Higher Order Conditions in Nonlinear Optimal Control

Thesis by

Christos Gavriel

In Partial Fulfilment of the Requirements

for the Degree of

Doctor of Philosophy

Department of Electrical and Electronic Engineering
Imperial College London
(2011)

To my parents

Abstract

The most widely used tool for the solution of optimal control problems is the Pontryagin Maximum Principle. But the Maximum Principle is, in general, only a necessary condition for optimality. It is therefore desirable to have supplementary conditions, for example second order sufficient conditions, which confirm optimality (at least locally) of an extremal arc, meaning one that satisfies the Maximum Principle.

Standard second order sufficient conditions for optimality, when they apply, yield the information not only that the extremal is locally minimizing, but that it is also locally unique. There are problems of interest, however, where minimizers are not locally unique, owing to the fact that the cost is invariant under small perturbations of the extremal of a particular structure (translations, rotations or time-shifting). For such problems the standard second order conditions can never apply.

The first contribution of this thesis is to develop new second order conditions for optimality of extremals which are applicable in some cases of interest when minimizers are not locally unique. The new conditions can, for example, be applied to problems with periodic boundary conditions when the cost is invariant under time translations.

The second order conditions investigated here apply to normal extremals. These extremals satisfy the conditions of the Maximum Principle in normal form (with the cost multiplier taken to be 1). It is, therefore, of interest to know when the Maximum Principle applies in normal form. This issue is also addressed in this thesis, for optimal control problems that can be expressed as calculus of variations problems. Normality of the Maximum Principle follows from the fact that, under the regularity conditions developed, the highest time derivative of an extremal arc is essentially bounded.

The thesis concludes with a brief account of possible future research directions.

Declaration

As required by the College, I hereby declare that this thesis is the result of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are appropriately referenced.

Christos Gavriel
Imperial College London
31 May 2011

Acknowledgments

First and foremost I would like to express my gratitude to my supervisor, Professor Richard Vinter for always finding the time and the motivation to discuss my project despite his very busy schedule. His enthusiasm has been inspirational and so has been his patience in teaching me and turning me into a researcher.

I would like to thank my examiners, Professor Malcolm Smith and Professor Alessandro Astolfi, for reviewing my thesis. Their constructive comments and corrections have certainly improved the quality of this work.

My time as a PhD student at the Control and Power group would not have been the same without my friends at Imperial College. I wish to thank Mario Sassano, Can Ozelci, Andri Alekou and Aris Kountouriotis who have made my stay such an enjoyable one. A further thanks to Angelos Georghiou and Ioannis Demetriou for our great pub conversations and Formula 1 excursions. A thank you goes to my friends in Cyprus, as well, Vassilis Erotokritou, Andreas Theopemptou and Christopher Kazantzis for adapting their schedules to mine whenever I was back home. A special thanks is due to Marios Valiantis for his generous advice.

I am greatly indebted to Clea for being loving, caring and understanding during what, at times, seemed to be an endless quest. It would have been much harder without her.

Last but not least, I want to thank my parents Gavriel and Eleni for their loving support and guidance through all these years. Without them both I would not be where I am today.

Contents

Contents **6**

List of Figures **8**

Chapter 1. Overview **9**

1.1 Outline and Contributions 13

1.2 Mathematical Background 15

Chapter 2. Non-unique Local Minimizers: a Nonlinear Programming Perspective **21**

2.1 Introduction 22

2.2 Standard Second Order Sufficient Conditions 23

2.3 Refined Second Order Sufficient Conditions 24

2.4 An Abstract Framework 30

2.5 Conclusion 40

Chapter 3. Non-unique local minimizers in optimal control **41**

3.1 Introduction 42

3.2 Second Order Conditions in Optimal Control: A Review 44

3.2.1 Uniqueness of Local Minimizers 48

3.3 Second Order Sufficient Conditions for Non-unique Local Minimizers 48

3.3.1 Reformulation of the Optimal Control Problem 49

3.3.2 Statement of the New Sufficient Conditions 51

3.3.3 Proof of Theorem 3.4 52

3.4 Application I: Periodic Optimal Control 63

3.4.1 Problem Statement 64
3.4.2 Speyer's Sailboat Problem 66
3.5 Application *II*: Other Endpoint Constraints 70
3.6 Conclusion 74

Chapter 4. Regularity of Minimizers for Second Order Problems **75**
4.1 Introduction 76
4.2 Conditions for Minimizer Regularity 78
4.2.1 Minimizer Regularity for First Order Problems 78
4.2.2 Minimizer Regularity for Second Order Problems 79
4.3 New Conditions for $W^{2,\infty}$ Regularity 80
4.4 Proof of Theorems 4.4 and 4.5 82
4.5 Conclusion 89

Chapter 5. Regularity of Minimizers for Higher Order Problems **90**
5.1 Introduction 91
5.2 Standard Conditions for $W^{N,\infty}$ Regularity 92
5.3 New Conditions for $W^{N,\infty}$ Regularity 93
5.4 Proof of Theorems 5.2 and 5.3 95
5.4.1 Proof of Theorem 5.2 98
5.4.2 Proof of Theorem 5.3 101
5.5 Conclusion 104

Chapter 6. Conclusions **105**
6.1 Summary 105
6.2 Future Research Directions 107

Bibliography **109**

List of Figures

1.1 Contributions of the thesis. 14

2.1 Family of constant cost extremals. 25
2.2 Plot of a function with non-unique local minimizers. 28
2.3 An intuitive interpretation of Lemma **2.9.** 35

3.1 Time invariance of periodic solution. 64
3.2 State plot 67
3.3 Control plot 68
3.4 Solution to Riccati equation 68
3.5 State plot 72
3.6 Control plot 72
3.7 Solution to Riccati equation 73

Chapter 1

Overview

Many engineering applications are required to meet objectives concerning efficiency, speed and tracking. These goals can naturally be expressed, in mathematical form, as optimization problems or, in the case of dynamical systems, as optimal control problems. From an engineering point of view then, it is important to develop methodologies that help solve such problems. But this is not merely engineering practice. It appears that an optimality principle also governs our daily routines; whether it involves finding the route that minimizes journey times or a strategy that maximizes monetary returns.

There are two main approaches to solving optimal control problems. On one hand, there is the dynamic programming approach [8] which, for continuous time systems, leads to the Hamilton-Jacobi-Bellman (HJB) equation [10] which is a partial differential equation. Solving the HJB equation, to obtain the solution to the optimal control problem is not an easy task in general. On the other hand, there is the classical approach in which candidate optimal solutions are first computed (provided a solution exists) and then the optimal solution is chosen amongst those candidates. This thesis is primarily concerned with the second approach and in particular the development of criteria for identifying minimizers, namely second order sufficient conditions of optimality.

The birth of optimal control can be traced back to the late 17^{th} century, even though problems concerning the shortest path between two points had been studied earlier by the ancient Greeks [62]. It is believed that interest was ignited in 1696 and Johann Bernoulli's challenge to solve a classical problem in what is now known as the field of calculus of variations. The so called "Brachistochrone problem", which in Greek means shortest time problem, involved finding the fastest path of a particle moving under gravity in a vertical plane between two points A and B. The challenge was accepted by many great minds of the time whose solutions marked the birth of optimal control.

Even though the study of classical problems in the calculus of variations has undoubt-

edly generated many of the tools that are still used today in optimal control, optimal control problems are significantly 'richer' mathematically. Indeed, calculus of variations problems can be thought of as special cases of optimal control problems because the presence of dynamical constraints in the latter restricts the set of allowed velocities. Put differently, the state of the system at time t, $x(t)$, in optimal control, is dictated by the ordinary differential equation

$$\frac{d}{dt}x(t) = f(t, x(t), u(t))$$

comprising a possibly nonlinear function f and function u. In this setting, a calculus of variations problem corresponds to the special case where $f(t, x(t), u(t)) = u(t)$.

