

OPTIMALITY CONDITIONS AND SENSITIVITY RELATIONS IN DYNAMIC OPTIMIZATION

by

Hannah Rachel Babad

Thesis submitted for the degree of
Doctor of Philosophy
of the University of London

January 1991

Department of Electrical Engineering
IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE
University of London

ABSTRACT

An optimal multiprocess problem is a dynamic optimization problem in which there is a collection of control systems coupled through constraints in the end-points of the constituent trajectories and through the cost function. Optimality conditions for multiprocess problems posed over a finite interval have already been derived. However, multiprocess problems in which one of the component processes operates over an infinite time interval arise, for example, in mathematical economics problems. The question then is whether general necessary conditions can be derived for such situations. In this thesis new necessary conditions in the form of a maximum principle are obtained for multistage problems defined over the infinite time horizon. The theory is then applied to a specific economics example in which an explicit dependence on the intermediate time (taken as a choice variable) is incorporated into the integrands of the cost functionals. The hypotheses under which these necessary conditions apply are mild and easily verified.

The thesis also contains an investigation of the relationship between necessary conditions (in the form of the Pontryagin maximum principle) and dynamic programming. The main results obtained are that, provided some kind of normality hypothesis (which in certain circumstances is verifiable) is satisfied, there is a relationship between the costate $p(\cdot)$ of the maximum principle, the Hamiltonian function, H , and some verification function, W ,

$$(H(t, x^*(t), u^*(t), p(t)), -p(t)) \in \partial^{\vee} W(t, x^*(t))$$

along a trajectory (x^*, u^*) which is both necessary and sufficient for that trajectory to be optimal. Results of this nature have been termed 'sensitivity relations' in the literature.

The 'generalized gradient' $\partial^{\vee} W$, in terms of which this relationship is expressed, is the superdifferential. The superdifferential provides a more precise sensitivity relation than the Clarke generalized gradient of the earlier literature on this topic. While sensitivity relations involving superdifferentials have appeared elsewhere, only free end-point problems have been treated. The main contribution of the thesis on sensitivity relationships is that we establish validity of the superdifferential sensitivity relation in the presence of general end-point constraints.

ACKNOWLEDGEMENTS

I am deeply grateful to my supervisor, Dr. R. B. Vinter, for all the friendly help, constructive advice and careful guidance I received during the course of my research project.

Many and sincere thanks.

I would also like to thank Dr. J. D. L. Rowland for his helpful comments and suggestions.

My parents have supported and encouraged me throughout all my studies and especially during my research.

IT IS TO MY PARENTS THAT I DEDICATE THIS THESIS, WITH LOVE.

CONTENTS

Chapter 1. INTRODUCTION AND PREVIEW	6
1.1 Introduction	6
1.2 Optimal Multiprocesses	7
1.3 Infinite Horizon Optimal Control Problems	8
1.4 Infinite Horizon Multistage Problems	10
1.5 Necessary Conditions and Applications of the Theory	12
1.6 Sufficient Conditions, Dynamic Programming and the Maximum Principle	13
1.7 Verification Theory	15
1.8 Minimum Cost Sensitivity Relations	17
1.9 Outline of Thesis	18
Chapter 2. BACKGROUND THEORY	19
2.1 Introduction	19
2.2 Generalized Differentiation	20
2.3 The Generalized Gradient and its Relation to Strict Derivatives	22
2.4 The Generalized Gradient and its Relation to Superdifferentials	24
2.5 Normal and Tangent Cones	25
2.6 The Maximum Principle	27
2.7 Multiprocess Theory	29
Chapter 3. AN INFINITE HORIZON, MULTISTAGE OPTIMIZATION PROBLEM	34
3.1 Introduction	34
3.2 Statement of the Problem	35
3.3 An Infinite Horizon Multiprocess Maximum Principle	37
3.4 Proof of Theorem 3.3.1	38
3.5 An Application of the Infinite Horizon Theory	49

Chapter 4. A NECESSARY AND SUFFICIENT SENSITIVITY RELATION FOR CONSTRAINED END-POINT PROBLEMS	58
4.1 Introduction	58
4.2 Preliminary Results	63
4.3 The ‘Strengthened Normality’ Condition	64
4.4 The Main Result	66
4.5 Properties of the Cost and Value Functions	68
4.6 Proof of Theorem 4.4.1 in Fixed End-Point Case	75
4.7 Proof of Theorem 4.4.1 for an End-Point Constraint Set	85
 Chapter 5. CONCLUSION	 89
 References	 92

INTRODUCTION AND PREVIEW

1.1 Introduction

Three central themes are present in the study of all optimization problems:

- 1) existence - is there an admissible arc which solves the problem?
- 2) necessary conditions - what conditions are there which help identify solutions?
- 3) sufficient conditions - what conditions are there to indicate that a given admissible arc (which may have been identified with the help of the necessary conditions) is in fact a solution of the problem?

In this thesis we concentrate on aspects of the latter two themes as they relate to dynamic optimization.

Dynamic optimization is the study of optimization problems where the underlying choice variables are functions of time and where explicit reference is made to the derivatives of these time functions. Optimal control theory which is a branch of dynamic optimization had its origins in the 1950's when engineers, motivated in particular by aerospace problems, became interested in ways of controlling a system governed by a set of differential equations so that a given performance index could be minimized. The optimal control problem can be formulated as follows:

The name 'control' is given to a function $u(\cdot)$ belonging to some class of functions (say that of measurable functions) mapping $[a, b]$ to a specified subset U of $\mathbb{R}^m$. Each control u is associated with an arc x (called the state) by means of the differential equation and initial condition

$$\dot{x}(t) = \phi(x(t), u(t)), \quad x(a) = x_0, \quad (1.1.1)$$

where ϕ is given and where the point x_0 is free to be chosen within a given set C_0 . The problem is then to choose u and x_0 (and hence x) so as to minimize a given functional of the type

$$f(x(b)) + \int_a^b F(x(t), u(t)) dt \quad (1.1.2)$$

subject to $x(b)$ lying in a given set C_1 . Note that U , F and ϕ may depend on t . Also a state constraint $g(t, x(t)) \leq 0$ may be admitted. The functional (1.1.2) is known variously in the literature as the performance index, performance criterion, objective function or cost function.

A large number of physical systems can be modelled by differential equations in

which some parameter values (which we interpret as ‘controls’) can be chosen within a certain range; this, combined with the fact that techniques have been developed which are practical for both computation and implementation of optimal controls, means that the theory can be applied to a wide range of fields, including, problems in economic theory.

From the strictly mathematical viewpoint, the optimal control problem subsumes the classical calculus of variations and conversely, many optimal control problems can be reformulated as standard variational problems with inequality and equality constraints, although this does not apply in all cases. However, the smoothness requirements of classical variational theory are restrictive for optimal control theory and much recent research has been aimed at weakening these requirements.

1.2 Optimal Multiprocesses

One of the best known results in optimal control theory is the set of necessary conditions for optimality derived by Pontryagin et al. [P1], known as the (Pontryagin) maximum principle. Research has been directed at weakening the hypotheses under which it is known to be true and formulating solutions to non-standard optimal control problems (e.g. where the dynamics are described by functional or partial differential equations). Recently the work has been extended to cover problems where the data are no longer assumed continuously differentiable in the state variable; in these cases optimality conditions are expressed in terms of ‘generalized gradients’, which are defined in the next chapter. A proof of the non-smooth version of the maximum principle is given in Clarke [C3]. A further extension of this work, by Clarke and Vinter [C8], is the development of a maximum principle for ‘multiprocesses’. Both the non-smooth maximum principle and the multiprocess maximum principle are stated in Chapter 2.

Multiprocess theory originated in a bid to solve a variety of non-standard problems in dynamic optimization. The common feature of these problems is that they involve a collection of dynamical equations with control term (i.e. control systems) which are coupled through constraints in the end-points of the constituent state trajectories and through a function of these boundary values in the cost. Examples of these problems arise in flight mechanics, optimal investment, routing problems over variable domains and plasma control, but until now ad-hoc techniques for each individual case were required for studying the necessary conditions for optimality.

As an example, we may consider ‘impulse control’ problems, which arise in

optimal investment and inventory control. Here, the standard optimal control problem is modified to permit the state trajectories to be discontinuous at a finite number of times and the jump times (choice variables) together with the end-states of the jumps appear in the cost and are constrained to lie in some closed set. There is also a constraint on the number of jumps. Necessary conditions of optimality are available but are limited to a narrow class of cost functions and constraints as far as the jumps are concerned, and apply only when the constraint on the number of jumps is inactive. Alternatively, we can view impulse control problems as an instance of an optimal multiprocess problem. The restrictions of trajectories, in the original problem, to intervals between jumps are interpreted as component trajectories in the optimal multiprocess problem.

Even when the theory is specialised to treatment of the standard optimal control problem, significant advances on known results ensue. For free time optimal control problems, the Pontryagin maximum principle provides a boundary condition on the maximized Hamiltonian, evaluated along an optimal trajectory. This has previously been proved only under the assumption that the data are continuous in the time variable. Multiprocess theory, by contrast, yields a version of the standard boundary condition when the data are merely measurable in the time variable, in which case it is replaced by an inclusion involving the ‘essential values’ of the maximized Hamiltonian (see Chapter 2).

The theory of optimal multiprocesses is a general theory in line with the tradition of research whose aim is the unification of optimality conditions for as wide a range of optimization problems as possible (in calculus of variations, control and mathematical programming). Nevertheless, it is targeted specifically at problems in dynamic optimization and thus the checking of hypotheses is an easier task for multiprocess theory than for the more general theories just referred to. Specifically, it turns out that generating optimality conditions for specific classes of problems is usually a simple matter of calculating normal cones (see Chapter 2).

In this thesis we investigate multiprocess problems (in particular two-stage problems) in which one of the component processes operates over an infinite time interval and we discover how multiprocess theory readily lends itself to this combined area of research.

1.3 Infinite Horizon Optimal Control Problems

The motivation for studying deterministic continuous time optimal control

problems defined over unbounded time intervals stems primarily from the economic and biological sciences, where it is natural to formulate models of this type. The reason for doing so is that, often, it is not meaningful to place a bound on the time horizon, when we consider the evolution of the state of a given economy or species.

The earliest studies of problems of this type were probably those of F. Ramsey [R1] in 1928 and H. Hotelling [H3] in 1931. Ramsey, in his theory of saving, modelled a capital accumulation process by considering a dynamic optimization model defined on an infinite time horizon, while Hotelling modelled a problem of non-renewable resource extraction, likewise, over the infinite horizon.

From the 1960's, applications of the optimal control formalism to the modelling and analysis of dynamic economic systems have flourished. In 1965, Kurz [K2] and Stoleru [S2] simultaneously initiated this research by applying Pontryagin's maximum principle to finite horizon optimal economic growth problems. Then shortly afterwards, in 1969, Dorfman [D2] (among others) presented an economic interpretation of Pontryagin's theorem. A significant development was the introduction by von Weizsäcker [W2] in 1965 of the overtaking optimality concept which provided, for the first time, a means of dealing with a general class of infinite horizon optimal control problems with unbounded cost criterion. This was in response to the question of whether divergence of the cost integral indicates that no optimal policy exists; the answer was found to be that this is not necessarily the case.

Other developments that played an important role in the development of this theory were, the establishment of a continuous time 'turnpike theorem' for optimal control problems by Cass [C2], existence results for overtaking and weakly overtaking optimal solutions in discrete time by Brock [B3] and a systematic treatment of continuous time infinite horizon optimal control problems by Arrow and Kurz in 1970 [A2].

All of these historical developments, treated in the economics literature, occurred in parallel with the widening of the theory of dynamic optimization in automatic control and applied mathematics to include the infinite horizon. An example of this was the linear regulator problem where the linear structure of the dynamics coupled with a quadratic performance index permitted the development of a link between infinite horizon optimization and stability analysis. In this case the optimally controlled system was found to be asymptotically stable.

The maximum principle for infinite horizon optimal control is discussed by Pontryagin and his associates, but only briefly, and their requirement that the cost functional be convergent is quite restrictive. However, a major turning point occurred

when Halkin, in 1974 [H1], extended Pontryagin's maximum principle to an infinite horizon thus stimulating a great deal of research in this area of optimal control. For a systematic account of the recent developments in deterministic infinite horizon optimal control theory see Carlson and Haurie [C1] and the many references cited there.

1.4 Infinite Horizon Multistage Problems

Further motivation for considering multistage problems (defined over the infinite horizon), as opposed to single-stage problems, is that dynamic optimization problems which arise in economic theory often do so in the form of two-stage optimal control problems, i.e. the performance index takes the following general form

$$J = \int_{t_0}^{t_1} L_1(t) dt + \int_{t_1}^{t_f} L_2(t) dt \quad (1.4.1)$$

where t_1 will be called the switching or intermediate time and the final time t_f can be either finite or infinite. The significance of this class of problems is that the corresponding optimality conditions can be obtained by a relatively simple generalization of results for regular optimal control problems.

Infinite horizon multistage problems arise in at least two areas of economics research: the adjustment cost literature and the exhaustible resource literature. The former area of research is concerned with studying the investment decision process of a firm where, to account for delays in adjustment of capital stock, costs are built into the optimization process. An example of this is where the acquisition of new capital goods is subject to an exogenous delivery lag (see e.g. Maccini [M1], Nickel [N1], Rossana [R2]). In these applications, the intermediate time is the capital goods delivery lag and this may appear in the cost functions if the purchase price of capital depends on the delivery lag. The latter area of research is concerned with natural resource management (see e.g. Hoel [H2], Dasgupta et al. [D1]) where the intermediate time is the time at which alternative energy resources or new (or innovations of existing) technologies become available.

However, the assumption in these studies that the intermediate time is fixed and exogenous (i.e. it is taken as a given variable in the decision making process) restricts the application of the model. We see this in both the first example, where control can be exercised over delivery lags, depending on the choice of supplier by the firm, and in the second example, where control can be exercised over the time at which alternative resources are used.

Necessary conditions to allow for an endogenous switch-point, (i.e. it is a choice variable in the decision process) thereby extending the applicability of the model, have been obtained through classical methods for the following two cases, namely, where the integrands of the functionals do not depend on the intermediate time (see Tomiyama [T1]) and where an explicit dependence upon this switch-point is incorporated into the integrands (see Tomiyama and Rossana [T2]). These authors establish necessary conditions only under severe, and somewhat restrictive, hypotheses - hypotheses which we will show can be relaxed.

The question we now set out to answer in this thesis is whether general necessary conditions, in the spirit of Halkin's classic work, can be derived for multistage, infinite horizon optimal control problems via the use of multiprocess theory as the latter appears to be ideally suited to applications of this kind. It turns out that the answer is in the affirmative.

We consider a problem of the form

$$\text{minimize } \int_{t_0}^{t_1} L_1(t, x(t), u(t)) dt + \int_{t_1}^{\infty} L_2(t, x(t), u(t)) dt \quad (1.4.2)$$

and aim to derive a set of necessary conditions for optimality which is valid over the infinite horizon and which can be extended to include problems of the more general form

$$\text{minimize } \sum_{i=1}^{i=n} \int_{t_{i-1}}^{t_i} L_i(t, x_i(t), u_i(t)) dt + \int_{t_n}^{\infty} L_{n+1}(t, x_{n+1}(t), u_{n+1}(t)) dt. \quad (1.4.3)$$

However, difficulties can arise in a number of areas. Firstly, we note that the notion of 'optimality' for the infinite horizon case must be carefully defined. This is because, although the definition of optimality is unambiguous over a finite horizon, several non-equivalent definitions of optimality are available over the infinite horizon (see e.g. Carlson and Haurie [C1] and Stern [S1]). The logical implications among the different definitions are also discussed in these two references. Since, in proving the necessary conditions, Halkin uses the weakest of the various definitions in the literature, we may conclude that the infinite horizon maximum principle is valid for any of the other definitions of optimality which have been employed.

A second problem area concerns the transversality conditions. Halkin [H1] notes that there is a tendency to assume that results for the finite horizon case can be carried over to the infinite horizon case by replacing evaluations of quantities at the terminal time with evaluations of the limit of the same quantities as the time tends to infinity.

However, Halkin then shows the limitation of this approach by providing a well-known counterexample where the transversality conditions for the infinite horizon case cannot be obtained as the limit of the transversality conditions in the finite horizon case.

Corresponding to a dynamic economizing problem of allocation over time, where the state variable has the interpretation of an economic variable, there is a corresponding dual problem of valuation over time where the costate variable (see Chapter 2) has the interpretation of a price - a shadow price - or marginal cost. Thus more concretely, in economics terminology, the transversality condition in finite horizon optimal control without a constraint on the final state is that the final value of the shadow price vector (costate) is zero, i.e. one more unit of any good at final time gives no additional value to the criterion. However, in the infinite horizon, one more unit of a good, at any time, changes the whole future and the zero value of the state becomes a limit property which is not necessarily verified. In the literature, studies of the infinite horizon transversality property have been made but these only give results in special cases, some of which we shall mention in a later chapter.

1.5 Necessary Conditions and Applications of the Theory

Our contribution to the theory of multiprocess problems is that we derive a set of necessary conditions of optimality applicable for an infinite horizon. This is achieved by:

- a) showing that an optimal solution for the given infinite horizon problem is still optimal when the data of that problem are restricted to a fixed time interval with a terminal condition on the state, and
- b) using this property to apply the finite horizon multiprocess maximum principle over strictly increasing time intervals in order to construct a function which will be a suitable candidate for the costate function.

We then apply the theory to two-stage problems (discussed in Tomiyama and Rossana [T2]) of the form

$$\text{minimize } \int_{t_0}^{t_1} L_1(t, x_1(t), u_1(t), t_1) dt + \int_{t_1}^{\infty} L_2(t, x_2(t), u_2(t), t_1) dt. \quad (1.5.1)$$

This is, in effect, the task of determining an optimal control strategy for multistage problems where the intermediate times are included in the choice variables. Standard optimality conditions of control theory, which accommodate changes in the dynamics only at fixed times, cannot be applied directly to such problems.

These difficulties can sometimes be overcome by the following techniques where,

to begin with, we solve a family of problems involving fixed times and then minimize over the times to determine the optimum values (two-phase optimization) or, alternatively, we reduce the free times to fixed times by transformation of the time variable. Thus we see that, in general, two-phase optimization consists of ‘freezing’ certain choice parameters thereby reducing the problem to one which can be solved by application of the classical maximum principle and then minimizing a value function (the minimum cost as a function of the ‘frozen’ parameter values) to determine the optimum parameter values.

However, the drawbacks to these techniques are that two-phase optimization requires tractable formulae describing the value function relative to the intermediate times (which cannot be done for complicated problems) and that time transformation techniques are typically applicable only when certain regularity conditions are imposed on the data. This rules out significant applications.

Optimal multiprocess theory overcomes these drawbacks in that, firstly, no explicit reference is made to value functions and thus implementation of optimization procedures will not depend on availability of such formulae and, secondly, in that it treats dynamic optimization problems where the data are merely measurable in the time variable. This is an important feature of the theory.

Although Tomiyama and Rossana establish necessary conditions for the two-stage problem described by (1.5.1) they do so under the assumptions that the data and the cost functions are continuously differentiable in the x , t and t_1 variables and also under a smoothness assumption on a certain ‘value function’ introduced to decouple the stages. This latter condition is difficult to verify in general. On the other hand, by viewing trajectory segments between intermediate times as component trajectories, we arrive at an optimal multiprocess problem for which the new theory supplies optimality conditions under mild, verifiable hypotheses expressed in terms of the data themselves. Presence of the intermediate time t_1 in the cost is simply dealt with by state augmentation techniques. Here, then, we immediately see the advantages of applying multiprocess theory instead of classical methods of analysis.

1.6 Sufficient Conditions, Dynamic Programming and the Maximum Principle

We now turn our attention to sufficient conditions for optimality and we explore, in particular, the relationship of the maximum principle to dynamic programming and the attempt to characterize the value function (1.6.2) as a solution to the Hamilton-

Jacobi-Bellman equation (1.6.3). (The use of the asterisk in the remaining sections of this chapter and throughout the thesis denotes optimal quantities, unless otherwise stated).

Let us consider, for example, the following free final state optimal control problem:

$$\text{minimize } \int_0^1 L(t, x(t), u(t)) dt + g(x(1)) \quad (1.6.1)$$

subject to

$$\begin{aligned} \dot{x}(t) &= f(t, x(t), u(t)) \text{ a.e. } t \in [0, 1], \\ x(0) &= x_0, \\ u(t) &\in \Omega \text{ a.e. } t \in [0, 1]. \end{aligned}$$

Here $L: \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$, $f: \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ and $g: \mathbb{R}^n \rightarrow \mathbb{R}$ are given functions, $x_0 \in \mathbb{R}^n$ is a given point and $\Omega \subset \mathbb{R}^m$ a given set.

Dynamic programming (initiated by Bellman [B1]) is a general methodology for solving dynamic optimization problems. It centres on the properties of the value function, which is defined as the infimum cost of the problem under consideration, taken as a function of the initial time and state. In this case, the value function $V: [0, 1] \times \mathbb{R}^n \rightarrow \mathbb{R}$ is taken to be

$$V(\tau, \xi) := \inf \left\{ g(x(1)) + \int_{\tau}^1 L(t, x(t), u(t)) dt : \dot{x}(t) = f(t, x(t), u(t)), u(t) \in \Omega, x(\tau) = \xi \right\}, \quad (1.6.2)$$

where the underlying time interval $[0, 1]$ is replaced by $[\tau, 1]$ and the initial state x_0 by ξ . Whenever the value function is continuously differentiable it satisfies the Hamilton-Jacobi-Bellman equation, namely

$$V_t(t, x) + \min_{u \in \Omega} \{ V_x(t, x) \cdot f(t, x, u) + L(t, x, u) \} = 0 \text{ for } (t, x) \in [0, 1] \times \mathbb{R}^n \quad (1.6.3)$$

and boundary condition $V(1, x) = g(x)$ for $x \in \mathbb{R}^n$. (1.6.4)

We find, furthermore, that if (x^*, u^*) is a minimizing control process for the problem, then

$$V_t(t, x^*(t)) + V_x(t, x^*(t)) \cdot f(t, x^*(t), u^*(t)) + L(t, x^*(t), u^*(t)) = 0, \text{ a.e.} \quad (1.6.5)$$

The dynamic programming approach is to solve the Hamilton-Jacobi-Bellman equation for V and to extract information about (x^*, u^*) from (1.6.5) (see e.g. Fleming and Rishel [F1]).

The question arises: how are the maximum principle and dynamic programming related? Many authors have advanced heuristic arguments suggesting the following

relationship

$$\left(H(t, x^*(t), u^*(t), p(t)), -p(t) \right) = \nabla V(t, x^*(t)) \quad (1.6.6)$$

where (x^*, u^*) is the optimal control process under consideration, $p(t)$ is the costate function and H is the Hamiltonian (see Chapter 2).

The drawback is that the hypotheses under which these relationships can be made rigorous are very severe, especially the assumption that the value function is smooth since, to quote Pontryagin et al. [P1] “it must be noted that the assumption on the continuous differentiability of the functional does not hold in the simplest cases”. Here, by ‘functional’ Pontryagin means the value function.

How do we overcome the difficulty that V need not be continuously differentiable? Since we can expect V to be locally Lipschitz continuous (refer to Chapter 2) it is natural to replace $-p(t) = V_x(t, x^*(t))$ by $-p(t) \in \partial_x V(t, x^*(t))$, where ∂_x denotes the partial generalized gradient of Clarke (see Chapter 2). This has been proved by Clarke and Vinter [C6] who show that for a very large class of non-smooth free end-point problems a costate function $p(\cdot)$ can be chosen to satisfy the inclusion.

Clarke and Vinter also show that the inclusion $-p(t) \in \partial_x V(t, x^*(t))$ is valid for some costate function even in the presence of a right end-point constraint set. Their approach is to substitute some ‘verification function’ (see below) for $V(\cdot, \cdot)$ and to replace the terminally constrained problem by an equivalent free end-point problem with a non-smooth penalty term. This reduces the original problem to one which can be solved using existing techniques.

1.7 Verification Theory

We pause here in order to explain the term verification function. Verification functions arise in a body of theory, referred to as ‘verification theory’, which provides sufficient conditions of optimality expressed in terms of the Hamilton-Jacobi equation.

Consider Problem (1.6.1) when we supplement the constraints by the end-point constraint $x(1) \in C$. The idea underlying the verification technique may be simply expressed. It is:

“Suppose that (x^*, u^*) is an admissible process for the problem, and suppose $\phi(t, x)$ is a continuously differentiable solution of the Hamilton-Jacobi equation

$$\frac{\partial \phi}{\partial t}(t, x) + \min_{u \in \Omega} \left(\frac{\partial \phi}{\partial x}(t, x) \cdot f(t, x, u) + L(t, x, u) \right) = 0 \text{ for } (t, x) \in [0, 1] \times \mathbb{R}^n, \quad (1.7.1)$$

and satisfies the boundary condition

$$\phi(1, x) = g(x) \text{ for } x \in C. \quad (1.7.2)$$

Then if

$$\phi(0, x_0) = g(x^*(1)) + \int_0^1 L(t, x^*, u^*) dt, \quad (1.7.3)$$

it follows that (x^*, u^*) is minimizing”.

ϕ is known as a verification function since it ‘verifies’ the (local) optimality of (x^*, u^*) .

However, the smoothness condition imposed on ϕ is restrictive, in that for many problems no such ϕ exists. Therefore, to raise the technique to a greater level of generality and, in particular, to allow functions ϕ which are non-smooth, we make use of the (extended) Hamilton-Jacobi equation which incorporates the generalized gradient of ϕ , namely we allow ϕ to be a locally Lipschitz solution of

$$\min\{\alpha - H(t, x, -\beta) : (\alpha, \beta) \in \partial\phi(t, x)\} = 0. \quad (1.7.4)$$

Here $\partial\phi$ denotes the Clarke generalized gradient of ϕ . H is the Hamiltonian function

$$H(t, x, p) = \text{Max}\{p \cdot f(t, x, u) : u \in \Omega\}.$$

In the case ϕ is smooth, $\partial\phi$ becomes $(\frac{\partial\phi}{\partial t}, \frac{\partial\phi}{\partial x})$, and (1.7.4) reduces to (1.7.1).

The extended verification technique, by allowing the flexibility of ‘non-smooth’ verification functions ϕ , applies to a wide range of problems.

Of course, there is a close relationship between verification functions and the value function. The value function satisfies equations (1.7.2) and (1.7.3) in the definition of verification function and we can expect equation (1.7.1) to be satisfied at least on neighbourhoods where V is Lipschitz continuous. The difference is that there can exist a (Lipschitz continuous) verification function for (x^*, u^*) even when V is ill-behaved (has domain with empty interior, say).

We return to our earlier discussion. Following inclusion of the terminal constraint $x(1) \in C$, there is a real possibility that the value function will not even be Lipschitz continuous on a neighbourhood of the optimal trajectory, hence the approach used by Clarke and Vinter [C6] is that of verification theory, where, as previously mentioned, some verification function is substituted for the value function. Using ‘exact penalty’ techniques for this constrained end-point problem, they show that as well as providing a sufficient condition for local optimality of a trajectory, the inclusion is also necessary under a strict normality hypothesis, i.e. when the first order optimality conditions yield non-trivial information. This is in line with recent research using non-smooth analysis which has aimed at weakening the requirements on verification functions to the point where, under mild and verifiable hypotheses, conditions are provided which are necessary as well as sufficient for optimality.

1.8 Minimum Cost Sensitivity Relations

As a next step, it is natural to conjecture that in place of (1.6.6) we can write:

$$\left(H(t, x^*(t), u^*(t), p(t)), -p(t) \right) \in \partial V(t, x^*(t)) \quad \text{for all } t \in [0, 1] \quad (1.8.1)$$

where ∂V denotes the generalized gradient of V . This has indeed been proved by Vinter [V1] for non-smooth free end-point problems. The new result improves on an earlier result [C6] which merely interpreted the generalized gradient of V in the x variable, and gave an ‘almost everywhere’ condition. Recently, Vinter and Wolenski [V3] have shown that inclusion (1.8.1) holds in the case where the data are measurable in the time variable.

A refinement of the above inclusion involves using the superdifferential of V , $\partial^+ V(t, x(t))$, instead of the generalized gradient of V . It was shown by Subbotina [S3] that standard first order necessary conditions coupled with a strengthened minimum cost sensitivity inclusion expressed in terms of the superdifferential of the value function are *sufficient* as well as necessary for optimality for a special class of problems with smooth data. Since $\partial^+ V \subset \partial V$ (see Chapter 2), the use of the superdifferential strengthens the necessary conditions of (1.8.1) as it allows the costate to be more accurately located than does the use of the generalized gradient and thus provides a more precise sensitivity relation.

A feature of Vinter’s [V1] and Subbotina’s [S3] results is that they treat a problem with a free right end-point. The question arises, can we develop the theory to obtain a sensitivity relation in both variables for a constrained right end-point?

The above discussion suggests that we expect the answer to be in the affirmative but that the sensitivity inclusion will not necessarily involve the value function. This is because the value function may be too irregular or have an unsuitable domain, especially in the presence of end-point constraints. Therefore we shall seek to use some other well-defined function in the inclusion and hence our motivation for attempting to formulate a result using verification theory.

Regarding ‘sensitivity relations’ we consider, in this thesis, the following problem with purely integral cost and a terminal constraint set:

$$\text{minimize } \int_0^1 L(t, x(t), u(t)) dt,$$

subject to

$$\dot{x}(t) = f(t, x(t), u(t)), \text{ a.e. } t \in [0, 1],$$

$$x(0) = x_0, x(1) \in C,$$

$$u(t) \in U, \text{ a.e. } t \in [0, 1].$$

The essential idea is to introduce a definition of ‘verification function’ which enables us to construct it as the ‘value function’ of some associated optimization problem. In this way existence of a verification function is assured.

Our main new result in this area is that, provided a normality hypothesis is satisfied (which in certain circumstances is verifiable), there is a relationship between the costate $p(\cdot)$ of the maximum principle and the superdifferential of some verification function W ,

$$\left(H(t, x^*(t), u^*(t), p(t)), -p(t) \right) \in \partial^* W(t, x^*(t)) \quad (1.8.2)$$

along a trajectory $x^*(\cdot)$, which (coupled with the maximum principle and a verification theorem) is *both necessary and sufficient for that trajectory to be optimal*.

We deal with the presence of an end-point constraint set in two stages :

- a) we prove that the theorem holds in the fixed end-point case, i.e. $x(1)=x_1$, via a penalty function technique,
- b) we modify the original problem so that it can be incorporated into a fixed end-point problem, via an extension of the underlying time interval.

1.9 Outline of Thesis

We now briefly comment on the structure of the thesis. We present in Chapter 2, as a reference source, the definitions and basic results of non-smooth analysis which we shall require throughout the work. We also state the maximum principle for the non-smooth case and its extension for multiprocesses. In Chapter 3 we state and prove new necessary conditions for optimality for multistage problems defined over the infinite horizon, and provide an application of the theory to a problem in economics. In Chapter 4 we present a detailed proof of the new sensitivity relation for general constrained end-point problems. We conclude in Chapter 5 with a summary of the results and suggestions for further work in these areas.

