\delta s_2\right]$$

$$\delta x(s_2) = A(s_2, s_1+r_1)\ \delta x(s_1+r_1)$$

$$\delta x(s_1+r_1) = A(s_1+r_1, s_1)\left[\delta x(s_1) + \right.$$

$$\left. + \int_{s_1}^{s_1+r_1} A(s_1,s)\left\{\underline{f}\,(v) - \underline{f}\,(u)\right\} ds\right]$$

$$\therefore\ \delta x(s_2+\delta s_2) = A(s_2+\delta s_2, s_2)\left[\underline{f}(v(s_2)) - \underline{f}(u(s_2))\right]\delta s_2 +$$

$$+ A(s_2+\delta s_2, s_1)\left[\delta x(s_1) + \int_{s_1}^{s_1+r_1} A(s_1,s)\left\{\underline{f}(v) - \underline{f}(u)\right\} ds\right]$$

Treating all the intervals in a similar fashion,

$$\delta x(s_f) = \sum_i A(s_f,s_i)\int_{s_i}^{s_i+r_i} A(s_i,s)\left\{\underline{f}(v) - \underline{f}(u)\right\} ds + \sum_j A(s_f,s_j)\left[\underline{f}\,(v(s_j)) - \underline{f}\,(u(s_j))\right]\delta s_j \qquad 6.10$$

6.5b is adjoint to 6.9 in the limit $\delta x \to 0$, for then $x^* \to x$, therefore

$$p(s_2) = p(s_1)\ A(s_1,s_2).$$

$$\therefore\ p(s_f).\ x(s_f) = \sum_i \int_{s_i}^{s_i+r_i} p(s).\left[\underline{f}(v(s)) - \underline{f}(u(s))\right] ds + \sum_j p(s_j).\left[\underline{f}\,(v(s_j)) - \underline{f}\,(u(s_j))\right]\delta s_j \qquad 6.11$$

In view of 6.5 a, every member on the right hand side of 6.11 is non-negative,

$$\therefore\ p(s_f).\ \delta x(s_f) \geqq 0 \qquad 6.12$$

$\delta x(s_f)$ is any vector from the terminal point of the trajectory $(x(s),\ u(s))$ constructed according to 6.5, directed toward the set $Q(x_o, s_f)$, for $x_1(s_f) = x(s_f) + \delta x\ (s_f)$ represents any point close to $x(s_f)$ that is reachable in the same (generalized) time. From the manner of construction 6.12 applies for any $s < s_f$,

$$\therefore \quad p(s).\ \delta x(s) \geq 0 \tag{6.13}$$

where $\delta x(s) = x_1(s) - x(s)$.

$p\ (s)$ can be interpreted as the normal to a hyperplane supporting $Q(x_o, s)$ at $x(s)$, and we have the interesting result that the reachable set is convex in the neighbourhood of the point on the trajectory 6.5, which is evidently a boundary point. This result should be more rigorously proven by an application of Lyapunov's theorem on the range of a vector measure (cf. the use of this theorem in similar cases by LaSalle 42, Halkin 7) but our cruder construction demonstrates the geometric picture sufficiently clearly.

We may show too, that under certain restrictions the trajectory constructed with the help of 6.5 is optimal, for, suppose both $x_1(v,s)$ and $x(u,s)$ reach the same point x_2, the former at $s = s_2$, the latter at $s = s_1$,

then $$\delta x(s_2) = x_2 - x(s_2),$$

and $$x(s_2) = x_2 + \int_{s_1}^{s_2} \underline{f}\ (x,u)ds$$

6.13 gives
$$p(s_2).\ \delta x(s_2) = -p(s_2).\ \int_{s_1}^{s_2} \underline{f}(x,u)\ ds \geq 0 \tag{6.14}$$

Since $\| \delta x \|$ is small we may suppose $\delta s = s_2 - s_1$ to be small, and 6.14 becomes

$$-\ p(s_2).\ \ \underline{f}(s_2)\,\delta s \geq 0$$

The initial value of $p.f$ (see 6.5 c) is -1, and it was shown in 4.3.1 that $H = p.f$ is constant. 6.15 then implies

$$\delta s = s_2 - s_1 \geq 0$$

and the time taken to any point along $x(u,s)$ is less than that along any other neighbouring path.

6.4 Sufficient Conditions

The above is far from being a proof of sufficiency, for it compares only trajectories that are close over their entire range, and it is possible that a trajectory through the same two points but not uniformly close to $x(s)$ gives an even better performance. However, when the system $f(x,u)$ is linear in x the analysis applies even when δx is not small, and the result, apart from the assumption made between 6.14 and 6.15, becomes more significant. More satisfactory and more extensive sufficiency proofs have been obtained (Lee 43, Neustadt 44).

Our purpose, however, was not to obtain sufficiency proofs, but to establish an interpretation of the optimal trajectory vis a vis the reachable 'wavelets'. If an optimum exists it will be provided by 6.5, for an optimal trajectory is certainly locally optimal, and 6.13 informs us that the reachable sets $Q(x_0,w(s))$ are convex in the region of the boundary points at which they meet the isotim. If the isotim has a normal at that point it will coincide with p, the wavelet normal, but it is possible that the wavelet has a smooth boundary and the isotim does not.

It was remarked that 6.5 can provide a trajectory that is not optimal, but nevertheless the optimal trajectory must be constructed according to 6.5, the minimum principle. The paradox is resolved by noting the far from obvious fact that two trajectories from x_0 with different initial values $p_1(0),p_2(0)$, might intersect at some point x_1. The limit approached by x_1 as $p_2(0) - p_1(0) \rightarrow 0$ is a focal point for extremals from x_0, and beyond this point the trajectories cease to provide true minima of cost. The geometric significance of the situation has been described in various ways. Lanczos (28 p.272) describes it as corresponding to a reduction in dimension of the wavefront; Yashilev (41) shows by example, but without analytic discussion, that isotims of different value coincide at such points; Bliss(5) states that the trajectories satisfying the minimum principle (or its classical equivalent) meet an envelope at that point. No doubt these characterizations are all equivalent, and imply that the trajectories, on reaching the boundary of the set $Q(x_o,w)$, are tangential to it, and for cost greater than w return to the interior of $Q(x_o,w)$, thus arriving at points which can also be reached with cost equal to (or less than) w, while still satisfying the minimum principle. Under these conditions p cannot represent a normal to the reachable set boundary.

A complete account of this situation is lacking, and, more important an easily computable criterion to judge whether or not it occurs. A promising approach is afforded by the fact that through every point x_c there passes an n-parameter family of trajectories constructed according to 6.5. the parameter being the initial value p_o. The trajectories are solutions of the equations

$$\begin{aligned} \dot{x} &= f\,(x,u(x,p)) \\ \dot{p} &= -pf_x\,(x,u(x,p)) \end{aligned} \qquad 6.16$$

whose right hand sides are piecewise differentiable with respect to x,p, for f is assumed to be differentiable for x,u; u, expressed as $u(x,p)$ as a result of the minimization operation under constraints of the type 2.4, 2.5, is piecewise differentiable. Thus according to the results obtained in section 3.2.3, the partial derivatives

$$\partial x/\partial p_o \quad ; \partial p/\partial p_o$$

will exist if the boundaries between the various regions of state space are differentiable manifolds.

We are concerned with the matrix $\partial x / \partial p_o$. At any time the difference between two neighbouring trajectories approaches

$$\frac{\partial x}{\partial p_o} \delta p_o$$

as $\delta p_o \to 0$, so that if trajectories do meet, it can only be because

$$\left| \frac{\partial x}{\partial p_o} \right| = 0 \qquad 6.17$$

Of course this refers to trajectories meeting within a region of differentiability of $u(x)$, for a reduction of dimentionality, often incurred in the transition to a different region, means that trajectories originally distinct must meet; this is not the situation alluded to here. It would seem necessary only to ensure that in each region there is no focal point (or, 'conjugate point' - a distinction is made by Bliss (5, p.170)) of the initial point, say x_i, of that region. There is no need to retain the derivatives with respect to p_o, which would entail consideration of the discontinuities at boundaries (cf. 3.2.3), but merely the partials $\partial x / \partial p_i$, where p_i is the value of p at the point x_i where the trajectory enters that particular region. The transformation suitable to each local subspace will ensure that the matrix of derivatives remains square.

The equations giving the partials are the linearized versions of 6.16,viz.

$$\begin{aligned} \dot{z} &= (f_x + f_u u_x) z + f_u u_p w \\ \dot{w} &= -w f_x - p \left[f_{xx} + f_{xu} u_x \right] z - p f_{xu} u_p w \end{aligned} \qquad 6.18$$

where $z = \partial x / \partial p_o$, $w = \partial p / \partial p_o$, the initial values being $z(o) = 0$, $w(0) =$ unit matrix. The focal point occurs where $|z| = 0$.

This is not a rigorous derivation, nor has it been shown that satisfaction of the minimum principle together with the non-occurrence of a focal point is sufficient for optimality, nevertheless, if this analogy with the classical results proves to be valid, this provides a useful computational test.

There are a number of situations in which the technique of the minimum principle breaks down. It may be that between given points no optimal trajectory exists, a focal point intervening; or that an optimal trajectory is isolated and cannot be embedded in a field (abnormality); or that minimization of $H(x,p,u)$ does not provide a unique value of control (singular, non - normal, problems). The problems of normality (in the classical sense), accessibility and focal points are evidently closely connected, if not in their mathematical formulation, at least in physical meaning, for they are all concerned with the question whether, given an optimal trajectory to a certain point, such trajectories can be constructed to all points in a sufficiently small neighbourhood of it. Recent work has also connected these problems with that of singularity. The examination of all these problems is in its infancy, but something can be gleaned from references 10,11,22,49, 50, and probably a thorough study of Caratheodory's work on these topics in a 'classical ' context would throw a great light on the matter (18).

Chapter 7 APPLICATIONS

Certain practical applications are suggested by the concept of a field of optimal trajectories. Few of these are new, but their significance becomes much clearer in the light of the constructions we have made.

7.1 Solution of the Equations of Optimality.

7.1.1 Initial value approximation.

The set of 2n differential equations for p and x from which the control is constructed are notoriously difficult to solve, involving boundary values at two - or, as in state - constrained problems even more - points. The obvious method of solution (Kipiniak 45 p.95) is to compute a number of members of the field corresponding to a variety of boundary values, gaining a reasonable approximation to the values required for the unique trajectory satisfying all the given conditions. This technique is doomed to failure, for variations of the boundary values of p have quite unpredictable effects on the trajectories; the smallest change in $p(o)$ can produce wild fluctuations in $x(t)$, or, on the other hand it may be that a trajectory cannot be persuaded to budge even by the most provocative variations of $p(o)$. The task of choosing, a priori, values of $p(o)$ that will give a trajectory in the region of interest would drive the most phlegmatic temper to distraction.

This is a problem which has not been studied in its own right, though more will be said on the matter, but we may note first of all that it is usually more practicable to compute the field of trajectories whose members all satisfy the same initial condition than to construct the field that we have been considering hitherto, whose members all satisfy the terminal conditions. It is possible to repeat the entire theory of fields and isotims for this reversed situation with no modification other than in the definition of the optimal cost function $J(x)$, which is now 'cost so far' rather than 'cost to go';

at

$$J(x(t)) = \int_0^t L(x,u)\,dt$$

The isotims will envelope the reachable sets (or wavelets) on their concave instead of convex side. The equations associated with the one field will be the same as those for the other, except for a difference in sign of the vector J_x , and the one trajectory which satisfies both the specified initial and terminal conditions will be a member of both fields.

If the two fields could be superimposed it would be found that along this unique trajectory common to both, the isotims of one field are osculatory to those of the other, the sum of the two values being constant, equal to the total cost for that trajectory. An example of this is the disturbances issuing from two point sources in still water; the ripples from each meet tangentially along the straight line joining them. In this example the circular waves are isochrones.

Given a system $\dot{x} = f\,(x,u)$ and a cost function $\int L(x,u)\,dt$, the initial point has a unique reachable set $Q(x_o,w)$ for a given w; i,e. a set of points reachable from x_o with cost w.

For the field of trajectories issuing from x_o, the boundary of this set is identical with the isotim of value w, if x_o represents the entire initial set. $p(t)$ is a normal to such a set and the space of the p - vectors can be regarded as a linear tangent space dual to that of the contravariant vectors $\dot{x}$. (The duality between p and $\dot{x}$ extends very deeply into the basis of the calculus of variations - cf. Rund 17 p18 Courant & Hilbert 40 vol 1 p 234 Gelfand & Fomin 27 p211. Pearson 46) Since the boundary of Q is not necessarly closed the totality of normals for all its points do not span the entire dual tangent space, but only a certain cone in it, corresponding to the cone of directions swept out by $\dot{x}$ under the constraints $\dot{x} = f(x,u)$.