The single most significant step in optimal control was, arguably, made in 1962 by the Soviet group of L. S. Pontryagin and his collaborators in the form of the "Pontryagin Maximum Principle " [13]. They presented a set of conditions that when satisfied (numerically or even analytically in favourable circumstances) supplied a solution to the optimal control problem. It has to be said that even though each of the conditions of the Maximum Principle can independently be traced back to earlier results in the calculus of variations, it was the first time that they were put together and formulated for optimal control problems as we know them today.

However, a solution obtained via the Maximum Principle is not, in general, an optimal solution. This is because they are necessary conditions for optimality; this means that an optimal solution satisfies these conditions but not the converse. Hence they cannot, in general, guarantee optimality but they can provide candidate optimal solutions. To confirm optimality, necessary conditions must either be strengthened or supplemented by additional tests. This brings us to the first topic of this thesis: the study of supplementary conditions (sufficient conditions) that, when satisfied, verify that a candidate solution is optimal.

Second order sufficient conditions, such as those studied in this thesis, assume that the first order conditions of the Maximum Principle are satisfied in 'normal' form. In this case the Maximum Principle can provide information about the minimizer. But while normality is typically directly hypothesised in the study of second order sufficient conditions this is not always the case. First order necessary conditions give useful information provided the minimizer belongs to an 'accessible' class of functions. In other words, they supply meaningful information about a minimizer provided the hypotheses under which necessary conditions have been derived and the hypotheses of existence of a solution are compatible. Regularity theory attempts to bridge this gap. Even though general results are difficult to obtain, studying special cases gives an insight about the nature of the conditions that need to be imposed on the underlying problem. This brings us to the second objective of this

thesis: to develop conditions under which minimizers have certain regularity properties (are 'well behaved') for problems in the calculus of variations, which in turn can be used to deduce validity of the Maximum Principle in normal form.

Apart from confirming optimality of solutions, sufficient conditions play an important role in the computation of optimal solutions because they determine the speed of convergence of numerical algorithms. In recursive quadratic programming [2, 5], for example, each new estimate for the state depends on the solution of a quadratic approximation of the original problem. Second order sufficient conditions determine the value of this estimate and hence the speed at which it converges to the optimal solution [22]. They are also central to sensitivity analysis [16]; perturbations and disturbances in system data are mathematically modelled as additional parameters in the system dynamics, boundary constraints and control or state constraints. For these problems, second order sufficient conditions need to be satisfied for all varied parameters close to the nominal ones [45].

The field of sufficient conditions for optimality is by no means a recent development. In fact, there is a vast amount of literature on this topic extending over at least the last 30 years. This can be partly attributed to the fact that sufficient conditions have found their way into a very broad spectrum of problems ranging from engineering to economics [7]. Therefore, it is difficult to give an account of all sufficient conditions present in the literature; instead a technical review of earlier results relevant to the optimal control problem studied here with accompanying mathematical background are provided in each subsequent chapter.

Linear quadratic (LQ) problems concern optimal control problems with quadratic integral cost and linear dynamics. LQ-problems with fixed initial state have been studied comprehensively in the past as control problems in their own right, since the seminal paper of Kalman on the optimal regulator [40] and the Riccati equation. But they are also particularly important here because problems of this kind with linear boundary constraints arise in the derivation of second order optimality conditions for control problems with nonlinear dynamics and mixed endpoint constraints [61].

Another important class of problems concerns optimal control problems with general integral cost, nonlinear dynamics and constraints on the state, the control or both. These problems have been extensively studied in the literature with authors using different tools to derive sufficient conditions. One approach is to extend known results from the calculus of variations. Known second order necessary and sufficient conditions in the calculus of variations [31] are based on the non-existence of a conjugate point on the time interval of interest. In particular, for a problem posed over the time interval $[0, T]$ non-existence of a conjugate point to T on the open set $(0, T)$ is a necessary condition while non-existence on $[0, T)$ is a sufficient condition known as the Jacobi condition. The latter has been used in

the derivation of second order sufficient conditions in optimal control by Mayne [51], when the control set is open, and later generalised by Zeidan et al. in [72–74, 77]. Zeidan, for example, considers optimal control problems with fixed initial condition and no convexity hypotheses on the control set.

Maurer and Pickenhain treat the same class of problems with more general endpoint constraints. Their approach in [49] differs from Zeidan's because they use an L^2 estimate for the deviation of the cost from the nominal one and obtain conditions by applying a direct sufficiency criterion, the Hamilton-Jacobi (HJ) inequality (a generalisation of the Hamilton-Jacobi equation). Strong second order sufficient conditions for the HJ inequality to hold are then obtained and hinge upon the existence of a solution to a certain Riccati equation satisfying boundary inequality constraints. Weaker conditions (conditions that are more easily satisfied) are then obtained requiring that the conditions on the Riccati equation are satisfied only on the relevant subspace. The conditions developed, when specialised to the special boundary constraints treated by Zeidan, reproduce the same result. The same line of research is followed by Malanowski et al. in [39] who develop sufficient conditions for problems with additional pure state constraints. Similar techniques are also employed by Sorger in [57] but his results are not as general as Maurer because the latter consider problems with mixed control-state constraints and more general boundary conditions.

Other special classes of problems, such as control constrained optimal control problems for example, have been studied by Dunn [23, 24] in the formulation of which the dynamics take a special form and include a quadratic control term. Problems for which the control parameter enters linearly into the dynamics (bang-bang control) have been recently examined in [47]. Minimum time optimal control problems are to be found in [46, 48].

A common theme of the second order sufficient conditions mentioned is that they have been developed on the assumption that the optimal control problem at hand has a locally unique solution. The question arising now is whether there are physical problems that, when formulated as optimal control problems, have a non-unique local solution.

Planetary orbits, the rhythm of the heart and other processes in nature follow a cyclic pattern. Not long ago, for certain autonomous problems such as fuel economy in vehicle cruise [32, 58] and chemical processes [37], for example, it was determined that steady state operation can be improved by cyclic operation. Guardabassi, Bittanti et al. have developed tests that can confirm whether steady state operation can be improved by cycling [34]. A distinct feature of these periodic solutions, however, is that they are not locally unique. Consequently, the 'standard' second order sufficient conditions cannot be applied to identify minimizing solutions. For these problems, special sufficient conditions need to be considered to identify an optimal periodic solution and these have mainly been

developed by Speyer and his collaborators.

What appears to be lacking in the literature, so far, is a set of sufficient conditions that can be applied to a more general class of problems than periodic control problems (such as those mentioned earlier) for the case when a minimizer is not locally unique. It is precisely this gap that the thesis sets to investigate.

1.1 Outline and Contributions

This section outlines in more detail the contributions of the thesis. Figure 1.1 shows where each chapter fits in terms of the classical approach to solving an optimal control problem. Chapters 2 and 3 concern the development of refined second order sufficient conditions of optimality. The results of these chapters can be found in [28] and in a forthcoming paper [30]. By looking at a calculus of variations problem, Chapters 4 and 5 derive conditions under which minimizers have bounded highest derivatives and ensure that first order necessary conditions 'make sense'. These results are presented in [29] and [27] respectively. A summary of the main contributions of each chapter is given below.

- Chapter 2 introduces the issue of non-unique local minimizers in a nonlinear programming context. Here, well-known second order sufficient conditions of optimality are rendered not applicable for problems with locally non-unique solutions. New, refined conditions are then developed which also apply in these cases. A similar concept is then extended to optimization problems with equality constraints, posed over functional spaces. The formulation of such problems is seen to resemble that of an optimal control problem when a pair consisting of an initial state and a control is associated with an element in a Hilbert space. This association will prove to be vital in the following chapter.