BACKGROUND THEORY

2.1 Introduction

In this chapter we present, for ease of reference, the essential background theory required for this thesis. We begin with the main definitions and theorems of non-smooth analysis used throughout this work. Unless otherwise stated, the notation follows that of Clarke [C3] where a more detailed presentation of this theory may be found.

In the second section, we replace the ordinary derivative by the generalized gradient since it can be defined for a very general class of functions. In particular, we consider real-valued ‘locally Lipschitz’ functions defined on a Banach space X , whose elements x we refer to as vectors or points, whose norm we denote $\|x\|$ and whose open unit ball is denoted $\mathbb{B}$.

In the third section, we consider, in particular, the relationship between the generalized gradient and the strict derivative of a function at a given point. This theory will be required when developing a necessary and sufficient sensitivity relation for a constrained end-point optimal control problem in a later chapter.

Similarly, in the fourth section, we define the superdifferential of a function, give an alternate characterization and indicate its relation to the generalized gradient.

In Section 5 we present results of the geometric theory of normal and tangent cones associated with the generalized gradient. Again, proofs of all the results may be found in Clarke [C3], unless otherwise stated.

In Section 6 we provide a statement of the basic optimal control problem, together with a set of necessary conditions for an admissible process to be optimal – the famous Pontryagin Maximum Principle. We also define the concept of normality for an optimal control problem.

In the seventh section we give a basic introduction to the theory of optimal multiprocesses developed by Clarke and Vinter [C7,C8], which extends the notion of the classical Pontryagin maximum principle. We define a multiprocess, state a maximum principle for optimal multiprocesses and prove some basic identities used in the derivation of normal cones. Analysis of the latter is required to enable us to obtain necessary conditions for specific multiprocess problems.

2.2 Generalized Differentiation

2.2.1 The Lipschitz Condition

We recall that X is a normed space (by definition a Banach space is a complete normed space) and begin our discussion with the following definition:

Definition 2.2.1: Let Y be a subset of X . A function $f: Y \rightarrow \mathbb{R}$ is said to satisfy a Lipschitz condition (of rank K) on Y if, for some non-negative scalar K ,

$$|f(x) - f(x')| \leq K \|x - x'\| \quad \text{for all points } x, x' \text{ in } Y. \quad (2.2.1)$$

We say that f is Lipschitz (of rank K) near x if, for some $\epsilon > 0$, f satisfies a Lipschitz condition (of rank K) on the set $x + \epsilon\mathbb{B}$, i.e. on the open ball of radius ϵ about x .

A function which is Lipschitz near a point is not necessarily differentiable there and in order to study differential properties of f , we need the following definition of a directional derivative. We note that it differs from the classical definition of directional derivative in that the base point of the difference quotient varies.

2.2.2 The Generalized Directional Derivative

Definition 2.2.2: Let f be Lipschitz near x and let v be any other vector in X . The generalized directional derivative of f at x in the direction v , denoted $f^\circ(x; v)$ is given by

$$f^\circ(x; v) = \lim_{\substack{y \rightarrow x \\ t \downarrow 0}} \sup \frac{f(y + tv) - f(y)}{t} \quad (2.2.2)$$

where $y \in X$ and t is a positive scalar.

The directional derivative $f^\circ(x; v)$ is a well-defined finite quantity since the difference quotient is bounded above by $K|v|$, in view of the Lipschitz condition. Furthermore, it is positively homogeneous and subadditive as a function of v , namely $f^\circ(x; \lambda v) = \lambda f^\circ(x; v)$ and $f^\circ(x; v + w) \leq f^\circ(x; v) + f^\circ(x; w)$, and hence allows us to develop a definition for the generalized gradient of f at x (see below).

2.2.3 The Generalized Gradient

The Hahn-Banach theorem (see e.g. [R3]) states that any positively homogeneous and subadditive functional on X majorizes some linear functional on X . Therefore there exists at least one linear functional $\zeta: X \rightarrow \mathbb{R}$ such that for all $v \in X$, $f^\circ(x; v) \geq \langle \zeta, v \rangle$. The functional ζ is bounded and belongs to the dual space X^* of continuous linear functionals on X . Hence we may define a non-empty set, $\partial f(x)$, the *generalized gradient* of f at x , as follows:

Definition 2.2.3: *The generalized gradient of f at x , denoted by $\partial f(x)$ is the subset of X^* given by*

$$\partial f(x) = \{\zeta \in X^*: f^\circ(x; v) \geq \langle \zeta, v \rangle \text{ for all } v \text{ in } X\}. \quad (2.2.3)$$

It can be shown by considering the properties of f° that $\partial f(x)$ is a non-empty, convex, weak*-compact subset of X^* and that the following result holds. Recall that X^* is the 'dual space' of continuous linear functionals on X .

Proposition 2.2.4: $f^\circ(x; v) = \max\{\langle \zeta, v \rangle: \zeta \in \partial f(x)\}.$ (2.2.4)

Thus we see that knowing $f^\circ(x; v)$ is equivalent to knowing $\partial f(x)$.

2.2.4 Support Functions

The above is an example of a general fact: closed convex sets are characterized by their *support functions*. We now state the definition of a support function and describe, without proof, some properties of support functions (see e.g. [C3]) as these are required later.

Definition 2.2.5: *The support function of a non-empty subset C of X is the function $S_C: X^* \rightarrow \mathbb{R} \cup \{+\infty\}$ defined by*

$$S_C(\xi) := \sup\{\langle \xi, x \rangle: x \in C\}. \quad (2.2.5)$$

The function measures how much the set C extends in the given direction ξ . If C is closed and convex, then C is the only closed convex set with support function S_C . We can state the following propositions, in which X is taken to be $\mathbb{R}^n$ and therefore X and its dual X^* are equal.

Proposition 2.2.6: *Let $C \subset X$ be a closed convex set. Then*

$$x \in C \text{ iff } S_C(p) \geq \langle p, x \rangle, \text{ for all } p \in X.$$

This result reiterates that $f^\circ(x; \cdot)$ is the support function of $\partial f(x)$ since we can write

$$\xi \in \partial f(x) \text{ iff } f^\circ(x; v) \geq \langle \xi, v \rangle, \text{ for all } v \text{ in } X.$$

Proposition 2.2.7: *Let $C \subset X$ be a bounded subset, then S_C is everywhere finite and Lipschitz continuous.*

We also require the definition of convex hull.

Definition 2.2.8: The convex hull, $\text{co } Q$, of a subset $Q \subset \mathbb{R}^n$ is the intersection of all convex sets which contain Q .

The closure of Q is denoted by $\bar{\text{co}} Q$.

2.3 The Generalized Gradient and its Relation to Strict Derivatives

It is natural to examine the relationship between the generalized gradient $\partial f(x)$ of the locally Lipschitz continuous function $f: X \rightarrow \mathbb{R}$ at $x \in X$, and other notions of derivative. Certainly $\partial f(x)$ is a generalization of the usual gradient, since when f is smooth (continuously differentiable), $\partial f(x)$ reduces to the singleton set $\{\nabla f(x)\}$. Now suppose that $f: X \rightarrow \mathbb{R}$ is convex. Recall that this means

$$f(\lambda u + (1-\lambda)u') \leq \lambda f(u) + (1-\lambda)f(u')$$

for all $u, u' \in X$ and $\lambda \in [0,1]$. It can be shown that f is locally Lipschitz continuous, so the generalized gradient $\partial f(x)$ of f at x is defined. It can also be shown that

i) $\partial f(x)$ coincides with the subdifferential of convex analysis, that is, the set of vectors ξ in X such that $f(x+u) - f(x) \geq \langle u, \xi \rangle$ for all u in X ,

and

ii) $f^\circ(x; v)$ coincides with the directional derivative $f'(x; v)$ for each v .

We shall also be interested in generalized gradients of ‘strictly differentiable’ functions. We begin by defining the strict derivative of a function at a given point. Note that $\mathcal{L}(X, Y)$ denotes the space of continuous linear functionals from X to Y .

Definition 2.3.1: Let F map X to another Banach space Y . F is said to admit a strict derivative at x , an element of $\mathcal{L}(X, Y)$ denoted $D_S F(x)$, provided that for each v ,

$$\lim_{\substack{x' \rightarrow x \\ t \downarrow 0}} \frac{F(x' + tv) - F(x')}{t} = \langle D_S F(x), v \rangle \quad (2.3.1)$$

and provided the convergence is uniform for v in compact sets (the latter is automatic if F is Lipschitz near x).

Proposition 2.3.2: Let F map a neighbourhood of x to Y , and let ξ be an element of $\mathcal{L}(X, Y)$. The following are equivalent:

- a) F is strictly differentiable at x and $D_S F(x) = \xi$.
- b) F is Lipschitz near x , and for each v in X one has

$$\lim_{\substack{x' \rightarrow x \\ t \downarrow 0}} \frac{F(x' + tv) - F(x')}{t} = \langle \xi, v \rangle.$$

As a corollary we have that if F is continuously differentiable at x , then F is strictly differentiable at x and therefore Lipschitz near x .

Specializing once again to real-valued functions, we note the following relationship between the strict derivative and the generalized gradient, which is proved in Clarke [C3].

Proposition 2.3.3: *If the function $f:X \rightarrow \mathbb{R}$ is strictly differentiable at $x \in X$, then f is Lipschitz near x and $\partial f(x) = \{D_s f(x)\}$.*

Conversely, if f is Lipschitz near x and $\partial f(x)$ reduces to a singleton $\{\xi\}$, then f is strictly differentiable at x and $D_s f(x) = \xi$.

We require this result to enable us to define a strictly differentiable ‘penalty function’ when later dealing with constrained end-point optimal control problems.

Let us now assume that $X = \mathbb{R}^n$ as this provides us with additional properties of generalized gradients. In this case it follows that the space X and its dual X^* are equal, so that $\partial f(x)$ is now a subset of $\mathbb{R}^n$.

We can give an equivalent definition of the generalized gradient in this case, based on Rademacher’s theorem which states that a locally Lipschitz function is differentiable almost everywhere in the sense of Lebesgue measure.

Theorem 2.3.4: *Let f be Lipschitz near x , and suppose S is any set of Lebesgue measure 0 in $\mathbb{R}^n$. Let Ω_f denote the set of points at which f fails to be differentiable. Then*

$$\partial f(x) = \text{co} \left\{ \lim \nabla f(x_i) : x_i \rightarrow x, x_i \notin S, x_i \notin \Omega_f \right\}. \quad (2.3.2)$$

(i.e. consider any sequence x_i , converging to x , while avoiding both S and the points at which it is not differentiable, and such that the sequence $\nabla f(x_i)$ converges; then the convex hull of all such limit points is $\partial f(x)$).

This characterization is often simpler to use than the original definition of generalized gradients in finite dimensions.

Many calculus rules of classical analysis have some kind of parallel for generalized gradients of locally Lipschitz functions. We note in particular:

Proposition 2.3.5: *Let $f_1, \dots, f_k: X \rightarrow \mathbb{R}$ be functions which are Lipschitz near a given point x . Then*

a) for any scalars c_1, c_2 ,

$$\partial(c_1 f_1 + c_2 f_2)(x) \subseteq c_1 \partial f_1(x) + c_2 \partial f_2(x)$$

- b) $\partial(f_1 f_2)(x) \subseteq f_1(x)\partial f_2(x) + f_2(x)\partial f_1(x)$
 c) if f_1 takes on a local minimum or maximum at x , then $0 \in \partial f_1(x)$.

2.4 The Generalized Gradient and its Relation to Superdifferentials

We begin with the definition of the superdifferential.

Definition 2.4.1: The superdifferential of a function $\phi: \mathbb{R}^k \rightarrow \mathbb{R}$ at a point $y \in \mathbb{R}^k$, written $\partial^\vee \phi(y)$ is defined to be the set

$$\partial^\vee \phi(y) := \left\{ p \in \mathbb{R}^k : \limsup_{y' \rightarrow y} \left((\phi(y') - \phi(y) - p \cdot (y' - y)) / |y' - y| \right) \leq 0 \right\}. \quad (2.4.1)$$

This set may be empty even when ϕ is a convex function but $\partial^\vee \phi(y)$ reduces to the familiar gradient $\phi_y(y)$ when ϕ is C^1 in a neighbourhood of y .

The following alternative characterization of the superdifferential is easy to check:

Lemma 2.4.2: Let $\phi: \mathbb{R}^k \rightarrow \mathbb{R}$ and $y \in \mathbb{R}^k$ be given. Then $\partial^\vee \phi(y)$ is the set of vectors p such that, for an arbitrary vector $z \in \mathbb{R}^k$

$$\limsup_{\substack{\epsilon \downarrow 0 \\ z' \rightarrow z}} \frac{\phi(y + \epsilon z') - \phi(y)}{\epsilon} \leq p \cdot z. \quad (2.4.2)$$

This leads to the following relationship between superdifferentials $\partial^\vee \phi$ and the generalized gradient $\partial \phi$.

Corollary 2.4.3: Suppose that the function $\phi: \mathbb{R}^k \rightarrow \mathbb{R}$ is Lipschitz continuous in a neighbourhood of the point $y \in \mathbb{R}^k$. Then

$$\partial^\vee \phi(y) \subset \partial \phi(y). \quad (2.4.3)$$

Proof: The definition of the set $\partial^\vee \phi(y)$ implies that for an arbitrary vector $p \in \partial^\vee \phi(y)$ and arbitrary $z \in \mathbb{R}^k$,

$$\liminf_{\epsilon \downarrow 0} \epsilon^{-1} (\phi(y + \epsilon z) - \phi(y)) \leq p \cdot z. \quad (2.4.4)$$

Multiplying across this inequality by -1 , introducing the function $\psi(y) := -\phi(y)$, and estimating the resulting $\limsup$ gives

$$\limsup_{\substack{\epsilon \downarrow 0 \\ y' \rightarrow y}} \epsilon^{-1} (\psi(y' + \epsilon z) - \psi(y')) \geq (-p) \cdot z. \quad (2.4.5)$$

The left side here is the Clarke generalized directional derivative $D^\circ\psi(y;z)$. Hence we have shown that

$$D^\circ\psi(y;z) \geq (-p) \cdot z \quad (2.4.6)$$

for arbitrary z . This implies that $(-p) \in \partial(-\phi)(y)$ by the definition of generalized gradient.

Hence $-p \in -\partial\phi$ (see Proposition 2.3.5(a)) and therefore $p \in \partial\phi$ and the result is proved. $\square$

The inclusion (2.4.3) can be strict since e.g. for $\phi(x) = |x|$, $\partial\phi(x=0)$ is the whole interval $[-1, +1]$, while $\partial^\vee\phi(x=0)$ is empty.

2.5 Normal and Tangent Cones

We again assume X to be a Banach space. We show how the generalized gradient is related to the geometric concepts associated with classical calculus and how it can be used, in conjunction with the distance function, to develop concepts of tangency and normality to an arbitrary set C that do not require smoothness. We begin by considering the following definition.

2.5.1 The Distance Function

Definition 2.5.1: Let C be a non-empty subset of X . Then its distance function, i.e. $d_C(\cdot): X \rightarrow \mathbb{R}$ is defined by

$$d_C(x) = \inf \{ \|x - c\| : c \in C \}. \quad (2.5.1)$$

The function d_C is non-smooth, however it is Lipschitz continuous with Lipschitz constant 1 on X , i.e.

$$|d_C(x) - d_C(y)| \leq \|x - y\|, \text{ for all } x, y \in X. \quad (2.5.2)$$

2.5.2 Tangent Cones

The distance function is used to define tangent vectors to a set. Let x be a point in C . We have that a vector v in X is tangent to C at x if $d_C^\circ(x; v) = 0$ hence we may define the tangent cone as follows:

Definition 2.5.2: The tangent cone $T_C(x)$ to C at a point x is defined as the set of all tangents to C at x , namely

$$T_C(x) := \{ v \in X : d_C^\circ(x; v) = 0 \}. \quad (2.5.3)$$

This condition specifies a closed, convex cone.

2.5.3 Normal Cones

A natural candidate for normal cone is the one obtained from $T_C(x)$ by polarity. We define:

Definition 2.5.3: *The (Clarke) normal cone to C at x is defined as*

$$N_C(x) := \left\{ \xi \in X^* : \langle \xi, v \rangle \leq 0 \text{ for all } v \text{ in } T_C(x) \right\}. \quad (2.5.4)$$

In the case that C is a closed, convex set the above definition is equivalent to the standard one, namely, a (non-zero) normal vector is the normal to a hyperplane which separates x and the relative interior of C :

Definition 2.5.4: $\xi \in X^*$ lies in the normal cone to C at x if and only if $\langle \xi, x' - x \rangle \leq 0$ for all $x' \in C$.

It can be shown that $N_C(x)$ is the closed convex cone generated by $\partial d_C(x)$, i.e.

$$N_C(x) = \text{cl} \left\{ \bigcup_{\lambda \geq 0} \lambda \partial d_C(x) \right\} \quad (2.5.5)$$

where cl denotes weak*-closure. The importance of the distance function in optimization is due to its role in the ‘penalty function’ techniques which we make use of in a later chapter. It is possible to define T_C and N_C directly without making use of the distance function, via the following results.

2.5.4 Characterization of $T_C(x)$

Theorem 2.5.5: *An element v of X is tangent to C at x iff, for every sequence x_i in C converging to x and sequence t_i in $(0, \infty)$ decreasing to 0, there is a sequence v_i in X converging to v such that $x_i + t_i v_i \in C$ for all i .*

The corollary which follows is made use of when calculating normal cones in multiprocess theory. The first equality follows readily from the characterization of the tangent cone given in the theorem, and the second follows by polarity.

Corollary 2.5.6: *Let $X = X_1 \times X_2$, where X_1, X_2 are Banach spaces, and let $x \in C_1 \times C_2$, where C_1, C_2 are subsets of X_1, X_2 respectively. Let $x = (x_1, x_2)$, then*

$$T_{C_1 \times C_2}(x) = T_{C_1}(x_1) \times T_{C_2}(x_2) \quad (2.5.6)$$

$$N_{C_1 \times C_2}(x) = N_{C_1}(x_1) \times N_{C_2}(x_2). \quad (2.5.7)$$

To define $N_C(x)$ directly we again assume $X=\mathbb{R}^n$ (i.e. we consider the finite dimensional case). We require the notion of a perpendicular.

Definition 2.5.7: A non-zero vector v is perpendicular to C at a point x in C (denoted $v \perp C$ at x) if $v = x' - x$, where the point x' has unique closest point x in C .

(Equivalently, $v = x' - x$, where there is a closed ball centred at x' which meets C only at x .)

The alternate characterization of normal vectors to sets in $\mathbb{R}^n$ is given below. It is useful in many particular instances and is the geometric analogue of Theorem 2.3.4.

2.5.5 Characterization of Normal Vectors

Theorem 2.5.8: Let x belong to $cl\ C$. Then

$$N_C(x) = \overline{co} \left\{ \lambda \lim \frac{v_i}{|v_i|} : \lambda \geq 0, v_i \perp C \text{ at } x_i, x_i \rightarrow x, v_i \rightarrow 0 \right\}. \quad (2.5.8)$$

i.e. $N_C(x)$ is the closed convex cone generated by the origin and the set

$$\left\{ \lim \frac{v_i}{|v_i|} : v_i \perp C \text{ at } x_i, x_i \rightarrow x, v_i \rightarrow 0 \right\}.$$

The two analytic concepts for functions (f° and ∂f) and the two geometric concepts for sets (T_C and N_C) are dual (obtainable from each other via Proposition 2.2.4 and by polarity respectively). They are also linked to each other via the *epigraph* of a function f , denoted $\text{epi } f$, which is defined to be the set $\{(x, r) \in X \times \mathbb{R} : f(x) \leq r\}$ i.e. the set of points on or above the graph of f . The following relationships can be shown to hold :

$$T_{\text{epi } f}(x, f(x)) = \text{epi } f^\circ(x; \cdot) \quad (2.5.9)$$

$$\partial f(x) = \left\{ \xi : (\xi, -1) \in N_{\text{epi } f}(x, f(x)) \right\}. \quad (2.5.10)$$

The last relationship reflects the familiar calculus result that if f is differentiable at x , the vector $[f'(x), -1]$ is normal to the graph of f at the point $(x, f(x))$.

All the functions considered have been locally Lipschitz since it is a reasonable and usually verifiable hypothesis in a wide range of applications.

2.6 The Maximum Principle

Optimal processes may be characterized through the *maximum principle* developed by Pontryagin et al. [P1] which provides a set of necessary conditions for a process to be minimizing, similar to the Euler equation of the Calculus of Variations.

2.6.1 The Optimal Control Problem

The problem under consideration is :

$$\text{Minimize } \int_0^1 L(t, x(t), u(t)) dt + g(x(1))$$

subject to

$$\begin{aligned} \dot{x}(t) &= f(t, x(t), u(t)), & \text{a.e. } t \in [0, 1], \\ x(0) &\in C_0, \quad x(1) \in C_1, \\ u(t) &\in \Omega, & \text{a.e. } t \in [0, 1]. \end{aligned}$$

Here $L: \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$, $f: \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$, $g: \mathbb{R}^n \rightarrow \mathbb{R}$ are given functions and $C_0, C_1 \subset \mathbb{R}^n$, $\Omega \subset \mathbb{R}^m$ are given sets.

Control functions $u(\cdot): [0, 1] \rightarrow \mathbb{R}^m$ are taken to be measurable functions such that $u(t) \in \Omega$ a.e. $t \in [0, 1]$.

A state trajectory $x(\cdot): [0, 1] \rightarrow \mathbb{R}^n$ (corresponding to a control function $u(\cdot)$) is a Lipschitz continuous function satisfying the differential equation $\dot{x}(t) = f(t, x(t), u(t))$, a.e. $t \in [0, 1]$.

Minimization is conducted over ‘admissible processes’. An ‘admissible process’ is a pair $(x(\cdot), u(\cdot))$ in which $x(\cdot)$ is a state trajectory, $u(\cdot)$ is a corresponding control function and $x(\cdot)$ also satisfies the end-point constraints $x(0) \in C_0$, $x(1) \in C_1$.

A minimizing admissible state trajectory or process is said to be ‘optimal’.

2.6.2 The Pontryagin Maximum Principle

It is convenient to state hypotheses in terms of the ‘augmented velocity function’ $\tilde{f}: [0, 1] \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^{n+1}$ where $\tilde{f}(t, x, u) = \begin{bmatrix} L(t, x, u) \\ f(t, x, u) \end{bmatrix}$.

We wish to ensure that $t \mapsto \tilde{f}(t, x(t), u(t))$ is measurable for each measurable function $u(\cdot)$ and continuous function $x(\cdot)$ and the appropriate weakest measurability requirement which guarantees this is $\mathcal{L} \times \mathcal{B}$ measurability. A function $\mathcal{M}: \mathbb{R} \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ is $\mathcal{L} \times \mathcal{B}$ measurable if it is measurable with respect to $\mathcal{L} \times \mathcal{B}$, where $\mathcal{L} \times \mathcal{B}$ denotes the σ -algebra of subsets of $[0, 1] \times \mathbb{R}^m$ generated by product sets $M \times N$, where M is a Lebesgue measurable subset of $[0, 1]$ and N is a Borel subset of $\mathbb{R}^m$.

We define the Hamiltonian function as

$$H(t, x, u, p, \lambda) = p \cdot f(t, x, u) - \lambda L(t, x, u).$$

Theorem 2.6.1 (Pontryagin maximum principle): *Let $(x^*(\cdot), u^*(\cdot))$ be an optimal process for the given problem. It is assumed that*

• for each $x \in \mathbb{R}^n$, $\tilde{f}(\cdot, x, \cdot)$ is $\mathcal{L} \times \mathcal{B}$ measurable where $\mathcal{L}$ denotes the class of Lebesgue subsets in $\mathbb{R}$ and $\mathcal{B}$ denotes the class of Borel subsets in $\mathbb{R}^m$,

• there exists a function $K(\cdot) \in L^1$ such that

$$\|\tilde{f}(t, x, u) - \tilde{f}(t, y, u)\| \leq K(t)\|x - y\|, \quad \text{a.e. } t \in [0, 1], x, y \in \mathbb{R}^n \text{ and } u \in \Omega,$$

• g is locally Lipschitz continuous,

• graph Ω is a compact set,

• C_0 and C_1 are closed sets.

Then there exists a number $\lambda \in \{0, 1\}$ and a Lipschitz continuous function $p(\cdot): [0, 1] \rightarrow \mathbb{R}^n$ such that

$$\lambda + |p(1)| > 0,$$

$$-\dot{p}(t) \in \partial_x H(t, x^*(t), u^*(t), p(t), \lambda) \quad \text{a.e.} \quad (2.6.1)$$

$$H(t, x^*(t), u^*(t), p(t), \lambda) = \max_{v \in \Omega} H(t, x^*(t), v, p(t), \lambda) \quad \text{a.e.} \quad (2.6.2)$$

and

$$p(0) \in N_{C_0}(x^*(0)), \quad -p(1) \in N_{C_1}(x^*(1)) + \lambda \xi \quad \text{for some } \xi \in \partial g(x^*(1)). \quad (2.6.3)$$

The function $p(\cdot)$ is called the *costate function* and λ the *cost multiplier*. The differential equation (2.6.1) is referred to as the *costate differential equation* and the constraints (2.6.3) as the *transversality conditions*. Condition (2.6.2) is the *maximization of the Hamiltonian* condition.

Here, $\partial_x H$ denotes the Clarke generalized gradient of H in the second argument. We note that the theorem stated above extends the classical maximum principle to problems where the data are non-smooth in the state variable. For a proof of this, see Clarke [C3].

When the data are continuously differentiable in the state variable, ∂H and ∂g are the single valued functions $p \cdot f_x(t, x, u) - \lambda L_x(t, x, u)$ and g_x respectively. In this case the optimality conditions take the form:

$$-\dot{p}(t) = p(t) \cdot f_x(t, x^*(t), u^*(t)) - \lambda L_x(t, x^*(t), u^*(t)) \quad \text{a.e.} \quad (2.6.4)$$

$$H(t, x^*(t), u^*(t), p(t), \lambda) = \max_{v \in \Omega} H(t, x^*(t), v, p(t), \lambda) \quad \text{a.e.} \quad (2.6.5)$$

and

$$p(0) \in N_{C_0}(x^*(0)), \quad -p(1) \in N_{C_1}(x^*(1)) + \lambda g_x(x^*(1)) \quad (2.6.6)$$

2.7 Multiprocess Theory

The notation of this section follows that of Clarke and Vinter (see [C7]). An optimal multiprocess problem, as defined by Clarke and Vinter, is a dynamic optimization problem involving a collection of control systems coupled through

constraints in the end-points of the constituent state trajectories and through the cost function.

The development of a maximum principle for multiprocesses extends the applicability of the classical Pontryagin maximum principle and permits the solution of a wider range of problems.

2.7.1 Multiprocesses

Consider positive integers k and n_i, m_i ($i=1, \dots, k$), subsets U^i of $\mathbb{R} \times \mathbb{R}^{m_i}$ ($i=1, \dots, k$), subsets X^i of $\mathbb{R} \times \mathbb{R}^{n_i}$ ($i=1, \dots, k$) and functions $\phi_i: \mathbb{R} \times \mathbb{R}^{n_i} \times \mathbb{R}^{m_i} \rightarrow \mathbb{R}^{n_i}$.

Since frequent reference is made to points in product spaces, and to products of product spaces, we denote a point $((a_1, b_1, \dots), (a_2, b_2, \dots), \dots, (a_k, b_k, \dots))$ by $\{a_i, b_i, \dots\}$.

Definition 2.7.1: A multiprocess is a k -tuple $\{\tau_0^i, \tau_1^i, y_i(\cdot), w_i(\cdot)\}$ each element in which comprises left and right endpoints τ_0^i, τ_1^i of a closed interval, an absolutely continuous function $y_i(\cdot): [\tau_0^i, \tau_1^i] \rightarrow \mathbb{R}^{n_i}$ and a measurable function $w_i(\cdot): [\tau_0^i, \tau_1^i] \rightarrow \mathbb{R}^{m_i}$ such that

$$\begin{aligned} \dot{y}_i(t) &= \phi_i(t, y_i(t), w_i(t)) & a.e. \ t \in [\tau_0^i, \tau_1^i] \\ w_i(t) &\in U^i & a.e. \ t \in [\tau_0^i, \tau_1^i] \\ y_i(t) &\in X^i & \forall \ t \in [\tau_0^i, \tau_1^i] \end{aligned}$$

(X^i denotes $\{x: (t, x) \in X^i\}$, etc.).

The individual elements are component processes, the $y_i(\cdot)$'s component trajectories, the $w_i(\cdot)$'s component control functions and the intervals $[\tau_0^i, \tau_1^i]$ component time intervals. The processes or state trajectories are *admissible* if the $y_i(\cdot)$'s also satisfy given endpoint constraints.

2.7.2 Optimal multiprocess problem

Let $L_i: \mathbb{R} \times \mathbb{R}^{n_i} \times \mathbb{R}^{m_i} \rightarrow \mathbb{R}$ ($i=1, \dots, k$) and $f: E \rightarrow \mathbb{R}$ be given functions in which $E = \prod_i (\mathbb{R} \times \mathbb{R} \times \mathbb{R}^{n_i} \times \mathbb{R}^{m_i})$ and let $\Lambda \subset \left\{ \{\tau_0^i, \tau_1^i, y_0^i, y_1^i\} \in E \mid \tau_1^i \geq \tau_0^i, \ i=1, \dots, k \right\}$ be a given closed constraint set. The optimal multiprocess problem is:

$$\text{Minimize } f(\{\tau_0^i, \tau_1^i, y_i(\tau_0^i), w_i(\tau_0^i)\}) + \sum_i \int_{\tau_0^i}^{\tau_1^i} L_i(t, y_i(t), w_i(t)) dt$$

over multiprocesses $\{\tau_0^i, \tau_1^i, y_i(\cdot), w_i(\cdot)\}$ which satisfy $\{\tau_0^i, \tau_1^i, y_i(\tau_0^i), y_i(\tau_1^i)\} \in \Lambda$.

2.7.3 Essential Values

For free time dynamic optimization problems, we can expect optimality

conditions to incorporate conditions on the boundary values of functions constructed from the costate variables. When the data are merely measurable in the time variable, the elementary definition of boundary value, namely evaluation at a point, cannot be adopted, basically because it is not robust under the limiting arguments employed in the derivation of optimality conditions. Instead we must interpret the boundary values as ‘essential values’. Similarly when the data are discontinuous in the time variable, the standard condition on the maximized Hamiltonian function is replaced by an inclusion involving the ‘essential values’ of the function.

Let $I \subset \mathbb{R}$ be an open interval, t a point in I , and $g: I \rightarrow \mathbb{R}^k$ an essentially bounded measurable function.

The set of essential values of g at t , denoted $\text{ess}_{\rightarrow t} g(s)$ is defined as follows.

Definition 2.7.2: $\gamma \in \text{ess}_{\rightarrow t} g(s)$ if and only if, for any $\epsilon > 0$, the following set has positive Lebesgue measure:

$$\{s: t - \epsilon < s < t + \epsilon, |\gamma - g(s)| < \epsilon\}.$$

If a point lies in $\text{co } \text{ess}_{\rightarrow t} g(s)$ we say it is a convex essential value of g at t . If the function g has finite left and right limits at t (written $g(t^-)$ and $g(t^+)$) then $\text{ess}_{\rightarrow t} g(s) = \{g(t^-), g(t^+)\}$. It is clearly the case that, if g is continuous at t , then $\text{ess}_{\rightarrow t} g(s) = \{g(t)\}$, i.e. the essential value is merely the value of the function.