In particular, $p(dt)$ is the normal to the infinitesimal wavelet $Q(x_o, dw)$.

If dw is very small, $p(dt)$ is a good approximation to the initial value p_o. Evidently there is only a restricted range of values of p_o for which the corresponding trajectory is a number of the field at all, quite apart from consideration of the terminal conditions. The actual construction of the infinitesimal wavelet, or better, the admissible cones $\dot{x}_o$, p_o is not directly available, but some relevant information may be gleaned by direct application of the fundamental inequality

$$H(x_o, p_o, u) \leq H(x_o, p_o, v) \qquad 7.1$$

where u is optimal and v is any admissible control.

This may be used in various ways. For example, if the minimum is stationary for u, then we have

$$H_{uu}(x_o, p_o) \geq 0 \qquad 7.2$$

which gives an immediate constraint for p_o. Again, $u(x,p)$, derived as a result of minimising H, may be substituted for u in H.

$$H(x_o, p_o) \leq H(x_o, p_o, v) \qquad 7.3$$

which, by direct inspection, substituting possible values of v, can give useful information. Another interesting relation arises out of the fact that the minimum value of H is constant along a trajectory, thus, expressing values at t_1 t_2 by suffixes $_1$, $_2$, and where $t_2 - t_1 = dt > 0$, small,

$$H(x_2, p_2, u_2) = H(x_1, p_1, u_1) \qquad 7.4$$

Using, for brevity, the Mayer form $H = p.f$, set

$$x_2 = x_1 + f(x_1, u_1)\, dt.$$
$$p_2 = p_1 - p_1 f_x(x_1, u_1)\, dt.$$

and expanding the left member of 7.4 we have

$$H(x_1, p_1, u_2) + H_x(x_1, p_1, u_2)\; f(x_1 u_1)\, dt -$$
$$- H_p(x_1, p_1, u_2)\; p_1 f_x(x_1, u_1)\, dt = H(x_1, p_1, u_1)$$

Recalling that $H = p.f$, this becomes

$$H(u_2) - H(u_1) + \left[p_1 f_x(u_2) f(u_1) - f(u_2)\, p_1 f_x(u_1) \right] dt = 0$$

Since

$$H(x_1, p_1, u_2) - H(x_1, p_1, u_1) \geq 0$$

there must hold

$$p_1 \left[f(u_2)\, f_x(u_1) - f(u_1)\, f_x(u_2) \right] \geq 0$$

the summation being according to

$$p_1 \left[f^j(u_2)\, f^i_{x^j}(u_1) - f^j(u_1)\, f^i_{x^j}(u_2) \right] \geq 0$$

Both u_1, u_2 are optimal, but at successive instants of time. This relation can be helpful when physical considerations dictate that u should be increasing or decreasing, or at switching instants, but it also supplies a further constraint for the choice of p_o, albeit a somewhat cumbersome one.

If the isotims for the original field (based on the terminal set)are convex, the relation

$$p_o \cdot (x(t_f) - x_o) \leq 0$$

holds, for p_o is the outward normal to the isotim at x_o. Unfortunately, it is not always known in advance when this applies, though conditions can be given for certain systems (LaSalle 42, Lee 43, Pearson 46), but when it is valid it can be very helpful.

Interesting information can be obtained by actually constructing an approximation to a small wavelet from x_o in the following way. Choose a small value dw of cost; let u take all possible values, giving corresponding time increments satisfying

$$dw = L(x,u)\, dt.$$

For each (u,dt) there will be some point

$$x = x_o + f(x,u)dt$$

$$= x_o + \frac{f(x,u)}{L(x,u)}\, dw$$

This is a set of n equations with m variables, the components of u ; it is equivalent to an (n-m)-dim hypersurface. The shape of this surface can indicate whether or not any awkward behaviour is to be expected: sharp corners may indicate sources of instability(cf. Kreindler 94).

Constructing similar curves for a further dw from points on the first set, and examining their envelope can be interesting, for it occasionally happens that the second set can be reached only from a restricted region of the first, suggesting that initially the optimal trajectories are confined to a very restricted cone of directions. Of course, this is only feasible for 2 or 3 dimensions, beyond which too much effort is involved to make these simple tests worthwhile.

It is impossible to predict in general how powerful any of these criteria are; sometimes they can limit the choice of p sensationally (Halkin 47), more often they reduce the region to little better than a half-space, and each condition turns out to be a repetition of the others. Certainly this does not amount to a systematic technique for approximating initial values, and the problem deserves considerably more attention.

7.1.2 Convergence schemes.

7.1.2.1 Convergence in solution space.

Once an approximate initial value of p is obtained in the manner described above, it must be improved upon, and a recursive scheme is suggested by a closer examination of the structure of the field. An incorrect solution represents a member of the field satisfying the specified initial but not terminal conditions. An improvement is gained by adjusting the initial value in such a way as to ensure a closer fit at the end point.

The terminal values may be written

$$x_f = x(x_o, p_o, t_f)$$

$$p_f = p(x_o, p_o, t_f)$$

being the solution of 2n differential equations. As discussed above, the partial derivatives

$$z = \frac{\partial x}{\partial p_o} \quad ; \qquad w = \frac{\partial p}{\partial p_o}$$

can be found, and used to implement a scheme such as Newton's method, or a hillclimbing technique, or some modification of these methods based upon the use of derivatives. A practical technique of this type has been developed by Levine (48), but is unfortunately subject to the usual handicap of such schemes—— the initial approximation must be sufficiently accurate to ensure convergence (Saaty and Bram 52 p58). There is, however, no doubt that the solution obtained is truly optimal, for it is quite clear if the sequence has converged to a false limit, which is usually a hazard in such schemes. (This is, of course, subject to the satisfaction of the sufficiency conditions). (Levine 53).

With the help of the transformation techniques of chapter 3 and the results relating to the partial derivatives with respect to initial conditions, such a technique should cope with state constrained problems and discontinuous controls. Consider, for example, a problem involving a switching surface described by a differentiable function

$$M\ (x(\tau),\ p(\tau)) = 0 \qquad 7.5$$

and a q'th order state boundary

$$C(x) = C^{(1)}(x) = \ldots = C^{(q-1)}(x) = 0 \qquad 7.6$$

The $2n^2$ equations 6.18 with initial conditions

$$z^i_j = \frac{\partial x^i(o)}{\partial p_j(o)} = 0 \; ; \quad w^i_j = \frac{\partial p_i(o)}{\partial p_j(o)} = \delta^i_j \qquad 7.7$$

provide the partial derivatives until the switching surface is reached. According to 3.27 the derivatives are discontinuous, requiring the addition oi terms of the form

$$\left[\dot{x}(\tau^-) - \dot{x}(\tau^+)\right] \frac{\partial \tau}{\partial p_o}$$

$$\left[\dot{p}(\tau^-) - \dot{p}(\tau^+)\right] \frac{\partial \tau}{\partial p_o} \qquad 7.8$$

at τ . τ_{p_o} can be found from 7.5, for

$$M_{p_o} = M_x z + M_p w + (M_x \dot{x} + M_p \dot{p})\tau_{p_o}$$

$$= 0$$

$$\therefore \quad \tau_{p_o} = - \frac{M_x z + M_p w}{M_x \dot{x}(\tau) + M_p \dot{p}(\tau)} \qquad 7.9$$

values being taken as M is approached from the left. τ_{p_o} becomes undefined if the trajectory approaches M tangentially, but in that case $\dot{x}$ and $\dot{p}$ are continuous and the discontinuity 7.8 is zero. With the addition of 7.8 at τ the solution of 6.18 continues normally for $t > \tau$ until the next jump occurs.

At $t = t_1$ the boundary $C(x) = 0$ is reached, and the remaining $q - 1$ conditions in 7.6 can be treated as terminal boundary values for the arc $0 \leq t \leq t_1$ and p_o altered accordingly until they are satisfied. Along the boundary a suitable transformation eliminates q components of x and of p , leaving the $(n-q) \times n$ matrices x_{p_o} , p_{p_o} . At a point of return $(t = t_2)$ to the interior the q rows of each matrix are reinstated, together with new variables $p(t_2)$ (cf.section 4.3.3), but the partial derivatives are now taken with respect to $p(t_2)$. The n^2 elements of each matrix z and w now comprise nq derivatives with respect to $p(t_2)$, and $n(n-q)$ with respect to p_o. The $n + q$ variables p_o, $p(t_2)$ must be adjusted to ensure satisfaction of the $n + q$ conditions at the boundary and terminal point.

This is only a brief sketch of the procedure----it is not possible to give a complete recipe for solving problems in a straightforward way, for each raises its own peculiar problems requiring endless modification and refinement. The amount of work involved in solving these problems is daunting in the extreme.

Another technique of boundary-value approximation-----too well-known to require repetition here-----is Neustadt's method (60) applying convexity properties of the reachable set for linear time-optimal problems. With the transformation indicated in 6.2, every problem can be expressed in time-optimal form, but the convexity requirement is a real restriction. Where it applies, Neustadt's technique and the various modifications of it (61, 62) can be quite attractive, for they do not require a large number of additional differential equations.

7.1.2.2 Convergence in control space.

There are basically two lines of attack for the two-point boundary value problem of control. The first, discussed above, involves approximations of optimal trajectories; the other, of which there are many possibilities of variation, uses a sequence of non-optimal trajectories converging to the optimum. (Aoki 65, Bryson and Denham 66, Dreyfus 38, Kelley 67, Halkin 47). All these schemes fit into our geometric construction in the following way: a point $x(h)$ on an arbitrary admissible trajectory from x_o at which the accumulated cost is h, is interior to the reachable set for that value of cost. Thus, if the optimal values corresponding to h are x_h, p_h, we have, from 6.13,

$$p_h \cdot \left[x(h) - x_h\right] \geq 0$$

The object of the iterative process, whatever its technical details, is to decrease the value of this inequality.

One possibility is the following: the cost function is $\int L(x,u)dt$, the terminal conditions

$$T^k(x) = 0 \qquad k = 1, \ldots, r < n \qquad 7.10$$

A non-optimal trajectory will not, in general, satisfy 7.10, and we may construct an additional cost function $\frac{1}{2}\sum (T^k(x))^2$ which is to be minimised. The terminal value of p for this Mayer function is

$$p_i(t_f) = T^k T^k_{x^i} \qquad 7.11$$

Choose a nominal control $v_1(t)$, giving a trajectory $x(t)$, $0 \leq t \leq t_f$, where t_f is chosen either arbitrarily, or using one of the T^k as a stopping condition. This $x(t)$ is the basis of a new dynamic system

$$\dot{x} = f(x(t), u) \qquad 7.12$$

the right hand side being a function only of (t, u). For this system, an optimal trajectory can be constructed in reverse time from t_f, using the minimum principle, and yielding a control $v_1^*(t)$, which is optimal for the approximate system 7.12. The next control chosen for a forward integration is

$$v_2(t) = v_1(t) + c(t)v_1^*(t) \qquad 7.13$$

$c(t)$ being chosen in some way to ensure rapid convergence. This is only one possibility, but most techniques exhibit properties in common with this. It is not to be recommended as a practical scheme without a careful convergence analysis.

7.2 Feedback.

The techniques discussed above relating to the solution of the differential equations, also have direct relevance to the construction of feedback control schemes. The obvious, but crude, flooding technique, involving the construction of a skeleton field of optimal trajectories, all satisfying the specified terminal conditions, is subject to the difficulties of finding suitable boundary values for $p(t_f)$, and, at the present time, not a feasible technique in general, though if in special cases the field proved easy to construct, a suitable interpolation scheme could provide a reasonable approximation to the control. More promising techniques are based upon approximation of the isotims $J(x)$, for if their functional form is known, $u(x,J_x)$ will be given at every point.

Such schemes were proposed early in the development of optimal control, but without this geometric motivation, and involved the approximation of $J(x)$ by a quadratic function, (Merriam 51)

$$J(x) = a_o(t) + a_i(t)x^i + a_{ij}(t)x^i x^j \qquad 7.14$$

differential equations being found for the coefficients, which absorb the higher order non-linearities. The technique founders on the difficulties of determining boundary values for the coefficients, but where this can be satisfactorily done, useful results can be obtained, (Pearson 63, Davis 64), especially for the linear regulator problem for which 7.14 is a precise representation. (Kalman 9).

Another popular scheme, frankly local in character, is closely allied to the technique of the previous section involving partial derivatives with respect to boundary values, but here the basic field is constructed with reference to the terminal conditions, and the object is to obtain a scheme for correcting errors due to perturbation from the prescribed trajectory.