- Chapter 3 is concerned with optimal control problems with nonlinear dynamics and mixed boundary conditions. For such problems it is shown that not only can standard second order sufficient conditions of optimality confirm that a solution is optimal but they actually imply that it is unique. The concept of non-uniqueness in optimal control is introduced, a particular case of which comprises periodic optimal control problems. Refined second order sufficient conditions for optimal control problems with non-unique minimizers are obtained by reformulating the optimal control problem as a special case of the abstract optimization problem of Chapter 2. Examples from periodic optimal control serve to illustrate the effectiveness of the new conditions.

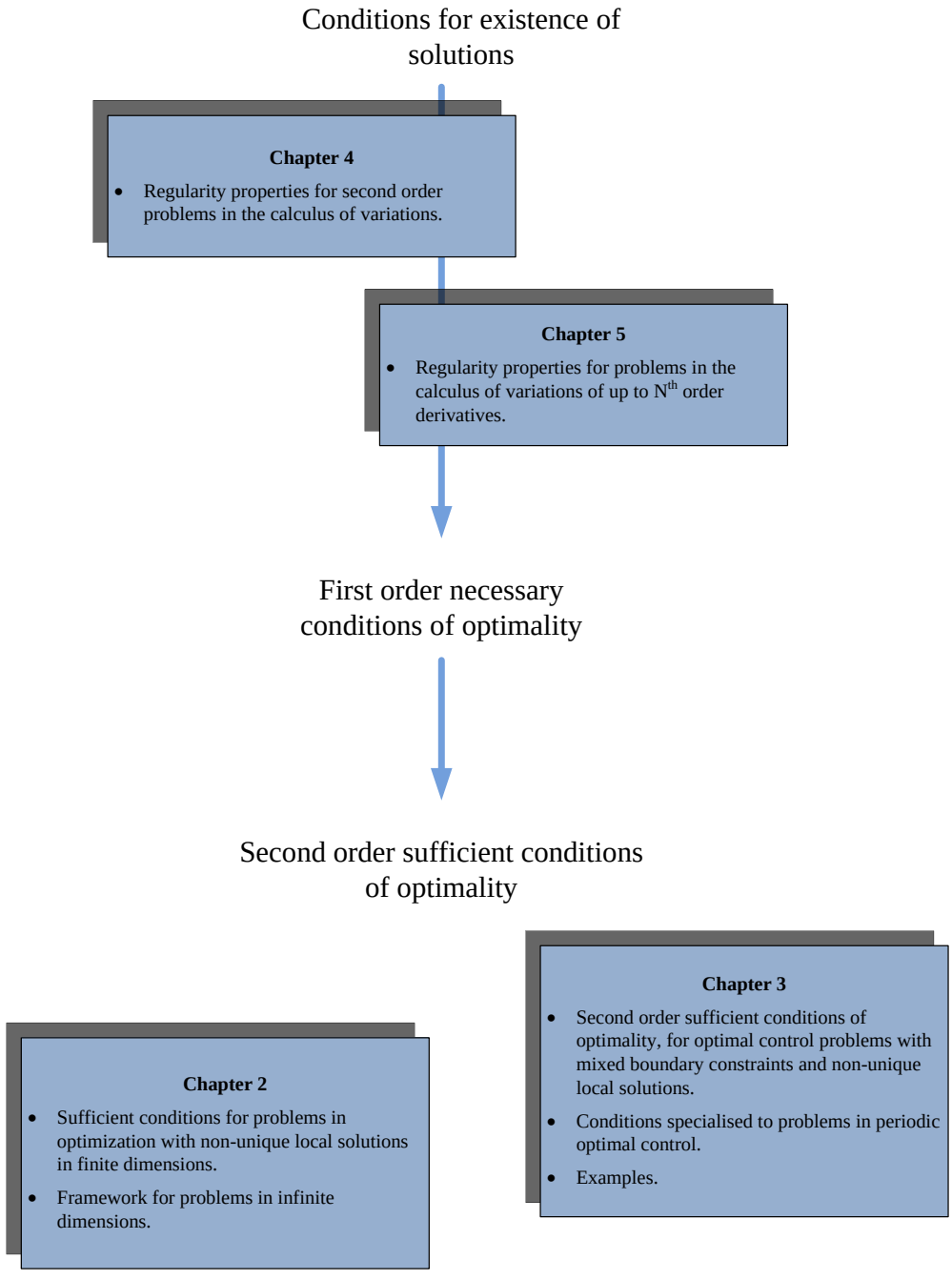


Figure 1.1: Contributions of the thesis.

- Chapter 4 examines second order problems in the calculus of variations with vector valued arcs. Problems of this nature are special cases of optimal control problems

with linear, time-invariant dynamics. Taking these problems as a starting point, this chapter sets to explore the conditions under which the hypotheses for existence and the hypotheses under which necessary conditions can be derived, are compatible. The main results comprise two theorems which establish boundedness of the second derivative of the minimizing arc under hypotheses that are less restrictive than the ones known in the literature.

- Chapter 5 focuses on deriving regularity properties for higher order problems in the calculus of variations. In contrast to Chapter 4, the problems studied here are non-autonomous, i.e. the integrand also depends explicitly on time t, and contain derivatives of the arc x of up to some arbitrary order N. In particular, it examines how the novel approach of Chapter 4 can be applied to such problems to derive more general results than what is already known.
- Chapter 6 brings the thesis to a conclusion with some remarks and possible future research directions.

1.2 Mathematical Background

The purpose of this section is to set the scene and introduce the main notation and definitions that will be employed throughout the thesis. It also serves to provide a brief summary of known results that link the calculus of variations problem to the optimal control one.

Notation. The Euclidean norm of a vector x is denoted by $|x|$. The unit closed ball in $\mathbb{R}^n$ is denoted by $\mathbb{B}^n$.

Definition. The essential supremum of a measurable function $x(.)$ on $[t_0, t_1]$ is defined as

$$\operatorname{ess\,sup} x = \inf \Big\{ \xi \in \mathbb{R} \mid \mu\left(\{t \mid |x(t)| > \xi\}\right) = 0,\ \forall t \in [t_0, t_1] \Big\},$$

where μ denotes the Lebesgue measure.

Definition. The infinite dimensional space $L^p([t_0, t_1]; \mathbb{R}^n)$ is defined as the space of all $n-$valued Lebesgue measurable functions $x : [t_0, t_1] \to \mathbb{R}^n$ that satisfy the property

$$\|x\|_{L^p} := \left(\int_{t_0}^{t_1} |x(t)|^p \, dt \right)^{1/p} < \infty,$$

where $1 \leq p < \infty$. The case $p = \infty$ is interpreted as the space of all $n-$valued Lebesgue measurable functions that are essentially bounded, in other words

$$\|x\|_{L^\infty} := \operatorname{ess\,sup} x < \infty.$$

Notation. $W^{m,p}([t_0,t_1];\mathbb{R}^n)$ is used to denote the space of functions the m^{th} derivative of which belongs to the space $L^p([t_0,t_1];\mathbb{R}^n)$.

Definition. The $n-$valued Lebesgue measurable function $t \to x(t)$ is absolutely continuous on $[t_0,t_1]$ (belongs to $W^{1,1}([t_0,t_1];\mathbb{R}^n)$) if and only if its derivative exists almost everywhere, is integrable and

$$x(t) = x(t_0) + \int_{t_0}^{t} \dot{x}(\tau)\, d\tau$$

for all $t \in [t_0,t_1]$.

Notation. $C^m([t_0,t_1];\mathbb{R}^n)$ will be used to denote the space of $m-$times continuously differentiable $n-$valued functions on the interval $[t_0,t_1]$.