2.7.4 A Maximum Principle for Multiprocesses

We now state the multiprocess maximum principle for the finite horizon case (see [C7]). To formulate the hypotheses it is convenient to introduce the augmented velocity functions $\tilde{\phi}_i = [\phi_i, L_i]$, $i=1, \dots, k$.

(H1) For each $x \in \mathbb{R}^{n_i}$, $\tilde{\phi}_i(\cdot, x, \cdot)$ is $\mathcal{L} \times \mathcal{B}$ measurable for $i=1, \dots, k$ where $\mathcal{L}$ denotes the class of Lebesgue subsets in $\mathbb{R}$ and $\mathcal{B}$ denotes the class of Borel subsets in $\mathbb{R}^{m_i}$.

(H2) U^i ($i=1, \dots, k$) are Borel measurable sets.

(H3) There exists a constant K such that for $i=1, \dots, k$

$$|\tilde{\phi}_i(t, y, w)| \leq K \text{ whenever } (t, y, w) \in \mathbb{R} \times X_i^i \times U_i^i \text{ and}$$

(H4) $|\tilde{\phi}_i(t, y, w) - \tilde{\phi}_i(t, y', w)| \leq K |y - y'|$ whenever $(t, y, w), (t, y', w) \in \mathbb{R} \times X_i^i \times U_i^i$.

(H5) f is locally Lipschitz continuous.

Define the Hamiltonian functions H_i , $i=1, \dots, k$ as

$$H_i(t, x, u, p, \lambda) = p \cdot \phi_i(t, x, u) - \lambda L_i(t, x, u).$$

Theorem 2.7.3 (Multiprocess maximum principle): Let $\{\tau_0^i, \tau_1^i, x_i(\cdot), u_i(\cdot)\}$ be an optimal multiprocess. Suppose that $\text{graph } \{x_i(\cdot)\} \subset \text{interior } \{X^i\}$ and that hypotheses (H1) to (H5)

are satisfied. Then there exist a real number λ equal to 0 or 1, real numbers $h_0^i, h_1^i, i=1, \dots, k$, and absolutely continuous functions $p_i: [\tau_0^i, \tau_1^i] \rightarrow \mathbb{R}^{n_i} i=1, \dots, k$ such that $\lambda + \sum_i |p_i(\tau_1^i)| > 0$ and

$$-\dot{p}_i(t) \in \partial_x H_i(t, x_i(t), u_i(t), p_i(t), \lambda) \quad \text{a.e. } t \in [\tau_0^i, \tau_1^i] \quad (2.7.1)$$

$$H_i(t, x_i(t), u_i(t), p_i(t), \lambda) = \max_{w \in U_i^t} H_i(t, x_i(t), w, p_i(t), \lambda) \quad \text{a.e. } t \in [\tau_0^i, \tau_1^i] \quad (2.7.2)$$

$$h_0^i \in \text{co ess}_{t \rightarrow \tau_0^i} \left\{ \sup_{w \in U_i^t} H_i(t, x_i(\tau_0^i), w, p_i(\tau_0^i), \lambda) \right\} \quad (2.7.3)$$

$$h_1^i \in \text{co ess}_{t \rightarrow \tau_1^i} \left\{ \sup_{w \in U_i^t} H_i(t, x_i(\tau_1^i), w, p_i(\tau_1^i), \lambda) \right\} \quad (2.7.4)$$

for $i=1, \dots, k$ and

$$\{-h_0^i, h_1^i, p_i(\tau_0^i), -p_i(\tau_1^i)\} \in N_\Lambda + \lambda \partial f. \quad (2.7.5)$$

The normal cone N_Λ and the generalized gradient ∂f are evaluated at the point $\{\tau_0^i, \tau_1^i, x_i(\tau_0^i), x_i(\tau_1^i)\}$.

Note that all the ingredients of the traditional maximum principle, namely costate functions $p_i(\cdot)$, costate differential equations and maximization of the Hamiltonians, which separate out into statements about individual component processes, are present in the multiprocess maximum principle. The fact that component processes in the optimal multiprocess problem are coupled through the end-point constraint and the cost function gives rise to a corresponding coupling of the component costate functions through their end-points via the multiprocess transversality condition

$$\{-h_0^i, h_1^i, p_i(\tau_0^i), -p_i(\tau_1^i)\} \in N_\Lambda + \lambda \partial f.$$

The formulation of the optimization problem incorporates the constraints $y_i(t) \in X_i^t$ mainly for the purpose of hypothesis refinement. A consequence of the theorem is that the assertions remain valid if $\{\tau_0^i, \tau_1^i, x_i(\cdot), u_i(\cdot)\}$ is merely a local solution to the given problem in the sense that it is minimizing with respect to all multiprocesses $\{\tau_0^i, \tau_1^i, y_i(\cdot), w_i(\cdot)\}$ that satisfy $\text{graph}\{y_i\} \subset \text{graph}\{x_i\} + \epsilon \mathbb{B}$.

2.7.5 Normal Cones

As we have just seen, we require analysis of generalized gradients and normal cones in order to obtain necessary conditions for the solution of multiprocess problems.

In this section we develop formulae for the normal cones of certain sets which arise in optimization, and which will be useful in the study of 'two-stage' problems later on. For simplicity we limit our attention to the convex case, though the following

formulae do have analogues for general closed sets.

Proposition 2.7.4: (i) For closed, convex sets $C_1 \subset \mathbb{R}^k$ and $C_2 \subset \mathbb{R}^l$, and points $a \in C_1$ and $b \in C_2$,

$$N_{C_1 \times C_2}(a, b) = N_{C_1}(a) \times N_{C_2}(b). \quad (2.7.6)$$

(ii) For any closed, convex subset $C \subset \mathbb{R}^k \times \mathbb{R}^l$ and a point $(a, b) \in C$,

$$N_{\{(x, y) | (x, y) \in C\}}((a, a, b)) \subset \{(p, q, r) | (p+q, r) \in N_C((a, b))\}. \quad (2.7.7)$$

Proof: (i) For any $(x, y) \in C_1 \times C_2$ it follows that $x \in C_1$ and $y \in C_2$.

Let $(p, q) \in N_{C_1 \times C_2}((a, b))$, then by the definition of a normal it implies that

$$p \cdot (x-a) + q \cdot (y-b) \leq 0 \text{ for all } (x, y) \in C_1 \times C_2.$$

If we fix $y=b$, then $p \cdot (x-a) \leq 0$ for all $x \in C_1$, i.e. $p \in N_{C_1}(a)$, (c.f. Definition 2.5.4).

Similarly, if we fix $x=a$, then $q \cdot (y-b) \leq 0$ for all $y \in C_2$, i.e. $q \in N_{C_2}(b)$.

Thus, if $(p, q) \in N_{C_1 \times C_2}((a, b))$ then $p \in N_{C_1}(a)$ and $q \in N_{C_2}(b)$.

Conversely, if $p \in N_{C_1}(a)$ and $q \in N_{C_2}(b)$, then we can immediately state that

$$p \cdot (x-a) + q \cdot (y-b) \leq 0 \text{ for all } x \in C_1 \text{ and all } y \in C_2.$$

This implies that $(p, q) \in N_{C_1 \times C_2}((a, b))$.

Hence the equality holds.

(ii) We use the result from part (i) in proving the second part of the proposition.

Let $C = C_1 \times C_2$ where $C_1 \subset \mathbb{R}^k$ and $C_2 \subset \mathbb{R}^l$.

In addition let $\tilde{C} = \tilde{C}_1 \times C_2$ where $\tilde{C}_1 \subset \mathbb{R}^k \times \mathbb{R}^k \times \mathbb{R}^l$.

Let $(p, q) \in N_{\tilde{C}_1}((a, a))$. Then by definition of a normal we have $(p, q) \cdot (x-a, x-a) \leq 0$, for all $(x, x) \in \tilde{C}_1$.

Thus $(p+q) \cdot (x-a) \leq 0$ and so $(p+q) \in N_{C_1}(a)$.

Let $r \in N_{C_2}(b)$ then $r \cdot (y-b) \leq 0$, for all $y \in C_2$.

From Proposition 2.7.4(i) we know that $N_{\tilde{C}} = N_{\tilde{C}_1} \times N_{C_2}$ thus (p, q, r) is a normal to the set $\tilde{C}$ at the point (a, a, b) . Furthermore, since $C = C_1 \times C_2$ it also follows that

$$N_{\{(x, y) \in C\}}((a, b)) = N_{\{x \in C_1\}}(a) \times N_{\{y \in C_2\}}(b).$$

Therefore $(p+q, r) \in N_{\{(x, y) \in C\}}((a, b))$.

Hence we obtain the required result

$$N_{\{(x, y) | (x, y) \in C\}}((a, a, b)) \subset \{(p, q, r) | (p+q, r) \in N_C((a, b))\}. \quad \square$$

Similar arguments to those employed in the above proof give

$$N_{\{(x, y, z) | (x, y, z) \in C\}}((a, a, b, b)) \subset \{(p, q, r, s) | (p+q, r+s) \in N_C((a, b))\}. \quad (2.7.8)$$

AN INFINITE HORIZON, MULTISTAGE OPTIMIZATION PROBLEM

3.1 Introduction

Dynamic optimization problems arise, especially in the economic and biological sciences, which are defined over an infinite time horizon. In the deterministic, continuous time case, the performance index is described by an improper integral which may not necessarily converge; however it is still possible to obtain necessary conditions for optimality in the form of a maximum principle as in Halkin's classic paper in this field [H1]. We now consider whether we may also derive general necessary conditions for the solution of infinite horizon *multistage* optimal control problems.

An optimal multiprocess problem is a dynamic optimization problem involving a collection of control systems coupled through constraints in the endpoints of the state trajectories and through the cost function. The development of a maximum principle for multiprocesses (see Clarke and Vinter [C8]) extends the notion of the classical Pontryagin maximum principle and permits the solution of a wider range of problems.

In line with our theme of extending the applicability of the theory, the main contribution in this chapter is to develop a maximum principle for multiprocesses over an infinite time horizon. We consider the simplest example of a multistage problem, namely a two-stage optimal control problem, in which one of the component processes operates over an infinite time horizon:

$$J = \int_{t_0}^{t_1} L_1(t, x(t), u(t)) dt + \int_{t_1}^{\infty} L_2(t, x(t), u(t)) dt.$$

Extension to other problems exhibiting greater complexity is described in later sections. We reformulate the problem in terms of multiprocesses and then state and prove necessary conditions for a solution to be optimal.

The proof is in two parts:

- 1) We take the infinite horizon problem and restrict the data of the problem to a fixed time interval with a terminal condition on the state. We prove that an optimal solution for the infinite horizon problem is still optimal when the problem is restricted as described.
- 2) We then show that if an optimal solution satisfies the first result we can prove general necessary conditions for optimality over the infinite horizon. This involves

applying the finite horizon multiprocess maximum principle over strictly increasing time intervals and using Ascoli's theorem to construct a nested family of subsequences of the costates with the required convergence properties. We are then able to construct a function which satisfies the necessary conditions of the maximum principle and is thus a candidate for the costate function. Since the proof method can be generalized to problems with more than two stages we have, in effect, a maximum principle for general multistage problems operating over an infinite time interval.

We show that this theory can be applied to two-stage optimal control problems of the form

$$J = \int_{t_0}^{t_1} L_1(t, x(t), u(t), t_1) dt + \int_{t_1}^{\infty} L_2(t, x(t), u(t), t_1) dt,$$

where, in addition to the intermediate time, t_1 , being a choice variable, there is an explicit dependence in the integrands of the cost functions on this intermediate time, (see Tomiyama and Rossana [T2]). We derive a set of necessary conditions for the general statement of the problem using multiprocess theory, after which we consider a problem from the field of mathematical economics. This is an example from the adjustment cost literature on the choice of optimal delivery lags. We refer to Tomiyama and Rossana [T2] whose discussion on this topic is based on a problem first posed by Maccini [M1].

The notation used in this chapter follows that of Section 2.7.

3.2. Statement Of The Problem

We consider the following performance index defined over the infinite horizon:

$$I = \int_{t_0}^{t_1} L_1(t, x(t), u(t)) dt + \int_{t_1}^{\infty} L_2(t, x(t), u(t)) dt \quad (3.2.1)$$

where $L_i: \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ ($i=1,2$) are given functions.

We wish to minimize (3.2.1) subject to the state equations

$$\dot{x}(t) = \begin{cases} f_1(t, x(t), u(t)) & \text{a.e. } t \in [t_0, t_1], \\ f_2(t, x(t), u(t)) & \text{a.e. } t \in [t_1, \infty), \end{cases} \quad (3.2.2)$$

$$\text{with } x(t) \in \begin{cases} X_t^1 & \forall t \in [t_0, t_1], \\ X_t^2 & \forall t \in [t_1, \infty). \end{cases} \quad (3.2.3)$$

We also wish the minimization to be subject to the control constraints

$$u(t) \in \begin{cases} U_t^1 & \text{a.e. } t \in [t_0, t_1], \\ U_t^2 & \text{a.e. } t \in [t_1, \infty), \end{cases} \quad (3.2.4)$$

and boundary conditions

$$x(t_0) \in C_0, (t_1, x(t_1)) \in C_1. \quad (3.2.5)$$

Here the $f_i: \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ ($i=1,2$) are given functions; $U^i \subset \mathbb{R} \times \mathbb{R}^m$, $X^i \subset \mathbb{R} \times \mathbb{R}^n$, $C_0 \subset \mathbb{R}^n$, $C_1 \subset \mathbb{R} \times \mathbb{R}^n$ are given sets, (X_t^i denotes $\{x: (t, x) \in X^i\}$, similarly for U_t^i); the state $x(\cdot)$ is absolutely continuous on $[t_0, \infty)$ and the control $u(\cdot)$ is measurable on $[t_0, \infty)$. The intermediate time t_1 is a choice variable, constrained to satisfy $t_0 \leq t_1 < \infty$. Conditions on f_i and L_i will be stated in Theorem 3.3.1 in the next section. Any triple (x, u, t_1) satisfying (3.2.1) to (3.2.5) is termed admissible. If, in addition, it is minimizing, it is termed optimal.

Notice that this can be set up as an 'infinite horizon' multiprocess problem (c.f. §2.7.2) with component dynamical equations $\dot{x}_1 = f_1$, $\dot{x}_2 = f_2$ (in which the x_i 's are 'segments' of the state trajectory, etc.), cost function

$$\sum_i \int_{\tau_0^i}^{\tau_1^i} L_i(t, x_i(t), u_i(t)) dt$$

and end-point constraint

$$(\tau_0^1, \tau_1^1, \tau_0^2, \tau_1^2, x_1(\tau_0^1), x_1(\tau_1^1), x_2(\tau_0^2), x_2(\tau_1^2)) \in \Lambda$$

on the multiprocesses $\{\tau_0^i, \tau_1^i, x_i(\cdot), u_i(\cdot)\}_{i=1,2}$ considered.

Here the set Λ is

$$\Lambda = \left\{ \left\{ \tau_0^i, \tau_1^i, x_i(\tau_0^i), x_i(\tau_1^i) \right\} : \tau_0^1 = t_0, \tau_1^1 = \tau_0^2 = t_1, \tau_1^2 = \infty, x_1(\tau_0^1) \in C_0, x_1(\tau_1^1) = x_2(\tau_0^2), x_2(\tau_1^2) \text{ free} \right\}. \quad (3.2.6)$$

There is no unique way to define 'optimality' when we are dealing with problems having infinite horizon. Issues are involved here of whether we require a control to give rise to convergence of the improper integral in the cost function and, if not, what is meant by the term 'optimal'. Our ultimate goal is to prove necessary conditions for 'optimal' controls in the form of a Maximum Principle, and it is natural then to adopt as definition of 'optimal control' as weak a notion of optimality as is compatible with the Maximum Principle we have in mind, for then this Maximum Principle will also serve as a necessary condition for other, stronger notions of optimality which other control theorists may care to employ. We shall work with a weak concept of optimality introduced by Halkin [H1], which does not require convergence of the improper integral in the cost function. Comparisons between this and other definitions of optimality relevant to infinite horizon problems are discussed in [S1].

DEFINITION 3.2.1: We say an admissible triple $(x^*(\cdot), u^*(\cdot), t_1^*)$ is 'weakly optimal' if, for every admissible triple (x, u, t_1) , every $T \in J$ ($J = [t_1, \infty)$), and every $\epsilon > 0$, there exists a $\tau \in J$ with $\tau \geq T$ for which

$$\int_{t_0}^{t_1} L_1(t, x(t), u(t)) dt + \int_{t_1}^{\tau} L_2(t, x(t), u(t)) dt \geq \int_{t_0}^{t_1^*} L_1(t, x^*(t), u^*(t)) dt + \int_{t_1^*}^{\tau} L_2(t, x^*(t), u^*(t)) dt - \epsilon. \quad (3.2.7)$$

3.3 An Infinite Horizon Multiprocess Maximum Principle

The main result in this chapter is the following set of necessary conditions on optimal triples for the infinite horizon problem.

Define the Hamiltonian function H_i ($i=1,2$) for each stage by

$$H_i(t, x, u, p, \lambda) = p \cdot f_i(t, x, u) - \lambda L_i(t, x, u)$$

and define the augmented velocity function $\tilde{f}_i$ by $\tilde{f}_i = (f_i, L_i)$ where $\tilde{f}_i: \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^{n+1}$ ($i=1,2$).

THEOREM 3.3.1: Let $(x^*(\cdot), u^*(\cdot), t_1^*)$ be the optimal triple for the infinite horizon problem of §3.2.

Assume that:

(H1) For each $x \in \mathbb{R}^n$, $\tilde{f}_i(\cdot, x, \cdot)$ is $\mathcal{L} \times \mathcal{B}$ measurable for $i=1,2$ where $\mathcal{L}$ denotes the Lebesgue subsets in $\mathbb{R}$ and $\mathcal{B}$ denotes the Borel subsets in $\mathbb{R}^m$.

(H2) U^i ($i=1,2$) are Borel measurable sets.

(H3) There exists a constant K such that for $i=1,2$

$$|\tilde{f}_i(t, x, u)| \leq K \text{ whenever } (t, x, u) \in \mathbb{R} \times X_i^i \times U_i^i \text{ and}$$

(H4) $|\tilde{f}_i(t, x, u) - \tilde{f}_i(t, y, u)| \leq K|x - y|$ whenever $(t, x, u), (t, y, u) \in \mathbb{R} \times X_i^i \times U_i^i$

and that all the hypotheses are satisfied.

Let graph $\{x_i(\cdot)\} \subset \text{interior}\{X^i\}$ for $i=1,2$.

Then there exist real numbers λ (equal 0 or 1), h_0, h_1 and a function $p(\cdot): [0, \infty) \rightarrow \mathbb{R}^n$, which is locally absolutely continuous on $[0, t_1^*)$, (t_1^*, ∞) such that $\lambda + \|p(\cdot)\|_{L^\infty} > 0$ and

$$\left. \begin{aligned} (i) \quad & -\dot{p}(t) \in \partial_x H_1(t, x^*(t), u^*(t), p(t), \lambda) & a.e. \ t \in [t_0, t_1^*] \\ & -\dot{p}(t) \in \partial_x H_2(t, x^*(t), u^*(t), p(t), \lambda) & a.e. \ t \in [t_1^*, \infty) \end{aligned} \right\} (3.3.1)$$

$$\left. \begin{aligned} (ii) \quad & H_1(t, x^*(t), u^*(t), p(t), \lambda) = \max_{w \in U_1^1} H_1(t, x^*(t), w, p(t), \lambda) & a.e. \ t \in [t_0, t_1^*] \\ & H_2(t, x^*(t), u^*(t), p(t), \lambda) = \max_{w \in U_2^2} H_2(t, x^*(t), w, p(t), \lambda) & a.e. \ t \in [t_1^*, \infty) \end{aligned} \right\} (3.3.2)$$

$$(iii) \quad h_0 \in \text{co}_{t \rightarrow t_1^*} \left\{ \sup_{w \in U_t^1} H_1(t, x^*(t_1^*), w, \lim_{t \uparrow t_1^*} p(t), \lambda) \right\} \quad \Bigg\} (3.3.3)$$

$$h_1 \in \text{co}_{t \rightarrow t_1^*} \left\{ \sup_{w \in U_t^2} H_2(t, x^*(t_1^*), w, \lim_{t \downarrow t_1^*} p(t), \lambda) \right\}$$

$$(iv) \quad p(t_0) \in N_{C_0}, (h_0 - h_1, -\lim_{t \uparrow t_1^*} p(t) + \lim_{t \downarrow t_1^*} p(t)) \in N_{C_1} \quad (3.3.4)$$

The normal cone N_{C_0} is evaluated at $x^*(t_0)$ and N_{C_1} is evaluated at $(t_1^*, x^*(t_1^*))$.

Equations (3.3.1) are the costate equations and equations (3.3.2) give the condition for the maximization of the Hamiltonian. The essential values, required for the evaluation of the normal cones, are given by (3.3.3) and the transversality conditions are given by (3.3.4).

3.4. Proof of Theorem 3.3.1

3.4.1 PRELIMINARY RESULTS

Before we proceed with the proof of the main theorem, we recall two results, namely Gronwall's lemma and Ascoli's theorem, which we now state for reference. Proofs of these results may be found in standard texts, e.g. [F1,R3] respectively.

LEMMA 3.4.1 (Gronwall's lemma): *Let $\alpha(\cdot):[0,1] \rightarrow \mathbb{R}$ be a bounded measurable non-negative function, let $y:[0,1] \rightarrow \mathbb{R}$ be a Lipschitz continuous function and let $k(t):[0,1] \rightarrow \mathbb{R}$ be a bounded measurable non-negative function such that*

$$\dot{y}(t) \leq \alpha(t)y(t) + k(t) \quad \text{a.e. } t \in [0,1],$$

then

$$y(t) \leq \left(y(0) + \int_0^t k(\tau) \exp\left\{-\int_0^\tau \alpha(r) dr\right\} d\tau \right) \exp\left\{\int_0^t \alpha(r) dr\right\} \quad \text{all } t \in [0,1].$$

This result provides us with a means of obtaining bounds to solutions of differential equations. The other result is the basic 'sequential compactness' theorem used to establish convergence of subsequences of state trajectories.

THEOREM 3.4.2 (Ascoli's theorem): *Let $\{x_i(\cdot):[0,1] \rightarrow \mathbb{R}^n\}$ be a sequence of functions which are uniformly bounded and uniformly Lipschitz continuous, i.e. there is a constant K such that*

(a) $\|x_i(t)\| \leq K$, all $t \in [0, 1]$, $i=1, 2, \dots$ and

(b) $\|x_i(t) - x_i(t')\| \leq K|t - t'|$, for all $t, t' \in [0, 1]$, $i=1, 2, \dots$

Then there exists a subsequence $\{x_{i_j}(\cdot)\}$ of $\{x_i(\cdot)\}$ such that

$$\max_{t \in [0, 1]} \|x_{i_j}(t) - \bar{x}(t)\| \rightarrow 0, \text{ as } j \rightarrow \infty,$$

for some function $\bar{x}(\cdot)$ which is Lipschitz continuous with Lipschitz rank at most K .

3.4.2 PROOF OF NECESSARY CONDITIONS

We obtain the maximum principle over the infinite horizon by combining two lemmas which we shall state and prove, but first we introduce some notation.

Given $T \in [t_1, \infty)$ and $\xi \in \mathbb{R}^n$, we denote by $P_{T, \xi}$ the finite horizon fixed end-point problem obtained by restricting the interval of integration in the infinite horizon problem to $[0, T]$, and by introducing the terminal condition $x(T) = \xi$, namely

$$\text{Minimize } \int_{t_0}^{t_1} L_1(t, x(t), u(t)) dt + \int_{t_1}^T L_2(t, x(t), u(t)) dt$$

subject to $t_0 \leq t_1 \leq T$,

$$\dot{x}(t) = \begin{cases} f_1(t, x(t), u(t)) & \text{a.e. } t \in [t_0, t_1], \\ f_2(t, x(t), u(t)) & \text{a.e. } t \in [t_1, \infty), \end{cases}$$

$$x(0) \in C_0, (t_1, x(t_1)) \in C_1, x(T) = \xi.$$

'Optimality' for the finite horizon problem $P_{T, \xi}$ is defined in the obvious way. Given an admissible triple (x, u, t_1) for the infinite horizon problem and also $T \in [t_1, \infty)$, we denote by $(x, u, t_1)_T$ a triple in which the first two arguments are the restrictions of the original state trajectory and control functions to $[0, T]$.

The first stage is to prove:

LEMMA 3.4.3: *If $(x^*(\cdot), u^*(\cdot), t_1^*)$ is weakly optimal for the infinite horizon problem then, given any $T > t_1^*$, $(x^*(\cdot), u^*(\cdot), t_1^*)_T$ is optimal for $P_{T, x^*(T)}$.*

Proof: Assume the assertion is false. Then there exist $T \geq t_1^*$, $\epsilon > 0$ and a triple $(\hat{x}, \hat{u}, \hat{t}_1)_T$ for $P_{T, x^*(T)}$, with $t_0 < \hat{t}_1 \leq T$, such that

$$\int_{t_0}^{\hat{t}_1} L_1(t, \hat{x}(t), \hat{u}(t)) dt + \int_{\hat{t}_1}^T L_2(t, \hat{x}(t), \hat{u}(t)) dt < \int_{t_0}^{t_1^*} L_1(t, x^*(t), u^*(t)) dt + \int_{t_1^*}^T L_2(t, x^*(t), u^*(t)) dt - \epsilon \quad (3.4.1)$$

Define an admissible triple (x, u, t_1) for the infinite horizon problem according to $t_1 = \hat{t}_1$ and

$$x(t) = \begin{cases} \hat{x}(t) & \text{a.e. } t \in [t_0, T] \\ x^*(t) & \text{a.e. } t \in [T, \infty) \end{cases}, \quad u(t) = \begin{cases} \hat{u}(t) & \text{a.e. } t \in [t_0, T] \\ u^*(t) & \text{a.e. } t \in [T, \infty) \end{cases}.$$

Then, for all $\tau \geq T$,

$$\begin{aligned} & \int_{t_0}^{\hat{t}_1} L_1(t, x(t), u(t)) dt + \int_{\hat{t}_1}^{\tau} L_2(t, x(t), u(t)) dt \\ &= \int_{t_0}^{\hat{t}_1} L_1(t, \hat{x}(t), \hat{u}(t)) dt + \int_{\hat{t}_1}^T L_2(t, \hat{x}(t), \hat{u}(t)) dt + \int_T^{\tau} L_2(t, x^*(t), u^*(t)) dt \\ &< \int_{t_0}^{t_1^*} L_1(t, x^*(t), u^*(t)) dt + \int_{t_1^*}^{\tau} L_2(t, x^*(t), u^*(t)) dt - \epsilon \end{aligned}$$

by inequality (3.4.1).

This contradicts the weak optimality of $(x^*(\cdot), u^*(\cdot), t_1^*)$. The lemma is proved. $\square$

The above result is clearly a variation on the Bellman optimality principle. In effect, we have shown that if (x^*, u^*, t_1^*) is an optimal solution for the infinite horizon problem then (x^*, u^*, t_1^*) is also optimal for any finite restriction of the problem.

The second stage is to prove:

LEMMA 3.4.4: *Let $(x^*(\cdot), u^*(\cdot), t_1^*)$ be an admissible triple. If, for any $T > t_1^*$, $(x^*(\cdot), u^*(\cdot), t_1^*)_T$ is optimal for $P_{T, x^*(T)}$, then the conclusions of the infinite horizon multiprocess maximum principle (Theorem 3.3.1) hold.*

Proof:

Step 1: Necessary Conditions for a Sequence of Finite Horizon Problems

Let $T^1, T^2, \dots$ be a strictly increasing sequence of real numbers in $[t_1^*, \infty)$ such that $T^j \rightarrow \infty$ as $j \rightarrow \infty$. For each $j \in \{1, 2, \dots\}$ we have that $(x^*(\cdot), u^*(\cdot), t_1^*)_{T^j}$ is optimal for the associated finite horizon fixed end-point problem, $P_{T^j, x^*(T^j)}$.

(A) Necessary conditions for the finite horizon problems.

We now apply necessary conditions of optimality to each of the above finite horizon

problems $P_{T^j, x^*(T^j)}$. From the finite horizon maximum principle for multiprocesses (see §2.7.4) we know that for each j , there exist real numbers λ^j equal 0 or 1, h_0^j , h_1^j , and $p^j: [t_0, T^j] \rightarrow \mathbb{R}^n$, absolutely continuous on $[t_0, t_1^*]$ and $[t_1^*, T^j]$ such that

$$\lambda^j + |p^j(t_0)| + |p^j(t_1^*)| > 0 \text{ and}$$

$$\begin{aligned} \text{(i)} \quad & -\dot{p}^j(t) \in \partial_x H_1(t, x^*(t), u^*(t), p^j(t), \lambda^j) && \text{a.e. } t \in [t_0, t_1^*] \\ & -\dot{p}^j(t) \in \partial_x H_2(t, x^*(t), u^*(t), p^j(t), \lambda^j) && \text{a.e. } t \in [t_1^*, T^j] \end{aligned} \quad \left. \vphantom{\begin{aligned} \text{(i)} \quad} \right\} (3.4.3)$$

$$\begin{aligned} \text{(ii)} \quad & H_1(t, x^*(t), u^*(t), p^j(t), \lambda^j) = \max_{w \in U_1^1} H_1(t, x^*(t), w, p^j(t), \lambda^j) \text{ a.e. } t \in [t_0, t_1^*] \\ & H_2(t, x^*(t), u^*(t), p^j(t), \lambda^j) = \max_{w \in U_2^2} H_2(t, x^*(t), w, p^j(t), \lambda^j) \text{ a.e. } t \in [t_1^*, T^j] \end{aligned} \quad \left. \vphantom{\begin{aligned} \text{(ii)} \quad} \right\} (3.4.4)$$

$$\text{(iii)} \quad h_0^j \in \text{co} \text{ess}_{t \rightarrow t_1^*} \left\{ \sup_{w \in U_1^1} H_1(t, x^*(t_1^*), w, \lim_{t \uparrow t_1^*} p^j(t), \lambda^j) \right\} \quad \left. \vphantom{\begin{aligned} \text{(iii)} \quad} \right\} (3.4.5)$$

$$h_1^j \in \text{co} \text{ess}_{t \rightarrow t_1^*} \left\{ \sup_{w \in U_2^2} H_2(t, x^*(t_1^*), w, \lim_{t \downarrow t_1^*} p^j(t), \lambda^j) \right\}$$

$$\text{(iv)} \quad p^j(t_0) \in N_{C_0}, (h_0^j - h_1^j, -\lim_{t \uparrow t_1^*} p^j(t) + \lim_{t \downarrow t_1^*} p^j(t)) \in N_{C_1}. \quad (3.4.6)$$

The normal cone N_{C_0} is evaluated at $x^*(t_0)$, N_{C_1} is evaluated at $(t_1^*, x^*(t_1^*))$.

Conditions (3.4.3) to (3.4.5) follow directly from the multiprocess maximum principle.

Notice that by asserting the p^j 's are absolutely continuous on the *closed* interval $[t_1^*, T^j]$ we have effectively taken the right-continuous version of the concatenation of the 'component' costate functions obtained from the multiprocess maximum principle.