A perturbation from the expected value of x indicates that the state point is on the path of a neighbouring trajectory, and the proper control is not $u(t)$ as computed, but $u(t)+\delta u(t)$. Since the optimal control is determined as a function $u(x,p)$, the relevant correction is

$$\delta u = u_x \delta x + u_p \delta p.$$

δx is the measured error, and δp is to be found. The difference between two trajectories can be traced back to different initial conditions, thus

$$\delta x = x_{x_o} \delta x_o, + x_{p_o} \delta p_o \qquad 7.15$$

and correspondingly

$$\delta p = p_{x_o} \delta x_o + p_{p_o} \delta p_o \qquad 7.16$$

The partial derivatives are evaluated in a way similar to that described in 7.1.2.1, but we require terminal conditions for all the four matrices in 7.15 , 7.16. We have

$$x_{x_o}(o) = p_{p_o}(o) = \text{unit matrix},$$

and at the terminal point there are n relations of the form

$$T\ (x(t_f),\ p(t_f)) = 0,$$

giving

$$(T_x x_{x_o} + T_p p_{x_o})\ \delta x_o + (T_x x_{p_o} + T_p p_{p_o})\ \delta p_o = 0$$

from which both $x_{p_o}(o)$ and $p_{x_o}(o)$ can be found, but requires the solution of a linear two-point boundary value problem. Then

$$\delta u = \left[u_x + u_p \left\{ p_{x_o}(x_{x_o})^{-1} + p_{p_o}(x_{p_o})^{-1} \right\} \right] \delta x$$

This indicates the bare bones of the scheme, of which several versions have been published, differing in detail but the same in essence. The usefulness of such a plan is very limited, for feedback of this nature is required when the system does not follow the planned course. This only occurs when the differential equation is not a sufficiently accurate representation of the physical system, or in the presence of unpredictable disturbances. This scheme is based on a deterministic system which is assumed to be correct, and therefore cannot deal with either of those situations, except in rare cases, when perturbations are impulsive, the system being deterministic over the intervals between them, or when initial conditions are not accurately known. Even in theses cases it cannot be used with any confidence in the absence of an estimate of the error involved in the linearization.

7.3 Education

Although every subject must be taught and learnt, the educative possibilities of any new study are invariably the most neglected. The contribution to theory or to practical application is always noted, but the question whether the ideas are straightforward or easy to grasp, is ignored. This may be of little concern to the experienced scholar, but to teachers and students it is crucial. In basing the theory upon physical rather than mathematical principles, and developing it along geometric and not analytic lines, this thesis attempts to contribute to engineering education - as much an 'application' as is any practical or computational technique.

Chapter 8 CONCLUSION

The geometric approach to the study of the optimal behaviour of differential engineering systems discloses properties which are obscured by other methods. In cases which can be satisfactorily handled in other ways no great improvements are to be expected, and our constructive conditions for optimality are precisely the same, for solutions in open regions of state space, as the familiar ones given by Pontryagin and derivable via classical arguments; but even here geometric considerations show promise of providing powerful tools for the numerical solution of the differential equations. The examples in Appendix D indicate the possibilities, but a systematic attack on theoretical aspects of the peculiar difficulties of these two-point boundary value problems still awaits treatment.

For problems involving restricted state space the concept of local optimal subspaces offers distinct advantages over other approaches. First, it does not treat bounded problems as a different species, but applies a uniform treatment to all problems, the boundary being regarded as a natural part of the background to the problem rather than as an externally imposed constraint. Second, it illuminates certain matters, which, approached in other ways, have been the source of much confusion. Third, the specification of the boundary leads to simplification of the problem in that region by virtue of a reduction in the dimension of the system, constrasted with the increased complexity incurred by other techniques. This type of simplification is not an accident of technique; it is fundamental to the approach, and should be expected whenever constraints appear in a problem of any type.

Lagrange's multipliers, as used in ordinary minimization problems, have the effect of introducing extra artificial degrees of freedom to compensate for the restrictions imposed by constraints, as an alternative to elimination of variables. It should be a matter for surprise, though we have become immune to it, that a restriction on the mode of behaviour of a system should not lead to simplification through excluding many alternative possibilities, rather than complication. When the problem is not simplified it may be an indication that the method of treatment is not ideal.

The multipliers familiar from the calculus of variations are introduced to ensure compatibility with the dynamic system. The equivalent variables in our treatment obviously are not amenable to this interpretation, but may be said to ensure compatibility of the solution with the constraint of optimality, and , pursuing the analogy, they increase the apparent complexity of the system while its freedom is reduced from that corresponding to unspecified control in a system of n first order equations, to that consonant with a set of n completely defined second order equations.

The insights gained by 'arguments by analogy', of which the above is a simple example, are a reminder that no method or viewpoint stands on its own, independent of others. The discussion in Chapter 5 demonstrated that each approach illuminated the problem in ways which, by their very nature, were outside the scope of other methods. Every method must admit the shortcomings of its own merits - the brighter the light, the stronger the shadow that it casts - and the temptation must be avoided of adopting one consistent viewpoint to the neglect of others.

An important fact common to all the available techniques is that despite talk of fields of trajectories the necessary conditions appear in the form of differential equations, the solution of which is a time function appertaining to only one trajectory. Although it is a primary aim of control theory to obtain feedback controls, they cannot arise from any of these techniques, and it is perhaps a little deceptive to present the discussion in terms of $u(x)$ rather than $u(t)$. Only in the case of a one-dimensional system can an explicit feedback control be obtained, by eliminating J_x from the equation $H(x,J_x) = 0$, but in other cases the scalar t cannot be replaced by the vector x as an argument of u.

It seems unlikely that this problem can be overcome by anything less drastic than a complete reformulation of the problem, for if the system is given in the time - like form of a differential equation, and the cost function is expressed as a time integral, the solution cannot be expected to emerge as a space-like function. Probably, what is required is a different form of system description, symmetric in all variables rather than giving special prominence to time. This is a fundamental issue in system theory and little effort has been put into it.

The conclusions that we are forced to are not far removed from the argument of the introduction: the problem we have been discussing is too limited, with its first order ordinary differential equations and scalar cost function - certainly it can no longer be regarded as <u>the</u> problem of optimal control - and it is time to call a truce to the vast effort being expended on it. Perhaps the only aspect of it which re.lly must be dealt with is the problem of constructing fields of extremals. If this could be easily done a great deal of information would be immediately available about the structure of the system, and feedback schemes would not be far behind. This all hinges on the boundary value problem, for which the techniques suggested in Chapter 7 and demonstrated in Appendix D open a door to more thorough treatment.

For the more general, and perhaps more pressing problems we must find a better method of representing systems, and more reasonable measurements of performance: the aim is a framework which will support a theory of feedback control of multivariable systems in accordance with flexible performance demands.

Nothing has been said of non-deterministic systems, adaptive-learning systems, or information-seeking systems, or.... but that is another story.

Appendix A. Physics in Control Theory.

The engineer and the physicist are occupied with two sides of what is essentially the same problem. 'Control' is equivalent to 'order': a process that is not entirely chaotic is, in a sense, controlled, and the physical laws express the principles governing the control action. For the physicist, the system is given and observable, and he must deduce the underlying principles. Being generalizations from empirical evidence, these are always subject to doubt. The engineer, on the other hand, is furnished with the principle, and he turns it into a practical programme for implementation. To perceive order is physics: to impose it, engineering. Nowhere does this correlation appear so vividly as in the treatment of the optimal control problem. Analogies with natural dynamic systems abound, and while we can go no further here than pointing to superficial likenesses, they are sufficiently interesting to touch upon. If, as Koestler maintains (85 p201), the essence of discovery is the marriage of previously unrelated frames of reference, it is more than likely that the further pursuit of these analogies will yield fruitful results, and it is worth suggesting areas which may prove rich, and some where analogies break down under scrutiny.

The most obvious branch of physics in this context is analytical mechanics, which suggests itself by virtue of the minimum and variational principles which underlie the science. The function H = L(x,u) + p.f(x,u) is evidently a Hamiltonian function, and

$$\begin{aligned} \underline{L} &= \underline{L}_{\dot{x}} . \dot{x} - H \\ &= p.(\dot{x} - f) - L \end{aligned}$$

is a Lagrangian. The canonical form in which the differential equations of optimality are expressed is derived from the equivalent form in dynamics, but beyond these formal analogies little has been done. Deeper parallels are perhaps not to be sought, for the very concepts of 'particle' and 'mass' are lacking in control theory, precluding the direct use of such concepts as kinetic energy, momentum, etc. Nevertheless, certain techniques could be formally applied; Poisson-bracket techniques (Whittaker 86 p.308) might occasionally provide extra integrals of the motion, but are rarely of great help. Transformation theory (86 p.283) gives interesting ideas, but to transform a particular system into a more convenient form requires too much ingenuity and luck to make it a reliably useful tool. The optical analogue of dynamics has already been applied in chapter 6.

A more promising source of ideas is continuum mechanics, though the status of variational or minimum principles is uncertain in this area; some writers grant them axiomatic status (e.g.,Edelen 87), others are scathing in their criticism, claiming that such principles are arbitrary, not sufficiently fundamental (Truesdell 71 p.595), or lacking in physical meaning (Kilmister 88 p. 49). This last point is interesting. In a natural system for which a variational principle can be found, a suggested variation can be effected only by forcing the motion to be other than what it in fact is, using constraints. In that case we are dealing with a different system, and comparisons are invalid. Borrowing control ideas, it might be possible to represent the natural system as an optimal version of a more general system in which some variables correspond to the control; this system might have an interesting physical interpretation, for, as we shall see, a similar situation does arise in thermodynamics.

Just as the minimum principle can be regarded as less than fundamental in physics, so can it be given the same inferior status in some contexts of control theory. This paradoxical situation arises when dealing with fields of optimal trajectories, for all members of the field are 'equally optimal' and the concept of optimality, being a comparative property, loses its force. Thus the concept was not applied in this thesis until chapter 4, and it would have been entirely possible to derive all the properties of optimal trajectories except the inequality relations without its use. The minimizing property is crucial to a 'constructive' theory (an engineer's job), but not to an investigation of optimal systems as such (a physicist's) where a relation such as

$$L(x,u) + J_x \cdot f(x,u) = 0 \qquad \text{A.1}$$

is more important.

A.1 could be treated as a conservation law in a field theory of optimal control. $J(x)$ itself suggests a potential, and $J_x \cdot f$ would be a rate of work in the potential field, $L(x,u)$ representing some deformation power. Further conservation laws are provided byLiouville's theorem, of which the most familiar form states that the 'volume', $\int dx\, dp$, is an invariant of the motion when the points constituting it move in accordance with a canonical system of equations. The transformations associated with a reduction in the dimension of state space do not affect the canonical form of the equations or the validity of this theorem. More general forms of such invariants are given by Synge (68 p.173).

The line integral $\int J_x \cdot dx$ is independent of path, (naturally, suitable local transformations must be made, respecting the dimension of each region), indicating that the vector field J_x is lamellar (Ericksen 89 p.824). This type of property invites further analysis along the lines of tensor field theory, the relevance of which is evident from results obtained, for example, for discontinuities and shocks: the only possible jump in a lamellar field is normal to a surface, the tangential components remaining continuous (71 p.494). This result we obtained here by the special arguments of chapter 3. More interesting results might be gained by analysing the vector field $u(x)$ along these lines.

Thermodynamics offers even more intriguing possibilities. One approach to this subject is via the caloric equation of state (71 p.619)

$$e = e(\eta, v)$$

η being a scalar (entropy) and v a state vector, representing physical properties of the system. e is the internal energy. Thermodynamic tensions are defined as

$$q = e_v \quad , \quad \theta = e_\eta$$

θ being temperature. Hence we have

$$de = \theta \, d\eta + q dv$$

and

$$\dot{e} = \theta \, \dot{\eta} + q\dot{v} \, .$$

Inequalities, such as

$$de \leq \theta \, d\eta + q dv \qquad \text{A.2}$$

rule out certain non-equilibrium states.

The formal similarity to equation A.1, and even its more general inequality form, is startling, but not so close as to be immediately translatable. What is particularly interesting is the possibility of variation to unstable states, admitted by A.2 , which surely bear some relation to the non-physical variations of mechanical systems noted above. Truesdell's comment (71 p.659): "In a theory where mechanical phenomena are of primary interest, it may be natural to seek and impose a requirement of universal stability, but in a theory aiming to determine criteria of stability of equilibrium it is more natural to include the theoretical possibility of unstable states", proclaims , for physics, the distinction between optimal and non-optimal behaviour in control theory, and could almost be a reply to Kilmister's strictures mentioned above.

A theorem of Carathéodory (75) applied to the problem of adiabatic accessibility of equilibrium states has been used in a control context in the problem of accessibility by extremals (11, 22) but otherwise the rapprochement between control theory and physics is not yet under way. In control one might expect to obtain global results for the solution space, on which feedback schemes might be based, but useful developments cannot be guaranteed. The outlook for physics is brighter, and we can hope for a unified classical field theory in which control concepts play a basic role and minimum principles regain a fundamental, though not axiomatic status,(which should satisfy everyone, even such prophets of 'variationalism' as Lanczos (28)), serving to distinguish actual from theoretically conceivable behaviour. Such developments must come from physicists rather than from engineers, and judging from the present state of the dialogue between the disciplines, the revolution will be a long time coming. Indeed, we might hope to delay it until the feedback concept has quite ousted open-loop methods, for, recalling the metaphysical excitement caused by the mildly teleological variational principle in the eighteenth century, the mind boggles at the thought of letting loose the idea of the universe as an open-loop control system!