Definition. Let X and Y be topological spaces. A function $f : X \to Y$ is said to be *strongly continuous* if and only if for each subset $A \subset X$

$$f(\bar{A}) \subset f(A),$$

where $\bar{A}$ denotes the closure of the set A.

Notation. The graph of a set U is denoted by $Gr\, U$.

Definition. Given a scalar valued Lipschitz continuous function f defined on an open subset $\mathcal{O}$ of $\mathbb{R}^n$ and a point $\bar{x} \in \mathcal{O}$, denote by $\partial f(\bar{x})$ the (set valued) subdifferential of f at $\bar{x}$, as defined, for example, in [54] or [64], synonymously the 'limiting subdifferential' in [42]:

$$\begin{aligned} \partial f(\bar{x}) \;=\; \Big\{ & \xi \in \mathbb{R}^n \mid \exists\ \xi_i \to \xi, x_i \to \bar{x} \text{ such that for each } i \\ & \limsup_{x \to x_i} |x - x_i|^{-1} [\xi_i \cdot (x - x_i) - (f(x) - f(x_i))] \le 0 \Big\}. \end{aligned}$$

Notation. Given a subset $X \subset \mathbb{R}^n$ the notation $\operatorname{co} X$ means the convex hull of X.

Definition. Given a convex subset $Y \subset \mathbb{R}^n$ the normal cone at a point $\bar{x} \in Y$ is defined as

$$N_Y(\bar{x}) := \Big\{ \xi \mid \xi \cdot (x - \bar{x}) \le 0\ \forall x \in Y \Big\}.$$

For a more general definition of cones (proximal normal cone, limiting normal cone) see [64, Section 1.8].

Notation. The shorthand "a.e." will be used to mean almost everywhere.

Finally, it is noted that whenever the space dimensionality and the underlying time interval are clear from the context, the extra parameters $([t_0,t_1];\mathbb{R}^n)$ are dropped.

Consider the so-called basic problem in the calculus of variations,

$$(BV)\begin{cases} \text{minimize } J(x) = \displaystyle\int_0^T L(t, x(t), \dot{x}(t))\, dt \\ \text{over arcs } x \in W^{1,1}([0,T];\mathbb{R}^n) \\ \text{subject to} \\ x(0) = x_0 \quad x(T) = x_T \end{cases}$$

where $T > 0$ and the function $L : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is C^2. The definition that is used for an absolutely continuous function is the following.

Under the additional assumption that there exists a constant $k > 0$ such that

$$|L(t, x, v) - L(t, x', v')| \leq k(|x - x'| + |v - v'|), \tag{1.1}$$

for all $t \in [0, T]$, $x, v, x', v' \in \mathbb{R}^n$ a well-known first order necessary condition for optimality, referred to as Euler condition, can be derived. It asserts that given a minimizing arc $\bar{x}$ the condition

$$\frac{d}{dt} L_v(t, \bar{x}(t), \dot{\bar{x}}(t)) = L_x(t, \bar{x}(t), \dot{\bar{x}}(t)) \tag{1.2}$$

is satisfied point-wise, where the subscripts x and v denote the partial derivatives of L with respect to the x and $\dot{x}$ variables respectively. The integral version of (1.2) is commonly known as the Euler-Lagrange condition and is expressed as

$$-\int_0^t L_x(\tau, \bar{x}(\tau), \dot{\bar{x}}(\tau))\, d\tau + L_v(t, \bar{x}(t), \dot{\bar{x}}(t)) = c \tag{1.3}$$

for some constant c. It follows from (1.1) that L_x is integrable which in turn implies that $t \to L_v$ is absolutely continuous. Expressed in terms of the (absolutely continuous) *adjoint arc*

$$p(t) = L_v(t, \bar{x}(t), \dot{\bar{x}}(t))$$

(1.3) is written as

$$\begin{pmatrix} \dot{p}(t) \\ p(t) \end{pmatrix} = \nabla_{x,v} L(t, \bar{x}(t), \dot{\bar{x}}(t)) \quad \text{a.e. } t \in [0, T]. \tag{1.4}$$

Analogous conditions to (1.2) and (1.4) for second order problems, i.e. problems for which the integrand contains derivatives of x up to second order like those that will be presented in Chapter 4, are

$$L_x(t, \bar{x}(t), \dot{\bar{x}}(t), \ddot{\bar{x}}(t)) - \frac{d}{dt} L_v(t, \bar{x}(t), \dot{\bar{x}}(t), \ddot{\bar{x}}(t)) + \frac{d^2}{dt^2} L_w(t, \bar{x}(t), \dot{\bar{x}}(t), \ddot{\bar{x}}(t)) = 0 \quad \text{a.e. } t \in [0, T]$$

and

$$\begin{pmatrix} \dot{p}_1(t) \\ \dot{p}_2(t) + p_1(t) \\ p_2(t) \end{pmatrix} = \nabla_{x,v,w} L(t, \bar{x}(t), \dot{\bar{x}}(t), \ddot{\bar{x}}(t)) \quad \text{a.e. } t \in [0,T]$$

where the subscript w denotes the partial derivative of L with respect to the $\ddot{x}$ variable, and p_1 and p_2 are $n-$dimensional absolutely continuous arcs on $[0,T]$. For general N^{th} order problems, which will be examined in Chapter 5, the Euler equation is given by [21]

$$L_x(t, \bar{x}(t), \dot{\bar{x}}(t), \ldots, \bar{x}^{(N)}(t)) + \sum_{i=1}^{N} (-1)^i \frac{d^i}{dt^i} L_{x^{(i)}}(t, \bar{x}(t), \dot{\bar{x}}(t), \ldots, \bar{x}^{(N)}(t)) = 0$$

in which the superscript (i) denotes the i^{th} derivative of x.

To justify the study of problems in the calculus of variations, in terms of deriving regularity results, consider (BV) and the optimal control problem

$$(OC) \begin{cases} \text{minimize } g(x(0), x(T)) \\ \text{over measurable functions } u : [0,T] \to \mathbb{R}^m \text{ and arcs } x \in W^{1,1}([0,T]; \mathbb{R}^n) \\ \text{subject to} \\ \dot{x}(t) = f(t, x(t), u(t)) \text{ and } u(t) \in U \quad \text{a.e.} \\ (x(0), x(T)) \in C, \end{cases}$$

in which the data comprise a number $T > 0$, functions $f : [0,T] \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ and $g : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ and sets $C \subset \mathbb{R}^n \times \mathbb{R}^n$ and $U \subset \mathbb{R}^m$. First order necessary conditions of optimality for (OC) are stated below.

Theorem 1.1. [64] (Maximum Principle) *Assume that $(\bar{x}, \bar{u})$ is minimizing for (OC) and that*

(i) g is continuously differentiable,

(ii) C is a closed set,

(iii) f is continuous, $f(t, ., u)$ is continuously differentiable for each (t, u) and there exists $\epsilon > 0$ and $k(.) \in L^1([0,T]; \mathbb{R})$ such that

$$|f(t,x,u) - f(t,x',u)| \leq k(t)|x - x'|$$

for all $x\,, x' \in \bar{x}(t) + \epsilon \mathbb{B}^n$ and $u \in U$ almost everywhere, and

(iv) $Gr\, U$ is a Borel set.