(B) Computation of Normal Cones.

From this point in the proof, we employ the notation $r(t^-)$, $r(t^+)$ to denote left and right limits of the function $r(\cdot)$ defined on an interval containing the point t .

We may write the multiprocess transversality condition as

$$\left\{ (h_0^j, p^j(t_0), -p^j(t_1^{*-})), (-h_1^j, p^j(t_1^{*+}), -p^j(T^j)) \right\} \in N_{\tilde{\Lambda}} \left((t_1^*, x^*(t_0), x^*(t_1^*)), (t_1^*, x^*(t_1^*), x^*(T^j)) \right)$$

$$\text{where } \tilde{\Lambda} = \left\{ ((t, y, z), (t', y', z')) \in \mathbb{R}^{2(1+n+n)} : y \in C_0, t = t', z = y', z' = x^*(T^j) \right\}.$$

In view of Proposition 2.7.4(i)

$$p^j(t_0) \in N_{C_0}, \quad (3.4.7)$$

and

$$(h_0^j, -p^j(t_1^{*-}), -h_1^j, p^j(t_1^{*+})) \in N_D(t_1^*, x^*(t_1^*), t_1^*, x^*(t_1^*))$$

$$\text{where } D := \left\{ (t, y, t', y') : (t, y) = (t', y'), (t, y) \in C_1 \right\}.$$

By Proposition 2.7.4(ii),

$$(h_0^j, -p^j(t_1^{*-}), -h_1^j, p^j(t_1^{*+})) \in \left\{ (\alpha, \gamma, \alpha', \gamma') : (\alpha + \alpha', \gamma + \gamma') \in N_{C_1}(t_1^*, x^*(t_1^*)) \right\}.$$

It follows that

$$\left(h_0^j - h_1^j, -p^j(t_1^{*-}) + p^j(t_1^{*+}) \right) \in N_{C_1}(t_1^*, x^*(t_1^*)). \quad (3.4.8)$$

(3.4.7) and (3.4.8) are the required transversality conditions satisfied along the sequence of 'finite horizon' problems we consider.

Step 2: Conclusion - Infinite Horizon Necessary Conditions

It remains to derive the necessary conditions of optimality for the infinite horizon problem, by passage to the limit.

Note that the condition

$$\lambda^j + |p^j(t_0)| + |p^j(t_1^*)| > 0$$

can be normalized to

$$|\lambda^j| + |p^j(t_0)| + |p^j(t_1^*)| = 1,$$

by scaling the multipliers.

(A) Asymptotic Normalization

For each j , we have $\lambda^j = 0$ or 1 . By extracting a subsequence, we can arrange that the $\{\lambda^j\}$ converge to a limit. We call this limit λ . Clearly $\lambda = 0$ or 1 .

By further extraction of subsequences, we can also arrange that $p^j(t_0) \rightarrow \bar{p}$ and that $p^j(t_1^{*-}) \rightarrow \bar{p}$, for points $\bar{p}$ and $\bar{p}$ satisfying

$$\lambda + |\bar{p}| + |\bar{p}| = 1.$$

(B) Construction of costate function $q(\cdot)$

We now consider the costate functions and construct a costate which satisfies the

necessary conditions over the infinite horizon.

(1) We show that, for each k , the sequence of functions $p^j(\cdot)$, $j=k, k+1, \dots$ are uniformly bounded and uniformly Lipschitz continuous on $[t_0, t_1^*)$ and the fixed subinterval $[t_1^*, T^k]$. From the maximum principle we deduce existence of bounded, measurable functions such that

$$-\dot{p}^j(t) = p^j(t) \cdot \gamma^j(t) - \lambda \mu^j(t) \text{ a.e. } t \in [t_1^*, T^k]$$

Thus for $j=k, k+1, \dots$

$$|\dot{p}^j(t)| \leq |p^j(t)| |\gamma^j(t)| + \lambda |\mu^j(t)|.$$

Hence, via Gronwall's lemma, $|p^j(t)| \leq p^j(t_1^*) D_1 + D_2$ for all $t \in [t_1^*, T^k]$, for some constants D_1, D_2 . By the normalization condition, $|p^j(t_1^*)| \leq 1$ for all j . Hence $|p^j(t)| \leq D$, for all $t \in [t_1^*, T^k]$, for some constant D .

Thus the functions $p^j(\cdot)$ are uniformly bounded over $[t_1^*, T^k]$. We deduce from the differential equations for the $p^j(\cdot)$'s that they are also uniformly Lipschitz continuous on $[t_1^*, T^k]$. Similar arguments are valid over the first time interval $[t_0, t_1^*)$.

(2) Ascoli's theorem now permits us to construct a nested family of subsequences of $p^j(t)$ over $[t_1^*, \infty)$ as follows:

Consider the interval $[t_1^*, T^1]$. We know the functions $p^j(\cdot)$, $j=1, 2, \dots$ are uniformly bounded and uniformly Lipschitz continuous on the subinterval $[t_1^*, T^1]$. Therefore, by application of Ascoli's theorem, we can choose a subsequence of functions from the sequence $p^j(t)$, $j=1, 2, \dots$, which we relabel as $p^{1,n}(t)$, such that

$$\max_{t \in [t_1^*, T^1]} \|p^{1,n}(t) - q^1(t)\| \rightarrow 0 \text{ as } n \rightarrow \infty \text{ for some Lipschitz continuous function } q^1(t).$$

Now we consider the interval $[t_1^*, T^2]$. We know that over this interval the $p^j(\cdot)$ $j=2, 3, \dots$ are uniformly bounded and uniformly Lipschitz continuous. Ascoli's theorem tells us that we can choose a subsequence of $p^{1,n}(t)$, which we relabel $p^{2,n}(t)$, whose elements are selected from $p^2(\cdot), p^3(\cdot), \dots$ such that the subsequence $p^{2,n}(t)$ converges uniformly over $[t_1^*, T^2]$, namely,

$$\max_{t \in [t_1^*, T^2]} \|p^{2,n}(t) - q^2(t)\| \rightarrow 0 \text{ as } n \rightarrow \infty \text{ for some Lipschitz continuous function } q^2(t).$$

Repeated application of this process results in a *nested* family of subsequences of $p^j(t)$, namely,

$$(p^{1,n}(\cdot))_{n=1}^{\infty} \supset (p^{2,n}(\cdot))_{n=1}^{\infty} \supset \dots$$

with the properties:

- i) for fixed k , terms in the sequence $(p^{k,n}(\cdot))_{n=1}^{\infty}$ are selected from $(p^j(\cdot))_{j=k}^{\infty}$ where $p^j(\cdot)$ is the original sequence, and
- ii) for each k , $p^{k,n}(t) \rightarrow q^k(t)$ uniformly as $n \rightarrow \infty$ over $[t_1^*, T^k]$ where $q^k(t)$ is Lipschitz continuous.

Furthermore, since the sequences are nested $q^k(\cdot)$ and $q^j(\cdot)$ coincide on $[t_1^*, T^k]$ for $j \geq k$. We can similarly show that, since the $p^j(\cdot)$, $j=1,2,\dots$ are uniformly bounded and uniformly Lipschitz continuous over $[t_0, t_1^*)$, a subsequence can be chosen which converges to some Lipschitz continuous function $q^0(t)$.

(3) Now define the function $q(\cdot): [t_0, t_1^*) \cup [t_1^*, \infty) \rightarrow \mathbb{R}^n$ as follows:

$$q(t) = \begin{cases} q^0(t) & t \in [t_0, t_1^*) \\ q^k(t) & t \in [t_1^*, \infty) \text{ where } k \text{ is such that } T^k > t. \end{cases}$$

Evidently, the definition of $q(\cdot)$ on $[t_0, \infty)$ is independent of k . Take any $t > t_1^*$, then t lies in some interval $[t_1^*, T^k]$ where k is the minimum value such that $T^k > t$ and, as we have just shown,

$$q^k(t) = \lim_{n \rightarrow \infty} p^{k,n}(t) \text{ over } [t_1^*, T^k].$$

We know by construction of the subsequences, that each $p^{k,n}(t)$ ($n=1,2,\dots$) satisfies the necessary conditions (3.4.3) to (3.4.6) over $[t_1^*, T^k]$.

We now proceed to show that $q^0(t)$ and q^k satisfy the necessary conditions.

(i) Consider $-p^{k,n}(t) = p^{k,n}(t) \cdot \bar{f}_x(t, x(t), u(t)) - \lambda \bar{L}_x(t, x(t), u(t))$, for some $\bar{f}_x(\cdot)$, $\bar{L}_x(\cdot)$.

Since each $p^{k,n}$ is absolutely continuous we can integrate both sides to give

$$-p^{k,n}(t) = p^{k,n}(t_1^*) + \int_{t_1^*}^t p^{k,n}(s) \cdot \bar{f}_x(s, x(s), u(s)) ds - \lambda \int_{t_1^*}^t \bar{L}_x(s, x(s), u(s)) ds.$$

By the uniform convergence of $p^{k,n}$ over $[t_1^*, T^k]$ we can take limits of both sides, to give

$$-q^k(t) = q^k(t_1^*) + \lim_{n \rightarrow \infty} \int_{t_1^*}^t p^{k,n}(s) \cdot \bar{f}_x(s, x(s), u(s)) ds - \lambda \int_{t_1^*}^t \bar{L}_x(s, x(s), u(s)) ds.$$

We know that $\bar{f}_x$ is integrable over $[t_1^*, T^k]$, $\bar{f}_x \leq \alpha(t) \in L^1[t_1^*, T^k]$ and the $p^{k,n}$ are uniformly bounded over $[t_1^*, T^k]$. Let $\tilde{p}^{k,n}$ be the uniform bound for the sequence $p^{k,n}$ over $[t_1^*, T^k]$, then

$$p^{k,n}(s) \cdot \bar{f}_x(s, x(s), u(s)) < |\alpha(s)| \tilde{p}^{k,n} \text{ for all } p^{k,n} \text{ over } [t_1^*, T^k].$$

Furthermore,

$$p^{k,n}(s) \cdot \bar{f}_x(s, x(s), u(s)) \rightarrow q^k(s) \cdot \bar{f}_x(s, x(s), u(s)).$$

Hence,

$$\lim_{n \rightarrow \infty} \int_{t_1^*}^t p^{k,n}(s) \cdot \bar{f}_x(s, x(s), u(s)) ds = \int_{t_1^*}^t q^k(s) \cdot \bar{f}_x(s, x(s), u(s)) ds$$

by the Lebesgue Convergence Theorem, (see eg. [R3]).

Thus for each k , $q^k(\cdot)$ satisfies condition (3.4.3), similarly for $q^0(\cdot)$. Therefore, by construction of $q(\cdot)$ it satisfies the costate equation over $[t_0, \infty)$.

(ii) We have, for each $p^{k,n}(t)$, $n=1,2,\dots$ and a given $w \in U^2$, that

$$H(t, x^*(t), u^*(t), p^{k,n}(t), \lambda) \geq H(t, x^*(t), w, p^{k,n}(t), \lambda) \text{ a.e. } t \in [t_1^*, T^k].$$

Since the inequality holds along the sequence then it holds for the limiting value of $p^{k,n}$ since H is continuous.

Therefore $H(t, x^*(t), u^*(t), q^k(t), \lambda) \geq H(t, x^*(t), w, q^k(t), \lambda)$.

Since this holds for any $w \in U^2$ it holds as a maximum over all $w \in U^2$, i.e.

$$H(t, x^*(t), u^*(t), q^k(t), \lambda) = \max_{w \in U^2} H(t, x^*(t), w, q^k(t), \lambda).$$

This holds for each k and similarly holds for q^0 , therefore $q(\cdot)$ satisfies the maximization of the Hamiltonian condition (3.4.4) over $[t_0, \infty)$.

(iii) We have that for any fixed k , (3.4.5) holds when $p^{k,n}(\cdot)$ replaces p^j for $n=1,2,\dots$. Passing to the limit (this is justified by the upper semi-continuity of the convex essential value [L1, Lemma 1.2]) we see that (3.4.5) remains valid when q replaces p^j .

(iv) We know that for $j=1,2,\dots$ $p^j(\cdot)$ satisfies $p^j(t_0) \in N_{C_0}(x^*(t_0))$. In view of the upper semi-continuity of the normal cone to a convex set, passage to the limit is valid, thus $q(t_0) \in N_{C_0}$.

Similarly, we can deduce that $(h_0 - h_1, -q(t_1^{*-}) + q(t_1^{*+})) \in N_{C_1}$.

(v) It is evident too, that $q(t_0) = \bar{p}$ and $q(t_1^{*+}) = \bar{\bar{p}}$ (see “(A)” of step 2). So

$$\lambda + |\bar{p}| + |\bar{\bar{p}}| = 1.$$

Thus we have constructed a costate $q(\cdot)$ that satisfies the necessary conditions over the infinite horizon. $\square$

Remarks:

1) We have obtained necessary conditions, in the form of a Multiprocess Maximum Principle, for (x^*, u^*, t_1^*) to be a solution to the infinite horizon problem using the concept of weak optimality. This definition is the weakest of the various definitions in

the literature. It therefore follows that a solution to the infinite horizon problem using any of the (stronger) notions of optimality will also satisfy the Multiprocess Maximum Principle (see [S1]).

2) We emphasize that the formulation of the problem makes no statement concerning the limiting values of either $x^*(T)$ or of $\int_{t_1}^T L_2(t, x^*(t), u^*(t)) dt$ as $T \rightarrow \infty$.

3) The theory can obviously be extended to cover any performance index of the form

$$\int_{t_0}^{t_1} L_1(t, x(t), u(t)) dt + \int_{t_1}^{t_2} L_2(t, x(t), u(t)) dt + \dots + \int_{t_n}^{\infty} L_{n+1}(t, x(t), u(t)) dt.$$

where transversality conditions for the intermediate times $t_1, \dots, t_n$ can be obtained by defining constraint sets $C_1, \dots, C_{n+1}$ in a similar fashion to the C_1 of the theorem.

4) The proof techniques used above can be applied to more general multiprocess problems of the form

$$\int_{\tau_0^1}^{\tau_1^1} L_1(t, x_1(t), u_1(t)) dt + \int_{\tau_0^2}^{\tau_1^2} L_2(t, x_2(t), u_2(t)) dt + \dots + \int_{\tau_0^k}^{\infty} L_k(t, x_k(t), u_k(t)) dt$$

in which we do not assume coincidence of the interval end points.

5) Research has been aimed at formulating transversality conditions for infinite horizon problems. We briefly mention some results.

Weitzman [W1] considered infinite horizon convex programming models with a free end-point. He obtained a set of necessary and sufficient conditions for optimality, one of which is a transversality condition in the limit at infinity, namely

$$\lim_{t \rightarrow \infty} p_t \cdot x_t^* = 0,$$

where x_t^* is the optimal trajectory from instant $t-1$ to t and p_t is a price row vector. This can be interpreted as saying that in an optimal program, $\{x_t^*, u_t^*\}$, where u_t^* is the optimal gain (utility) from $t-1$ to t , the present value of unused capital must eventually go to zero. However, setting to zero the limit value of the optimal capital stocks, although often economically intuitive, requires special assumptions, in particular a non-empty interior condition and moreover, Weitzman's limit property is not satisfied by optimal growth paths in the undiscounted growth model.

In the analogous continuous time problem, many authors have assumed that

$$\lim_{t \rightarrow \infty} p(t) \cdot x(t) = 0,$$

the main justification for its use being that it looks like the 'natural generalization' of

the finite horizon transversality condition of Pontryagin et al. [P1].

Benveniste and Scheinkman [B2] study concave, variational calculus type models of the form

$$\text{maximize } \int_0^{\infty} L(t, x, \dot{x}) dt$$

subject to $(x, \dot{x}) \in T$, $x(0) = x_0$, where T is a convex subset of $\mathbb{R}^{2n}$ and L is a continuous function which is concave in $(x, \dot{x})$ for every $t \geq 0$. These are interpreted as dynamic models of capital accumulation. The condition $\lim_{t \rightarrow \infty} p(t) \cdot x(t) = 0$ has long been recognised as sufficient for optimality under certain assumptions, but Benveniste and Scheinkman [B2] prove the necessity of this condition under a set of similar assumptions. They interpret $p(t) \cdot x(t)$ as a linear approximation to the present value of $x(t)$ and hence, as before, the transversality condition means that the present value of leftover capital is zero. However, they acknowledge that in the continuous time analogue their results are not as general as Weitzman's.

An extension of this work by Araujo and Scheinkman [A1] proves similar results while allowing state constraints to be binding so that a wider class of economic problems may be addressed by the theory.

A study by Aubin and Clarke [A3] of the following infinite horizon problem:

$$\text{minimize } \int_0^{\infty} e^{-\delta t} g(x(t), u(t)) dt$$

in which the cost functional is locally Lipschitz and the dynamics linear and/or convex indicates that there is a sensitive relationship between the growth rates of the cost functional and the costate variables. They obtain the transversality conditions

$$\int_0^{\infty} e^{(q-1)\delta t} |p(t)|^q dt < \infty, \quad \int_0^{\infty} e^{(q-1)\delta t} |\dot{p}(t)|^q dt < \infty, \quad (3.4.9)$$

where $q \in \mathbb{R}$ is suitably defined, and

$$\lim_{t \rightarrow \infty} e^{(q-1)\delta t} |p(t)|^q = 0. \quad (3.4.10)$$

Based on the results of their paper, the necessary conditions may be expected to hold in the strong form (i.e. 'normally', without the presence of a possibly vanishing multiplier) provided the 'discount rate' δ is sufficiently large. Furthermore, in the infinite horizon case, growth (or dual) conditions such as (3.4.9) are more natural than pointwise transversality conditions such as (3.4.10) (which are simply consequences of (3.4.9)). The role of the infinite horizon and the discount factor are discussed in detail in T. C. Koopmans [K1].

Gray and Salant, in a discussion paper [G1], develop transversality conditions for optimal programs in discrete time infinite horizon problems, firstly in which the value of the control variable may be restricted but the value of the state is not and secondly where the state variable, rather than the control, is restricted. They show that optimal solutions to the Halkin counterexample, which belongs to the former class of problems, do indeed satisfy this new condition, namely

$$\lim_{t' \rightarrow \infty} \beta^{t'} p(t') [f_{u(t^*)} f_{x(t^*+1)} f_{x(t^*+2)} \dots f_{x(t')}] = 0. \quad (3.4.11)$$

This condition involves the product of the limiting extension of the standard condition

$$\lim_{t' \rightarrow \infty} \beta^{t'} p(t') = 0$$

where β is a discount factor ($\beta=1$ in the Halkin example) and the bracketed terms that represent the cumulative change in the state variable generated by 'unreversed' perturbations. A perturbation is a variation in a proposed program. In an 'unreversed' perturbation the value of the control is not altered after some t^* (such that $u(t^*)$ is interior to the control set) in order to restore the state variable to its original path. The new necessary condition then states that either equation (3.4.11) is satisfied; or the value function is not defined at $u(t^*)$, or the series of derivatives describing the costs associated with an unreversed perturbation over the infinite horizon is not uniformly convergent when expressed as a function of $u(t^*)$.

Michel [M3] has shown that for a general discounted problem it is possible to prove another property, namely, the maximized Hamiltonian converges to zero as time goes to infinity. (This result may be obtained with the Hamilton-Jacobi equation but then requires differentiability of the value function – see Section 1.6 for definitions of these terms). He also shows that in certain circumstances this implies the transversality condition. However, the problem formulation only deals with smooth data.

Currently, Michel [M4] has attempted to fill a gap in the literature by establishing the transversality condition in a general form applying to concave discrete time infinite horizon problems. The transversality condition can be interpreted as an exhaustion property of the limit value of the optimal state variables and is obtained in both strong ($\liminf_{i \rightarrow \infty} p_i \cdot (x_i - x_i^*) \geq 0$) and weak ($\limsup_{i \rightarrow \infty} p_i \cdot (x_i - x_i^*) \geq 0$) forms, corresponding to the definitions of optimality used. The work is an attempt to remove the need for a special argument in each case but further studies are required for other general classes of problems.

3.5 An Application of the Infinite Horizon Theory

In this section we consider an example of the application of the infinite horizon multiprocess maximum principle, namely, a two-stage optimal control problem where the switch-point is a choice variable and appears as an argument in the integrands of the performance index. Previously, necessary conditions for problems of this type have been obtained by Tomiyama and Rossana [T2] over the finite horizon. Using a multiprocess formulation we show that general necessary conditions can be obtained over the infinite horizon under relaxed hypotheses.

The problem considered by Tomiyama and Rossana [T2] is to find a triple (x^*, u^*, t_1^*) that maximizes the performance index

$$J = \int_{t_0}^{t_1} L_1(t, x(t), u(t), t_1) dt + \int_{t_1}^{\infty} L_2(t, x(t), u(t), t_1) dt$$

subject to the state equations

$$\dot{x}(t) = \begin{cases} f_1(t, x(t), u(t), t_1) & \text{a.e. } t \in [t_0, t_1] \\ f_2(t, x(t), u(t), t_1) & \text{a.e. } t \in [t_1, \infty), \end{cases} \quad (3.5.1)$$

the boundary condition

$$x(t_0) \in X_0, \quad (3.5.2)$$

and the control constraint

$$u(t) \in \begin{cases} U_1 \subset \mathbb{R}^m & \text{a.e. } t \in [t_0, t_1] \\ U_2 \subset \mathbb{R}^m & \text{a.e. } t \in [t_1, \infty). \end{cases} \quad (3.5.3)$$

In the original formulation, the functions f_i , L_i ($i=1,2$) are at least continuously differentiable in x , t and t_1 and continuous in u ; X_0 is a non-empty closed subset of $\mathbb{R}^n$ and U_i ($i=1,2$) are closed subsets of $\mathbb{R}^m$.

A triple which achieves the minimum of the performance index over triples satisfying the constraints of this problem is said to be optimal. Note that t_1 is a choice variable to be chosen optimally and that the asterisk denotes optimal quantities. It is assumed that an optimal control triple (x^*, u^*, t_1^*) exists where $t_0 \leq t_1^* < \infty$.

We remove explicit t_1 dependence by state augmentation. An equivalent formulation is as follows:

$$\text{Minimize } \int_{t_0}^{t_1} \tilde{L}_1(t, x_1(t), u_1(t), x_3(t)) dt + \int_{t_1}^{\infty} \tilde{L}_2(t, x_2(t), u_2(t), x_4(t)) dt \quad (3.5.4)$$

subject to

$$\begin{aligned}
\dot{x}_1(t) &= f_1(t, x_1(t), u_1(t), x_3(t)) & \text{a.e. } t \in [t_0, t_1] \\
\dot{x}_2(t) &= f_2(t, x_2(t), u_2(t), x_4(t)) & \text{a.e. } t \in [t_1, \infty) \\
\dot{x}_3(t) &= 0 & t \in [t_0, t_1] \text{ where } x_3(t_1) = t_1 \\
\dot{x}_4(t) &= 0 & t \in [t_1, \infty) \text{ where } x_4(t_1) = t_1
\end{aligned} \tag{3.5.5}$$

$$\begin{aligned}
&\text{and } u_1(t) \in U_1^1 \text{ a.e. } t \in [t_0, t_1], \quad x_1(t) \in X_1^1 \text{ all } t \in [t_0, t_1], \\
&\quad u_2(t) \in U_2^2 \text{ a.e. } t \in [t_1, \infty), \quad x_2(t) \in X_2^2 \text{ all } t \in [t_1, \infty),
\end{aligned} \tag{3.5.6}$$

$$\text{where } x_1(t_0) \in C_0, (t_1, x_1(t_1)) \in C_1. \tag{3.5.7}$$

The minimization is conducted over multiprocesses

$\left\{ (t_0, t_1, x_1(\cdot), x_3(\cdot), u_1(\cdot)), (t_1', t_2, x_2(\cdot), x_4(\cdot), u_2(\cdot)) \right\}$
subject to the constraint set Λ , defined by:

$$\begin{aligned}
\Lambda = & \left\{ (t_0, t_1, x_1(t_0), x_1(t_1), x_3(t_0), x_3(t_1)), (t_1', t_2, x_2(t_1'), x_2(t_2), x_4(t_1'), x_4(t_2)) : t_1 = t_1', t_2 = \infty, \right. \\
& \left. x_1(t_0) \in C_0, x_1(t_1) = x_2(t_1'), x_3(t_0) \text{ free}, x_3(t_1) = t_1 = x_4(t_1'), x_4(t_2) \text{ free} \right\}.
\end{aligned} \tag{3.5.8}$$

Here $\tilde{L}_i: \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \rightarrow \mathbb{R}$ and $f_i: \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \rightarrow \mathbb{R}^n$ ($i=1,2$) are given functions ($\tilde{L}_i = -L_i$) and $U^i \subset \mathbb{R} \times \mathbb{R}^m$, $X^i \subset \mathbb{R} \times \mathbb{R}^n$ ($i=1,2$) are given sets.

$C_0 \subset \mathbb{R}^n$, $C_1 \subset \mathbb{R} \times \mathbb{R}^n$ are given closed endpoint constraint sets.

The intermediate time t_1 is a choice variable, constrained to satisfy $t_0 \leq t_1 < \infty$ where t_0 is fixed. Note that the state is continuous at the switch-point i.e. $x_1(t_1) = x_2(t_1)$.

It is important to realise, however, that by reformulating the two-stage problem as a multiprocess problem we can relax the smoothness conditions imposed on the functions to ones of Lipschitz continuity. The following theorem provides necessary conditions of optimality over the infinite horizon for this particular problem.

Define the Hamiltonian function, H , for each stage by

$$H_i(t, x, u, p, t_1) = p_i \cdot f_i(t, x, u, t_1) + p_{i+2} \cdot f_{i+2}(t, x, u, t_1) + \lambda L_i(t, x, u, t_1) \quad i=1,2$$

where $p(t)$ is the costate.

(Note that this reduces to $H_i = p_i \cdot f_i + \lambda L_i$, $i=1,2$, since $f_3 = f_4 \equiv 0$).

Define the augmented velocity function $\tilde{f}_i = (f_i, L_i)$ $i=1,2$ where $\tilde{f}_i: \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^{n+1}$.

THEOREM 3.5.1: *Let $((x_1^*, u_1^*, x_3^*), (x_2^*, u_2^*, x_4^*))$ solve the reformulated problem.*

Assume that:

(H1) For each $(x, t_1) \in \mathbb{R}^n \times \mathbb{R}$, $\tilde{f}_i(\cdot, x, \cdot, t_1)$ is $\mathcal{L} \times \mathcal{B}$ measurable for $i=1,2$ where $\mathcal{L}$ denotes the Lebesgue subsets in $\mathbb{R}$ and $\mathcal{B}$ denotes the Borel subsets in $\mathbb{R}^m$.

(H2) U^i ($i=1,2$) are Borel measurable sets.

(H3) There exists a constant K such that for $i=1,2$

$$|\tilde{f}_i(t, x, u, t_1)| \leq K \text{ whenever } (t, x, u, t_1) \in \mathbb{R} \times X_i^1 \times U_i^1 \times \mathbb{R} \text{ and}$$

(H4) $|\tilde{f}_i(t, x, u, t_1) - \tilde{f}_i(t, y, u, t_1)| \leq K|\bar{x} - \bar{y}|$ (where $\bar{x}=(x, t_1)$, $\bar{y}=(y, t_1)$) whenever

$$(t, x, u, t_1), (t, y, u, t_1) \in \mathbb{R} \times X_i^1 \times U_i^1 \times \mathbb{R}$$

and that all the hypotheses are satisfied. Let graph $\{x_i(\cdot)\} \subset \text{interior}\{X^i\}$ for $i=1,2$. Then there exist real numbers λ (equal 0 or 1), h_1 , h_2 and absolutely continuous functions $p_1(\cdot): [t_0, t_1^*] \rightarrow \mathbb{R}^n$, $p_2(\cdot): [t_1^*, \infty) \rightarrow \mathbb{R}^n$, $p_3(\cdot): [t_0, t_1] \rightarrow \mathbb{R}^n$, $p_4(\cdot): [t_1^*, \infty) \rightarrow \mathbb{R}^n$, such that

$$\lambda + |p_1(t_0)| + |p_2(t_1^*)| + |p_3(t_0)| + |p_4(t_1^*)| > 0 \text{ and}$$

$$\begin{aligned} (i) \quad & -\dot{p}_1(t) \in \partial_{x_1} H_1(t, x_1^*(t), u_1^*(t), p_1(t), x_3^*(t)) \quad a.e. \ t \in [t_0, t_1^*] \\ & -\dot{p}_2(t) \in \partial_{x_2} H_2(t, x_2^*(t), u_2^*(t), p_2(t), x_4^*(t)) \quad a.e. \ t \in [t_1^*, \infty) \\ & -\dot{p}_3(t) \in \partial_{x_3} H_1(t, x_1^*(t), u_1^*(t), p_1(t), x_3^*(t)) \quad a.e. \ t \in [t_0, t_1^*] \\ & -\dot{p}_4(t) \in \partial_{x_4} H_2(t, x_2^*(t), u_2^*(t), p_2(t), x_4^*(t)) \quad a.e. \ t \in [t_1^*, \infty) \end{aligned} \quad \left. \vphantom{\begin{aligned} (i) \quad} \right\} (3.5.9)$$

$$\begin{aligned} (ii) \quad & H_1(t, x_1^*(t), u_1^*(t), p_1(t), x_3^*(t)) = \max_{w \in U_1^1} H_1(t, x_1^*(t), w, p_1(t), x_3^*(t)) \quad a.e. \ t \in [t_0, t_1^*] \\ & H_2(t, x_2^*(t), u_2^*(t), p_2(t), x_4^*(t)) = \max_{w \in U_2^2} H_2(t, x_2^*(t), w, p_2(t), x_4^*(t)) \quad a.e. \ t \in [t_1^*, \infty) \end{aligned} \quad \left. \vphantom{\begin{aligned} (ii) \quad} \right\} (3.5.10)$$

$$\begin{aligned} (iii) \quad & h_1 \in \text{co ess}_{t \rightarrow t_1^*} \left\{ \sup_{w \in U_1^1} H_1(t, x_1^*(t_1^*), w, p_1(t_1^*), x_3^*(t_1^*)) \right\} \\ & h_2 \in \text{co ess}_{t \rightarrow t_1^*} \left\{ \sup_{w \in U_2^2} H_2(t, x_2^*(t_1^*), w, p_2(t_1^*), x_4^*(t_1^*)) \right\} \end{aligned} \quad \left. \vphantom{\begin{aligned} (iii) \quad} \right\} (3.5.11)$$

$$(iv) \quad p_1(t_0) \in N_{C_0}, \quad (3.5.12)$$

$$(h_1 - h_2 + p_3(t_1^*) + p_4(t_1^*), -p_1(t_1^*) + p_2(t_1^*)) \in N_{C_1}. \quad (3.5.13)$$

The normal cone N_{C_0} is evaluated at $(x_1^*(t_0))$, N_{C_1} is evaluated at $(t_1^*, x_1^*(t_1^*))$.