Appendix B. Transformation of the Auxiliary Equations.

The transformation will be carried out for the equations of the Mayer problem, avoiding the inclusion of L_x which obviously transforms without difficulty.

The notation is that coordinates in the z - system are denoted by primed indices, those in the x - system are unprimed.

The vectors q, g, correspond to p, f, respectively, thus

$$q_{r'} = A^k_{r'} p_k \qquad g^{s'} = A^{s'}_r f^r$$

where

$$A^a_{b'} = \frac{\partial x^a}{\partial z^b}$$

We write $\frac{\partial}{\partial x^r}$ as ∂_r ; $\frac{\partial}{\partial z^s}$ as $\partial_{s'}$.

The equations, in x - coordinates, are

$$\dot{p}_i + p_j \partial_i f^j = 0 \qquad \text{B.1}$$

Now

$$\begin{aligned} \dot{p}_i &= \frac{d}{dt}(A^{r'}_i q_{r'}) \\ &= \partial_s(A^{r'}_i) f^s q_{r'} + A^{r'}_i \dot{q}_{r'} \\ &= \partial_s(A^{r'}_i) A^s_{r'} q_{r'} g^{r'} + A^{r'}_i \dot{q}_{r'} \end{aligned} \qquad \text{B.2}$$

and

$$\begin{aligned} p_j \partial_i f^j &= A^{r'}_j q_{r'} \partial_i (A^j_{s'} g^{s'}) \\ &= A^{r'}_j q_{r'} A^{t'}_i \left[\partial_{t'}(A^j_s) g^{s'} + A^j_{s'} \partial_{t'} g^{s'}\right] \end{aligned} \qquad \text{B.3}$$

Adding B.2 and B.3 , and performing the usual manipulations of tensor calculus, we have

$$A^{r'}_i\left[\dot{q}_{r'} + q_{s'} \partial_{r'} g^{s'}\right] + q_{r'} g^{r'}\left[A^s_{r'} \partial_s A^{r'}_i + A^{r'}_j A^{t'}_i \partial_{t'} A^j_{r'}\right] = 0$$

The term in brackets in the second member is equal to

$$\begin{aligned} &A^s_{r'} \partial_i (A^{r'}_s) + A^{r'}_s A^{t'}_i \partial_{t'} (A^s_{r'}) \\ &= \partial_i (A^s_{r'} A^{r'}_s) \\ &= 0 \ . \end{aligned}$$

Thus,

$$\dot{p}_i + p_j \partial_i (f^j) = A_i^{r'} \left[\dot{q}_{r'} + q_{s'} \partial_{r'} g^{s'} \right] \qquad \text{B.4}$$

showing that the expression on the left of B.1 does in fact transform as a covariant vector, and maintains its value of zero.

In an r-dim. space, suppose $A_i^{r'}$ to be $n \times n$ and chosen such that $f^{r+1} = \ . \ . \ . = f^n = 0$. Using indices

$$k, s = 1, \ldots, r \ ; \quad m, t = r+1, \ldots n$$

the left hand side of B.4 becomes

$$\dot{p}_s + p_k \partial_s f^k + p_m \partial_s f^m = 0$$
$$\dot{p}_t + p_k \partial_t f^k + p_m \partial_t f^m = 0$$

The second equation can be ignored, for p_t is undefined. The expression ${}_s f^m$ is the component of a gradient in a direction parallel with the tangent space to which p_m is normal. The product $p_m \ {}_s f^m$ must be zero, leaving the r-dim. vector equation

$$\dot{p}_s + p_k \partial_s f^k = 0 \ .$$

Appendix C. Discrete Derivation of Auxiliary Equations.

The basic technique for the classical problem was given by Cicala (92). The operations of minimisation and integration can be reversed, for the Lagrange problem, only if they are independent. Obviously they cannot be independent, for successive values of x are connected by the dynamic equations, but if this constraint is included using Lagrange multipliers, the interchange is permissible.

$$\int_0^{t_f} L(x,u)dt = \underset{\substack{\Delta t_i \to 0 \\ N \to \infty}}{\mathrm{lt}} \sum_{i=1}^{N} L(x_i , u_i)\Delta t_i ,$$

where $\sum_{i=1}^{N} t_i = t_f$.

$$\min \int L\, dt = \mathrm{lt} \sum_i \min \Big[L(x_i , u_i)\Delta t_i - p_i(x_{i+1} - x_i - f(x_i , u_i)\Delta t_i) \Big]$$

Minimising with respect to x and u at each instant, since the value of u is independent of its value at any other time, we have simply

$$\min_{u_i} \Big[L(x_i , u_i) + p_i f(x_i , u_i) \Big] ,$$

whereas for x_i , assuming a stationary minimum,

$$\Big[L_{x_i} + p_i f_{x_i} \Big] \Delta t_i - p_{i-1} + p_i = 0$$

for each i , and in the limit $\Delta t_i \to 0$,

$$\dot{p} = -L_x - pf_x .$$

Appendix D. Examples.

We shall discuss several problems from the point of view developed in the text. The object is not to obtain solutions——such was never the purpose of this thesis——but to demonstrate certain points which are all the clearer for being exemplified by familiar and simple cases. These are taken, either directly, or in modified form, from Pontryagin (1), Bryson and Denham (39), Dreyfus (38), and Kipiniak (45). Some valuable examples, not all reproduced here, of problems with analytic solutions, designed to establish the validity of Bellman's partial differential equation, and the relation between J_x and Pontryagin's $\psi(t)$ (our $p(t)$) are to be found in Fuller 95. 96.

Example 1. Second order, linear, time-optimal.

Following the specification of section 2.3 we have

$$\dot{x}^1 = x^2 \qquad \dot{x}^2 = u$$

U defined by constraints $B_1 : (-u - 1) \leq 0 \qquad B_2 : (u - 1) \leq 0$

Initial and terminal sets $S : (C_1 , C_2) \qquad T : (0 , 0)$

cost function $\int_0^{t_f} dt \equiv t_f$.

Using the Lagrange form,

$$H = p_1 x^2 + p_2 u = -1$$

The dimensionality test (section 5.1.1) indicates that since

$$(H + qB)_{uu}$$

is always singular, there may be regions of dimension less than 2 , i.e., the field may degenerate to single trajectories.

The minimum principle gives

$$\dot{p}_1 = 0 \qquad \dot{p}_2 = -p_1 \qquad u = \pm 1$$

This problem has a familiar analytic solution,

$$\left.\begin{aligned} x^1 &= c_1 + c_2 t \pm \tfrac{1}{2} t^2 \\ x^2 &= c_2 \pm t \end{aligned}\right\} \text{ for } u = \pm 1 \text{ respectively ;}$$

the switching curves are given by

$$M(x) = x^1 \pm \tfrac{1}{2}(x^2)^2 = 0 \qquad \text{D.1}$$

on which $u = \mp 1$ respectively, and are the 1-dim. manifolds alluded to. The isotim value is equal to the optimal time to go; for points on M

$$J(x) = \pm x^2 \quad (u = \mp 1) , \qquad \text{D.2}$$

and for other points, say c, the time to reach M is given by

$$(c_1 + c_2 t \pm \tfrac{1}{2}t^2) \pm \tfrac{1}{2}(c_2 \pm t)^2 = 0$$

$$\therefore \; t = \mp c_2 \pm (\tfrac{1}{2}c_2{}^2 \mp c_1)^{\frac{1}{2}} \qquad (u = \pm 1) \qquad \text{D.3}$$

and at the switching point,

$$\begin{aligned} x^2 &= c_2 \pm t \\ &= (\tfrac{1}{2}c_2{}^2 - c_1)^{\frac{1}{2}} \text{ or } -(\tfrac{1}{2}c_2^2 + c_1)^{\frac{1}{2}} \end{aligned} \qquad \text{D.4}$$

for $u = +1 , -1$ respectively, before switching.

D.2, D.3, D.4 give, since c is any point,

$$J(x) = \begin{cases} -x^2 + 2(\tfrac{1}{2}(x^2)^2 - x^1)^{\frac{1}{2}} & u = +1 , -1 \\ x^2 + 2(\tfrac{1}{2}(x^2)^2 + x^1)^{\frac{1}{2}} & u = -1 , +1 \end{cases} \qquad \text{D.5}$$

Evidently $J(x)$ is continuous. In the 2-dim. regions we have

$$J_{x^1} = \begin{matrix} -(\tfrac{1}{2}(x^2)^2 - x^1)^{-\frac{1}{2}} \\ -(\tfrac{1}{2}(x^2)^2 + x^1)^{-\frac{1}{2}} \end{matrix} \qquad J_{x^2} = \begin{matrix} -1 + x^2(\tfrac{1}{2}(x^2)^2 - x^1)^{-\frac{1}{2}} \\ 1 - x^2(\tfrac{1}{2}(x^2)^2 - x^1)^{-\frac{1}{2}} \end{matrix} \qquad \text{D.6}$$

which are not continuous and not defined on M.

Following the arguments of section 3.2.3 we introduce new vectors y , J_y such that one component of $\dot{y}$ is tangent to M , the other normal.

Thus $\dot{y} = A\dot{x}$ $\quad A = \{a_{ij}\}$ $\quad i, j = 1, 2.$ D.7

Choose $\dot{y}^2 = M_x \cdot \dot{x}$

$$\therefore \quad M_{x^1} = 1 = a_{21} \qquad a_{22} = M_{x^2} = \pm x^2 .$$

$\dot{y}^1$, $\dot{y}^2$ are mutually orthogonal,

$$\therefore \quad a_{11}(x^2)^2 \pm a_{12}x^2 = 0$$

and $\det A = 1$ D.8

$$\therefore \quad \pm a_{11}x^2 - a_{12} = 1$$

giving

$$A = \begin{bmatrix} \pm \frac{1}{2x^2} & -\frac{1}{2} \\ 1 & \pm x^2 \end{bmatrix} \qquad \text{D.9}$$

corresponding to $u = \mp 1$ respectively, on M.

$$\therefore \quad A^{-1} = \begin{bmatrix} \pm x^2 & \frac{1}{2} \\ -1 & \pm \frac{1}{2x^2} \end{bmatrix} . \qquad \text{D.10}$$

Using $J_y = J_x A^{-1}$, D.11

together with D.6 and D.1 we have, for the upper branch of the switching curve $(u = -1)$,

$$\begin{aligned} J_{y^1} &= -(\tfrac{1}{2}(x^2)^2 - x^1)^{-\frac{1}{2}}x^2 + 1 - x^2(\tfrac{1}{2}(x^2)^2 - x^1)^{-\frac{1}{2}} \\ &= -1 \end{aligned}$$

and

$$\begin{aligned} J_{y^1} &= -(\tfrac{1}{2}(x^2)^2 + x^1)^{-\frac{1}{2}}x^2 - 1 + x^2(\tfrac{1}{2}(x^2)^2 + x^1)^{-\frac{1}{2}} \\ &= -1 \end{aligned}$$

as M is approached from one side or the other. J_{y^1} is thus uniqely defined.

But

$$\begin{aligned} J_{y^2} &= -(\tfrac{1}{2}(x^2)^2 - x^1)^{-\frac{1}{2}}\tfrac{1}{2} + \frac{1}{2x^2}\left[-1 + x^2(\tfrac{1}{2}(x^2)^2 - x^1)^{-\frac{1}{2}}\right] \\ &= -1/2x^2 \end{aligned}$$

and

$$\begin{aligned} J_{y^2} &= -(\tfrac{1}{2}(x^2)^2 + x^1)^{-\frac{1}{2}}\tfrac{1}{2} + \frac{1}{2x^2}\left[1 - x^2(\tfrac{1}{2}(x^2)^2 + x^1)^{-\frac{1}{2}}\right] \\ &= \infty \end{aligned}$$

Thus J_{y^2} is not defined, but since $\dot{y}^2 = 0$,

$$H = J_y \cdot \dot{y} = -1 \qquad \text{D.12}$$

In principle, the entire problem could be treated in the transformed spaces. D.11 (inverted), D.12 and D.9 give

$$H = J_{y^1}(\pm 1 - u)\tfrac{1}{2} + J_{y^2}(1 \pm u)x^2 = -1 ,$$

then.

$$\dot{J}_{y^1} = -J_{y^2}(1 \pm u)\, \partial x^2/\partial y^1$$

$$= -J_{y^2}(1 \pm u)\, a_{21}^{-1}$$

$$= J_{y^2}(1 \pm u)$$

$$\dot{J}_{y^2} = J_{y^2}(1 + u)\frac{1}{2x^2}$$

$$u = -\mathrm{sign}(\pm x^2 J_{y^2} - \tfrac{1}{2}J_{y^1})$$

The transformation was one of tangent spaces, not the state space itself, and is x - dependent. The equations for $\dot{y}$ are irrelevant, and what should be treated is the set

$$\dot{x}^1 = x^2 \qquad \dot{x}^2 = u$$

$$\dot{q}_1 = q_2(1 \pm u) \qquad \dot{q}_2 = q_2(1 + u)/2x^2$$

$$u = -\mathrm{sign}(\pm x^2 q_2 - \tfrac{1}{2}q_1)$$

After switching, the $\dot{q}_2$ equation is discarded, and $\dot{q}_1 = 0$.