Then there exists an arc $p \in W^{1,1}([0,T];\mathbb{R}^n)$ *and* $\lambda \geq 0$*, not both zero, such that*

$$\begin{aligned} -\dot{p}(t) &= f_x^T(t,\bar{x}(t),\bar{u}(t))p(t) \quad a.e. \\ f(t,\bar{x}(t),\bar{u}(t))\cdot p(t) &= \max_{u\in U} f(t,\bar{x}(t),u)\cdot p(t) \quad a.e. \\ (p(0),-p(T)) &= \lambda\nabla g(\bar{x}(0),\bar{x}(T)) + \eta \end{aligned}$$

for some $\eta \in N_C(\bar{x}(0),\bar{x}(T))$ *.*

Notice that (BV) can be expressed as an optimal control problem by writing,

$$\begin{cases} \text{minimize } y(T) \\ \text{over arcs } (y,x) \in W^{1,1}([0,T];\mathbb{R}^{n+1}) \text{ and measurable functions } u:[0,T]\to\mathbb{R}^n \\ \text{subject to} \\ \dfrac{d}{dt}\begin{pmatrix} y(t) \\ x(t) \end{pmatrix} = \begin{pmatrix} L(t,x,u) \\ u \end{pmatrix} \quad \text{a.e. } t\in[0,T] \\ \begin{pmatrix} y(0) \\ x(0) \\ x(T) \end{pmatrix} = \begin{pmatrix} 0 \\ x_0 \\ x_T \end{pmatrix}. \end{cases}$$

Also note that if $\bar{x}$ (with derivative $\dot{\bar{x}}$) is a minimizer for (BV) then $\bar{u} = \dot{\bar{x}}$ and $\bar{y}(t) = \int_0^t L(\tau,\bar{x}(\tau),\bar{u}(\tau))\,d\tau$ is a minimizer for this optimal control problem.

To confirm that in these circumstances the Maximum Principle is valid in normal form, consider the Maximum Principle conditions: there exist an adjoint $n+1$ valued arc partitioned as (q_0,q) and $\lambda \geq 0$, not all zero, such that

$$\begin{aligned} \dot{q}_0(t) &= 0 \\ -\dot{q}(t) &= q_0(t)L_x(t,\bar{x}(t),\dot{\bar{x}}(t)) \\ \max_u\{q_0(t)L(t,\bar{x}(t),u(t)) + q(t)\cdot u(t)\} &= q_0(t)L(t,\bar{x}(t),\dot{\bar{x}}(t)) + q(t)\cdot\dot{\bar{x}}(t) \\ (q(0),-q(T)) &= \text{free} \\ q_0(0) &= \text{free} \\ -q_0(T) &= \lambda. \end{aligned}$$

General conditions are known for the existence of minimizers for (BV), such as Tonelli's hypotheses for example

$$\begin{cases} L(.,.,.) \text{ is continuous,} \\ L(t,x,v) \text{ is convex in } v \text{ for fixed } (t,x), \\ L(t,x,v) \geq -\beta|v| + \theta(|v|) \text{ for all } (t,x) \text{ and } v, \end{cases}$$

where $\beta > 0$ and $\theta(.) : \mathbb{R}^+ \to \mathbb{R}^+$ has the property that $\lim_{s\to\infty} \theta(s)/s \to \infty$.

Let $\bar{x}$ be a minimizer (which under such hypotheses is known to exist). It is natural then to seek information about $\bar{x}$ by applying the Maximum Principle. But this is not possible, in general, because the Tonelli hypotheses do not imply validity of the Maximum Principle in relation to $\bar{x}$. Suppose, however, that it can be shown (under appropriate additional hypotheses) that $\dot{\bar{x}}$ is essentially bounded. Then it can be shown that the Maximum Principle is valid, valid indeed in normal form. This establishes the link between establishing essential boundedness of $\dot{\bar{x}}$ and applicability of the Maximum Principle in normal form.

Chapter 2

Non-unique Local Minimizers: a Nonlinear Programming Perspective

Second order sufficient conditions of optimality play an important role in optimization and optimal control theory. They constitute supplementary tests that are applied to confirm whether a candidate solution (obtained either via the application of necessary conditions or other methods) is indeed an optimal one. This chapter is devoted to establishing refined second order sufficient conditions of optimality both in nonlinear programming and in the form of a more abstract framework. These refined conditions are then shown to be applicable to a broader class of problems than the standard ones.

Of particular interest here are optimization problems that yield non-unique local solutions with analogues in optimal control. It is known that standard second order sufficient conditions of optimality not only provide the information that a putative solution is optimal but also that it is locally unique. For many problems of interest, however, this is not the case. For these problems the standard sufficient conditions are not applicable and an optimal solution cannot be identified with such means.

The chapter begins with a study of problems in nonlinear programming since, due to their finite dimensional nature, they provide a helpful intuitive interpretation of the more abstract framework proposed to deal with the issue of locally non-unique minimizers. Section 2.1 states the standard nonlinear programming problem while Section 2.2 introduces the new problem for which a local solution is non-unique. Refined second order sufficient conditions, that deal with this problem, are then provided in Section 2.3. An infinite dimensional optimization problem, with non-unique local solutions is then considered with corresponding second order sufficient conditions developed in Section 2.4. The significance of these conditions is that they form the basis for deriving second order sufficient conditions for problems in optimal control.

2.1 Introduction

The first part of the chapter is motivational and presents preliminary concepts by focusing on the nonlinear programming problem

$$(P_0)\begin{cases}\text{minimize } f(x) \text{ over } x \in X\\ \text{subject to}\\ g(x)=0,\end{cases}$$

where X is a convex subset of $\mathbb{R}^n$, $f:\mathbb{R}^n \to \mathbb{R}$ is the cost function and $g:\mathbb{R}^n \to \mathbb{R}^m$ represents m equality constraints. For problem (P_0), it is assumed that f and g are twice differentiable.

Definition. A point x is called *admissible* if it satisfies the constraints $g(x)=0$.

Definition. A solution $\bar{x}$ to (P_0) is referred to as a *local minimizer* if there exists $\epsilon > 0$ such that

$$f(x) \geq f(\bar{x}) \tag{2.1}$$

over all admissible $x \in \bar{x} + \epsilon\mathbb{B}^n$.

Definition. A solution $\bar{x}$ is referred to as a *strict local minimizer* if (2.1) is replaced with a strict inequality for all admissible $x \in (\bar{x} + \epsilon\mathbb{B}^n) \setminus \{\bar{x}\}$.

Definition. A point $\bar{x}$ is *regular* with respect to the linearized constraints if

$$\alpha^T \nabla g(\bar{x}) = 0$$

implies

$$\alpha = 0.$$

Regularity plays an important role in nonlinear programming because it implies that the constraint set near the point $\bar{x}$ can be approximated by the tangent cone defined by linear approximations to the constraint functions involved. The best known set of necessary conditions applicable to regular minimizers are the Kuhn-Tucker conditions [41]. These are conditions which any minimizing solution $\bar{x} \in X$ satisfies; put differently, these conditions can be used to generate candidate locally optimal solutions. Central to this analysis is the Lagrangian function defined as

$$L(x,\lambda) := f(x) + \lambda^T g(x),$$

where $\lambda \in \mathbb{R}^m$ is the constraint multiplier. These conditions are summarized in Theorem 2.1.

Theorem 2.1. (Kuhn-Tucker conditions) *Let $\bar{x}$ be a minimizer for problem (P_0) and suppose $\bar{x}$ is a regular point for the constraints. Then there exists $\lambda \in \mathbb{R}^m$, such that*

$$\nabla_x L(\bar{x}, \lambda) \;=\; \nabla f(\bar{x}) + \lambda^T \nabla g(\bar{x}) \;=\; 0 \tag{2.2}$$

$$g(\bar{x}) \;=\; 0. \tag{2.3}$$

Proof. [43, Section 10.6]. ■

2.2 Standard Second Order Sufficient Conditions

Second order sufficient conditions of optimality for (P_0) centre on the quadratic approximation of the Lagrangian function at the point $(\bar{x}, \lambda)$ obtained either directly, by applying necessary conditions such as Theorem 2.1, or by numerical methods. A standard second order sufficiency criterion [43, Chapter 10] is stated below.