Remarks:

1) Note that $((x_1^*, u_1^*, x_3^*), (x_2^*, u_2^*, x_4^*))$ is the solution to the given problem where $x_1^* = x^*|_{[t_0, t_1^*]}$, $u_1^* = u^*|_{[t_0, t_1^*]}$, $x_3^* = t_1^*|_{[t_0, t_1^*]}$ and $x_2^* = x^*|_{[t_1^*, \infty)}$, $u_2^* = u^*|_{[t_1^*, \infty)}$, $x_4^* = t_1^*|_{[t_1^*, \infty)}$.

2) Note that if we maximize the performance index directly as in [T2], the usual maximization condition of the Hamiltonian is replaced by the minimization condition, and the infimum of H_i replaces the supremum in the essential value inclusions.

Proof: Conditions (3.5.9) to (3.5.12) follow from a direct application of our infinite horizon multiprocess maximum principle (Theorem 3.3.1) to this particular two-stage problem.

We derive the conditions for a normal cone at C_1 . This follows closely from Step 1(B) in the proof of Lemma 3.4.4.

We may write the multiprocess transversality condition as

$$\left\{ (h_1, -p_1(t_1^*), p_3(t_1^*), -h_2, p_2(t_1^*), p_4(t_1^*)) \right\} \in N_{\tilde{\Lambda}} \left(t_1^*, x_1^*(t_1^*), x_3^*(t_1^*), t_1^*, x_2^*(t_1^*), x_4^*(t_1^*) \right)$$

$$\text{where } \tilde{\Lambda} := \left\{ (t, y, z, t', y', z') : t=t'=z=z', y=y', (t, y) \in C_1 \right\}.$$

By Proposition 2.7.4(ii) we obtain

$$(h_1, -p_1(t_1^*), p_3(t_1^*), -h_2, p_2(t_1^*), p_4(t_1^*)) \in \left\{ (\alpha, \beta, \gamma, \alpha', \beta', \gamma') : (\alpha + \alpha' + \gamma + \gamma', \beta + \beta') \in N_{C_1}(t_1^*, x_1^*(t_1^*)) \right\}.$$

It follows that

$$(h_1 - h_2 + p_3(t_1^*) + p_4(t_1^*), -p_1(t_1^*) + p_2(t_1^*)) \in N_{C_1}(t_1^*, x_1^*(t_1^*)).$$

Thus we obtain the transversality condition on the set C_1 at the optimal switch-point as required. $\square$

We note that this condition yields further information if we do not incorporate a constraint set, C_1 , in our problem formulation. Then $(t_1^*, x_1^*(t_1^*)) \in \mathbb{R} \times \mathbb{R}^n$ and by theory of normal cones, $h_1 - h_2 + p_3(t_1^*) + p_4(t_1^*) = 0$ and $-p_1(t_1^*) + p_2(t_1^*) = 0$. Thus $p_1(t_1^*) = p_2(t_1^*)$, hence the costates corresponding to the state trajectories $x_1(\cdot)$ and $x_2(\cdot)$ are continuous at the switch-point, as expected. The condition $h_2 - h_1 = p_3(t_1^*) + p_4(t_1^*)$ specifies the non-continuity of the Hamiltonians at the switch-point due to the explicit t_1 dependence of the functions and corresponds to the result obtained via the classical maximum principle (see [T2]). The terms h_1, h_2 correspond to the optimal Hamiltonian values at the optimal switch-point and $p_3(t_1^*), p_4(t_1^*)$ can be obtained by solving the costate equations together with the boundary conditions for $p_3(\cdot)$ and $p_4(\cdot)$.

This analysis shows that the transversality conditions derived classically for the two-stage problem can be recovered, without imposing such severe conditions on the functions involved, via the application of the multiprocess maximum principle.

Having established the optimality conditions for the general two-stage control problem we now apply the optimality criteria to a problem adapted by Tomiyama and Rossana [T2] from an original formulation by Maccini [M1]. This is the development of a dynamic optimization model that deals with the effect of delivery lags on the demand for investment by a firm in the case where the delivery lag is a choice variable in the firm's decision process. (In Maccini's example the delivery lag is an exogenous variable, i.e. it is beyond the firm's control and assumed fixed for all time).

A delivery lag is the period of time which elapses between the placement of an order for a capital good and its subsequent delivery. Previous research has considered dynamic optimization models that deal with the effect of delivery lags from the point of view of the supply side of the market (see e.g. Maccini [M2]). However, a complete understanding of the effect of these delays requires some knowledge of the optimal behaviour on the buying side as in the example we now consider. We seek to obtain some qualitative results about the optimal choice of delivery lag.

We state the optimization problem as follows:

$$\begin{aligned} \text{maximize } J = & \int_0^v \{P F(K_0, L(t)) - W L(t)\} e^{-rt} dt \\ & + \int_v^\infty \{P [F(K(t), L(t)) - R(I(t)) e^{rv} - B(I(t))] - W L(t) - G_v I(t)\} e^{-rt} dt, \end{aligned} \quad (3.5.13)$$

subject to $\dot{K}(t) = I(t)$, $K(t) = K_0 > 0$, $0 \leq t \leq v$, $G_v = G_0 a(v)$, where

G_v = nominal purchase price of capital goods (i.e. price of capital goods on delivery at $t=v$),

G_0 = purchase price at time $t=0$ (immediate delivery),

$K(t)$ = stock of capital goods,

$I(t)$ = deliveries of capital goods,

$L(t)$ = labour hired by firm,

$F(K(t), L(t))$ = neo-classical production function,

$R(I(t))$ = planning costs, $B(I(t))$ = installation costs,

P = output price, r = discount rate,

v = delivery lag, W = money wage rate.

The capital stock for a typical firm is defined by the accumulation equation $\dot{K}(t) = I(t)$ which states that the rate of change of the firm's capital stock equals current deliveries of capital goods. Investment expenditures may be identified with the value of deliveries of capital goods since we assume that payment for the goods is made at the time of delivery. Investment demand is identified with the placement of an order for a capital good.

The initial conditions include the firm's capital stock from time zero to time v . The behaviour of the firm's capital stock during this time period is determined by orders placed prior to the planning date (taken as time zero).

The planning and installation costs are associated respectively with placing orders for, and taking delivery of, capital goods. Thus the functions $R(\cdot)$ and $B(\cdot)$ have properties that are standard in the adjustment cost literature:

$$R(0)=0, \quad R'(I(t)) \leq 0 \text{ as } I(t) \leq 0, \quad R''(I(t)) > 0,$$

$$B(0)=0, \quad B'(I(t)) \leq 0 \text{ as } I(t) \leq 0, \quad B''(I(t)) > 0.$$

Similarly, the production function $F(K(t), L(t))$, which is concave, has the standard properties:

$$F_K, F_L > 0, \quad F_{KK}, F_{LL} < 0, \quad F_{KK}F_{LL} - F_{LK}^2 > 0,$$

$$\lim_{L \rightarrow 0} F_L = 0, \quad \lim_{L \rightarrow \infty} F_L = \infty, \quad \lim_{K \rightarrow 0} F_K = 0, \quad \lim_{K \rightarrow \infty} F_K = \infty.$$

We express the purchase price of capital goods G_v in terms of a spot price G_0 , associated with the immediate delivery of capital goods, and a factor of proportionality $a(v)$. The latter depends on the delivery lag with

$$0 < a(v) < 1, \quad a'(v) < 0,$$

$$1 < a(v) < \infty, \quad a'(v) > 0,$$

$$a(v) = 1, \quad a'(v) = 0,$$

$$a(0) = 1,$$

and captures any discounts (when $a'(v) < 0$) or premia (when $a'(v) > 0$) imposed by suppliers if the firm takes forward delivery of new capital.

We assume that $a''(v) > 0$ so that cash flow for $v \leq t < \infty$ is concave in v . The interval $0 \leq t < v$ is a finite time where the firm's capital stock is fixed and the firm can only engage in more limited decision making, namely, vary output by its choice of labour input. A choice of delivery lag enables the firm to choose the length of this short run. The second interval is when deliveries may be taken.

Using the notation of multiprocess theory, we can rewrite the problem as:

$$\text{maximize } \int_{t_0}^v \mathcal{F}_1(t, K_1(t), L_1(t)) dt + \int_v^\infty \mathcal{F}_2(t, K_2(t), L_2(t), I_2(t), K_3(t)) dt \quad (3.5.14)$$

subject to

$$\left. \begin{aligned} \dot{K}_1(t) &= I_1(t) & t \in [t_0, v] \\ \dot{K}_2(t) &= I_2(t) & t \in [v, \infty) \\ \dot{K}_3(t) &= 0 & t \in [v, \infty) \text{ where } K_3(v) = v \end{aligned} \right\} (3.5.15)$$

and initial condition $K_1(t) = K_0 > 0, \quad 0 \leq t \leq v$.

$$\text{We set } K_1(t_0) \in C_0, \quad (v, K_1(v)) \in C_1, \quad (3.5.16)$$

where $C_0 \subset \mathbb{R}^n$, $C_1 \subset \mathbb{R} \times \mathbb{R}^n$, are given closed endpoint constraint sets.

The maximization is conducted over multiprocesses

$$\left\{ (t_0, t_1, K_1(\cdot), L_1(\cdot), I_1(\cdot)), (t'_1, t_2, K_2(\cdot), K_3(\cdot), L_2(\cdot), I_2(\cdot)) \right\}$$

which are subject to the following constraint set Λ ,

$$\Lambda = \left\{ (t_0, t_1, K_1(t_0), K_1(t_1)), (t'_1, t_2, K_2(t'_1), K_2(t_2), K_3(t'_1), K_3(t_2)): t_1 = t'_1, \right. \\ \left. t_2 = \infty, K_1(t_0) \in C_0, K_1(t_1) = K_2(t'_1), t'_1 = K_3(t'_1), K_3(t_2) \text{ free} \right\}. \quad (3.5.17)$$

The delivery lag v is a choice variable, constrained to satisfy $t_0 \leq v < \infty$ where t_0 is fixed.

Note that the state is continuous at the switch-point i.e. $K_1(v) = K_2(v)$.

We define the Hamiltonian function for each stage:

$$H_1(t, K_1, L_1, I_1, p_1, v) = Pe^{-rt} \{ F(K_1, L_1) - (W/P)L_1 + p_1 I_1 \}$$

$$H_2(t, K_2, L_2, I_2, p_2, v) = Pe^{-rt} \{ F(K_2, L_2) - R(I_2)e^{rv} - B(I_2) - (W/P)L_2 - (G_v/P)I_2 + p_2 I_2 \}$$

where $p_i(t)$ are costate variables that have the usual shadow price interpretation, (see Maccini [M1]).

Let $\{(K_1^*(\cdot), L_1^*(\cdot), I_1^*(\cdot), v^*), (K_2^*(\cdot), L_2^*(\cdot), I_2^*(\cdot), v^*)\}$ solve the above problem. Then by Theorem 3.5.1, there exist real numbers λ (equal 0 or 1), h_1 , h_2 and absolutely continuous functions $p_1(\cdot): [t_0, v^*] \rightarrow \mathbb{R}^n$, $p_i(\cdot): [v^*, \infty) \rightarrow \mathbb{R}^n$ ($i=2,3$) such that $\lambda + |p_1(t_0)| + |p_2(v^*)| + |p_3(v^*)| > 0$ and

$$(i) \quad \begin{aligned} -\dot{p}_1(t) &= \partial_{K_1} H_1(t, K_1^*(t), L_1^*(t), I_1^*(t), p_1(t)) & \text{a.e. } t \in [t_0, v^*] \\ -\dot{p}_2(t) &= \partial_{K_2} H_2(t, K_2^*(t), L_2^*(t), I_2^*(t), p_2(t), K_3^*(t)) & \text{a.e. } t \in [v^*, \infty) \\ -\dot{p}_3(t) &= \partial_{K_3} H_2(t, K_2^*(t), L_2^*(t), I_2^*(t), p_2(t), K_3^*(t)) & \text{a.e. } t \in [v^*, \infty) \end{aligned} \quad \left. \vphantom{\begin{aligned} -\dot{p}_1(t) = \partial_{K_1} H_1(t, K_1^*(t), L_1^*(t), I_1^*(t), p_1(t)) \\ -\dot{p}_2(t) = \partial_{K_2} H_2(t, K_2^*(t), L_2^*(t), I_2^*(t), p_2(t), K_3^*(t)) \\ -\dot{p}_3(t) = \partial_{K_3} H_2(t, K_2^*(t), L_2^*(t), I_2^*(t), p_2(t), K_3^*(t)) \end{aligned}} \right\} (3.5.18)$$

$$(ii) \quad \begin{aligned} H_1(t, K_1^*(t), L_1^*(t), I_1^*(t), p_1(t)) &= \min_{L_1, I_1 \in \mathbb{R}} H_1(t, K_1^*(t), L_1, I_1, p_1(t)) \text{ a.e. } t \in [t_0, v^*] \\ H_2(t, K_2^*(t), L_2^*(t), I_2^*(t), p_2(t), v^*(t)) &= \min_{L_2, I_2 \in \mathbb{R}} H_2(t, K_2^*(t), L_2, I_2, p_2(t), K_3^*(t)) \\ &\text{a.e. } t \in [v^*, \infty) \end{aligned} \quad \left. \vphantom{\begin{aligned} H_1(t, K_1^*(t), L_1^*(t), I_1^*(t), p_1(t)) = \min_{L_1, I_1 \in \mathbb{R}} H_1(t, K_1^*(t), L_1, I_1, p_1(t)) \text{ a.e. } t \in [t_0, v^*] \\ H_2(t, K_2^*(t), L_2^*(t), I_2^*(t), p_2(t), v^*(t)) = \min_{L_2, I_2 \in \mathbb{R}} H_2(t, K_2^*(t), L_2, I_2, p_2(t), K_3^*(t)) \text{ a.e. } t \in [v^*, \infty) \end{aligned}} \right\} (3.5.19)$$

$$(iii) \quad \begin{aligned} h_1 &= \text{co ess}_{t \rightarrow v^*} \left\{ \inf_{L_1, I_1 \in \mathbb{R}} H_1(t, K_1^*(v^*), L_1, I_1, p_1(v^*)) \right\} \\ h_2 &= \text{co ess}_{t \rightarrow v^*} \left\{ \inf_{L_2, I_2 \in \mathbb{R}} H_2(t, K_2^*(v^*), L_2, I_2, p_2(v^*), K_3^*(v^*)) \right\} \end{aligned} \quad \left. \vphantom{\begin{aligned} h_1 = \text{co ess}_{t \rightarrow v^*} \left\{ \inf_{L_1, I_1 \in \mathbb{R}} H_1(t, K_1^*(v^*), L_1, I_1, p_1(v^*)) \right\} \\ h_2 = \text{co ess}_{t \rightarrow v^*} \left\{ \inf_{L_2, I_2 \in \mathbb{R}} H_2(t, K_2^*(v^*), L_2, I_2, p_2(v^*), K_3^*(v^*)) \right\} \end{aligned}} \right\} (3.5.20)$$

$$(iv) \quad p_1(t_0) \in N_{C_0}; \left(h_1 - h_2 + p_3(v^*), -p_1(v^*) + p_2(v^*) \right) \in N_{C_1}. \quad (3.5.21)$$

The normal cone N_{C_0} is evaluated at $(K_1^*(t_0))$, N_{C_1} is evaluated at $(v^*, K_1^*(v^*))$.

As before, if we remove the constraint set C_1 , we obtain

$$p_1(v^*) = p_2(v^*) \quad (3.5.22)$$

i.e. the costates are continuous at the optimal lag v^* and

$$h_1 - h_2 = p_3(v^*), \quad (3.5.23)$$

which is the additional switch-point condition owing to the explicit v dependence in the integrand of the cost function over the second time interval.

Thus optimality criteria for the problem are given by:

for $0 \leq t \leq v^*$,

$$F_L(K_0, L_1(t)) = W/P, \quad (3.5.24)$$

for $v^* \leq t < \infty$,

$$F_L(K_2(t), L_2(t)) = W/P, \quad (3.5.25)$$

$$p_2(t) = R'(I_2(t))e^{rv^*} + B'(I_2(t)) + G_v/P, \quad (3.5.26)$$

$$\dot{p}_2(t) = -F_K(K_2(t), L_2(t)) + rp_2, \quad (3.5.27)$$

$$\lim_{t \rightarrow \infty} p_2(t)e^{-rt}K_2(t) = 0. \quad (3.5.28)$$

A full interpretation of these criteria is provided in Maccini [M1].

We note that the problem is formulated in a smooth setting and that we have assumed continuity of the state variable at the switch-point and a constant capital stock over the first time interval. (This latter condition does not allow for variations in capital stock which may occur as a result of orders placed prior to the planning date). In fact, since K_1 is constant, we have that $\dot{p}_1 = 0$. However, the analysis of multiprocess theory remains valid without such assumptions.

Owing to the additional structure imposed on the problem by Tomiyama and Rossana, condition 3.5.23 gives the following integral matching condition:

$$\begin{aligned} & R(I_2(v^*))e^{rv^*} + B(I_2(v^*)) + (G_v/P)I_2(v^*) - p_2(v^*)I_2(v^*) \\ &= - \int_{v^*}^{\infty} [R(I_2(t))re^{rv^*} + (G_0/P)a'(v^*)I_2(t)] e^{-r(t-v^*)} dt. \end{aligned} \quad (3.5.29)$$

By imposing quadratic form assumptions as in the economics literature, we can simplify this expression. We assume that

$$F(K, L) = c_0K + (c_1/2)K^2 + d_0L + (d_1/2)L^2 + e_1KL,$$

$$c_0, d_0, e_1 > 0, \quad c_1, d_1 < 0, \quad e_1^2 - c_1d_1 < 0,$$

$$R(I) = (s_1/2)I^2, \quad B(I) = (q_1/2)I^2, \quad s_1, q_1 > 0.$$

Using these assumptions and eliminating p_2 from (3.5.29) via (3.5.26), we obtain

$$\left((s_1/2)e^{rv^*} + (q_1/2) \right) I_2(v^*)^2 = - \int_{v^*}^{\infty} \left[(s_1/2) I_2(t)^2 r e^{rt} + (G_0/P) a'(v^*) I_2(t) \right] e^{-r(t-v^*)} dt. \quad (3.5.30)$$

This condition indicates that the firm should choose its delivery lag such that the planning and installation costs at the delivery time are equal to the discounted benefits and costs attached to resources used in placing orders for goods and any discounts or premia associated with the purchase price of capital goods for future delivery. This condition also shows that for the delivery lag to be non-zero, if $I(t)$ is positive (negative), then $a'(v^*)$ must be negative (positive). Finally, we note that variations in the parameters (G_0, P, r, W) also affect the optimal delivery lag.

A NECESSARY AND SUFFICIENT SENSITIVITY RELATION FOR CONSTRAINED END-POINT PROBLEMS

4.1 Introduction

In this chapter we consider the following optimal control problem, which we label (P).

$$\begin{aligned} &\text{Minimize } \int_0^1 L(t, x(t), u(t)) dt, \\ &\text{subject to} \\ &\quad \dot{x}(t) = f(t, x(t), u(t)) \quad \text{a.e. } t \in [0, 1], \\ &\quad u(t) \in U \quad \text{a.e. } t \in [0, 1], \\ &\quad x(0) = x_0, \quad x(1) \in C. \end{aligned} \tag{P}$$

The data for problem (P) are functions $f: \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ and $L: \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$, a subset $U \subset \mathbb{R}^m$, an n -vector x_0 and a non-empty subset $C \subset \mathbb{R}^n$.

The minimization is conducted over ‘admissible processes’ which are pairs of functions $(x(\cdot), u(\cdot))$ in which $x(t)$ is an absolutely continuous function which is a solution to the differential equation $\dot{x}(t) = f(t, x(t), u(t))$ and satisfies the boundary conditions $x(0) = x_0$, $x(1) \in C$, and $u(t)$ is a corresponding measurable function satisfying $u(t) \in U$ almost everywhere. A process which achieves the minimum is called a minimizing process.

We assume that the data satisfy the following hypotheses in which the ‘augmented velocity function’, $\tilde{f}(\cdot, \cdot): [0, 1] \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^{n+1}$ is

$$\tilde{f}(t, x, u) := \text{col}[L(t, x, u), f(t, x, u)].$$

(H1) $\tilde{f}(t, x, u)$ is measurable in t for fixed $(x, u) \in \mathbb{R}^n \times \mathbb{R}^m$, and continuous in u for fixed $(t, x) \in [0, 1] \times \mathbb{R}^n$,

(H2) there exists a function $\alpha(\cdot) \in L^1$ such that, for all $x, y \in \mathbb{R}^n$, $w \in U$, a.e. $t \in [0, 1]$,

$$|\tilde{f}(t, x, w) - \tilde{f}(t, y, w)| \leq \alpha(t) |x - y|,$$

(H3) there exists $k(\cdot) \in L^\infty$ such that, for all $x \in \mathbb{R}^n$, $w \in U$, a.e. $t \in [0, 1]$,

$$|\tilde{f}(t, x, w)| \leq k(t) [1 + |x(t)|],$$

(H4) U is compact for almost every $t \in [0, 1]$ and $\tilde{f}(t, x, U) \subset [0, 1] \times \mathbb{R}^n \times \mathbb{R}^m$, $C \subset \mathbb{R}^n$ are closed convex sets.

We are concerned with a rigorous exploration of the interrelation between two important branches of optimal control theory, namely the Pontryagin maximum principle and dynamic programming.

We recall that the maximum principle provides a set of necessary conditions for an admissible process to be optimal and can be expressed as follows:

THEOREM 4.1.1 (Maximum Principle): *Assume that the data satisfy hypotheses (H1) to (H4) and that $\tilde{f}(t, x, u)$ is continuously differentiable in the x variable. Let $(x^*(\cdot), u^*(\cdot))$ be a minimizing process. Then there exists a number $\lambda \in \{0, 1\}$ and an absolutely continuous function $p(\cdot): [0, 1] \rightarrow \mathbb{R}^n$ such that $\lambda + |p(1)| \neq 0$,*

$$-\dot{p}(t) = p(t) \cdot f_x(t, x^*(t), u^*(t)) - \lambda L_x(t, x^*(t), u^*(t)) \quad \text{a.e. } t \in [0, 1], \quad (4.1.1)$$

$$H(t, x^*(t), u^*(t), p(t), \lambda) = \max_{u \in U} H(t, x^*(t), u, p(t), \lambda) \quad \text{a.e. } t \in [0, 1], \quad (4.1.2)$$

and

$$-p(1) \in N_C(x^*(1)). \quad (4.1.3)$$

In the above H is the Hamiltonian function

$$H(t, x, u, p, \lambda) = p \cdot f(t, x, u) - \lambda L(t, x, u)$$

and $N_C(x)$ denotes the (Clarke) normal cone to the closed set C at a point $x \in C$. Recall that the (Clarke) normal cone is a generalization to closed sets of the usual definition of normal cone expressed in terms of 'outward' normal vectors to a convex set. In the absence of the 'continuous differentiability' assumption we can replace equation (4.1.1) by $-\dot{p}(t) \in \partial H(t, x^*(t), u^*(t), p(t), \lambda)$.

Here ∂H denotes the generalized gradient of H in the second argument, namely $\partial H(t, x, u, p, \lambda) = \overline{\text{co}} \left\{ \lim_i \nabla H(t, \cdot, u, p, \lambda) \right\}$ when the 'closed convex hull' operation $\overline{\text{co}}$ is taken over all sequences $\{x_i\}$ along which the (ordinary) gradient $\nabla H(t, \cdot, u, p, \lambda)$ exists and for which $x_i \rightarrow x$ and $\{\nabla H(t, \cdot, u, p, \lambda)\}$ has a limit. For smooth data, ∂H is the single-valued function $p \cdot f_x(t, x, u) - \lambda L_x(t, x, u)$, as in the statement of the theorem. A proof of Theorem 4.1.1 may be found, for example, in Fleming and Rishel [F1]. For a proof of the extension of the classical maximum principle of Pontryagin to cases where the data is non-smooth, see Clarke [C3].

Dynamic programming is based on properties of the value function $V: [0, 1] \times \mathbb{R}^n \rightarrow \mathbb{R}$. For $(s, \xi) \in [0, 1] \times \mathbb{R}^n$, $V(s, \xi)$ is defined to be the infimum cost of a problem related to (P) in which the underlying time interval $[0, 1]$ is replaced by $[s, 1]$ and the initial state x_0 by ξ , i.e.

$$V(s, \xi) := \inf \left\{ \int_s^1 L(t, x(t), u(t)) dt : \dot{x}(t) = f(t, x(t), u(t)), u(t) \in U, x(s) = \xi, x(1) \in C \right\}. \quad (4.1.4)$$

Suppose for the time being that the constraint set C is all of $\mathbb{R}^n$ (thus the right end-point is effectively unconstrained), then hypotheses (H1) to (H4) ensure that the evaluation of V involves taking an infimum over a non-empty set, where this infimum is bounded below—thus V is well-defined. Heuristic arguments have been advanced suggesting a relationship of the form:

$$(h(t), -p(t)) = \nabla V(t, x^*(t)),$$

where $h(t)$ is the Hamiltonian function evaluated along $x^*(t)$ and its associated costate $p(t)$:

$$h(t) = H(t, x^*(t), u^*(t), p(t), \lambda).$$

However, such arguments, depending as they do on the differentiability of V , cannot be made rigorous in a general setting since the differentiability assumption is not justified.

As we recall from the introductory chapter, weaker versions of this relationship have been proved, an example of which is provided by the following theorem (due to Vinter and Wolenski [V3]):

THEOREM 4.1.2: *Assume that the data satisfy hypotheses (H1) to (H4) and assume that $C = \mathbb{R}^n$. Let $(x^*(\cdot), u^*(\cdot))$ be a minimizing process. Then there exists an absolutely continuous function $p(\cdot): [0, 1] \rightarrow \mathbb{R}^n$ such that the pair $(\lambda=1, p(\cdot))$ satisfies conditions (4.1.1) to (4.1.3) of the maximum principle and also satisfies the inclusion*

$$(h(t), -p(t)) \in \partial V(t, x^*(t)) \quad \text{a.e. } t \in [0, 1]. \quad (4.1.5)$$

Note that hypothesis (H3) ensures that V is locally Lipschitz continuous so that the generalized gradient ∂V of V is defined.

The theorem asserts that *one* costate can always be chosen to satisfy the ‘minimum cost sensitivity inclusion’ (4.1.5). Examples of control problems are known where certain costates fail to satisfy the inclusion.

The above result, which relates the maximum principle to dynamic programming, improves upon that obtainable by the methods of [V1] which require $\tilde{f}$ to be continuous with respect to the time variable and upon that of [C6] which merely establishes “ $-p(t) \in \partial_x V$ ” and supplies no information about $h(t)$. A similar result which is both necessary *and* sufficient for optimality in a smooth setting has been proved by Subbotina [S3] who uses the superdifferential of V , $\partial^+ V(t, x(t))$, (refer to Definition 2.4.1) instead of the generalized gradient of V .

We now turn our attention to problem (P) where C is some given closed, convex constraint set. The difficulty now is that V may possibly be discontinuous, and its

domain (points in (t,x) space which can be driven to the terminal constraint set at time $t=1$) may be a small subset of $[0,1] \times \mathbb{R}^n$. Thus relationship (4.1.5), which is expressed in terms of a Lipschitz continuous function defined on some 'tube' about $x^*(\cdot)$, does not even make sense. We shall circumvent the difficulty by establishing the relationship (4.1.5), not in terms of the value function, but instead in terms of a function, namely a verification function, having related, but often nicer, properties.

As we have discussed in our introductory chapter, verification functions arise in Hamilton-Jacobi theory. The theory, as it relates to problem (P), contains two main elements:

- (i) verification theorems and
- (ii) properties of the value function.

These are linked by the Hamilton-Jacobi equation or its related inequality.

We again recall (from Chapter 1) that a verification theorem (or the Hamilton-Jacobi verification technique) is a sufficient condition for an arc to be minimizing - this is expressed in terms of 'verification functions' which are solutions to the Hamilton-Jacobi equation (suitably defined) and its associated boundary condition. However, the value function also provides a solution to the Hamilton-Jacobi equation in the appropriate sense.

Recent research has shown that the requirements on verification functions to be classical C^1 solutions to the Hamilton-Jacobi equation may be weakened using non-smooth analysis, to include functions ϕ which are locally Lipschitz continuous on $[0,1] \times \mathbb{R}^n$ and which are taken to be extended solutions to the Hamilton-Jacobi equation (see e.g. Clarke and Vinter [C5]).

Vinter and Wolenski [V4] have proved a sufficient condition of optimality in terms of locally Lipschitz verification functions, for a class of free end-point problems with terminal cost $g(x(1))$. They use a class of verification functions which are defined as lower-Dini solutions of the Hamilton-Jacobi inequality, i.e. functions which satisfy

$$\min_{u \in U} d^- \phi((t, x(t)); (1, f(t, x(t), u(t)))) \geq 0 \quad (4.1.6)$$

with boundary condition $\phi(1, x) = g(x)$ for all $x \in \mathbb{R}^n$. Here $d^- \phi(\eta; \omega)$ denotes the lower-Dini derivative of ϕ in the direction ω :

$$d^- \phi(\eta; \omega) := \lim_{\substack{h \downarrow 0 \\ u \rightarrow \omega}} \inf h^{-1} \{ \phi(\eta + hu) - \phi(\eta) \}. \quad (4.1.7)$$

The points of interest here are that their results are valid for data which are measurable in the time variable and furthermore, since the value function serves as a verification

function, we are assured that optimality of a putative minimizer can be confirmed by an appropriate verification function.

Their approach is based on the following assertions (also proved in [V4]):

- that the value function is indeed a verification function and is identifiable as the maximal element in the set of verification functions;
- and moreover,
- that the value function is a lower-Dini solution of a suitably defined, extended Hamilton-Jacobi equation which replaces the weaker inequality condition.

We use these developments as a starting point to obtain a minimum cost sensitivity relation between the costate $p(\cdot)$ of the maximum principle and the superdifferential $\partial^{\vee}W$ of some verification function W , namely,

$$(h(t), -p(t)) \in \partial^{\vee}W(t, x^*(t)), \text{ where } (x^*(\cdot), u^*(\cdot)) \text{ is an optimal process.}$$

We show furthermore, that this inclusion, together with the conditions in the Pontryagin Maximum Principle, are *sufficient* (as well as necessary) for optimality.

This extends results from the earlier literature which are expressed in terms of the value function, as it permits the use of a broader class of functions which are well-defined even in the presence of end-point constraints. Proof of the theorem hinges on consideration of appropriate auxiliary control problems and is in two parts.

In the first part we treat a fixed end-point problem. The critical idea here is that the fixed right end-point constraint can be replaced by a strictly differentiable penalty function under the strict normality hypothesis (H) of Section 4.3. The most challenging step is to demonstrate existence of a verification function under (H); we show that the value function of the penalized problem is a suitable candidate as the verification function for the original problem (P).

In the second part we relax the restriction that C comprises a single point. Reduction of a general convex constraint set to a single point by extending the underlying time interval of the problem is involved.

The function $\tilde{f}(t, x, u)$ is permitted to be merely measurable in the time variable; the methods in [S3] are more restrictive since they involve taking time derivatives of f and consequently treat only problems with data smooth in the time variable. $\tilde{f}(t, x, u)$ is required, however, to be continuously differentiable in the state variable.

We note, finally, that optimality conditions involving Hamilton-Jacobi theory (but not involving sensitivity inclusions) have been previously obtained for non-normal control problems (i.e. those in which the maximum principle provides trivial information about the optimal process) see e.g. Vinter and Mendoza [V2]. Their results

are less convenient than ours however, because they are expressed in terms of a sequence of locally Lipschitz continuous functions (instead of the single function we employ).