Of course, this is not proposed as a practical technique, for M is not known until the problem is solved, but it demonstrates the principle.

If we propose to solve this problem numerically, the ideas of chapter 7 come into play.

Suppose the initial point is $(-1 , 0)$; it is necessary to determine the initial values of p . Knowing that $u(0) = \pm 1$, we may apply 7.4a , obtaining

$$p_1(0)(u(dt) - u(0) \geq 0$$

which gives no information unless there is a switch at $t = 0$. This relation can be applied at $(t, t+dt)$ as well as at $(0, dt)$, with the result that if u switches $-1 \rightarrow +1$, $p_1 = \text{const.} \geq 0$, and if the reverse, $p_1 \leq 0$.

Since the isotims for this problem are known to be convex (Neustadt 60) 7.4b applies, giving

$$p_1(0)(1) \leq 0$$

and setting $H(0) = -1$ we have

$$p_2(0) = \pm 1 \; ; \quad u(0) = \mp 1$$

which considerably reduces the range of search for $p(0)$.

To construct an infinitesimal reachable set, note that for e small,

$$x^1(e) = -1 \qquad\qquad x^2(e) = u(0)e$$
$$x^1(2e) = -1 + u(0)e^2 \qquad x^2(2e) = (u(0) + u(e))e \qquad\qquad \text{etc.}$$

Some points reachable in these two stages are shown in Fig. D.1, for pairs $u(0)$, $u(e)$ from $-1 \leq u \leq +1$, giving a closed curve. Since the normals to this set can point in all directions, no further information is obtainable for $p(0)$.

Fig. D.2 shows wavelets fanning out from points on the boundary of the reachable set for 2e. The envelope of these gives the boundary of the reachable set for 3e, and approximates to the familiar shape of the isochrones for this system.

<u>Example</u> 1a. Bang-bang : state constraint.

To the above problem add the constraint

$$C(x) = x^2 - .5 \leq 0$$

The boundary represents a 1-dim. manifold----a single trajectory----for which we have an explicit form, and can therefore transform to $\dot{y}$, J_y on it.

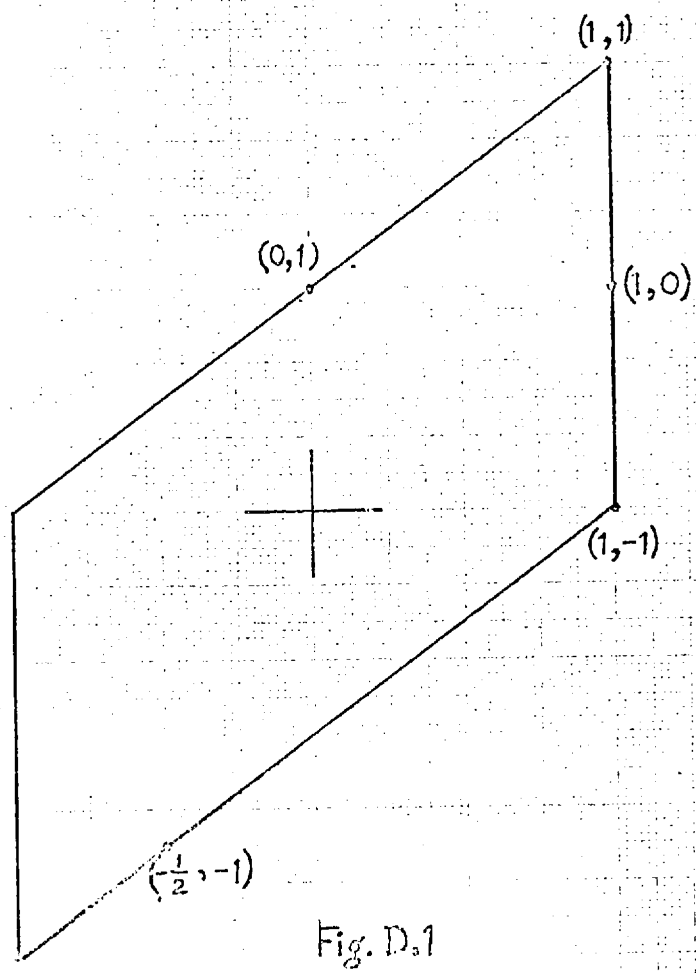


Fig. D.1

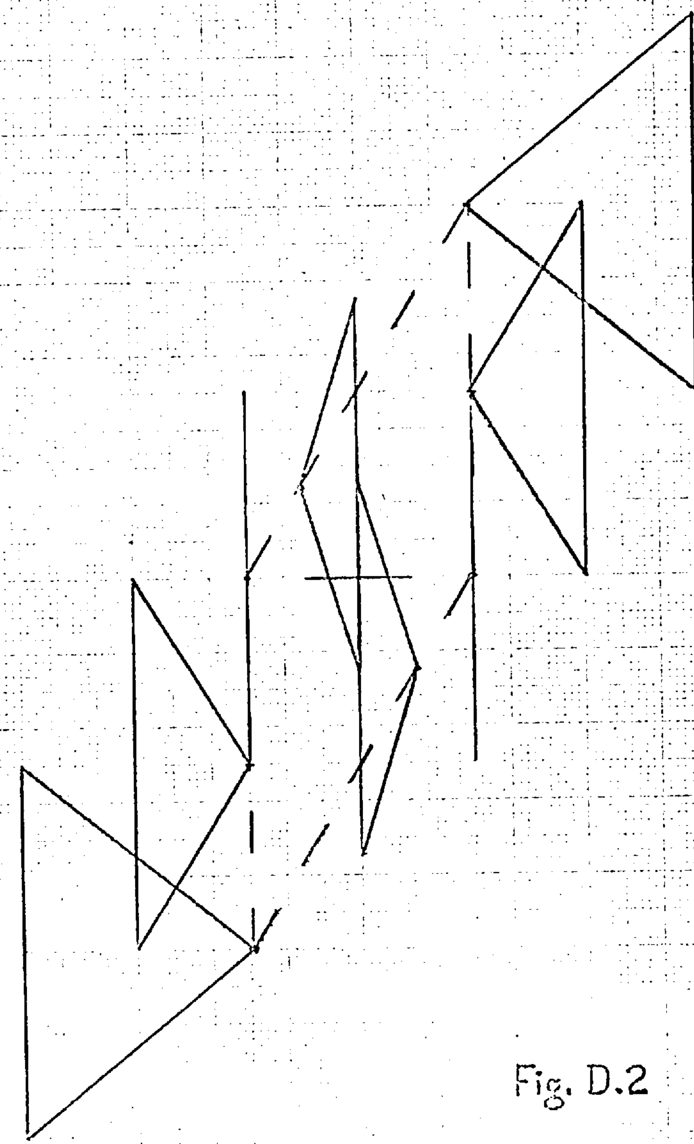


Fig. D.2

As before, choose

$$\dot{y}^2 = C_{x^1}\,\dot{x}^1 + C_{x^2}\,\dot{x}^2 = u$$

The conditions that the Jacobian of the transformation be unity, and that $\dot{y}^1$ be orthogonal to $\dot{y}^2$, give

$$\dot{y}^1 = x^2 \qquad\qquad \dot{y}^2 = u$$

which is where we came in ! We are now assured that p_1 is continuous throughout, and p_2 may be ignored on the boundary.

The trajectory comprises three parts: i, iii interior, ii on the boundary. It is usually a good plan in such cases to deal with the interior arcs first, as far as possible, fitting in the boundary arc later. For the final arc we have the special argument of section 4.3.5: the trajectory leaves the boundary at the same point as the unique unconstrained trajectory which touches the boundary at only one point. It is important to recognize that this condition, whether we can use it explicitly or not, implies that, for 2-dim. problems, the final arc can be solved independently of the rest of the problem.

In this case that arc is easily found, as Fig.D.3 shows, the point at which it leaves the boundary being $x^1 = -.125$, but if it were not readily obtained we would proceed as follows. To investigate the point of exit from the boundary we use 4.25:

$$p_1(x^2 - x^2) - p_2 u = 0$$

Evidently, either $u = 0$ or $p_2 = 0$, but since $u = -\text{sign}\, p_2$, $\dot{p}_2$ must be zero in either case. In addition,

$$H = p_1 x^2 + p_2 u = -1,$$

$\therefore$ $p_1 = -2$, and since $\dot{p}_2 = -p_1$, we have $u = -1$, which gives all the