Theorem 2.2. *Let $f, g \in C^2$. Suppose $\bar{x}$ is a regular point and that conditions* (2.2) *and* (2.3) *are satisfied. Furthermore, suppose that*

$$\nabla^2_{xx} L(\bar{x}, \lambda) \;=\; \nabla^2 f(\bar{x}) + \sum_{j=1}^{m} \lambda_j \nabla^2 g_j(\bar{x})$$

is a positive definite matrix over the subspace defined by

$$S \;:=\; \{y \in\; X \mid \nabla g(\bar{x})\, y \;=\; 0\}.$$

Then $\bar{x}$ is a strict local minimizer for (P_0).

A key feature of sufficient conditions such as those of Theorem 2.2 is that not only do they provide information that a candidate solution is optimal but also that this minimizer is locally unique. For a number of problems, however, the solution is not locally unique and consequently the standard second order sufficient conditions cannot be used to identify minimizing solutions.

To illuminate the difficulties that arise when minimizers are not locally unique, suppose that there exists an open set $\mathcal{A} \subset \mathbb{R}^k$ containing the origin, a C^1 function $\xi : \mathbb{R}^k \to \mathbb{R}^n$ and a family of constant cost extremals $\big\{\xi(\alpha) \,|\, \alpha \in \mathcal{A}\big\}$, with associated α-parameterised

Lagrange multiplier $\lambda^\alpha \in \mathbb{R}^m$ such that

$$(H_S)\begin{cases} \xi(0) = \bar{x}, \\ f\big(\xi(\alpha)\big) = f(\bar{x}) \quad \text{for all } \alpha \in \mathcal{A}, \\ g(\xi(\alpha)) = 0 \quad \text{for all } \alpha \in \mathcal{A}. \end{cases}$$

It can be easily deduced from the assertions of Theorem 2.2 that such a second order sufficient condition is not applicable to problems for which the minimizer belongs to a set of extremals with identical cost.

To see exactly how this turns out to be the case, observe that for any $\epsilon > 0$ there exists a number $\bar{\alpha} > 0$ such that, defining $\mathcal{A} := 0 + \bar{\alpha}\mathbb{B}^k$,

$$f\big(\xi(\alpha)\big) - f(\bar{x}) = 0$$

for all admissible $\xi(\alpha)$ and $\alpha \in \mathcal{A}$. Consider a sequence $\alpha_i \to 0$, where $\alpha_i \in \mathcal{A}$ for each i. Using Taylor's expansion and taking into account first order necessary conditions, there exists some continuity modulus $\theta(.) : \mathbb{R}^+ \to \mathbb{R}^+$ with the property $\lim_{s\downarrow 0} \theta(s) = 0$ such that

$$|(\xi(\alpha_i) - \bar{x})^T \nabla^2_{xx} L(\bar{x}, \lambda^0)(\xi(\alpha_i) - \bar{x})| \leq \theta(|\xi(\alpha_i) - \bar{x}|)|\xi(\alpha_i) - \bar{x}|^2 \tag{2.4}$$

$$|\nabla g_l(\bar{x})(\xi(\alpha_i) - \bar{x})| \leq \theta(|\xi(\alpha_i) - \bar{x}|)|\xi(\alpha_i) - \bar{x}| \tag{2.5}$$

for $l = 1, 2, \ldots, m$. Now divide equation (2.4) by $|\xi(\alpha_i) - \bar{x}|^2$ and (2.5) by $|\xi(\alpha_i) - \bar{x}|$ and define

$$y_i := \frac{(\xi(\alpha_i) - \bar{x})}{|(\xi(\alpha_i) - \bar{x})|}.$$

Extracting a subsequence $y_i \to \bar{y}$, relabelling and taking the limit as $i \to \infty$, yields

$$\begin{aligned} \bar{y}^T \nabla^2_{xx} L(\bar{x}, \lambda^0)\bar{y} &= 0 \\ \nabla g_l(\bar{x})\bar{y} &= 0 \quad \text{for } l = 1, 2, \ldots, m. \end{aligned}$$

The fact that $|\bar{y}| = |y_i| = 1$ confirms the earlier observation that the condition requiring strict positivity of $\nabla^2_{xx} L(\bar{x}, \lambda^0)$ can never be satisfied. To address this issue, new, refined conditions need to be developed.

2.3 Refined Second Order Sufficient Conditions

To obtain conditions for problems the minimizer of which belongs to a set of extremals satisfying (H_S) in circumstances when the minimizers are not locally unique, the require-

ment that the Hessian be strictly positive on the tangent space S needs to be modified. This section examines how considering a smaller subspace, S', gives rise to refined conditions that can also be applied to the class of problems in question. In cases where the an extremal $\bar{x}$ is a strict local minimizer, the standard conditions are then automatically recovered. A refined second order sufficient condition is now given, in which $\nabla^2_{xx} L(\bar{x}, \lambda)$

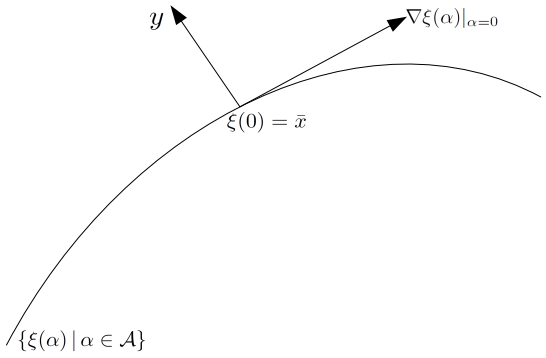


Figure 2.1: Family of constant cost extremals.

is required to be positive definite only on the subspace S' of S:

$$S' := \{y \in S \mid \nabla\xi^T(\alpha)|_{\alpha=0}\, y = 0\}.$$

Figure 2.1 illustrates the manifold of constant cost extremals. The extremal point $\bar{x}$ is the nominal candidate minimizer and $\nabla\xi(\alpha)|_{\alpha=0}$ is the gradient at $\bar{x}$ with respect to α.

Theorem 2.3 (**Refined second order sufficient conditions**). *Suppose that $f, g \in C^2$ and that $\xi(.) \in C^1$. Assume that $\bar{x}$ can be embedded in a family of constant cost extremals, given by $\{\xi(\alpha) \mid \alpha \in \mathcal{A}\}$, such that (HS) hold. Furthermore,*

(H_1) *the $k \times n$ matrix $\nabla\xi^T(0)$ has full row rank,*

(H_2) *there exists an $\alpha-$parameterised multiplier $\lambda^\alpha \in \mathbb{R}^m$ such that*

$$\nabla_x L(\xi(\alpha), \lambda^\alpha) = 0$$

for every $\alpha \in \mathcal{A}$,

(H_3) *the matrix*

$$\nabla^2_{xx} L(\bar{x}, \lambda^0)$$

is positive definite on S'.

Then, there exists $\epsilon > 0$ such that,

$$f(x) \geq f(\bar{x})$$

for all $x \in \bar{x} + \epsilon\mathbb{B}^n$ satisfying $g(x) = 0$.

The following two lemmas play an important role in the proof of Theorem 2.3.

Lemma 2.4 (Application of the Implicit Function Theorem). *Let $\xi : \mathbb{R}^k \to \mathbb{R}^n$ be a C^1 function such that $\xi(\alpha)|_{\alpha=0} = \bar{x}$ and the rows of $\nabla\xi^T(0)$ be linearly independent. Then, there exists an $n-$dimensional neighbourhood of $\bar{x}$, i.e. $(\bar{x}+\epsilon_1\mathbb{B}^n)$, a $k-$dimensional neighbourhood of $\alpha = 0$, i.e. $(0 + \epsilon_2\mathbb{B}^k)$, and a unique C^1 function $\eta(.) : (\bar{x} + \epsilon_1\mathbb{B}^n) \to (0 + \epsilon_2\mathbb{B}^k)$ such that*

$$\eta(\bar{x}) = 0$$

and

$$\nabla\xi^T(0)\big(x - \xi \circ \eta(x)\big) = 0$$

for every $x \in \bar{x} + \epsilon_1\mathbb{B}^n$.