4.2. Preliminary Results

The following results (a straightforward extension of results of Vinter and Wolenski [V4] to allow for an integral cost) express precisely the ideas discussed in the previous section. We shall make use of them in the proof of the main theorem of the chapter.

DEFINITION 4.2.1: Suppose $\phi: [0,1] \times \mathbb{R}^n \rightarrow \mathbb{R}$. Then ϕ is called a verification function (for the problem (P)) if:

- (a) $\phi(\cdot, \cdot)$ is locally Lipschitz continuous,
- (b) $\phi(1, \xi) \leq 0$ for all $\xi \in C$,
- (c) there exists $J \subseteq [0,1]$ of measure zero so that for all $t \notin J$ and $\xi \in \mathbb{R}^n$,

$$\min_{u \in U} \left\{ d^- \phi((t, x); (1, f(t, x, u))) + L(t, x, u) \right\} \geq 0. \quad (4.2.1)$$

We now quote two theorems for which we assume that $C = \mathbb{R}^n$ i.e. the right end-point is unconstrained. However, as we shall see in the next section the ideas do extend to the constrained end-point problem.

The first theorem in this section asserts that the value function V is indeed a verification function and moreover that it is the maximal element in the set of verification functions. Furthermore, V has the additional property that the inequality in (4.2.1) can be replaced by the stronger Hamilton-Jacobi equality.

THEOREM 4.2.2: Suppose that $C = \mathbb{R}^n$ and suppose that the data satisfy the basic hypotheses. Then V is a verification function that satisfies, for each $(t, \xi) \in [0,1] \times \mathbb{R}^n$,

$$V(t, \xi) = \max \left\{ \phi(t, \xi) : \phi \text{ is a verification function} \right\}. \quad (4.2.2)$$

Moreover, V satisfies the Hamilton-Jacobi equation in the following form using lower-Dini derivatives: for $t \notin J$ and $\xi \in \mathbb{R}^n$, we have

$$\min_{u \in U} \left\{ d^- \phi((t, x); (1, f(t, x, u))) + L(t, x, u) \right\} = 0. \quad (4.2.3)$$

Proof: see [V4].

The second theorem in this section provides a set of necessary and sufficient conditions of optimality expressed in terms of verification functions.

THEOREM 4.2.3 (Hamilton-Jacobi Verification Technique): *Suppose that $C=\mathbb{R}^n$ and suppose that the data satisfy the basic hypotheses. Then $(x^*(\cdot), u^*(\cdot))$ is an optimal solution to (4.2.1) if and only if there exists a verification function ϕ such that*

$$\phi(0, x_0) = \int_0^1 L(t, x^*(t), u^*(t)) dt. \quad (4.2.4)$$

Proof: see [V4].

4.3 The 'Strengthened Normality' Condition

A key role in the new optimality conditions of this section is played by a nondegeneracy condition, strengthened normality. In this section we define this concept and give a sufficient condition for strengthened normality in terms of controllability of a linearized system. The following definition relates to problems with data continuously differentiable in the x variable.

Strengthened Normality Hypothesis (H)

Problem (P) has a unique minimizer $(x^*(\cdot), u^*(\cdot))$, there is a unique function $p(\cdot): [0,1] \rightarrow \mathbb{R}^n$ which is absolutely continuous and satisfies

$$-\dot{p}(t) = p(t) \cdot f_x(t, x^*(t), u^*(t)) - L_x(t, x^*(t), u^*(t)) \quad \text{a.e. } t \in [0,1] \quad (4.3.1)$$

$$H(t, x^*(t), u^*(t), p(t)) = \max_{u \in U} H(t, x^*(t), u, p(t)) \quad \text{a.e. } t \in [0,1] \quad (4.3.2)$$

$$-p(1) \in N_C(x^*(1)) \quad (4.3.3)$$

and, for almost every $t \in [0,1]$, $u \mapsto H(t, x^*(t), u, p(t))$ has a unique minimizer over U .

One situation where strengthened normality is automatically satisfied is when there is a unique optimal control and $C=\mathbb{R}^n$. For then $p(\cdot)$ is the unique solution to the differential equation:

$$-\dot{p}(t) = p(t) \cdot f_x(t, x^*(t), u^*(t)) - L_x(t, x^*(t), u^*(t)) \quad \text{a.e. } t \in [0,1]$$

$$-p(1) = 0.$$

When C is not the whole space, $p(\cdot)$ need not be unique (even when there is a unique optimal control), because of the latitude which may be available to us in choice of $p(\cdot)$ to satisfy the boundary condition

$$-p(1) \in N_C(x^*(1)).$$

We now explore the extreme case of this situation when $C=\{x_1\}$, for some $x_1 \in \mathbb{R}^n$. This corresponds to a problem with fixed right end-point; now the transversality condition is $-p(1) \in \mathbb{R}^n$ and so conveys no information about $p(\cdot)$. Here a controllability type criterion can be given for strengthened normality.

DEFINITION 4.3.1: Let $(x^*(\cdot), u^*(\cdot))$ be a process. We say the linearized system about $x^*(\cdot)$ is controllable if $\mathfrak{R} = \mathbb{R}^n$, where

$$\mathfrak{R} := \left\{ y(1) : y(\cdot) \text{ satisfies } \dot{y}(t) = A(t)y(t) + B(t)u(t) \text{ a.e. } t \in [0, 1], y(0) = 0 \text{ for some } u(\cdot) \in L^2([0, 1]; \mathbb{R}^m) \right\}$$

where $A(t) := f_x(t, x^*(t), u^*(t))$ and $B(t) := f_u(t, x^*(t), u^*(t))$.

A well-known equivalent definition of local controllability is the following (see [B4]):

“Let z_0 be an arbitrary non-zero n -vector. Then $y(t) \equiv 0$ on $[0, 1]$, where $y(t)$ is the solution of

$$\dot{z}(t) = A^T(t)z(t), \quad z(0) = z_0,$$

and $y(t) = B^T(t)z(t)$.”

Although less intuitive, this is often the most convenient form of the ‘local controllability’ hypothesis for application purposes.

PROPOSITION 4.3.2: Suppose there is a unique minimizer $(x^*(t), u^*(t))$ for (P) , the linearized system is controllable and $u^*(t) \in \text{interior}\{U\}$ for all t , then (P) satisfies the strengthened normality hypothesis.

Proof: Suppose $u^*(t) \in \text{int}(U)$ and that U equals the space $\mathbb{R}^m$. Also suppose that the linearized system described by $\dot{x}(t) = f_x(t, x^*(t), u^*(t)) \cdot x(t) + f_u(t, x^*(t), u^*(t)) \cdot u(t)$ is controllable.

Assume that $p(\cdot)$ is non-unique. Let $f_x^T(t, x^*(t), u^*(t))$ and $f_u^T(t, x^*(t), u^*(t))$ denote the matrix transpose of $f_x(t, x^*(t), u^*(t))$ and $f_u(t, x^*(t), u^*(t))$ respectively. Then there exist $p_1(\cdot)$ and $p_2(\cdot)$ such that

$$-\dot{p}_1(t) = f_x^T(t, x^*(t), u^*(t))p_1(t) - L_x(t, x^*(t), u^*(t))$$

$$-\dot{p}_2(t) = f_x^T(t, x^*(t), u^*(t))p_2(t) - L_x(t, x^*(t), u^*(t)).$$

By interiority of $u^*(t)$ we have that

$$\frac{\partial}{\partial u}[p_1(t) \cdot f(t, x^*(t), u^*(t)) - L(t, x^*(t), u^*(t))] = 0$$

$$\frac{\partial}{\partial u}[p_2(t) \cdot f(t, x^*(t), u^*(t)) - L(t, x^*(t), u^*(t))] = 0.$$

Subtracting the equations gives

$$-(\dot{p}_1(t) - \dot{p}_2(t)) = f_x^T(t, x^*(t), u^*(t)) \cdot (p_1(t) - p_2(t))$$

$$(p_1(t) - p_2(t)) \cdot f_u(t, x^*(t), u^*(t)) = 0.$$

Defining $q(t) = p_1(t) - p_2(t)$, we then obtain the system of equations

$$-\dot{q}(t) = f_x^T(t, x^*(t), u^*(t)) q(t)$$

$$y(t) = f_u^T(t, x^*(t), u^*(t)) q(t),$$

where $y(\cdot)$ is the output. Since the linearized system described by $f_x(t, x^*(t), u^*(t))$ and $f_u(t, x^*(t), u^*(t))$ is controllable, we have, by the above observation, that $q(\cdot) \equiv 0$ i.e. $p_1(\cdot) \equiv p_2(\cdot)$ and the costate is unique. $\square$

4.4 The Main Result

This section provides a statement of the main results. Additional hypotheses are invoked (the process $(x^*(\cdot), u^*(\cdot))$ to which they make reference being determined by context).

(H5) For each $t \in [0, 1]$, $\tilde{f}(t, \cdot, u^*(t))$ is continuously differentiable in a neighbourhood of $x^*(t)$.

(H) (Strengthened normality hypothesis): There is a unique minimizer $(x^*(\cdot), u^*(\cdot))$ to problem (P), a unique costate $p(\cdot)$ (corresponding to the cost multiplier $\lambda=1$) associated with $(x^*(\cdot), u^*(\cdot))$ and, for almost every $t \in [0, 1]$, $u \mapsto H(t, x^*(t), u, p(t))$ has a unique minimizer over U .

THEOREM 4.4.1:

Let $(x^*(\cdot), u^*(\cdot))$ be any admissible process for the given problem. Assume that the data satisfy hypotheses (H1) to (H4).

i) (Necessity)

Assume additionally (H5) and the strengthened normality hypothesis (H). Suppose $(x^*(\cdot), u^*(\cdot))$ is minimizing. Then there exists a verification function $\phi(\cdot, \cdot): [0, 1] \times \mathbb{R}^n \rightarrow \mathbb{R}$ satisfying

$$\min_{u \in U} \left\{ d^- \phi((t, x), (1, f(t, x, u))) + L(t, x, u) \right\} = 0 \quad \text{a.e. } t \in [0, 1], \forall x \in \mathbb{R}^n \quad (4.4.1)$$

$$\phi(0, x_0) = \int_0^1 L(t, x^*(t), u^*(t)) dt \quad (4.4.2)$$

and, corresponding to any such verification function, there exists a unique absolutely continuous function $p(\cdot): [0, 1] \rightarrow \mathbb{R}^n$ such that

$$-\dot{p}(t) = p(t) \cdot f_x(t, x^*(t), u^*(t)) - L_x(t, x^*(t), u^*(t)) \quad \text{a.e. } t \in [0, 1], \quad (4.4.3)$$

$$H(t, x^*(t), u^*(t), p(t)) = h(t) \quad \text{a.e. } t \in [0, 1], \quad (4.4.4)$$

$$-p(1) \in N_C(x^*(1)) \quad (4.4.5)$$

and

$$(h(t), -p(t)) \in \partial^\vee \phi(t, x^*(t)) \quad \text{a.e. } t \in [0, 1] \quad (4.4.6)$$

(where $h(t) := \sup_{u \in U} \{ p(t) \cdot f(t, x^*(t), u) - L(t, x^*(t), u) \}$).

ii) (Sufficiency)

Let ϕ be any verification function $\phi(\cdot, \cdot): [0, 1] \times \mathbb{R}^n \rightarrow \mathbb{R}$. Suppose there exists a costate function $p(\cdot): [0, 1] \rightarrow \mathbb{R}^n$ satisfying

$$-\dot{p}(t) \in \partial_x H(t, x^*(t), u^*(t), p(t)) \quad \text{a.e. } t \in [0, 1], \quad (4.4.7)$$

$$H(t, x^*(t), u^*(t), p(t)) = h(t) \quad \text{a.e. } t \in [0, 1], \quad (4.4.8)$$

$$-p(1) \in N_C(x^*(1)) \quad (4.4.9)$$

and

$$(h(t), -p(t)) \in \partial^\vee \phi(t, x^*(t)) \quad \text{a.e. } t \in [0, 1] \quad (4.4.10)$$

in relation to ϕ , then $(x^*(\cdot), u^*(\cdot))$ is a minimizing process.

Remarks:

1) The theorem gives necessary conditions for the existence of a minimizing process through a class of verification functions (in relation to lower-Dini solutions of the Hamilton-Jacobi equation) and the equations of the maximum principle which are coupled with a sensitivity inclusion expressed in terms of the superdifferential of a verification function. The other result in Theorem 4.4.1 (which has no parallel in Theorem 4.1.2) is a set of sufficient conditions; even for non-smooth data, the ‘maximization of the Hamiltonian’ condition from the maximum principle and the strengthened sensitivity inclusion in terms of a verification function ϕ assure that the process (x^*, u^*) is minimizing.

2) ‘Normality’ is usually taken to mean that there exist costate functions associated with a non-zero ‘cost multiplier’ (see Chapter 2). Our ‘strengthened normality’ hypothesis requires not only existence of such costate functions, but uniqueness as well, as the controls range over all optimal controls for the optimization problem.

3) We note that the verification function of the theorem satisfies the stronger Hamilton-Jacobi equality. Furthermore, via equation (4.4.2), the theorem also incorporates results for the constrained end-point case that are similar to the Hamilton-Jacobi verification technique of the previous section.

4.5 Properties of the cost and value functions

Given an admissible process $(\bar{x}(\cdot), \bar{u}(\cdot))$ for (P), we can associate with it a cost function $\bar{V}: [0,1] \times \mathbb{R}^n \rightarrow \mathbb{R}$, as follows. For $(s, \xi) \in [0,1] \times \mathbb{R}^n$,

$$\bar{V}(s, \xi) := \int_s^1 L(t, x(t), \bar{u}(t)) dt \quad (4.5.1)$$

where $x(\cdot)$ is the solution to the differential equation

$$\dot{x}(t) = f(t, x(t), \bar{u}(t)) \quad \text{a.e. } t \in [s, 1] \quad (4.5.2)$$

with boundary condition

$$x(s) = \xi. \quad (4.5.3)$$

It is clear from this definition that $\bar{V}$ is the value function that would result from replacing the control constraint set U by the subset $\{\bar{u}(t)\}$ almost everywhere. A direct consequence of this observation and the principle of optimality is the following relationship between $\bar{V}$ and V (the value function as defined by equation (4.1.4)).

LEMMA 4.5.1: *For arbitrary $(s, \xi) \in [0, 1] \times \mathbb{R}^n$ we have*

$$V(s, \xi) \leq \bar{V}(s, \xi). \quad (4.5.4)$$

If $(\bar{x}(\cdot), \bar{u}(\cdot))$ is a minimizing process for (P) then, for any $s \in [0, 1]$

$$V(s, \bar{x}(s)) = \bar{V}(s, \bar{x}(s)). \quad (4.5.5)$$

We note a differentiability property of the value function V ; this follows directly from the fact that V is locally Lipschitz continuous (see [V4]) and that the composition of a Lipschitz continuous function and an absolutely continuous function is absolutely continuous.

PROPOSITION 4.5.2: *Assume hypotheses (H1) to (H4). Let $(\bar{x}(\cdot), \bar{u}(\cdot))$ be an arbitrary process. Then $t \mapsto V(t, \bar{x}(t))$ is an absolutely continuous function.*

The following proposition summarises the relevant differentiability properties of the cost function, $\bar{V}$.

PROPOSITION 4.5.3: *Let $(\bar{x}(\cdot), \bar{u}(\cdot))$ be an admissible process for problem (P), and take $\bar{V}(\cdot, \cdot): [0, 1] \times \mathbb{R}^n \rightarrow \mathbb{R}$ as defined in (4.5.1). Suppose that hypotheses (H1) to (H5) are satisfied (in relation to $(\bar{x}(\cdot), \bar{u}(\cdot))$). Let $t \in [0, 1]$ be a Lebesgue point of $s \mapsto \tilde{f}(s, \bar{x}(s), \bar{u}(s))$ and let $\{a_i\}$, $\{\tau_i\}$ and $\{\epsilon_i\}$ be arbitrary convergent sequences in $\mathbb{R}^n$, $\mathbb{R}$ and $\mathbb{R}^+$ respectively with limits a , τ and 0 respectively.*

Then

$$\epsilon_i^{-1} \left\{ \bar{V}(t + \epsilon_i \tau_i, \bar{x}(t) + \epsilon_i a_i) - \bar{V}(t, \bar{x}(t)) \right\} \rightarrow (h(t), -p(t)) \cdot (\tau, a) \text{ as } i \rightarrow \infty \quad (4.5.6)$$

where $p(\cdot)$ solves

$$-\dot{p}(t) = p(t) \cdot f_x(t, \bar{x}(t), \bar{u}(t)) - L_x(t, \bar{x}(t), \bar{u}(t)) \quad \text{a.e. } t \in [0, 1] \quad (4.5.7)$$

with boundary condition

$$-p(1) = 0, \quad (4.5.8)$$

and $h(t)$ is the function

$$h(t) = p(t) \cdot f(t, \bar{x}(t), \bar{u}(t)) - L(t, \bar{x}(t), \bar{u}(t)). \quad (4.5.9)$$

Proof of the proposition is preceded by a number of technical lemmas.

LEMMA 4.5.4: Let $(\bar{x}(\cdot), \bar{u}(\cdot))$ be an admissible process for problem (P) and let $\bar{V}(\cdot, \cdot): [0, 1] \times \mathbb{R}^n \rightarrow \mathbb{R}$ be as given by (4.5.1). Suppose that hypotheses (H1) to (H5) are satisfied (in relation to $(\bar{x}(\cdot), \bar{u}(\cdot))$). Take $t \in [0, 1)$ to be a Lebesgue point of $s \mapsto \tilde{f}(s, \bar{x}(s), \bar{u}(s))$ and let $\{\tau_i\}$ and $\{\epsilon_i\}$ be arbitrary convergent sequences in $\mathbb{R}$ and $\mathbb{R}^+$ respectively with limits τ and 0 respectively. Then

$$\lim_{i \rightarrow \infty} \epsilon_i^{-1} \left| \bar{V}(t + \epsilon_i \tau_i, \bar{x}(t + \epsilon_i \tau_i)) - \bar{V}(t + \epsilon_i \tau, \bar{x}(t) + \epsilon_i \tau f(t, \bar{x}(t), \bar{u}(t))) \right| = 0. \quad (4.5.10)$$

Proof: We know, by the absolute continuity of $\bar{x}(\cdot)$, that

$$\begin{aligned} \bar{V}(t + \epsilon_i \tau_i, \bar{x}(t + \epsilon_i \tau_i)) &= \bar{V}\left(t + \epsilon_i \tau_i, \bar{x}(t) + \int_t^{t + \epsilon_i \tau_i} \dot{\bar{x}}(s) ds\right) \\ &= \bar{V}\left(t + \epsilon_i \tau_i, \bar{x}(t) + \int_t^{t + \epsilon_i \tau_i} f(s, \bar{x}(s), \bar{u}(s)) ds\right) \end{aligned}$$

Furthermore, since t is a Lebesgue point of $s \mapsto f(s, \bar{x}(s), \bar{u}(s))$, then

$$\epsilon_i^{-1} \int_t^{t + \epsilon_i \tau_i} f(s, \bar{x}(s), \bar{u}(s)) ds \rightarrow \tau f(t, \bar{x}(t), \bar{u}(t)) \text{ as } i \rightarrow \infty, \tau_i \rightarrow \tau.$$

Hence, by local Lipschitz continuity, there exists $K > 0$ such that for each i ,

$$\begin{aligned} &\epsilon_i^{-1} \left\{ \left| \bar{V}(t + \epsilon_i \tau_i, \bar{x}(t + \epsilon_i \tau_i)) - \bar{V}(t + \epsilon_i \tau, \bar{x}(t) + \epsilon_i \tau f(t, \bar{x}(t), \bar{u}(t))) \right| \right\} \\ &= \epsilon_i^{-1} \left| \bar{V}\left(t + \epsilon_i \tau_i, \bar{x}(t) + \int_t^{t + \epsilon_i \tau_i} f(s, \bar{x}(s), \bar{u}(s)) ds\right) - \bar{V}\left(t + \epsilon_i \tau, \bar{x}(t) + \epsilon_i \tau f(t, \bar{x}(t), \bar{u}(t))\right) \right| \end{aligned}$$

$$\leq \epsilon_i^{-1} K \epsilon_i \left| \left(\tau_i, \epsilon_i^{-1} \int_{\tau_i}^{\tau_i + \epsilon_i \tau_i} f(s, \bar{x}(s), \bar{u}(s)) ds \right) - \left(\tau, \tau f(t, \bar{x}(t), \bar{u}(t)) \right) \right| \rightarrow 0 \text{ as } i \rightarrow \infty.$$

The assertion follows. $\square$

LEMMA 4.5.5: Let $(\bar{x}(\cdot), \bar{u}(\cdot))$ be an admissible process for problem (P) and let $\bar{V}(\cdot, \cdot): [0, 1] \times \mathbb{R}^n \rightarrow \mathbb{R}$ be as given by (4.5.1). Suppose that hypotheses (H1) to (H5) are satisfied (in relation to $(\bar{x}(\cdot), \bar{u}(\cdot))$). Let $\{\tau_i\}$, $\{a_i\}$ and $\{\epsilon_i\}$ be arbitrary convergent sequences in $\mathbb{R}^n$, $\mathbb{R}$ and $\mathbb{R}^+$ respectively with limits a , τ and 0 respectively. Then

$$\lim_{i \rightarrow \infty} \epsilon_i^{-1} |\bar{V}(t + \epsilon_i \tau_i, \bar{x}(t) + \epsilon_i a_i) - \bar{V}(t + \epsilon_i \tau, \bar{x}(t) + \epsilon_i a)| = 0 \quad (4.5.11)$$

Proof: By local Lipschitz continuity there exists $K > 0$ such that for each i ,

$$\frac{1}{\epsilon_i} |\bar{V}(t + \epsilon_i \tau_i, \bar{x}(t) + \epsilon_i a_i) - \bar{V}(t + \epsilon_i \tau, \bar{x}(t) + \epsilon_i a)| \leq \frac{1}{\epsilon_i} K \epsilon_i |(\tau_i, a_i) - (\tau, a)|.$$

The right-hand side of the above expression has limit zero as $i \rightarrow \infty$. The assertion follows. $\square$

LEMMA 4.5.6: Let $(\bar{x}(\cdot), \bar{u}(\cdot))$ be an admissible process and suppose that hypotheses (H1) to (H5) are satisfied (in relation to $(\bar{x}(\cdot), \bar{u}(\cdot))$). Let $\bar{x}(\cdot)$ solve $\dot{\bar{x}}(t) = f(t, \bar{x}(t), \bar{u}(t))$ and let $\bar{x}^{\epsilon_i}(\cdot)$ solve $\dot{\bar{x}}^{\epsilon_i}(t) = f(t, \bar{x}^{\epsilon_i}(t), \bar{u}(t))$ where $\bar{x}^{\epsilon_i}(0) = \bar{x}(0) + \epsilon_i \xi$, i.e. there is a perturbation in the initial condition.

Then $\epsilon_i^{-1}(\bar{x}^{\epsilon_i}(t) - \bar{x}(t))$ converges uniformly to $y(t)$ where $y(t)$ solves

$$\begin{aligned} \dot{y}(t) &= f_x(t, \bar{x}(t), \bar{u}(t)) y(t) \\ y(0) &= \xi. \end{aligned}$$

Proof: We know by the growth condition (hypothesis H3) and by the standard theory of ordinary differential equations that $\bar{x}^{\epsilon_i}(\cdot)$ is bounded and exists for each $\epsilon_i > 0$.

We wish to show that $(\bar{x}^{\epsilon_i}(t) - \bar{x}(t))$ is uniformly bounded.

By hypothesis H2 on the data we have that

$$|\dot{\bar{x}}^{\epsilon_i}(t) - \dot{\bar{x}}(t)| = |f(t, \bar{x}^{\epsilon_i}(t), \bar{u}(t)) - f(t, \bar{x}(t), \bar{u}(t))| \leq \alpha(t) |\bar{x}^{\epsilon_i}(t) - \bar{x}(t)| \text{ a.e. } t \in [0, 1].$$

We now pause to show that $|\frac{d}{dt}(\bar{x}^{\epsilon_i}(t) - \bar{x}(t))| = \frac{d}{dt}|\bar{x}^{\epsilon_i}(t) - \bar{x}(t)|$ a.e.

Setting $z(t) = \bar{x}^{\epsilon_i}(t) - \bar{x}(t)$, we have that

$$\frac{d}{dt}|z(t)| = \lim_{h \rightarrow 0} h^{-1} (|z(t+h)| - |z(t)|) = \begin{cases} \lim_{h \rightarrow 0} h^{-1} (z(t+h) - z(t)) & \text{if } z(t) > 0 \\ \lim_{h \rightarrow 0} -h^{-1} (z(t+h) - z(t)) & \text{if } z(t) < 0 \end{cases}$$

and $\frac{d}{dt}|z(t)| = \lim_{h \rightarrow 0} h^{-1}(|z(t+h)| - 0) = \lim_{h \rightarrow 0} |h^{-1}z(t+h)|$ if $z(t)=0$.

Thus, in general, at every point t where $z(t)$ and $|z(t)|$ are differentiable,

$$\left| \frac{d}{dt} z(t) \right| = \frac{d}{dt} |z(t)|.$$

But since $z(\cdot)$ and $|z(\cdot)|$ are Lipschitz continuous, they are differentiable almost everywhere and this relationship is satisfied almost everywhere.

Therefore we can apply Gronwall's lemma to obtain

$$|\bar{x}^{\epsilon_i}(t) - \bar{x}(t)| \leq |\bar{x}^{\epsilon_i}(0) - \bar{x}(0)| \exp \left\{ \int_0^t \alpha(s) ds \right\}, \quad (4.5.12)$$

where $|\bar{x}^{\epsilon_i}(0) - \bar{x}(0)| = \epsilon_i \xi$.

Thus we have shown that $\bar{x}^{\epsilon_i}(t) - \bar{x}(t)$ is uniformly bounded.

We know that the sequence $\{\epsilon_i\}$ tends to zero. Consequently the right-hand side of (4.5.12) tends to zero for all $t \in [0, 1]$ and therefore $\bar{x}^{\epsilon_i}(t)$ tends to $\bar{x}(t)$ uniformly as $\epsilon_i \rightarrow 0$.

Furthermore, we note that

$$|\epsilon_i^{-1}(\bar{x}^{\epsilon_i}(t) - \bar{x}(t))| \leq \xi \exp \left\{ \int_0^t \alpha(s) ds \right\}$$

and is therefore bounded for each i .

We now wish to show that $\epsilon_i^{-1}(\bar{x}^{\epsilon_i}(t) - x(t))$ converges to $y(t)$.

Using an exact first order Taylor expansion of the data about $\bar{x}(t)$, we have

$$\frac{f(t, \bar{x}^{\epsilon_i}(t), \bar{u}(t)) - f(t, \bar{x}(t), \bar{u}(t))}{\epsilon_i} = \frac{\Phi_i(t)(\bar{x}^{\epsilon_i}(t) - \bar{x}(t))}{\epsilon_i}$$

$$\text{where } \Phi_i(t) := \int_0^1 f_x(t, \bar{x}(t) + \lambda(\bar{x}^{\epsilon_i}(t) - \bar{x}(t)), \bar{u}(t)) d\lambda.$$

Hence

$$\frac{\bar{x}^{\epsilon_i}(t) - \bar{x}(t)}{\epsilon_i} = \frac{\Phi_i(t)(\bar{x}^{\epsilon_i}(t) - \bar{x}(t))}{\epsilon_i}.$$

Now define

$$y^{\epsilon_i}(t) = \frac{\bar{x}^{\epsilon_i}(t) - \bar{x}(t)}{\epsilon_i} \quad \text{then} \quad \dot{y}^{\epsilon_i}(t) = \frac{\dot{\bar{x}}^{\epsilon_i}(t) - \dot{\bar{x}}(t)}{\epsilon_i}.$$

We have that $y^{\epsilon_i}(\cdot)$ solves

$$\dot{y}^{\epsilon_i}(t) = \Phi_i(t) y^{\epsilon_i}(t) \quad \text{where } y^{\epsilon_i}(0) = \xi. \quad (4.5.13)$$

Now $\Phi_i(t) \rightarrow f_x(t, \bar{x}(t), \bar{u}(t))$ almost everywhere since $\bar{x}^{\epsilon_i}(t) \rightarrow \bar{x}(t)$ for all t .

We deduce from the differential equation (4.5.13) satisfied by the $y^{\epsilon_i}(\cdot)$'s that they are in fact uniformly Lipschitz continuous. Standard compactness arguments coupled with

the above convergence properties now yield

$$y^{\epsilon_i}(t) \rightarrow y(t) \text{ uniformly,}$$

where $y(\cdot)$ satisfies

$$\dot{y}(s) = f_x(t, \bar{x}(t), \bar{u}(t))y(t)$$

$$y(0) = \xi,$$

$$\text{i.e. } \frac{\bar{x}^{\epsilon_i}(t) - \bar{x}(t)}{\epsilon_i} \rightarrow y(t) \text{ uniformly.}$$

□

We now continue with our proof of the convergence properties of $\bar{V}$.