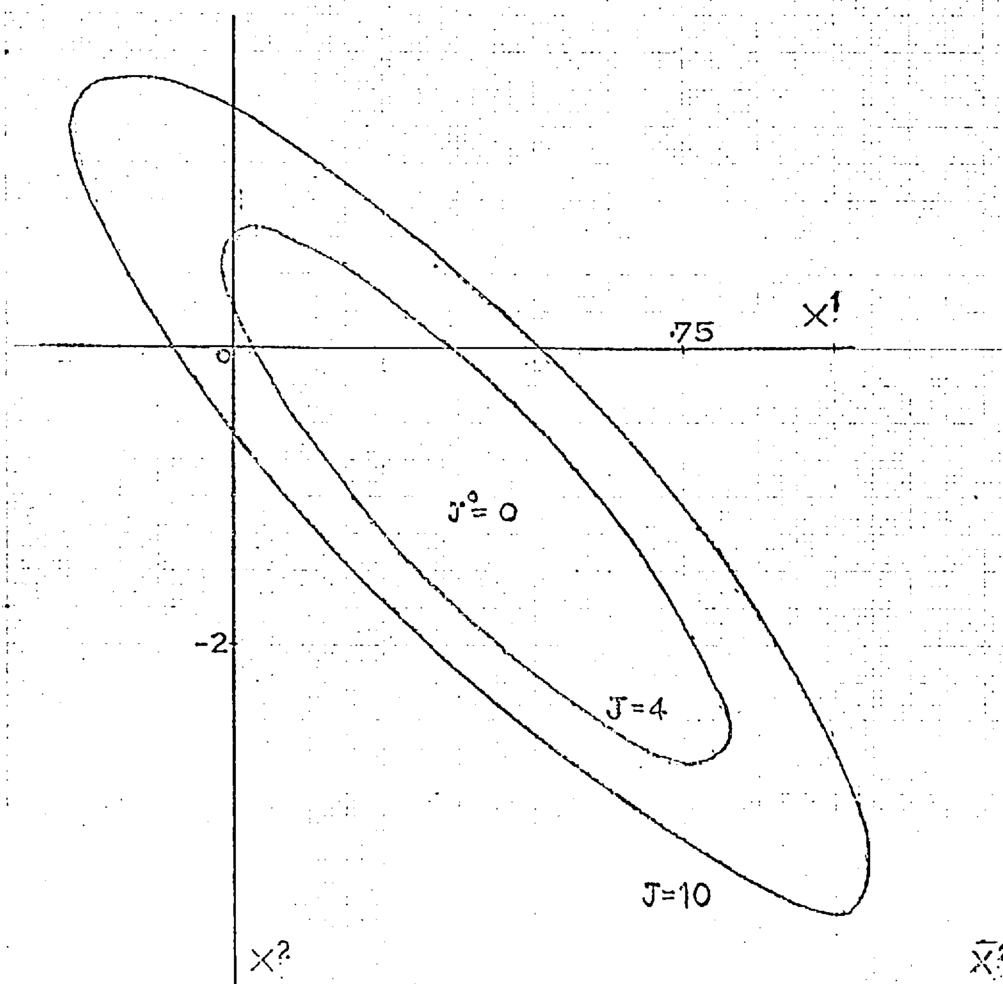


Fig. D.4

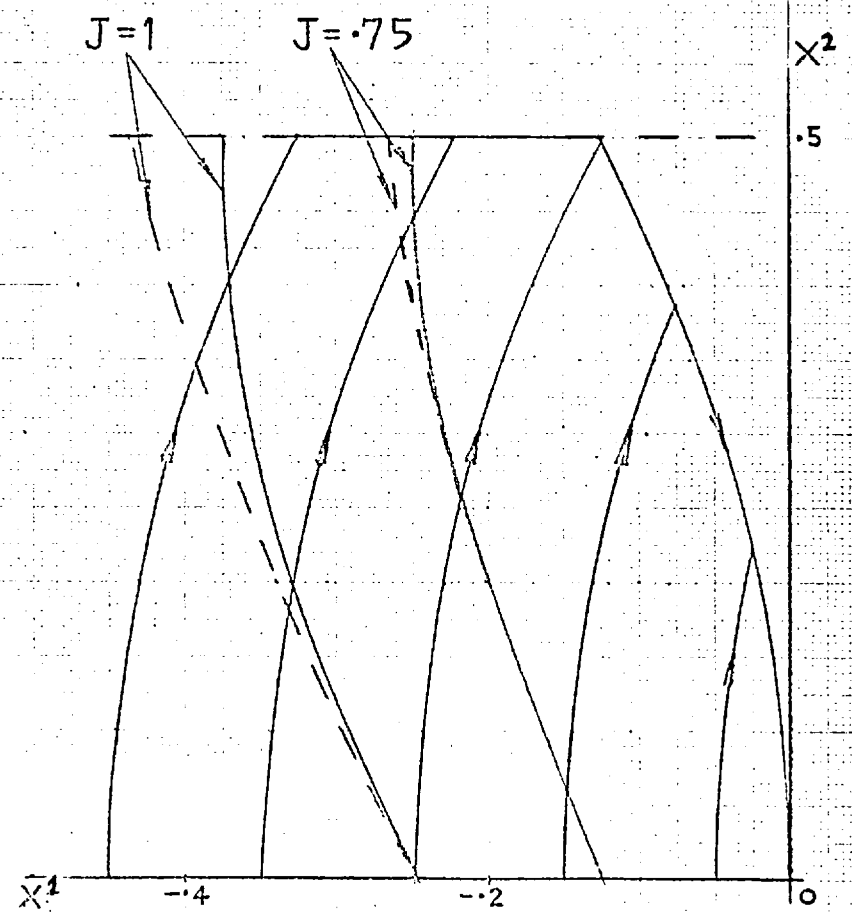


Fig. D.3

information required to determine the final arc. Similar use of the corner condition and the vanishing of $H + 1$ solves the initial arc too.

The value of $J(x)$ for points before the boundary is :

$$\text{Time on final arc} = .5$$

$$\text{Time on boundary} = -2(.125 + x_b^1),$$

where x_b^1 is point at which boundary is entered,

$$\text{Time to reach boundary} = .5 - x^2$$

$$\therefore \quad J(x) = .5 - 2x^1 + (x^2)^2 - x^2,$$

confirming that $J_{x^1} = -2.$

Fig. D.3 shows part of the field, together with isochrones, (those corresponding to the unconstrained problem in broken lines) exemplifying the situation discussed in section 4.3.5.

<u>Example 2</u> Linear Feedback.

With the same system as in Ex. 1, we use the cost function

$$\int_0^1 \frac{1}{2} (u)^2 \, dt$$

and the terminal set $T : (0, -1)$. Since the time t_f is explicitly given we treat this as a 3-dim. problem, adding $\dot{x}^0 = 1$, and we have

$$H = \frac{1}{2} (u)^2 + p_0 + p_1 x^2 + p_2 u = 0$$

$$\therefore \dot{p}_0 = \dot{p}_1 = 0 \qquad \dot{p}_2 = -p_1 \qquad u = -p_2$$

with $H_{uu} = 1$, ensuring a) that $H_u = 0$ gives a minimum, b) that the trajectories are always in a 3-dim. region, so that $p = J_x$.

This problem has an easy analytic solution: for an initial point $(0, c_1, c_2)$ we have

$$\begin{aligned} x^1(t) &= c_1 + c_2 t - (3c_1 + 2c_2 - 1)t^2 - (2c_1 + c_2 - 1)t^3 \\ x^2(t) &= c_2 - 2(3c_1 + 2c_2 - 1)t - 3(2c_1 + c_2 - 1)t^2 \\ u(t) &= -2(3c_1 + 2c_2 - 1) - 6(2c_1 + c_2 - 1)t \end{aligned} \qquad \text{D.13}$$

The isotim value is easily computed :

$$J(x(t)) = \frac{1}{2}\int_t^1 \left[2(3c_1 + 2c_2 - 1) + 6(2c_1 + c_2 - 1)t\right]^2 dt$$

giving J as a function of c, t, quadratic in c. For a given value of t, $x = x(c,t)$ is linear in c (cf. D.13) so that $J(x)$ is a quadratic function. Isotims at fixed times are shown in Fig.D.4; the ellipses are the intersections of $J(x) =$ const., $x^o =$ time = const. The surface $J(x) = 0$ containing the terminal point is a trajectory corresponding to $u = o$, for which $p_1 = p_2(t) = 0$, $c_1 = 1$ $c_2 = -1$. This is the value of u which minimises $\int u^2 dt$ regardless of the dynamic constraints, and in view of the interpretation of the multipliers p as the effort of maintaining the constraint, it is natural that p should be identically zero.

To find an approximate value $p(o)$ we might apply 7.4a, getting, as in Ex. 1.

$$p_1 (u(dt) - u(o)) \geqq 0,$$

but it gives no useful information. 7.4 b gives

$$-p_1 c_1 - p_2 (1 + c \quad) \leqq 0$$

which is more helpful. In addition,

$$H(o) = -\frac{1}{2} p_2^2 + p_o + p_1 c_2 = 0$$

reduces the search for initial values considerably, for we may set $p_o = 1$ if it is not zero.

To construct wavelets according to section 6.2 we form the system

$$x^{1'} = 2x^2/(u)^2 \qquad x^{2'} = 2/u \qquad x^{o'} = 2/(u)^2$$

The points reached from an initial point $c = (0, c_1, c_2)$ for all u, $-\infty \leq u \leq \infty$, $s = ds$ form a parabola. The wavelets issuing from the points of this parabola are again parabolae, the envelope of which is the boundary of the reachable set for $s = 2$ ds. (Fig.D.8)

If the envelope touches every parabola, then every initial value of u is a candidate for an optimal trajectory, but this can occur only if all pairs of neighbouring trajectories intersect.

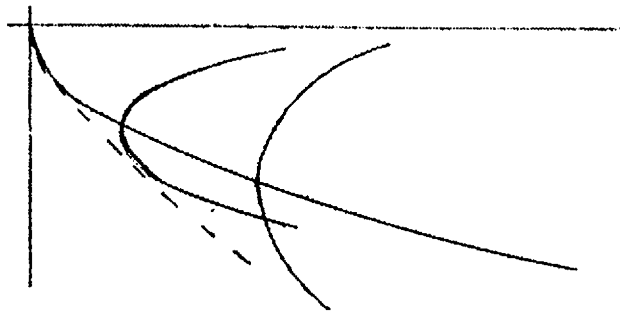


Fig. D.8

The parabola with origin at (c_1, c_2) has the form

$$x^1(ds) = c_1 + 2c_2\, ds / (u)^2$$
$$x^2(ds) = c_2 + 2ds/u$$

Thus two parabolae are

$$x^1 - c_1 = c_2\,(x^2 - c_2)^2 / 2\,ds$$
$$x^1 - b_1 = b_2\,(x^2 - b_2)^2 / 2\,ds$$

and their intersection has

$$c_1 - b_1 = (\, b_2\,(x^2 - b_2)^2 - c_2\,(x^2 - c_2)^2\,) / 2ds.$$

If such an intersection is possible ,

$$(c_2^2 - b_2^2)^2 \geq 2(b_2 - c_2)((b_2^3 - c_2^3)/(2ds) + b_1 - c_1)ds \qquad \text{D.14}$$

b,c are themselves points on the parabola whose origin is at the initial point for the problem, say a_1, a_2, and correspond to two values of control, say u, v,

$$\therefore\; b_1 = a_1 + 2a_2 dt / u^2$$
$$b_2 = a_2 + 2\,dt / u$$

and similarly for c, using v. Substituting into D.14 we obtain, after simplifying,

$$a_2^2 + 2a_2\,(dt - ds)\left(\frac{1}{u} + \frac{1}{v}\right) + \frac{4\,dt^2}{uv} = 0 \qquad \text{D.15}$$

If a_2 is not very small, a_2^2 dominates. If it is small (dt - ds) can be made as small as desired, and D.15 is positive if u,v are of the same sign.

Thus closely neighbouring parabolae will intersect, and the envelope will touch them all. No further restriction can be found for possible initial values of u or p.

Imposing a state-constraint $x^1 \leq c$ on this system, we have a second order boundary,

$$x^1 - c = 0; \quad x^2 = 0; \quad u \leq 0.$$

The transformed dynamic system will be identical with the original one, so that p_1, p_2 will not be defined on the boundary, but p_o is continuous and constant. The corner condition reduces to

$$p_2 u = 0,$$

implying continuity of u, and $H = 0$ gives $p_o = 0$. The inequality 7.4b gives $cp_1 \geq 0$ at exit from the boundary, and the problem now presents no difficulty to numerical solution.

<u>Example 3</u>. The Brachistochrone.

An interesting variation of the famous classical problem is furnished by the imposition of a state constraint.

$$\left.\begin{aligned} \dot{x}^1 &= V(x^2) \cos u \\ \dot{x}^2 &= -V(x^2) \sin u \end{aligned}\right\} \text{ where } V = (2g(x^2(0)-x^2)+V^2(0))^{\frac{1}{2}}$$

$$\min \int_0^{t_f} dt \equiv \min t_f$$

$$S: \; x = (0,6) \qquad V(0) = 1 \qquad T: \; x^1 = 6$$

$$X: \; x^2 + .5x^1 - 5 \geq 0$$

Dealing first with the unconstrained problem,

$$H = 1 + p_1 V \cos u - p_2 V \sin u ,$$

$$\text{and } H_{uu} = -p_1 V \cos u + p_2 V \sin u \neq 0$$

indicates that the space of optimal trajectories is 2- dim.

$$\dot{p}_1 = 0 \qquad \dot{p}_2 = -g(p_2 \sin u - p_1 \cos u)/V$$

$$\tan u = -p_2/p_1.$$

Setting $H_{uu} > 0$ gives $p_2 > 0$, and $H = 0$ gives $\cos u = kV(x^2)$, a convenient semi-feedback form.

The boundary value $p_2(t_f) = 0$ indicates that $u(t_f) = 0$, and therefore $k = 1 / V(t_f)$. From given points $(6, x^2)$ the dynamic equations can readily be integrated backwards to produce a field, part of which, together with isotims, is shown in Fig. D.5. It is interesting to notice that in this case transversality is equivalent to orthogonality. This occurs, in the classical problem, when

$$L(x,\dot{x}) = G(x)\left[\sum_i^1 (\dot{x}^i)^2\right]^{\frac{1}{2}}$$

(Rund 17 p.27) implying a locally Euclidean metric, which is the same as saying that the infinitesimal wavelets are spherical. For all possible u the dynamic equations of this system are, for fixed x, the parametric equations of a circle.

When considering the constrained problem we shall again find that the final sub - arc, from the boundary to the terminal set, can be isolated from the remainder of the trajectory. This will always occur when an n-dim. system has a q'th order boundary, and $n-q = 1$, for the boundary itself provides q conditions, the corner condition is the extra one required, and $H = 0$ provides for the unknown interval. At T there are always n conditions, and the $2n$ differential equations can be completely solved.