Proof. Define

$$F(x, \alpha) := \nabla\xi^T(0)\big(x - \xi(\alpha)\big).$$

Notice that $F(\bar{x}, 0) = 0$ and that

$$\nabla_\alpha F(\bar{x}, \alpha)|_{\alpha=0} = -\nabla\xi^T(\alpha)|_{\alpha=0}\nabla\xi(0)$$

is a full rank $k \times k$ matrix. Then, the assertions of the lemma follow automatically from the implicit function theorem [3]. ■

Lemma 2.5. *Suppose that $y^T\nabla^2_{xx}L(\bar{x}, \lambda^0)y > 0$ for all $y \in S'$. Then, there exists $\gamma > 0$ such that*

$$y^T\nabla^2_{xx}L(\bar{x}, \lambda^0)y \geq \gamma|y|^2 \quad \text{on } S'. \tag{2.6}$$

Furthermore, the converse is also true.

Proof.($\Leftarrow$) Assume that $y^T\nabla^2_{xx}L(\bar{x}, \lambda^0)y > 0$ for all $y \in S'$ and suppose to the contrary that for some sequence $\gamma_i \downarrow 0$ there exist points $y_i \in S'$ such that

$$y_i^T\nabla^2_{xx}L(\bar{x}, \lambda^0)y_i < \gamma_i|y_i|^2$$

for all i. Defining $s_i := \frac{y_i}{|y_i|}$ gives

$$s_i^T \nabla^2_{xx} L(\bar{x}, \lambda^0) s_i < \gamma_i,$$

where $|s_i| = 1$ and $s_i \in S'$. Since $\{s_i\}$ is a bounded sequence, extract a subsequence (without relabelling) converging to $\bar{s}$. Clearly, $|\bar{s}| = 1$. Now, taking the limit $i \to \infty$ and noting (2.6) gives

$$\bar{s}^T \nabla^2_{xx} L(\bar{x}, \lambda^0) \bar{s} = 0,$$

which is the required contradiction.
($\Rightarrow$) The converse is obvious. ■

Proof (of Theorem 2.3). Take $\epsilon_1, \epsilon_2 > 0$ such that the assertions of Lemma 2.4 hold. Let $\mathcal{A} = 0 + \epsilon_2 \mathbb{B}^k$ and consider an admissible point $x \in \bar{x} + \epsilon_1 \mathbb{B}^n$. By local non-uniqueness,

$$\begin{aligned} f(x) - f(\bar{x}) &= f(x) - f\big(\xi(\alpha)\big) \\ &= f(x) + (\lambda^\alpha)^T g(x) - [f\big(\xi(\alpha)\big) + (\lambda^\alpha)^T g\big(\xi(\alpha)\big)] \\ &= L(x, \lambda^\alpha) - L\big(\xi(\alpha), \lambda^\alpha\big). \end{aligned}$$

Expanding the RHS to second order and making use of (H_2) leads to

$$\begin{aligned} L(x, \lambda^\alpha) - L\big(\xi(\alpha), \lambda^\alpha\big) \;\geq\; & \frac{1}{2}\big(x - \xi(\alpha)\big)^T \nabla^2_{xx} L\big(\xi(\alpha), \lambda^\alpha\big)\big(x - \xi(\alpha)\big) \\ & - \theta(|x - \xi(\alpha)|)|x - \xi(\alpha)|^2 \end{aligned} \tag{2.7}$$

for all $\alpha \in \mathcal{A}$ and some function $\theta(.) : \mathbb{R}^+ \to \mathbb{R}^+$ with the property that $\lim_{s \downarrow 0} \theta(s) = 0$. Since (2.7) holds for all $\alpha \in \mathcal{A}$, the assertions of Lemma 2.4 yield

$$\begin{aligned} L(x, \lambda^\alpha) - L\big(\xi(\alpha), \lambda^\alpha\big) &= L(x, \lambda^{\eta(x)}) - L\big(\xi \circ \eta(x), \lambda^{\eta(x)}\big) \\ &\geq \frac{1}{2}\big(x - \xi \circ \eta(x)\big)^T \nabla^2_{xx} L\big(\xi \circ \eta(x), \lambda^{\eta(x)}\big)\big(x - \xi \circ \eta(x)\big) \\ &\quad - \theta(|x - \xi \circ \eta(x)|)|x - \xi \circ \eta(x)|^2 \end{aligned}$$

and

$$\nabla \xi^T(0)\big(x - \xi \circ \eta(x)\big) = 0$$

for some C^1 function $\eta(.)$. Now, in view of (H_3), the continuous dependence of $\nabla^2_{xx} L$ on

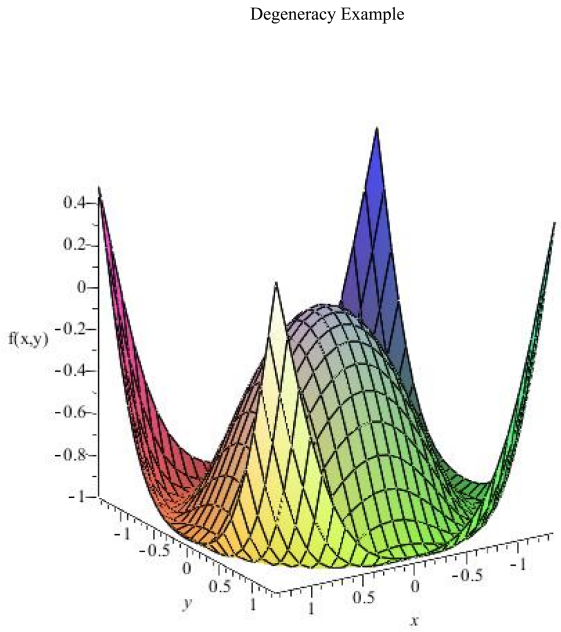


Figure 2.2: Plot of a function with non-unique local minimizers.

x and Lemma 2.5, it can be arranged (by reducing the size of ϵ_1 if necessary) that

$$\begin{aligned}\frac{1}{2}\big(x-\xi\circ\eta(x)\big)^T\nabla^2_{xx}L\big(\xi\circ\eta(x),\lambda^{\eta(x)}\big)\big(x-\xi\circ\eta(x)\big)-\theta(|x-\xi\circ\eta(x)|)|x-\xi\circ\eta(x)|^2\\ \geq\big(\frac{1}{2}\gamma-\theta(|x-\xi\circ\eta(x)|)\big)|x-\xi\circ\eta(x)|^2\\ >0\end{aligned}$$

for all $x\in\bar{x}+\epsilon_1\mathbb{B}^n$. Thus, the proof of the theorem is complete. ■

Example 2.6 illustrates a scenario in which the standard second order sufficient conditions can fail to provide useful information about a minimizer.