PROOF OF PROPOSITION 4.5.3

We can write

$$\begin{aligned} & \lim_{\epsilon_i \downarrow 0} \epsilon_i^{-1} \left\{ \bar{V}(t + \epsilon_i \tau_i, \bar{x}(t) + \epsilon_i a_i) - \bar{V}(t, \bar{x}(t)) \right\} \\ &= \lim_{\epsilon_i \downarrow 0} \epsilon_i^{-1} \left\{ \bar{V}(t + \epsilon_i \tau_i, \bar{x}(t + \epsilon_i \tau_i)) - \bar{V}(t, \bar{x}(t)) + \bar{V}(t + \epsilon_i \tau_i, \bar{x}(t) + \epsilon_i a_i) - \bar{V}(t + \epsilon_i \tau_i, \bar{x}(t + \epsilon_i \tau_i)) \right\} \\ &= \lim_{\epsilon_i \downarrow 0} \{A_i + B_i\} \end{aligned}$$

$$\text{where } A_i := \epsilon_i^{-1} \left\{ \bar{V}(t + \epsilon_i \tau_i, \bar{x}(t + \epsilon_i \tau_i)) - \bar{V}(t, \bar{x}(t)) \right\}$$

$$B_i := \epsilon_i^{-1} \left\{ \bar{V}(t + \epsilon_i \tau_i, \bar{x}(t) + \epsilon_i a_i) - \bar{V}(t + \epsilon_i \tau_i, \bar{x}(t + \epsilon_i \tau_i)) \right\}.$$

From the definition of $\bar{V}(\cdot, \cdot)$, (equation 4.5.1), we deduce that

$$\begin{aligned} A_i &= \epsilon_i^{-1} \left\{ \int_{t + \epsilon_i \tau_i}^t L(s, \bar{x}(s), \bar{u}(s)) ds - \int_t^t L(s, \bar{x}(s), \bar{u}(s)) ds \right\} \\ &= \epsilon_i^{-1} \left\{ - \int_t^{t + \epsilon_i \tau_i} L(s, \bar{x}(s), \bar{u}(s)) ds \right\}. \end{aligned}$$

Now since t is a Lebesgue point of $s \mapsto L(s, \bar{x}(s), \bar{u}(s))$, it follows that

$$\lim_{\substack{i \rightarrow \infty \\ \epsilon_i \rightarrow 0}} \epsilon_i^{-1} \tau_i^{-1} \int_t^{t + \epsilon_i \tau_i} L(s, \bar{x}(s), \bar{u}(s)) ds = L(t, \bar{x}(t), \bar{u}(t)).$$

Hence,

$$\lim_{i \rightarrow \infty} \epsilon_i^{-1} \left\{ \bar{V}(t + \epsilon_i \tau_i, \bar{x}(t + \epsilon_i \tau_i)) - \bar{V}(t, \bar{x}(t)) \right\}$$

$$\begin{aligned}
 &= \lim_{i \rightarrow \infty} \epsilon_i^{-1} \left\{ - \int_t^{t+\epsilon_i \tau_i} L(s, \bar{x}(s), \bar{u}(s)) ds \right\} \\
 &= -\tau L(t, \bar{x}(t), \bar{u}(t)).
 \end{aligned}$$

We turn next to analysing (B_i) in the limit. But first some new notation is required. Define the augmented state variable $\tilde{x}(\cdot):[0,1] \rightarrow \mathbb{R}^{n+1}$ by:

$$\tilde{x}(t) = \begin{bmatrix} x_{n+1}(t) \\ x(t) \end{bmatrix},$$

and define the functions $\tilde{F}: [0,1] \times \mathbb{R}^{n+1} \times \mathbb{R}^m \rightarrow \mathbb{R}^{n+1}$, $G: \mathbb{R}^{n+1} \rightarrow \mathbb{R}$, by:

$$\tilde{F}(t, \tilde{x}(t), \bar{u}(t)) = \begin{bmatrix} L(t, x(t), \bar{u}(t)) \\ f(t, x(t), \bar{u}(t)) \end{bmatrix}, \quad G(\tilde{x}) = x_{n+1}.$$

With this notation, we see that $\bar{V}(t, \xi) = G(\tilde{x}(1))$ where $\tilde{x}(\cdot)$ solves:

$$\dot{\tilde{x}}(s) = \begin{bmatrix} \dot{x}_{n+1}(s) \\ \dot{x}(s) \end{bmatrix} = \tilde{F}(s, \tilde{x}(s), \bar{u}(s)) \quad \text{a.e. } s \in [t, 1],$$

$$\tilde{x}(t) = \begin{bmatrix} 0 \\ \xi \end{bmatrix},$$

Furthermore, we define the system:

$$\dot{\hat{y}}(s) = \tilde{F}_{\tilde{x}} \cdot \hat{y}(s) \quad \text{a.e. } s \in [t, 1]$$

$$\hat{y}(t) = \begin{bmatrix} 0 \\ a - \tau f(t, \bar{x}(t), \bar{u}(t)) \end{bmatrix},$$

where $\tilde{F}_{\tilde{x}}$ is the $(n+1) \times (n+1)$ block matrix $\begin{bmatrix} 0 & L_x \\ 0 & f_x \end{bmatrix}$.

We denote by $X(1, t)$ the $(n+1) \times (n+1)$ transition matrix for $\dot{\hat{y}}(s) = \tilde{F}_{\tilde{x}} \cdot \hat{y}(s)$ which is uniquely defined by the following properties:

$$\frac{d}{d\sigma} X(\tau, \sigma) = -X(\tau, \sigma) \tilde{F}_{\tilde{x}}$$

$$X(\tau, \tau) = I \text{ where } I \text{ is the identity matrix.}$$

Let f denote $f(t, \bar{x}(t), \bar{u}(t))$. We have

$$\begin{aligned}
 (B_i) &= \frac{1}{\epsilon_i} \left\{ \bar{V}(t + \epsilon_i \tau_i, \bar{x}(t) + \epsilon_i a_i) - \bar{V}(t + \epsilon_i \tau_i, \bar{x}(t + \epsilon_i \tau_i)) \right\} \\
 &= \frac{1}{\epsilon_i} \left\{ \bar{V}(t + \epsilon_i \tau_i, \bar{x}(t) + \epsilon_i a_i) - \bar{V}(t + \epsilon_i \tau_i, \bar{x}(t) + \epsilon_i \tau f) \right\} + O(\epsilon_i)
 \end{aligned}$$

by Lemma 4.5.4

$$= \frac{1}{\epsilon_i} \left\{ \bar{V}(t+\epsilon_i\tau, \bar{x}(t)+\epsilon_i a) - \bar{V}(t+\epsilon_i\tau, \bar{x}(t)+\epsilon_i\tau f) \right\} + O(\epsilon_i)$$

by Lemma 4.5.5

$$= \frac{1}{\epsilon_i} \left\{ \bar{V}(t_i, y_i + \epsilon_i(a - \tau f)) - \bar{V}(t_i, y_i) \right\} + O(\epsilon_i)$$

where $y_i = \bar{x}(t) + \epsilon_i\tau f$, $t_i = t + \epsilon_i\tau$

$$= \frac{1}{\epsilon_i} \left\{ \bar{V}(t, \bar{x}(t) + \epsilon_i(a - \tau f)) - \bar{V}(t, \bar{x}(t)) \right\} + O(\epsilon_i)$$

$$= \frac{1}{\epsilon_i} \left\{ G(\tilde{x}^{\epsilon_i}(1)) - G(\tilde{x}(1)) \right\} + O(\epsilon_i)$$

by definition of G where $\tilde{x}^{\epsilon_i}(s)$ is the solution to the differential equation

$\dot{\tilde{x}}^{\epsilon_i}(s) = \tilde{F}(s, \tilde{x}(s), \bar{u}(s))$ a.e. $s \in [t, 1]$, with initial condition

$\tilde{x}^{\epsilon_i}(t) = [0, \bar{x}(t) + \epsilon_i(a - \tau f)]^T$

$$= \frac{1}{\epsilon_i} \left\{ G\left(\tilde{x}(1) + \frac{\tilde{x}^{\epsilon_i}(1) - \tilde{x}(1)}{\epsilon_i} \cdot \epsilon_i\right) - G(\tilde{x}(1)) \right\} + O(\epsilon_i)$$

$$= \frac{1}{\epsilon_i} \left\{ G(\tilde{x}(1) + \epsilon_i \tilde{y}^{\epsilon_i}(1)) - G(\tilde{x}(1)) \right\} + O(\epsilon_i)$$

where $\tilde{y}^{\epsilon_i}(s) := \frac{\tilde{x}^{\epsilon_i}(s) - \tilde{x}(s)}{\epsilon_i}$

$$= G_{\tilde{x}}(\tilde{x}(1) + O(\epsilon_i)) \cdot \tilde{y}^{\epsilon_i}(1) + O(\epsilon_i)$$

by the mean value theorem.

By Lemma 4.5.6 we are justified in passing to the limit. There results

$$\lim_{i \rightarrow \infty} B_i = G_{\tilde{x}}(\tilde{x}(1)) \tilde{y}(1)$$

$$= G_{\tilde{x}}(\tilde{x}(1)) X(1, t) \tilde{y}(t).$$

Define $\tilde{P}(t) = -G_{\tilde{x}}(\tilde{x}(1)) X(1, t)$.

Then $\dot{\tilde{P}}(t) = G_{\tilde{x}}(\tilde{x}(1)) X(1, t) \tilde{F}_{\tilde{x}}$ by the defining properties of the transition matrix,

i.e. $-\dot{\tilde{P}}(t) = \tilde{P}(t) \tilde{F}_{\tilde{x}}$.

Now let $\tilde{P}(t) = (q(t), p(t))$ where $q(\cdot): [0, 1] \rightarrow \mathbb{R}$, $p(\cdot): [0, 1] \rightarrow \mathbb{R}^n$.

Thus $-\dot{\tilde{P}}(t) = (-\dot{q}(t), -\dot{p}(t)) = (q(t), p(t)) \begin{bmatrix} 0 & L_x \\ 0 & f_x \end{bmatrix}$ which gives

$$-\dot{q}(t) = 0,$$

$$-\dot{p}(t) = q(t) \cdot L_x + p(t) \cdot f_x.$$

We know that $\tilde{P}(1) = -G_{\tilde{x}}(\tilde{x}(1))X(1,1) = [-1 \ 0 \ \cdots 0] I_{n+1 \times n+1}$

Hence $q(1) = -1$ and therefore $q(\cdot) \equiv -1$.

Thus $p(t)$ solves $-\dot{p}(t) = p(t) \cdot f_x(t, \bar{x}(t), \bar{u}(t)) - L_x(t, \bar{x}(t), \bar{u}(t))$
 $-p(1) = 0$.

We thus obtain the limiting value of expression (B_i) , namely

$$\begin{aligned} G_{\tilde{x}}(\tilde{x}(1))X(1,t)\tilde{y}(t) \\ &= -\tilde{P}(t) \cdot \tilde{y}(t) \\ &= -p(t) \cdot (a - \tau f(t, \bar{x}(t), \bar{u}(t))). \end{aligned}$$

Therefore, combining the results from (A_i) and (B_i) gives that at points t which are Lebesgue points of $s \mapsto \tilde{f}(s, \bar{x}(s), \bar{u}(s))$,

$$\begin{aligned} \lim_{\epsilon_i \rightarrow 0} \frac{1}{\epsilon_i} \{ &\bar{V}(t + \epsilon_i \tau_i, \bar{x}(t) + \epsilon_i a_i) - \bar{V}(t, \bar{x}(t)) \} \\ &= -\tau L(t, \bar{x}(t), \bar{u}(t)) - p(t) \cdot (a - \tau f(t, \bar{x}(t), \bar{u}(t))) \\ &= -p(t) \cdot a + (p(t) \cdot f(t, \bar{x}(t), \bar{u}(t)) - L(t, \bar{x}(t), \bar{u}(t))) \cdot \tau \\ &= (h(t), -p(t)) \cdot (\tau, a) \end{aligned}$$

where $h(t) = p(t) \cdot f(t, \bar{x}(t), \bar{u}(t)) - L(t, \bar{x}(t), \bar{u}(t))$.

This completes the proof of Proposition 4.5.3. $\square$

4.6 Proof of Theorem 4.4.1 in Fixed End-Point Case

The purpose of this section is to demonstrate the validity of Theorem 4.4.1 when the end-point constraint set takes the form:

$$C = \{x_1\}$$

for some $x_1 \in \mathbb{R}^n$. We consider then a special case of (P) in which the right end-point of the state trajectories considered are fixed.

The proof relies on the construction of an auxiliary problem for which we need the following cost function $\Phi(\xi)$ where

$$\Phi(\xi) := \inf \left\{ \int_0^1 L(t, x(t), u(t)) dt : \dot{x}(t) = f(t, x(t), u(t)), u(t) \in U, x(0) = x_0, x(1) = \xi \right\}. \quad (4.6.1)$$

Under the hypotheses, similar arguments to those employed by Clarke and Loewen [C4] show that Φ is Lipschitz continuous on some ball about $\xi = x_1$ and

$$\partial \Phi(\xi = x_1) \in \text{co}\{p(1) : p(\cdot) \in Q\}$$

where Q is the set of all costate functions (associated with a unity cost multiplier) corresponding to some optimal control for problem (P). But under our strengthened normality hypothesis the set contains only one element. It follows that Φ is Lipschitz

continuous on a neighbourhood of x_1 and is strictly differentiable at $\xi=x_1$ (see Proposition 2.3.3). Fix $\epsilon>0$ such that Φ is Lipschitz continuous on the ϵ -ball about x_1 .

For purposes of constructing a suitable verification function for (P) in this case, we introduce an auxiliary problem, labelled (P_{Aux}) , in which the right end-point constraint of the original problem is replaced by a penalty term ($\hat{\Phi}$) which is Lipschitz continuous in a neighbourhood of x_1 and strictly differentiable at the point x_1 itself.

$$\begin{aligned} & \text{Minimize } \int_0^1 L(t, x(t), u(t)) dt - \hat{\Phi}(x(1)) + K|x(1) - x_1|^2 + \hat{\Phi}(x_1) \\ & \text{subject to } \dot{x}(t) = f(t, x(t), u(t)) \quad \text{a.e. } t \in [0, 1], \\ & \quad u(t) \in U \quad \text{a.e. } t \in [0, 1], \\ & \quad x(0) = x_0, \\ & \text{where } \hat{\Phi}(\xi) = \begin{cases} \Phi(\xi) & \text{if } \|\xi - x_1\| < \epsilon \\ \Phi\left(x_1 + \frac{\xi - x_1}{\|\xi - x_1\|} \cdot \epsilon\right) & \text{if } \|\xi - x_1\| \geq \epsilon. \end{cases} \end{aligned} \quad (P_{Aux})$$

Here K is a positive number, whose value will be chosen later.

We define the value function for the auxiliary problem to be

$$V_A(s, \xi) := \inf \left\{ \int_s^1 L(t, x(t), u(t)) dt - \hat{\Phi}(x(1)) + K|x(1) - x_1|^2 + \hat{\Phi}(x_1) \right\} \quad (4.6.2)$$

where the infimum is taken over all processes $(x(\cdot), u(\cdot))$ subject to the constraints of Problem (P_{Aux}) except that the underlying time interval $[0, 1]$ is replaced by $[s, 1]$ and the initial state x_0 by ξ .

Our goal is to show that $V_A(s, \xi)$ is a verification function for the original problem, i.e. that it satisfies the conditions of Definition 4.2.1 and that it also satisfies equations (4.4.1) and (4.4.2) and is therefore a suitable candidate function for the ϕ of the theorem. However, before doing so we show that a minimizing solution to problem (P) is, in fact, also a minimizing solution to (P_{Aux}) and thus the two problems are indeed equivalent.

4.6.1 PRELIMINARY RESULTS

In what follows, we assume that the data satisfy hypotheses (H1) to (H5) and the strengthened normality hypothesis.

PROPOSITION 4.6.1: *A minimizer $(x^*(\cdot), u^*(\cdot))$ for the original problem (P) is a global*

minimizer for the auxiliary problem (P_{Aux}) provided K is sufficiently large.

Proof: Firstly, we show that for all admissible pairs $(x(\cdot), u(\cdot))$, there exists k such that

$$\left| \int_0^1 L(t, x(t), u(t)) dt - \hat{\Phi}(x(1)) + \hat{\Phi}(x_1) \right| \leq k. \quad (4.6.3)$$

From the assumptions on the data (refer to hypothesis (H3)),

$$|\tilde{f}(t, x(t), u(t))| \leq a + b|x(t)|,$$

therefore

$$|\dot{x}(t)| \leq a + b|x(t)|, \quad x(0) = x_0,$$

and by Gronwall's lemma this implies the existence of a constant K' such that $|x(t)| \leq K'$ for all t .

We also deduce the existence of K'' such that $|\hat{\Phi}(x(1))| \leq K''$ for all x , and so

$$\left| \int_0^1 L(t, x(t), u(t)) dt - \hat{\Phi}(x(1)) + \hat{\Phi}(x_1) \right| \leq a + bK' + 2K''.$$

Setting $k = a + bK' + 2K''$, we have shown that the bound exists for all admissible pairs.

Now choose K such that $K\epsilon^2 > 2k$.

We wish to show that with this choice of K , $(x^*(\cdot), u^*(\cdot))$, the minimizer for the original problem, is also a minimizer for the auxiliary problem. We introduce the following notation:

$$\begin{aligned} \text{Let } \eta(x, u) &= \int_0^1 L(t, x(t), u(t)) dt \text{ and let} \\ \hat{\eta}(x, u) &= \int_0^1 L(t, x(t), u(t)) dt - \hat{\Phi}(x(1)) + K|x(1) - x_1|^2 + \hat{\Phi}(x_1). \end{aligned}$$

Under the hypotheses, the auxiliary problem has a minimizer $(\hat{x}(\cdot), \hat{u}(\cdot))$. We show that $|x(1) - x_1| \leq \epsilon$.

Suppose to the contrary. Then $\hat{\eta}(\hat{x}, \hat{u}) > -k + K\epsilon^2 > k > \eta(x^*, u^*)$, (from (4.6.3) and since $K\epsilon^2 > 2k$) but by optimality of $(\hat{x}, \hat{u})$ we have that $\hat{\eta}(x^*, u^*) \geq \hat{\eta}(\hat{x}, \hat{u})$.

However,

$$\begin{aligned} \hat{\eta}(x^*, u^*) &= \int_0^1 L(t, x^*(t), u^*(t)) dt - \hat{\Phi}(x^*(1)) + K|x^*(1) - x_1|^2 + \hat{\Phi}(x_1) \\ &= \eta(x^*, u^*), \text{ since on the optimal trajectory } x^*(1) = x_1. \end{aligned}$$

This gives $\eta(x^*, u^*) = \hat{\eta}(x^*, u^*) \geq \hat{\eta}(\hat{x}, \hat{u}) > \eta(x^*, u^*)$. From this contradiction we conclude that $|x(1) - x_1| \leq \epsilon$.

By definition of Φ , $\eta(x^*, u^*) = \Phi(x_1)$, and also, for any (x, u) ,

$$\int_0^1 L(t, x(t), u(t)) dt + K|x(1) - x_1|^2 \geq \Phi(x(1)).$$

Since Φ and $\hat{\Phi}$ coincide on $x_1 + \epsilon \mathbb{B}$,

$$\eta(x^*, u^*) \leq \int_0^1 L(t, \hat{x}(t), \hat{u}(t)) dt - \Phi(\hat{x}(1)) + K|\hat{x}(1) - x_1|^2 + \Phi(x_1)$$

but $\eta(x^*, u^*) = \hat{\eta}(x^*, u^*)$ (since $x^*(1) = x_1$).

Hence we have obtained the result that $\eta(x^*, u^*) = \hat{\eta}(x^*, u^*) \leq \hat{\eta}(\hat{x}, \hat{u})$ where $\hat{\eta}(\hat{x}, \hat{u})$ is the minimum cost. Since $\hat{\eta}(x^*, u^*)$ cannot be less than the minimum cost, $(x^*(\cdot), u^*(\cdot))$ is thus a minimizer for the auxiliary problem. $\square$

We may write the dynamics in the form of the inclusion $\dot{x}(t) \in f(t, x(t), U)$ a.e. $t \in [0, 1]$ where the set on the right-hand side is $\{f(t, x(t), u) : u \in U\}$. We use this in the proof of the following lemma for which we recall some results from the theory of support functions and the characterization of convex sets, introduced in Chapter 2. We know that if $Q \subset \mathbb{R}^n$ is a closed convex set then, $x \in Q$ if and only if $\sup_{x \in Q} p \cdot x \geq p \cdot x$ for all $p \in \mathbb{R}^n$ (see Proposition 2.2.6).

LEMMA 4.6.2: *Let $(x(\cdot), u(\cdot))$ be any admissible process for P_{Aux} . Then*

$$\lim_{i \rightarrow \infty} \frac{1}{\epsilon_i} \int_t^{t+\epsilon_i} \tilde{f}(s, x(s), u(s)) ds \in \tilde{f}(t, x(t), U),$$

for any $t \in [0, 1)$ and any sequence $\{\epsilon_i\}$, $\epsilon_i \downarrow 0$, such that the limit on the left exists.

Proof: Since we assume $\tilde{f}(t, x(t), U)$ is a closed convex set, then by Proposition 2.2.6, we have that

$$\lim_{i \rightarrow \infty} \frac{1}{\epsilon_i} \int_t^{t+\epsilon_i} \tilde{f}(s, x(s), u(s)) ds \in \tilde{f}(t, x(t), U) \text{ if and only if}$$

$$\sup_{u \in U} p \cdot \tilde{f}(t, x(t), u) \geq \lim_{i \rightarrow \infty} \frac{1}{\epsilon_i} \int_t^{t+\epsilon_i} p \cdot \tilde{f}(s, x(s), u(s)) ds \text{ for all } p \in \mathbb{R}^n.$$

Conversely,

$$\lim_{i \rightarrow \infty} \frac{1}{\epsilon_i} \int_t^{t+\epsilon_i} \tilde{f}(s, x(s), u(s)) ds \notin \tilde{f}(t, x(t), U) \text{ if and only if there exists some } p \in \mathbb{R}^n \text{ such that}$$

$$\sup_{u \in U} p \cdot \tilde{f}(t, x(t), u) < \lim_{i \rightarrow \infty} \frac{1}{\epsilon_i} \int_t^{t+\epsilon_i} p \cdot \tilde{f}(s, x(s), u(s)) ds.$$

Now let us assume that the assertion of the lemma is not true. Then there exists $p \in \mathbb{R}^n$ and $\delta > 0$ such that, for all $u \in U$

$$p \cdot \tilde{f}(t, x(t), u) < \lim_{i \rightarrow \infty} \frac{1}{\epsilon_i} \int_t^{t+\epsilon_i} p \cdot \tilde{f}(s, x(s), u(s)) ds - \delta.$$

However,

$$\begin{aligned} & \lim_{i \rightarrow \infty} \frac{1}{\epsilon_i} \int_t^{t+\epsilon_i} p \cdot \tilde{f}(s, x(s), u(s)) ds \\ & \leq \lim_{i \rightarrow \infty} \frac{1}{\epsilon_i} \int_t^{t+\epsilon_i} \max_{u \in U} p \cdot \tilde{f}(s, x(s), u) ds \\ & = \lim_{i \rightarrow \infty} \frac{1}{\epsilon_i} \int_t^{t+\epsilon_i} \max_{u \in U} p \cdot \tilde{f}(t, x(t), u) ds \end{aligned}$$

since $t \mapsto \max_{u \in U} p \cdot \tilde{f}(t, x(t), u)$ is continuous

$$= \min_{u \in U} p \cdot \tilde{f}(t, x(t), u).$$

However, since $p \cdot \tilde{f}(t, x(t), u) < \lim_{i \rightarrow \infty} \frac{1}{\epsilon_i} \int_t^{t+\epsilon_i} p \cdot \tilde{f}(s, x(s), u(s)) ds - \delta$ for all u , then

$$p \cdot \tilde{f}(t, x(t), u) < \max_{u \in U} p \cdot \tilde{f}(t, x(t), u) - \delta \quad \text{for all } u.$$

This is obviously a contradiction and hence the lemma is proved. $\square$

In effect, under the hypotheses of the lemma there exists $u \in U$ such that

$$\lim_{i \rightarrow \infty} \frac{1}{\epsilon_i} \int_t^{t+\epsilon_i} \tilde{f}(s, x(s), u(s)) ds = \tilde{f}(t, x(t), u). \quad (4.6.4)$$

We deduce:

LEMMA 4.6.3: *Given any admissible process $(x(\cdot), u(\cdot))$ for (P_{Ax}) , any time $t \in [0, 1]$ which is a Lebesgue point for $s \mapsto f(s, x(s), u(s))$ and any sequence $\{\epsilon_i\}$, $\epsilon_i \downarrow 0$, then the function V_A satisfies*

$$\lim_{i \rightarrow \infty} \frac{1}{\epsilon_i} \left| V_A \left(t + \epsilon_i, x(t) + \int_t^{t+\epsilon_i} f(s, x(s), u(s)) ds \right) - V_A \left(t + \epsilon_i, x(t) + \epsilon_i f(t, x(t), u) \right) \right| = 0.$$

for some $u \in U$.

Proof: By Lipschitz continuity of $V_A(t, x(t))$ there exists $\mathfrak{K} > 0$ such that for each i ,

$$\begin{aligned} & \frac{1}{\epsilon_i} \left| V_A\left(t + \epsilon_i, x(t) + \int_t^{t+\epsilon_i} f(s, x(s), u(s)) ds\right) - V_A\left(t + \epsilon_i, x(t) + \epsilon_i f(t, x(t), u)\right) \right| \\ & \leq \frac{1}{\epsilon_i} \mathfrak{K} \epsilon_i \left| \frac{1}{\epsilon_i} \int_t^{t+\epsilon_i} f(s, x(s), u(s)) ds - f(t, x(t), u) \right| \rightarrow 0 \text{ as } i \rightarrow \infty \end{aligned}$$

by (4.6.4).

The assertion follows. $\square$

We make use of this result in the next section for demonstrating that the value function V_A satisfies the conditions of a verification function.

4.6.2 PROOF OF NECESSARY CONDITIONS

The value function satisfies condition (a) in Definition (4.2.1). We first need to demonstrate the validity of the Hamilton-Jacobi equality and the associated boundary condition for the fixed end-point case (namely $V_A(1, \xi) \leq 0$).

Take any $(t, x) \in [0, 1) \times \mathbb{R}^n$ and $u \in U$. Let $x(s)$, $t \leq s \leq 1$, solve $\dot{x}(s) = f(s, x(s), u(s) \equiv u)$ with $x(t) = x$. By the principle of optimality applied to the control policy $u(\cdot) \equiv u$ on $[t, 1]$, we have, for each i ,

$$V_A(t + \epsilon_i, x(t + \epsilon_i)) - V_A(t, x) + \int_t^{t+\epsilon_i} L(s, x(s), u) ds \geq 0$$

(the result is unaffected by the end-point constraint terms).

Therefore,

$$\begin{aligned} 0 & \leq \liminf_{i \rightarrow \infty} \left\{ \frac{1}{\epsilon_i} \left\{ V_A(t + \epsilon_i, x(t + \epsilon_i)) - V_A(t, x) \right\} + \frac{1}{\epsilon_i} \int_t^{t+\epsilon_i} L(s, x(s), u) ds \right\} \\ & = \liminf_{i \rightarrow \infty} \left\{ \frac{1}{\epsilon_i} \left\{ V_A\left(t + \epsilon_i, x + \int_t^{t+\epsilon_i} \dot{x}(s) ds\right) - V_A(t, x) \right\} + \frac{1}{\epsilon_i} \int_t^{t+\epsilon_i} L(s, x(s), u) ds \right\} \\ & = \liminf_{i \rightarrow \infty} \left\{ \frac{1}{\epsilon_i} \left\{ V_A\left(t + \epsilon_i, x + \epsilon_i f(t, x, u)\right) - V_A(t, x) \right\} + \frac{1}{\epsilon_i} \int_t^{t+\epsilon_i} L(s, x(s), u) ds \right\} \end{aligned}$$

by Lipschitz continuity

$$= \liminf_{\substack{i \rightarrow \infty \\ (\tau_i, a_i) \rightarrow (1, f)}} \left\{ \frac{1}{\epsilon_i} \left\{ V_A(t + \epsilon_i \tau_i, x + \epsilon_i a_i) - V_A(t, x) \right\} + \frac{1}{\epsilon_i} \int_t^{t+\epsilon_i} L(s, x(s), u) ds \right\}$$

by Lipschitz continuity and where the $\liminf$ has quantifiers $(\tau_i, a_i) \rightarrow (1, f(t, x, u))$

$$= d^- V_A(t, x; 1, f(t, x, u)) + L(t, x, u)$$

by definition of $d^- V_A$ and since t is a Lebesgue point of $s \rightarrow L(s, x(s), u)$.

It follows that

$$d^- V_A(t, x; 1, f(t, x, u)) + L(t, x, u) \geq 0.$$

Since u was an arbitrary point in U ,

$$\inf_{u \in U} d^- V_A(t, x; 1, f(t, x, u)) + L(t, x, u) \geq 0.$$

We now prove the reverse inequality. Again fix $(t, x) \in [0, 1] \times \mathbb{R}^n$. Let $u(\cdot): [t, 1] \rightarrow \mathbb{R}^m$ solve a variant of the auxiliary problem $(P_{A_{ux}})$, in which (t, x) replaces $(0, x_0)$. We have

$$0 = \lim_{i \rightarrow \infty} \frac{1}{\epsilon_i} \left\{ V_A(t + \epsilon_i, x(t + \epsilon_i)) - V_A(t, x) + \int_t^{t + \epsilon_i} L(s, x(s), u(s)) ds \right\}$$

by the Principle of Optimality

$$= \lim_{i \rightarrow \infty} \frac{1}{\epsilon_i} \left\{ V_A(t + \epsilon_i, x + \int_t^{t + \epsilon_i} f(s, x(s), u(s)) ds) - V_A(t, x) + \int_t^{t + \epsilon_i} L(s, x(s), u(s)) ds \right\}$$

$$= \lim_{i \rightarrow \infty} \left\{ \frac{1}{\epsilon_i} \left\{ V_A(t + \epsilon_i, x + \epsilon_i f(t, x, \bar{u})) - V_A(t, x) \right\} \right\} + L(t, x, \bar{u})$$

for some $\bar{u} \in U$, by Lemma 4.6.3

$$\geq \liminf_{i \rightarrow \infty} \left\{ \frac{1}{\epsilon_i} \left\{ V_A(t + \epsilon_i, x + \epsilon_i f(t, x, \bar{u})) - V_A(t, x) \right\} \right\} + L(t, x, \bar{u})$$

$$= d^- V_A(t, x; 1, f(t, x, \bar{u})) + L(t, x, \bar{u}) \text{ by definition of } d^- V_A$$

$$\geq \inf_{u \in U} d^- V_A(t, x; 1, f(t, x, u)) + L(t, x, u).$$

We have shown that

$$\inf_{u \in U} d^- V_A(t, x; 1, f(t, x, u)) + L(t, x, u) \leq 0.$$

Combining this with the earlier inequality we obtain

$$\inf_{u \in U} d^- V_A(t, x; 1, f(t, x, u)) + L(t, x, u) = 0.$$

This confirms that V_A satisfies equation (4.4.1), as well as the inequality in condition (c) of Definition 4.2.1.

To check the boundary condition at $x(1)=x_1$, we note

$$V_A(1, x_1) = \inf \left\{ \int_1^1 L(t, x, u) dt - \hat{\Phi}(x_1) + 0 + \hat{\Phi}(x_1) \right\} = 0.$$

So V_A certainly satisfies condition (b) of Definition 4.2.1.

Now take (x^*, u^*) to be the optimal process referred to in Theorem 4.4.1.

To verify equation (4.4.2) we have merely to note that (x^*, u^*) solves the auxiliary problem for initial data $(t=0, x=x_0)$ (see Proposition 4.6.1). By definition of V_A ,

$$V_A(0, x_0) = \int_0^1 L(t, x^*(t), u^*(t)) dt.$$

Thus equation (4.4.2) is verified and hence we have demonstrated that the value function for the auxiliary problem, which replaces the end-point constraint by a strictly differentiable penalty term, is a suitable choice of verification function for the theorem.

We finally wish to show that " $(h, -p) \in \partial^V V_A(t, x^*(t))$ " can be included in the optimality conditions satisfied by the optimal process (x^*, u^*) .

Take any point $t \in [0, 1]$ which is a Lebesgue point of $s \mapsto \tilde{f}(s, x^*(s), u^*(s))$ and arbitrary convergent sequences $a_i \rightarrow a$, $\tau_i \rightarrow \tau$ and $\epsilon_i \downarrow 0$, in $\mathbb{R}^n$, $\mathbb{R}$ and $\mathbb{R}^+$ respectively.

From Lemma 4.5.1 which states that $V_A(t, x(t)) \leq \bar{V}_A(t, x(t))$ where V_A is the value function and $\bar{V}_A$ is the cost function, we obtain, for each i ,

$$\frac{1}{\epsilon_i} \left\{ V_A(t + \epsilon_i \tau_i, x^*(t) + \epsilon_i a_i) - V_A(t, x^*(t)) \right\} \leq \frac{1}{\epsilon_i} \left\{ \bar{V}_A(t + \epsilon_i \tau_i, x^*(t) + \epsilon_i a_i) - \bar{V}_A(t, x^*(t)) \right\}.$$

However, by Proposition 4.5.3 the right-hand side has limit

$$(p(t) \cdot f(t, x^*(t), u^*(t)) - L(t, x^*(t), u^*(t))) \cdot \tau - p(t) \cdot a$$

where $p(t)$ satisfies the costate equation (4.4.3).