The 1 - dim. region is

$$C(x) = x^2 + .5x^1 - 5 = 0$$

on which the control is given by

$$C^{(1)}(x,u) = - V \sin u + .5 V \cos u = 0$$

Fig D.6

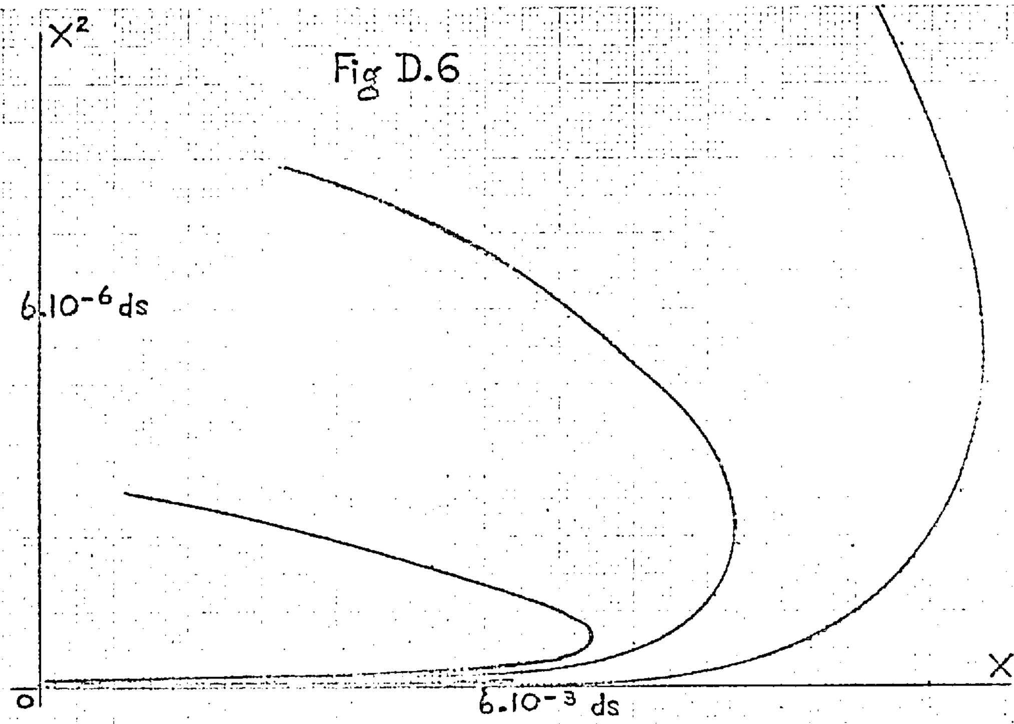
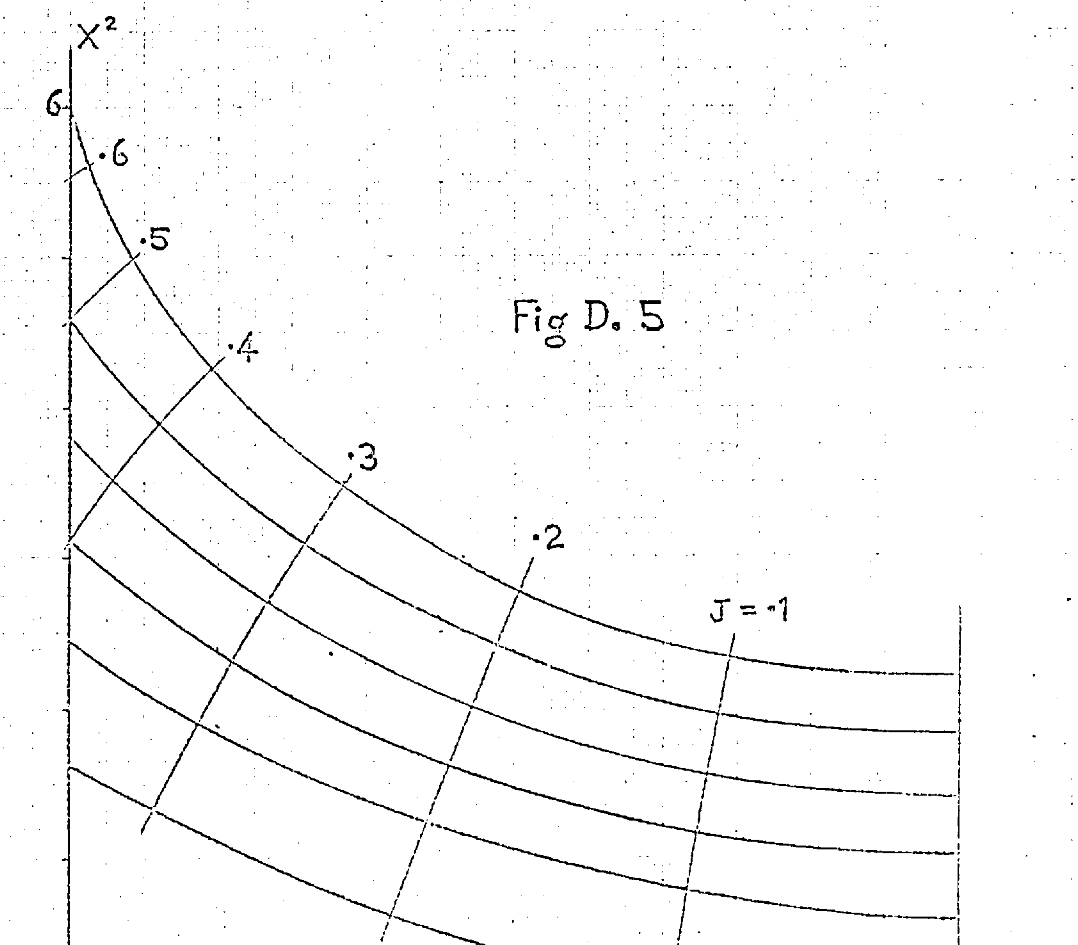


Fig D. 5



$$\therefore \tan u = .5.$$

Choosing a transformation $\dot{y} = A(x).\dot{x}$ with $\dot{y}^2 = 0$, we have

$$a_{21} = C_{x^1} = .5 \qquad a_{22} = C_{x^2} = 1$$

$\dot{y}^1$ is normal to $\dot{y}^2$,

$$\therefore \quad .5a_{11} V^2 \cos^2 u + a_{12} V^2 \sin^2 u = 0$$

$$\therefore \quad .4a_{11} + .2a_{12} = 0$$

det $A = 1$, $\therefore \quad a_{11} - .5a_{12} = 1$

$$A = \begin{bmatrix} .5 & -1 \\ .5 & 1 \end{bmatrix} \qquad A^{-1} = \begin{bmatrix} 1 & 1 \\ -.5 & .5 \end{bmatrix}$$

$$\begin{aligned} H &= 1 + p.f(x,u) \\ &= 1 + qA.f(x,u) \\ &= 1 + V(x^2)\, q_1\, (.5 \cos u + \sin u) + V(x^2) q_2 (.5 \cos u - \sin u) \end{aligned}$$

(on the boundary the coefficient of q_2 vanishes).

$$\dot{q}_1 = V_{x^2} \left[q_1 (.5 \cos u + \sin u) + q_2 (.5 \cos u - \sin u) \right] a^{-1}_{2i}$$

$$\therefore \dot{q}_1 = -\frac{g}{V} \left[q_1 (.5 \cos u + \sin u) + q_2 (.5 \cos u - \sin u) \right] (-.5)$$

$$\dot{q}_2 = -\frac{g}{V} \left[\quad '' \qquad\qquad '' \quad \right] (.5)$$

On the boundary these equations become

$$\dot{q}_1 = q_1 g / V(x^2)$$

$$\dot{q}_2 = - q_1 g / V(x^2)$$

though the second can be ignored.

The corner condition 4.25 gives

$$q_1 \left(\frac{2}{\sqrt{5}} - .5 \cos u - \sin u \right) - q_2 (.5 \cos u - \sin u) = 0, \qquad \text{D.16}$$

and $H_u = 0$ gives

$$q_1 (.5 \sin u - \cos u) - q_2 (-.5 \sin u - \cos u) = 0 \qquad \text{D.17}$$

Substituting for q_2 in D.16 gives

$$\sin u + 2\cos u - \sqrt{5} = 0$$

which has only one solution in $0 \leq u \leq \pi/2$, namely $u = \tan^{-1} .5$.
u is continuous, and, from D. 17, $q_2 = .6\, q_1$ at the point of exit from the boundary.

The same applies, of course, at the point of entry, and again, the initial arc can be solved in isolation, using the additional information $H = 0$, or $\sqrt{5} + 2Vq_1 = 0$.

For numerical solution the convexity of the isotims (Fig D.5) allows the application of 7.4b., which in this case is very useful, for if it is not satisfied the solutions of the equations oscillate wildly.

<u>Example 4</u> A Rocket problem.

Something of a coup is achieved by applying the reachable set technique to a problem posed by Kipiniak. The system is

$$\dot{x}^1 = \frac{-9.8}{(1+x^2)^2} + \frac{1}{1+\exp(-10t)}\left[10x^1 \exp(-10t) - \frac{2(x^1)^7}{(1+10x^2)^8} + u\right]$$

$$\dot{x}^2 = x^1$$

$$\min \int_o^{t_f} u^2\, dt$$

$$S: \quad x^1(o) = x^2(o) = 0$$

$$T: \quad x^2(t_f) = t_f^2 - t_f + .35 \qquad x^1(t_f) = 2t_f - 1$$

The transformed system $x' = f(x,u)/L(x,u)$ is, at $t = 0$,

$$x^{1'} = (.5u - 9.8)/(u)^2$$

$$x^{2'} = x^1/(u)^2$$

Allowing u to take all values $-\infty \leq u \leq +\infty$ the infinitesimal wavelet from $(0,0)$ remains on the x^1 axis (Fig D.6), $dx^1 = x^{1'}ds$ having a maximum of $.00638\, ds$ at $u = 39.2$. From a selection of points on this wavelet, infinitesimal wavelets can be constructed in the same way (Fig.D.6) and show that the set with its source at the extreme point $dx^1(0) = .00638\ ds$ includes all other sets. The boundary of the reachable set for 2 ds can only be attained from that point, suggesting that the initial value of control must be 39.2, regardless of terminal conditions. The optimal control is given by $u = -p_1/4$, giving an immediate value for $p_1(0)$, equal to -156.8.

Applying 7.4b we have

$$p_1(0)\left[\,5(-9.8 + .5\,u(ds)) - 5(-9.8 + .5\,u(0)\right] +$$
$$+\,p_2(0)\left[-9.8 + .5\,u(0) + 9.8 - .5u\,(ds)\right] \geq 0$$

$$\therefore\quad (u(ds) - u(0))(2.5p_1(0) - .5p_2(0)) \geq 0 \qquad \text{D. 18}$$

Now,
$$p_1' = \frac{1}{u^2}\left\{ \frac{p_1}{1+\exp(-10t)}\left[10\exp(-10t) - \frac{14\,(2x^1)6}{(1+10x^2)^8}\right] - p_2\right\}$$

$$\therefore\quad p_1'(0) = \frac{1}{u^2(0)}\left\{-5p_1 - p_2\right\}$$
$$= -4u'(o)$$

$$\therefore\quad u(ds) - u(o) = \frac{1}{4u^2(o)}(\,5p_1(o) - p_2(o)\,)$$

so that D.18 implies $\left|p_2(o)\right| \geq 784$. The sign of p_2 is not so easily determined from the equations, but physical considerations suggest that the initial boost of a rocket should be greater than the subsequent thrust, so that u' should be negative, implying $p_2(o)$ negative.

A set of trial trajectories was computed, using $p_1(o) = -156.8$ and $p_2(o) = -800$, decreasing in steps of 50. (Fig. D.7) Such was the sensitivity to the initial value that only one of these would have been a feasible initial approximation, the others not meeting the terminal set at all. A second trial run produced a solution close to the optimum, only requiring 'trimming' by a convergence process to any desired accuracy. Without these techniques the search for initial values would be very tedious.

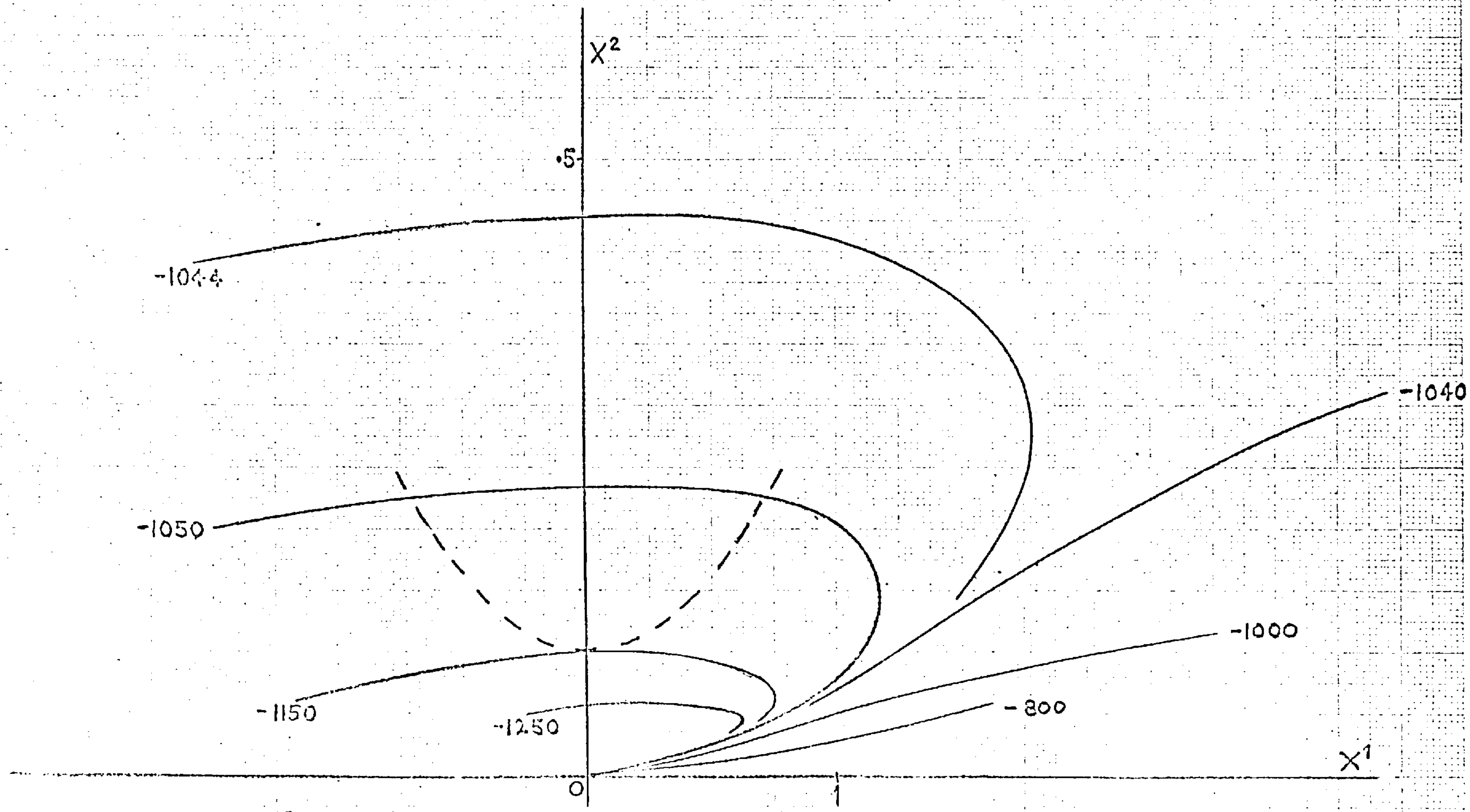


Fig. D.7

REFERENCES

1 Pontryagin, L., Poltyanskii, V., Gamkrelidze, R., Mishchenko, E. The Mathematical Theory of Optimal Processes. Interscience 1962.