Since the sequences are arbitrary, this gives

$$\begin{aligned} \limsup_{\substack{\epsilon_i \downarrow 0 \\ (\tau_i, a_i) \rightarrow (\tau, a)}} \frac{1}{\epsilon_i} \left\{ V_A(t + \epsilon_i \tau_i, x^*(t) + \epsilon_i a_i) - V_A(t, x^*(t)) \right\} \\ \leq (p(t) \cdot f(t, x^*(t), u^*(t)) - L(t, x^*(t), u^*(t)), -p(t)) \cdot (\tau, a). \end{aligned}$$

Since (τ, a) is arbitrarily obtained and since the set of Lebesgue points of an integrable function has full measure (and since the union of two sets of full measure is a set of full measure) then, by the characterization of the superdifferential (Lemma 2.4.2)

$$\left((p(t) \cdot f(t, x^*(t), u^*(t)) - L(t, x^*(t), u^*(t))), -p(t) \right) \in \partial^V V_A(t, x^*(t)) \quad \text{a.e. } t \in [0, 1].$$

To complete the proof of necessity, we apply the maximum principle. The solution to the costate equation (4.4.3) and the transversality condition (4.4.5) is unique under the

hypotheses, hence by Theorem 4.1.1 the function $p(t)$ must also satisfy the maximization of the Hamiltonian (equation (4.4.4)), i.e.

$$p(t) \cdot f(t, x^*(t), u^*(t)) - L(t, x^*(t), u^*(t)) = \max_{u \in U} p(t) \cdot f(t, x^*(t), u) - L(t, x^*(t), u) \text{ a.e.}$$

This is the last relationship to be verified. $\square$

4.6.3 PROOF OF SUFFICIENT CONDITIONS

Suppose (x^*, u^*) is a process satisfying equations (4.4.7) to (4.4.9). We now wish to show that $(h, -p) \in \partial^\vee \phi(t, x^*(t))$ is a sufficient condition for optimality in conjunction with the equations of the maximum principle.

We choose a point $t \in [0, 1]$ which is both a Lebesgue point for $s \mapsto \tilde{f}(s, x^*(s), u^*(s))$ i.e.

$$\frac{1}{\epsilon_i} \int_t^{t+\epsilon_i} \tilde{f}(s, x^*(s), u^*(s)) ds \rightarrow \tilde{f}(t, x^*(t), u^*(t)),$$

and is such that

$$(p(t) \cdot f(t, x^*(t), u^*(t)) - L(t, x^*(t), u^*(t)), -p(t)) \in \partial^\vee \phi(t, x^*(t)).$$

We assume, in addition, that $s \mapsto \phi(s, x^*(s))$ is differentiable at t , i.e. that the derivative $\frac{d\phi}{dt}(t, x^*(t))$ exists. The subset of points in $[0, 1]$ formed with these properties is a set of full measure. Then,

$$\frac{d}{dt} \phi(t, x^*(t)) = \lim_{\epsilon_i \downarrow 0} \frac{1}{\epsilon_i} \left\{ \phi\left(t + \epsilon_i, x^*(t) + \int_t^{t+\epsilon_i} f(s, x^*(s), u^*(s)) ds\right) - \phi(t, x^*(t)) \right\}$$

$$\leq \limsup \frac{1}{\epsilon_i} \left\{ \phi\left(t + \epsilon_i, x^*(t) + \epsilon_i a_i\right) - \phi(t, x^*(t)) \right\}$$

where the $\limsup$ has quantifiers $\epsilon_i \downarrow 0, (\tau_i, a_i) \rightarrow (1, f(t, x^*(t), u^*(t)))$

$$\leq (p(t) \cdot f(t, x^*(t), u^*(t)) - L(t, x^*(t), u^*(t)), -p(t)) \cdot (1, f(t, x^*(t), u^*(t)))$$

by the inclusion and by the characterization of the superdifferential (Lemma 2.4.2)

$$= p(t) \cdot f(t, x^*(t), u^*(t)) - L(t, x^*(t), u^*(t)) - p(t) \cdot f(t, x^*(t), u^*(t))$$

$$= -L(t, x^*(t), u^*(t)).$$

$$\text{Therefore } \frac{d}{dt} \phi(t, x^*(t)) + L(t, x^*(t), u^*(t)) \leq 0 \text{ a.e.} \quad (4.6.5)$$

However, choose any control function $u(\cdot)$ with corresponding state trajectory $x(\cdot)$ and take t which is a Lebesgue point of $s \mapsto f(s, x(s), u(s))$ and which is such that $s \mapsto \frac{d}{ds} \phi(s, x(s))$ is differentiable and $u(t) \in U$. Then

$$\frac{d\phi}{dt}(t, x(t)) = \lim_{\epsilon_i \downarrow 0} \frac{1}{\epsilon_i} \left\{ \phi(t + \epsilon_i, x(t + \epsilon_i)) - \phi(t, x(t)) \right\}$$

$$\begin{aligned}
&= \lim_{\epsilon_i \downarrow 0} \frac{1}{\epsilon_i} \left\{ \phi(t+\epsilon_i, x(t) + \int_t^{t+\epsilon_i} \dot{x}(s) ds) - \phi(t, x(t)) \right\} \\
&= \lim_{\epsilon_i \downarrow 0} \frac{1}{\epsilon_i} \left\{ \phi(t+\epsilon_i, x(t) + \epsilon_i f(t, x(t), u)) - \phi(t, x(t)) \right\} \quad \text{by Lipschitz continuity} \\
&= \liminf_{\epsilon_i \downarrow 0} \frac{1}{\epsilon_i} \left\{ \phi(t+\epsilon_i, x(t) + \epsilon_i f(t, x(t), u)) - \phi(t, x(t)) \right\} \\
&= d^- \phi(t, x(t); 1, f(t, x(t), u)) \quad \text{by the definition of lower-Dini derivative.}
\end{aligned}$$

We can write $\phi(1, x_1) - \phi(0, x_0) = \int_0^1 \frac{d\phi}{dt}(t, x(t)) dt$.

Recalling the definition of a verification function, we know that the following conditions hold:

$$\min_{u \in U} d^- \phi(t, x; 1, f(t, x, u)) + L(t, x, u) \geq 0 \quad \text{and} \quad \phi(1, x_1) = 0, \quad \text{so that}$$

$$d^- \phi(t, x; 1, f(t, x, u)) + L(t, x, u) \geq 0.$$

$$\text{Then } \phi(0, x_0) = - \int_0^1 d^- \phi(t, x; 1, f(t, x, u)) dt \leq \int_0^1 L(t, x(t), u(t)) dt. \quad (4.6.6)$$

Hence $\phi(0, x_0)$ is a lower bound on the cost for an arbitrary control function.

In particular, $\phi(0, x_0) \leq \int_0^1 L(t, x^*(t), u^*(t)) dt$ for the given process $(x^*(\cdot), u^*(\cdot))$.

We also know from our earlier result (inequality 4.6.5) that for $(x^*(\cdot), u^*(\cdot))$

$$\frac{d}{dt} \phi(t, x^*(t)) + L(t, x^*(t), u^*(t)) \leq 0, \quad \text{and so}$$

$$d^- \phi(t, x^*(t); 1, f(t, x^*(t), u^*(t))) + L(t, x^*(t), u^*(t)) \leq 0.$$

$$\text{This gives } \phi(0, x_0) = - \int_0^1 d^- \phi(t, x(t); 1, f(t, x, u)) dt \geq \int_0^1 L(t, x^*(t), u^*(t)) dt.$$

Therefore we have shown that $(x^*(\cdot), u^*(\cdot))$ satisfies

$$d^- \phi(t, x^*(t); 1, f(t, x^*(t), u^*(t))) + L(t, x^*(t), u^*(t)) = 0,$$

$$\text{and } \phi(0, x_0) = \int_0^1 L(t, x^*(t), u^*(t)) dt \quad \text{for the given } (x^*(\cdot), u^*(\cdot)).$$

$$\text{Since } \phi(0, x_0) = \int_0^1 L(t, x^*(t), u^*(t)) dt \leq \int_0^1 L(t, x(t), u(t)) dt, \quad \text{by inequality (4.6.6),}$$

$\phi(0, x_0)$ is the minimum cost and $(x^*(\cdot), u^*(\cdot))$ is the optimal (minimizing) pair. $\square$

4.7 Proof of Theorem 4.4.1 for an End-point Constraint Set

In this section we complete the proof of Theorem 4.4.1 by relaxing the requirement that the end-point constraint set be a single point. As earlier foreseen, the approach we take is to modify the original problem so that it can be incorporated into a fixed end-point problem. We do this by extending the time interval over which the problem is defined and making use of the convexity properties of the constraint set. The following representation of compact, convex sets is required.

LEMMA 4.7.1: *Let $C \subset \mathbb{R}^n$ be a compact convex set.*

Then $C = \left\{ \int_0^1 u(s) ds : u(\cdot) \text{ is measurable, } u(t) \in C \text{ a.e. } t \right\}$.

Proof: Denote the set on the right hand side by $\mathfrak{C}$. The inclusion $C \subset \mathfrak{C}$ follows from the definition of the sets (and can be shown by considering constant functions u).

To prove the reverse inclusion, we suppose, by way of contradiction, that there exists $\xi \in \mathfrak{C} \setminus C$. Let $u(\cdot)$ be a measurable function such that $u(t) \in C$, a.e., and

$$\xi = \int_0^1 u(t) dt.$$

Since C is closed and $\xi \notin C$, $\{\xi\}$ and C are strictly separated, i.e. there exists a vector p such that $p' \xi > \max_{c \in C} p' c$, hence

$$p' \int_0^1 u(t) dt > \max_{c \in C} p' c.$$

This implies

$$\int_0^1 (p' u(t) - \max_{c \in C} p' c) ds > 0,$$

but this is not possible since the integrand is non-positive everywhere. It therefore follows that $\mathfrak{C} \subset C$.

Hence the equality is proved. □

We recall the original problem which is:

$$\begin{aligned} & \text{minimize } \int_0^1 L(t, x(t), u(t)) dt, \\ & \text{subject to} \end{aligned}$$

$$\dot{x}(t) = f(t, x(t), u(t)) \text{ a.e. } t \in [0, 1],$$

$$u(t) \in U \text{ a.e. } t \in [0, 1],$$

$$\text{and } x(0) = x_0, x(1) \in C.$$

The constraint set C is a closed convex set.

We re-define it as a fixed end-point problem, labelled $(\mathcal{P})$, over the extended time interval $t \in [0, 2]$, as follows.

$$\begin{aligned} & \text{Minimize } \int_0^2 \mathcal{L}(t, x(t), u(t)) dt, \\ & \text{subject to } \dot{x}(t) = \begin{cases} f(t, x(t), u(t)) & \text{a.e. } t \in [0, 1] \\ v & \text{a.e. } t \in (1, 2] \end{cases} \quad \text{where } -v \in C, \\ & u(t) \in U \quad \text{a.e. } t \in [0, 2], \\ & x(0) = x_0, x(2) = 0, \\ & \mathcal{L}(t, x(t), u(t)) = \begin{cases} L(t, x(t), u(t)) & \text{a.e. } t \in [0, 1] \\ 0 & \text{a.e. } t \in (1, 2]. \end{cases} \end{aligned} \quad (\mathcal{P})$$

Let (x^*, u^*) be a minimizing process for the original problem (P).

$$\text{Let } \chi = \begin{cases} x^*(t) & 0 \leq t \leq 1 \\ x^*(1) - (t-1)x^*(1) & 1 < t \leq 2 \end{cases}$$

with corresponding controls

$$\begin{aligned} v &= \begin{cases} u^*(t) & 0 \leq t \leq 1 \\ \text{arbitrary} & 1 < t \leq 2 \end{cases} \\ \nu &= \begin{cases} \text{arbitrary} & 0 \leq t \leq 1 \\ -x^*(1) & 1 < t \leq 2 \end{cases} \end{aligned}$$

Then (χ, v, ν) solves $(\mathcal{P})$ (the fixed end-point problem which we derive from the original problem). We retain our assumptions on the data and, in particular, the normality hypothesis on the data is still valid as we now prove.

PROPOSITION 4.7.2: *If the original problem satisfies the strengthened normality hypothesis (H) then so does the derived problem $(\mathcal{P})$.*

Proof: From the maximum principle, the costate equations are:

$$-\dot{p}(t) = p(t) \cdot f_x(t, x(t), u(t)) - L_x(t, x(t), u(t)) \quad \text{a.e. } t \in [0, 1]$$

$$-\dot{p}(t) = 0 \quad \text{a.e. } t \in (1, 2].$$

We can therefore set $p(t) \equiv p(1)$ for $t \geq 1$.

The maximization of the Hamiltonians gives :

$$\begin{aligned}
\max_{u \in U} \{p(t) \cdot f(t, x^*(t), u) - L(t, x^*(t), u)\} &= p(t) \cdot f(t, x^*(t), u^*(t)) - L(t, x^*(t), u^*(t)) \quad 0 \leq t \leq 1 \\
\max_{v \in -C} \{-p(t) \cdot v\} &= -p(t) \cdot \nu \\
&= -p(t) \cdot x^*(1) \quad \text{since } \nu = -x^*(1) \\
&= -p(1) \cdot x^*(1).
\end{aligned}$$

From the above equation

$$-p(t) \cdot \nu \geq -p(t) \cdot v \quad \text{for } -v \in C.$$

$$\text{Therefore } -p(1)[x^*(1) + v] \leq 0$$

$$\Rightarrow -p(1) \cdot x^*(1) \leq p(1) \cdot v$$

$$\Rightarrow -p(1) \cdot x^*(1) \leq -p(1) \cdot (-v) \quad \text{for all } -v \in C.$$

$$\text{Hence } -p(1) \in N_C(x^*(1)).$$

By the normality hypothesis which we apply to the first time interval we have a unique $p(\cdot)$ defined over this interval. Since $p(\cdot)$ is constant over the second time interval and $-p(1) \in N_C(x^*(1))$, then by constancy and the transversality condition at $t=1$, $p(\cdot)$ is unique over the extended time interval. Thus we have shown that the derived problem satisfies the normality hypothesis. $\square$

We wish to construct a verification function for the original problem and do so via consideration of the derived problem $(\mathcal{P})$ and an associated auxiliary problem $(\mathcal{P}_A)$ which we shall introduce. To do this we need the following definition:

$$\Phi(\xi) := \text{Inf} \left\{ \int_0^2 \mathcal{L}(t, x(t), u(t)) dt \right\} \quad (4.7.1)$$

where the infimum is taken over all processes $(x(\cdot), u(\cdot))$ subject to the constraints of problem $(\mathcal{P})$ except that $x(2) = \xi$.

Under the hypotheses, Φ is Lipschitz continuous on some neighbourhood about $\xi=0$, and $\partial\Phi(\xi=0) \in \bar{co} \{p(2): p(\cdot) \in Q\}$ where Q is the set of all costate functions (associated with a unity cost multiplier) corresponding to some optimal control for problem $(\mathcal{P})$. Thus we conclude (by Proposition 2.3.3) that Φ is strictly differentiable at $\xi=0$.

Fix $\epsilon > 0$ such that Φ is Lipschitz continuous on the ϵ -ball about 0.

$$\text{Define } \hat{\Phi}(\xi) = \begin{cases} \Phi(\xi) & \text{if } |\xi| < \epsilon \\ \Phi\left(\frac{\xi}{|\xi|} \cdot \epsilon\right) & \text{if } |\xi| \geq \epsilon. \end{cases}$$

We introduce the auxiliary problem $(\mathcal{P}_A)$ in which the end-point constraint is replaced by a strictly differentiable penalty term:

$$\text{minimize } \int_0^2 \mathcal{L}(t, x(t), u(t)) dt - \hat{\Phi}(x(2)) + K|x(2)|^2 + \hat{\Phi}(0),$$

$$\text{subject to } \dot{x}(t) = \begin{cases} f(t, x(t), u(t)) & \text{a.e. } t \in [0, 1] \\ v & \text{a.e. } t \in (1, 2], \end{cases} \quad (\mathcal{P}_A)$$

$$\begin{aligned} u(t) &\in U & \text{a.e. } t \in [0, 2], \\ x(0) &= x_0. \end{aligned}$$

Here K is a positive number whose value may be chosen.

We now define the value function for the auxiliary problem to be

$$\varphi_A(s, \xi) := \inf \left\{ \int_s^2 \mathcal{L}(t, x(t), u(t)) dt - \hat{\Phi}(x(2)) + K |x(2)|^2 + \hat{\Phi}(0) \right\} \quad (4.7.2)$$

where the infimum is taken over all processes $(x(\cdot), u(\cdot))$ subject to the constraints of Problem $(\mathcal{P}_A)$ except that the underlying time interval $[0, 2]$ is replaced by $[s, 2]$ and the initial state x_0 by ξ .

We wish to show that φ_A is a verification function for the original problem, i.e. that it satisfies

$$\min_{u \in U} \left\{ d^- \varphi_A((t, x); (1, f(t, x, u))) + L(t, x, u) \right\} = 0 \quad \text{a.e. } t \in [0, 1], \forall x \in \mathbb{R}^n \quad (4.7.3)$$

$$\varphi_A(1, x) \leq 0 \quad \forall x \in C \quad (4.7.4)$$

and

$$d^- \varphi_A(t, x^*(t); (1, f(t, x^*(t), u^*(t)))) + L(t, x^*(t), u^*(t)) = 0 \quad \text{a.e. } t \in [0, 1]. \quad (4.7.5)$$

Firstly we claim, as before, that for K sufficiently large (x^*, u^*) is a global minimizer for the auxiliary problem. The proof follows exactly as for Proposition 4.6.1. Furthermore we may quote the results from Section 4.6.2 which will verify that equations (4.7.3) and (4.7.5) are satisfied. It remains to verify the boundary condition (4.7.4).

For our problem $\dot{x} = v$ a.e. $t \in [1, 2]$, $-v \in C$ and $x(2) = 0$.

By definition of C , $-\int_1^2 v(s) ds \in C$ and we also know that $x(1) \in C$.

Thus we can write $x(2) = x(1) - \int_1^2 v(s) ds$ and therefore if we take any $x(1) \in C$ there exists

$-v \in C$ driving the state to zero, i.e. there exists $-v \in C$ such that $x(2) = 0$.

However, the corresponding cost is zero by definition of the auxiliary problem where $\mathcal{L} = 0$ over $[1, 2]$. Thus the infimum cost, $\varphi_A(1, x) \leq 0$ for all $x \in C$.

Hence we have shown that φ_A is a suitable candidate function for the ϕ of the theorem.

Furthermore, we may quote the results stating the necessity and sufficiency of the inclusion in conjunction with the maximum principle and the verification theorem since the proof follows that of Sections (4.6.2) and (4.6.3). The main result, namely Theorem 4.4.1, is proved. $\square$

CONCLUSION

This thesis is primarily concerned with two problems:

- 1) can we develop necessary conditions for optimality over the infinite horizon for multistage optimal control problems,
- 2) can we develop a 'sensitivity relation' in terms of a verification function for an optimal control problem with an end-point constraint set?

Our motivation for research into the first topic stems from application of such theory to problems in the economic and biological sciences which may be formulated over an unbounded time interval. Knowing that optimality conditions for multiprocess problems posed over a finite interval have already been derived, we have attempted to extend the theory so that it is valid over an infinite time interval and have obtained necessary conditions for multistage problems defined over the infinite time horizon. These necessary conditions are in the form of a maximum principle. We note that these conditions depend neither on the limiting value of the state trajectory as the time tends to infinity, nor on the convergence of the integral of the cost functional.

The techniques we employ yield necessary conditions of optimality under mild and readily verifiable hypotheses. They yield stronger results than the 'two-phase' optimization methods typical of the mathematical economics literature.

However, as a consequence of the free end-point at infinity, although one might expect that the transversality condition is such that the limiting value of the costate function at infinity is zero, it has been shown that the transversality condition for the infinite horizon case cannot necessarily be obtained as the limit of this condition in the finite horizon case [H1]. Therefore, the statement of the maximum principle for general classes of problems with an infinite horizon does not include the transversality condition at infinity. We might hope, however, to derive transversality conditions at infinity in special cases. Aubin and Clarke [A3] do just this for an infinite horizon problem involving linear dynamics. An interesting area of future research will be to extend Aubin and Clarke's work to cover problems with non-linear dynamics.

We have applied the theory to a general formulation of a two-stage optimal control problem in which an explicit dependence on the intermediate time (taken as a choice variable) is incorporated into the integrands of the cost functionals. We have shown, in these cases, that necessary conditions may be more easily obtained using multiprocess theory than via classical methods. Furthermore, we have demonstrated the

versatility of the theory by applying it to a specific example from the mathematical economics literature concerning optimal delivery lags, where the time lag is a choice variable, and it is likely that further research will yield more applications of the theory in the field of mathematical economics. The theory might also be applied to problems in mathematical biology.

The next part of the thesis contains an investigation into the relationship between the maximum principle of optimal control theory and the dynamic programming technique. The dynamic programming approach to optimal control theory involves characterizing the value function V as a solution to the Hamilton-Jacobi equation. Recent research has proved the validity of a relationship in terms of the Clarke generalized gradient of V , for a large class of non-smooth, free end-point problems. Results of this nature have been termed 'sensitivity relations' in the literature.

The question which we have addressed in this thesis is whether we can develop such a relationship for an optimal control problem with an end-point constraint set in terms of a verification function.

Our motivation for using verification functions has been that we require functions with similar properties to the value function, but which are defined and locally Lipschitz on a suitable large domain, even in the presence of end-point constraints. Furthermore we have attempted to use a different type of 'generalized gradient', namely the superdifferential, and our motivation for doing so has been that it provides a more precise sensitivity relation than the Clarke generalized gradient of the earlier literature on this topic. We note that while sensitivity relations involving superdifferentials have appeared elsewhere [S3], only free end-point problems have been treated.

The results that we have obtained are, that provided some normality hypothesis is satisfied (we provide sufficient verifiable conditions for this to be so), there is a sensitivity relation between the costate $p(\cdot)$ of the maximum principle, the Hamiltonian function H and some verification function W along a trajectory (x^*, u^*) , which is both necessary and sufficient for that trajectory to be optimal. Moreover, we have shown that the sufficient conditions are valid in a non-smooth setting.

Although for most practical purposes results for a convex constraint set are adequate, a possible extension of this work is to attempt to generalize the theory so that it is valid for arbitrary closed end-point constraint sets. Similarly, the question arises as to whether the results are still valid when the initial point belongs to some constraint set as opposed to being fixed.

We conclude by summarising the main contributions of this thesis:

- firstly, we have provided necessary conditions for optimality for infinite horizon, multistage, dynamic optimization problems under mild hypotheses (which dispense with the need, present in earlier literature, to impose smoothness conditions on the data), and
- secondly, by establishing the validity of the superdifferential minimum cost sensitivity relation in the presence of general end-point constraints, we have provided new information about the relationship between the maximum principle and verification functions.

REFERENCES

- [A1] ARAUJO, A. and J. A. SCHEINKMAN, *Maximum Principle and Transversality Condition for Concave Infinite Horizon Economic Models*, Journal of Economic Theory, 30 (1983), 1-16.
- [A2] ARROW, K. J. and M. KURZ, *Public Investment, the Rate of Return and Optimal Fiscal Policy*, John Hopkins Press, Baltimore, 1970.
- [A3] AUBIN, J. P. and F. H. CLARKE, *Shadow Prices and Duality for a Class of Optimal Control Problems*, SIAM Journal of Control and Optimization, 17 (1979), 567-586.
- [B1] BELLMAN, R., *Dynamic Programming*, Princeton University Press, Princeton, N. J., 1957.
- [B2] BENVENISTE, L. M. and J. A. SCHEINKMAN, *Duality Theory for Dynamic Optimization Models of Economics: The Continuous Time Case*, Journal of Economic Theory, 27 (1982), 1-19.
- [B3] BROCK, W. A., *On Existence of Weakly Maximal Programmes in a Multi-Sector Economy*, Review of Economic Studies, 37 (1970), 275-280.
- [B4] BROCKETT, R. W., *Finite Dimensional Linear Systems*, Wiley, New York, 1970.
- [C1] CARLSON, D. A. and A. HAURIE, *Infinite Horizon Optimal Control - Theory and Applications*, Lecture Notes in Economics and Mathematical Systems, vol. 270, Springer-Verlag, Berlin Heidelberg, 1987.
- [C2] CASS, D., *Optimum Growth in an Aggregative Model of Capital Accumulation: A Turnpike Theorem*, Econometrica, 34 (1966), 833-850.
- [C3] CLARKE, F. H., *Optimization and Nonsmooth Analysis*, Wiley Interscience, New York, 1983.
- [C4] CLARKE, F.H. and P.D. LOEWEN, *The Value Function in Optimal Control: Sensitivity, Controllability, and Time-Optimality*, SIAM J. Control Optim., 24 (1986), 243-263.
- [C5] CLARKE, F. H. and R. B. VINTER, *Local Optimality Conditions and Lipschitzian Solutions to the Hamilton-Jacobi equation*, SIAM J. Control Optim., 21 (1983), 856-870.
- [C6] CLARKE, F. H. and R. B. VINTER, *The Relationship Between the Maximum Principle and Dynamic Programming*, SIAM J. Control Optim., 23 (1987), 1291-1311.
- [C7] CLARKE, F. H. and R. B. VINTER, *Applications of Optimal Multiprocesses*, SIAM J. Control Optim., 27 (1989), 1048-1071.
- [C8] CLARKE, F. H. and R. B. VINTER, *Optimal Multiprocesses*, SIAM J. Control Optim., 27 (1989), 1072-1091.
- [C9] CLARKE, F. H. and R. B. VINTER, *Regularity Properties of Optimal Controls*, Centre de Recherches Mathématiques, University of Montréal, CRM-1611, April 1989.
- [D1] DASGUPTA, P., R. J. GILBERT and J. E. STIGLITZ, *Invention and Innovation Under Alternative Market Structures: The Case of Natural Resources*, Review of Economic Studies, 49 (1982), 567-582.
- [D2] DORFMAN, R., *An Economic Interpretation of Optimal Control Theory*, The American Economic Review, 59 (1969), 817-831.

- [F1] FLEMING, W. H. and R. W. RISHEL, *Deterministic and Stochastic Optimal Control*, Springer-Verlag, Berlin, 1975.
- [G1] GRAY, J. A. and S. W. SALANT, *Transversality Conditions in Infinite Horizon Models*, International Finance Discussion Papers, Board of Governors of U. S. Federal Reserve, no. 172, Jan. 1981.
- [H1] HALKIN, H., *Necessary Conditions for Optimal Control Problems with Infinite Horizons*, *Econometrica*, 42 (1974), 267-272.
- [H2] HOEL, M., *Resource Extraction when a Future Substitute has an Uncertain Cost*, *Review of Economic Studies*, 45 (1978), 637-644.
- [H3] HOTELLING, H., *The Economics of Exhaustible Resources*, *Journal of Political Economy*, 39 (1931), 137-175.
- [K1] KOOPMANS, T. C., *Concepts of Optimality and Their Uses*, Nobel Lectures 1975, reproduced in *Amer. Econ. Rev.*, 67 (1977), 261-274.
- [K2] KURZ, M., *Optimal Paths of Capital Accumulation Under the Minimum Time Objective*, *Econometrica*, 33 (1965), 42-66.
- [L1] LOEWEN, P. D., F. H. CLARKE and R. B. VINTER, *Differential Inclusions with Free Time*, *Ann. Inst. Henri Poincaré*, 5 (1988), 573-593.
- [M1] MACCINI, L. J., *Delivery Lags and the Demand for Investment*, *Review of Economic Studies*, 40 (1973), 269-281.
- [M2] MACCINI, L. J., *On Optimal Delivery Lags*, *Journal of Economic Theory*, 6 (1973), 107-125.
- [M3] MICHEL, P., *On the Transversality Condition in Infinite Horizon Optimal Problems*, *Econometrica*, 50 (1982), 975-985.
- [M4] MICHEL, P., *Some Clarifications on the Transversality Condition*, *Econometrica*, 58 (1990), 705-723.
- [N1] NICKELL, S. J., *Uncertainty and Lags in the Investment Decisions of Firms*, *Review of Economic Studies*, 44 (1977), 249-263.
- [P1] PONTRYAGIN, L. S., V. G. BOLTYANSKII, R. V. GAMKRELIDZE and E. F. MISHCHENKO, *The Mathematical Theory of Optimal Processes* (English trans. L. W. Neustadt ed.), Interscience, John Wiley & Sons, New York, 1962.
- [R1] RAMSEY, F., *A Mathematical Theory of Saving*, *Economic Journal*, 38 (1928), 543-559.
- [R2] ROSSANA, R. J., *Delivery Lags and Buffer Stocks in the Theory of Investment by the Firm*, *Journal of Economic Dynamics and Control*, 9 (1985), 153-193.
- [R3] ROYDEN, H. L., *Real Analysis*, Collier-Macmillan, New York, 1988.
- [S1] STERN, L. E., *Criteria of Optimality in the Infinite-Time Optimal Control Problem*, *Journal of Optimization Theory and its Applications*, 44 (1984), 497-508.
- [S2] STOLERU, L. G., *An Optimal Policy for Economic Growth*, *Econometrica*, 33 (1965), 321-348.
- [S3] SUBBOTINA, N. N., *The Maximum Principle and the Super-differential of the Value Function*, *Problems of Control and Information Theory*, 18 (1989), 151-160.

- [T1] TOMIYAMA, K., *Two-Stage Optimal Control Problems and Optimality Conditions*, J. Econ. Dyn. Control, 9 (1985), 317-337.
- [T2] TOMIYAMA, K., and R. J. ROSSANA, *Two-Stage Optimal Control Problems with an Explicit Switch-Point Dependence: Optimality Criteria and an Example of Delivery Lags and Investment*, J. Econ. Dyn. Control, 13 (1989), 319-337.
- [V1] VINTER, R. B., *New Results on the Relationship Between Dynamic Programming and the Maximum Principle*, Mathematics of Control, Signals and Systems, 1 (1988), 97-105.
- [V2] VINTER, R. B., and L. A. MENDOZA, *Global Optimality Conditions for Nonnormal Control Problems*, IMA Journal of Mathematical Control and Information, 2 (1985), 241-250.
- [V3] VINTER, R. B., and P. R. WOLENSKI, *Coextremals and the Value Function for Control Problems with Data Measurable in Time*, Journal of Mathematical Analysis and Applications, 153 (1990), 37-51.
- [V4] VINTER, R. B. and P. R. WOLENSKI, *Hamilton Jacobi Theory for Optimal Control Problems with Data Measurable in Time*, SIAM J. Control Optim., to appear.
- [W1] WEITZMAN, M. L., *Duality Theory for Infinite Horizon Convex Models*, Management Science, 19 (1973), 783-789.
- [W2] von WEIZSÄCKER, C. C., *Existence of Optimal Programs of Accumulation for an Infinite Time Horizon*, Rev. Econ. St., 32 (1965), 85-104.