2 Chang, S. Optimal Control in Bounded Phase Space. Automatica 1 No.1 1963.

3 Bridgland, T. On the Existence of Optimal Feedback Controls. SIAM Journal on Control 1 No.3 1963.

4 Zadeh, L. Optimality and Non-scalar-valued performance Criteria. IEEE Trans. Aut. Control AC-8 Jan.1963.

5 Bliss, G. Lectures on the Calculus of Variations. University of Chicago 1946.

6 Halkin, H. The Principle of Optimal Evolution. Non-linear differential equations and non-linear mechanics. LaSalle and Lefschetz (eds.) Academic Press 1963.

7 Halkin, H. On the Necessary Condition for Optimal Control of Non-linear Systems. Journal d'Analyse Mathematique 12 1964.

8 Roxin, E. A Geometric Interpretation of Pontryagin's Maximum Principle. Non-lin.diff.eqns. (see 6).

9 Kalman, R. Contributions to the Theory of Optimal Control. Boletin de la Soc. Matematica Mexicana p.102 1960.

10 Kalman, R., Ho, Y. Narendra, K. Controllability of linear dynamical systems. Contributions to differential equations 1 No.2 p.189 Wiley 1963.

11 Hermann, R. The Accessibility Problem in Control Theory. Non-lin. diff.equations. (see 6).

12 Lefshetz, S. Differential Equations---Geometric Theory. Interscience 1962.

13 Russell, B. The Principles of Mathematics. Allen and Unwin 1956.

14 Veblen, O., Whitehead, A. Foundations of Differential Geometry. Cambridge University Tracts in Math. No.29 1932.

15 Berge, C. Topological Spaces. Oliver and Boyd 1963.

16 Coddington, E. and Levinson, N. Theory of Ordinary Differential Equations. McGraw-Hill 1955.

17 Rund, H. The Differential Geometry of Finsler Spaces. Springer. Berlin. 1959.

18 Caratheodory, C. Variationsrechnung und Partielle Differentialgleichungen erster Ordnung. Berlin 1935.

19 Bellman, R. and Dreyfus, S. Applied Dynamic Programming. Oxford 1962.

20 Johnson, c. and Gibson, J. Singular Solutions in Problems of Optimal Control. IEEE Trans. Aut. Control AC-8 Jan.1963.

21 Kelley, H. A Transformation Approach to Singular Subarcs in Optimal Trajectory and Control Problems. SIAM Journal on Control 2 No.2. 1964.

22 Hermes, H. Controllability and the Singular Problem . SIAM Journal on Control 2 No.2. 1964.

23 Snow, D. Singular Optimal Controls for a Class of Minimum-effort Problems. SIAM Journal on Control 2 No.2 1964.

24 Wonham, W. and Johnson, C. Optimal Bang-bang Control with Quadratic Performance Index. JACC Proc 1963.

25 Berkovitz, L. Variational Methods in Problems of Control and Programming. J.Math. Analysis and Applications 3 No.1 1961.

26. Troitskii, V. On Variational Problems of Optimization of Control Processes. Applied Math. and Mech.(P.M.M) 26 1962.

27 Gelfand, I. and Fomin, S. Calculus of Variations. Prentice-Hall, New Jersey 1963.

28 Lanczos, C. The Variational Principles of Mechanics. Toronto 1949.

29 Caratheodory, C. Gesammelte Math. Schriften. Beck (ed).

30 Osborn, H. On the Foundations of Dynamic Programming. J. Math. and Mech. 8 1959.

31 Bolza, O. Lectures on the Calculus of Variations. Dover 1961.

32 Forsyth, A. Calculus of Variations. Dover 1960.

33 Hadamard, J. Lecons sur le Calcul des Variations. Hermann, Paris 1910.

34 McShane, E. On Multipliers for Lagrange Problems. Amer. J. of Maths. 61 1936.

35 Bliss, G. and Underhill, The Minimum of a Definite Integral for Unilateral Variations in Space. Trans. Amer. Math. Soc. 15 1914.

36 Berkovitz, L. On Control Problems with Bounded State Variables. J. Math. Anal. and Appl. 5 No. 3 1962.

37 Berkovitz, L. and Dreyfus, S. The Equivalence of some necessary Conditions for Optimal Control in Problems with Bounded State Variables. Rand Corp. Memo. RM-3871-PR 1963.

38 Dreyfus, S. Variational Problems with Inequality Constraints. J. Math. Anal. and Appl. 4 No. 2 1962.

39 Bryson, A. and Denham, W. The Solution of Optimal Programming Problems with Inequality Constraints. Raytheon Co. report BR2121 1962.

40 Courant, R. and Hilbert, D. Methods of Mathematical Physics. Interscience 1962.

41 Yashilev, A. Necessary and Sufficient Conditions for the Optimality of Controlled Systems. Automation and Remote Control 25 No.10 1964.

42 LaSalle, J. The Time-optimal Control Problem. Contributions to the Theory of Non-linear Oscillations vol. 5.Princeton Univ. Press. 1960.

43 Lee, E. A Sufficient Condition in the Theory of Optimal Control. SIAM J. Control 1 No. 3 1963.

44 Neustadt, L. The Existence of Optimal Controls in the absence of Convexity Conditions. J. Math. Anal. and Appl.

45 Kipiniak, W. Dynamic Optimization and Control. M.I.T.and John Wiley 1961.

46 Pearson, J. Reciprocity and Duality in Control Programming Problems. J. Math. Anal. and Appl. 10 No. 2 1965.

47 Halkin, H. Method of Convex Ascent. Computing Methods in Optimization Problems. Balakrishnan and Neustadt (eds.) Academic Press 1964.

48 Levine, M. A Steepest Descent Method for Synthesizing Optimal Control Programmes. Proc. Symposium on Optimal Control. Imperial College 1964.

49 Thau, F.E. Optimal Time-control of Non-normal Linear Systems. International J. of Control 1 No. 4 1965.

50 Mitter, S. Ph.D. Thesis London Univ. 1965.

51 Merriam, C. Optimization Theory and the Design of Feedback Control Systems. McGraw-Hill 1963.

52 Saaty and Bram. Non-linear Mathematics. McGraw-Hill 1964.

53 Levine, M. Ph.D. Thesis London Univ. 1965.

54 Breakwell, J., Speyer, J., Bryson, A. Optimization and Control of Non-linear Systems using the Second Variation. SIAM J. on Control 1 No. 2 1963.

55 Kelley, H. Guidance Theory and Extremal Fields.
I.R.E. Trans. Aut. Control. 1962.

56 Jazwinski. Optimal Trajectories and Linear Control of Non-linear Systems. A.I.A.A. Journal 2 No. 8 1964.

57 Dreyfus, S. and Elliott, J. An Optimal Linear Feedback Guidance Scheme. Rand Corp. Memo. RM 3604 Pr 1963.

58 Kelley, H. Second Variation Test for Singular Extremals.
AIAA Journal 2 1964.

59 Dubovitskii, A. and Milyutin, A. Problems on the Extremum in the Presence of Limitations. Dokl. Akad. Nauk. 1963. National Lending Library translation SLA 63-18707.

60 Neustadt, L. Synthesis of Time-optimal Control Systems.
J. Math. Anal. and Appl. 1 1960.

61 Eaton, J. An Iterative Solution to Time-optimal Control.
J. Math. Anal. and Appl. 5 1962.

62 Fadden and Gilbert. Computational Aspects of the Time-optimal Control Problem. Computing Methods (see 47).

63 Pearson, J. Approximation Methods in Optimal Control.
J. Electronics and Control 13 1962.

64 Davis, N. Approximately Optimal Control of a Non-linear System.
M.Sc. Thesis London 1964.

65 Aoki, M. On a Successive Approximation Technique in solving some Control System Optimization Problems. J.Basic Eng. No.2 1963.

66 Bryson, A. and Denham, W. A Steepest Ascent Method for Solving Optimum Programming Problems. J. Appl. Mech. 29 No. 2 1962.

67 Kelley, H. Gradient Methods. Optimization Techniques (Leitmann, ed.). Academic Press 1963.

68 Synge, J. Classical Dynamics. Handbuch der Physik (Flugge, ed.). vol. 111/1. Springer 1960.

69 McKinsey, J., Sugar, A., Suppes, P. Axiomatic Foundations of Classical Particle Mechanics. J. Rational Mechanics and Anal. 2 1953.

70 Hamel, G. Die Axiome der Mechanik. Handbuch der Physik. vol V Springer 1927.

71 Truesdell, C.,, Toupin, R. The Classical Field Theories . Handbuch der Physik vol 111/1 Springer 1960.

72 Landsberg, P. Is Thermodynamics an Axiomatic Discipline? Bull. of the Institute of Physics and the Phys. Soc. June 1964.

73 Kilmister, C. The Principles of Mechanics. J. of Maths. and Mech. 12 1963.

74 Hadamard, J. The Psychology of Invention in the Mathematical Field. Dover 1954.

75 Caratheodory, C. Untersuchungen über die Grundlagen der Thermodynamik. Math. Ann. 67 1909.

76 Zade , L. and Descer, C. Linear System Theory. McGraw-Hill 1963.

77 Arbib, M. Automata Theory and Control Theory---a Rapprochement. Symposium on Optimal Control, Imperial College. London 1964,

78 Hestenes, M. Variational Theory and Optimal Control Theory. Computing Methods... (see 47).

79 Hestenes, M. A General Problem in the Calculus of Variations with Applications to Paths of Least Time. Rand Corp. Memo.RM-100 1949.

80 Pearson, J. Ph.D. Thesis. London Univ. 1963.

81 Valentine, F. The Problem of Lagrange with Differential Inequalities as Added Side Conditions. Contributions to the Calculus of Variations 1933-1937. University of Chicago 1937.

82 Guinn, T. On First - order Necessary Conditions for Variational and Control Problems. Dissertation, Univ. of California L.A. 1964

83 Liusternik & Sobolev, Elements of Functional Analysis. Unger 1961

84 Balakrishnan, A. An Operator - theoretic Formulation of a Class of Control Problems. SIAM J. on Control 1 No. 2 1963

85 Koestler, A. The Act of Creation. Hutchinson 1964

86 Whittaker, E. Analytical Dynamics Cambridge 1904

87 Edelen, D. The Structure of Field Space Univ. of California 1962

88 Kilmister, C. Hamiltonian Dynamics Longmans 1964

89 Ericksen, J. Tensor Fields Handbuch der Physik 111/i 1960

90 Markus, L. & Lee, E. On the Existence of Optimal Controls JACC 1961

91 Roxin, E. The Existence of Optimal Controls Mich. J. Math. 9 1962

92 Cicala, P. An Engineering Approach to the Calculus of Variations Levroto & Bella Torino 1957

93 Faulkner, F. Direct Methods Optimization Techniques (see 67)

94 Kreindler, E. Contributions to the Theory of Time - optimal Control J. Franklin Inst. 275 No.4 1963

95 Fuller, A. Study of an Optimum Non - linear Control System J. Electronics and Control 15 No.1 1963

96 Fuller, A. Further Study of an Optimum Non - linear Control System J. Electronics and Control 17 No.3 1964