Example 2.6. *Consider the optimization problem which is a special case of* (P_0) *with* $g\equiv 0$*,*

$$(P_1)\begin{cases}\text{minimize } f(x_1,x_2)=-\sin(x_1^2+x_2^2)\\ \text{over } (x_1,x_2)\in\mathbb{R}^2.\end{cases}$$

A plot of the objective function is shown in Figure 2.2 for values of x_1 *and* x_2 *between* $[-1.35,..,1.35]$*. Necessary conditions yield extremals at* $(0,0)$ *and* $\bar{x}_1^2+\bar{x}_2^2=n\pi+\frac{\pi}{2}$ *for* $n=0,1,2,...$ *Take the 'ring' family of extremals with* $n=0$ *(a minimizer) the Hessian*

matrix for which is

$$\begin{bmatrix} 4\bar{x}_1^2 & 4\bar{x}_1\sqrt{\frac{\pi}{2}-\bar{x}_1^2} \\ 4\bar{x}_1\sqrt{\frac{\pi}{2}-\bar{x}_1^2} & 4(\frac{\pi}{2}-\bar{x}_1^2) \end{bmatrix} \not> 0$$

for any (x_1, x_2) *in the family of extremals. Standard sufficient conditions, then, prove to be inconclusive. Now, take for example* $\bar{x} = (0, \sqrt{\frac{\pi}{2}})$. *Here,* $\alpha \in (-\pi, +\pi)$ *and*

$$\xi(\alpha) = \bar{x} + \begin{bmatrix} \sin\alpha \\ (\cos\alpha - 1) \end{bmatrix}.$$

Notice that (HS) *are satisfied. Furthermore,*

$$\nabla\xi(0) = \begin{bmatrix} \cos\alpha \\ -\sin\alpha \end{bmatrix}_{\alpha=0} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

According to the new conditions, checking positivity over the subspace

$$\{(y_1, y_2) \in \mathbb{R}^{2n} \mid y_1 = 0\}$$

gives

$$\begin{bmatrix} 0 & y_2 \end{bmatrix} \begin{bmatrix} 4\bar{x}_1^2 & 4\bar{x}_1\sqrt{\frac{\pi}{2}-\bar{x}_1^2} \\ 4\bar{x}_1\sqrt{\frac{\pi}{2}-\bar{x}_1^2} & 4(\frac{\pi}{2}-\bar{x}_1^2) \end{bmatrix}_{x=\bar{x}} \begin{bmatrix} 0 \\ y_2 \end{bmatrix} = 2\pi y_2^2 > 0$$

for all $y_2 \neq 0$ *and thus confirms optimality.*

Of course (P_1) can be very simply solved by re-defining the domain of the optimization problem: it can be reformulated as the problem of minimizing $\tilde{f}(\eta) = -\eta$ over the set of $\eta \in \mathbb{R}$ satisfying

$$\eta = \sin(x_1^2 + x_2^2) \tag{2.8}$$

for some $(x_1, x_2) \in \mathbb{R}^2$. Clearly the minimizing η is $\bar{\eta} = 1$ and (2.8) is satisfied by any x_1, x_2 such that $x_1^2 + x_2^2 = \pi/2$, when $\eta = \bar{\eta}$.

There is no need, in this case, for the application of Theorem 2.3. However, there are other situations where non-uniqueness of minimizers cannot be dealt with simply by problem reformulation and new analytical apparatus is required. Many of the most interesting examples appear to arise in an infinite dimensional setting, such as the one considered in the following section.

2.4 An Abstract Framework

The remainder of this chapter is devoted to the development of an abstract optimization framework for minimizer non-uniqueness and the derivation of new second order sufficient conditions for infinite dimensional problems. It needs to be clarified at the outset that only problems with equality constraints will be examined here because their formulation provides a convenient framework for deriving second order sufficient conditions for continuous time optimal control problems of interest. But more on this will follow in Chapter 3.

Note that generalizing sufficiency conditions applicable in a finite- dimensional setting to an infinite-dimensional one is not straight-forward. The fundamental difficulty here lies in the fact that compactness of closed, bounded subsets, which is guaranteed in a finite dimensional complete space, does not hold in infinite dimensions. Therefore, one can no longer extract subsequences on the unit sphere which converge to some point on the sphere. As a result, the indirect proof techniques which are typically used when proving sufficient conditions for optimality such as, for example [36, Chapter 3] and [35], are not applicable. The assertions of Lemma 2.5 are not automatically true. Indeed a counter example in [50] shows that the two conditions

(i) $y^T \nabla^2_{xx} L(\bar{x}, \lambda^0) y > 0,$

(ii) $y^T \nabla^2_{xx} L(\bar{x}, \lambda^0) y \geq \gamma |y|^2,$

are not equivalent.

Ways around this difficulty have been established by various authors, apparently by either imposing restrictions on the underlying space [14] or type of constraints [38] or considering a stronger form of the second order condition as in [50]. Other methods employed to avoid this difficulty involve exploiting finite dimensional aspects of the problem at hand, such as the finite dimensionality of the space of endpoint values in applications to optimal control for example.

The optimization problem considered here is the following.

$$(P_2) \begin{cases} \text{minimize } F \circ \Pi(z) \\ \text{over elements } z \in Z \\ \text{subject to} \\ G \circ \Pi(z) = 0, \end{cases}$$

where the data for this problem comprise: a Hilbert space $(Z, \|.\|_Z)$, functions $\Pi : Z \to \mathbb{R}^r$, $F : \mathbb{R}^r \to \mathbb{R}$ and $G : \mathbb{R}^r \to \mathbb{R}^l$. The "$\circ$" symbol is used to denote the composition of two functions; for example, $F \circ \Pi(z)$ means $F\big(\Pi(z)\big)$.

Even though details will follow in a later chapter, it suffices at this point to note that Problem (P_2) provides a convenient framework for studying refined optimality conditions for optimal control problems because one can identify with the element $z \in Z$ (and indeed the Hilbert space Z) a pair comprising an initial state x_0 and a control function u. Similarly, $\Pi(z)$ can be thought of as the evaluation of a state trajectory at its endpoints.

Definition. Let the closed unit ball in Z be denoted by $\mathbb{B}_Z$. Take a convex set containing the origin $\mathcal{D} \subset \mathbb{B}_Z$. An element $\bar{z}$ will be said to be a $\mathcal{D}-$local minimizer of (P_2) if $G \circ \Pi(\bar{z}) = 0$ and there exists $\delta > 0$ such that

$$F \circ \Pi(z) \geq F \circ \Pi(\bar{z})$$

for all $z \in \bar{z} + \delta\mathcal{D}$.

he notation $(\bar{z}, \bar{\nu})$ will be used to signify that $\bar{z}$ is a normal extremal associated with multiplier $\bar{\nu}$ where $\bar{\nu} \in \mathbb{R}^l$. Finally, the following notion of non-uniqueness is used.

Definition. A normal extremal $(\bar{z}, \bar{\nu})$ belongs to an *α-parameterised family of extremals* if there exists an open set $\mathcal{A} \subset \mathbb{R}^k$ containing the origin and $(\bar{z}, \bar{\nu})$ can be continuously embedded in a family of constant cost extremals $\{(z^\alpha, \nu^\alpha) \in Z \times \mathbb{R}^l \,|\, \alpha \in \mathcal{A}\}$ and $\alpha \to (z^\alpha, \nu^\alpha)$ is a mapping from $\mathcal{A} \to Z \times \mathbb{R}^l$. In line with earlier definitions interpret $(\bar{z}, \bar{\nu})$ as $(z^\alpha, \nu^\alpha)|_{\alpha=0}$ and for each $\alpha \in \mathcal{A}$,

(*i*) z^α is feasible and $F \circ \Pi(z^\alpha) = F \circ \Pi(\bar{z})$,

(*ii*) $Dh^\alpha(z^\alpha) = 0$,

where

$$h^\alpha(z) \;=\; (F + (\nu^\alpha)^T G) \circ \Pi(z^\alpha). \tag{2.9}$$

The following hypotheses will be invoked in the sequel, in which the Lagrangian plays a similar role to the finite dimensional case and $\theta(.) : \mathbb{R}^+ \to \mathbb{R}^+$ is some 'continuity modulus', i.e. a monotone increasing function such that $\lim_{s \downarrow 0} \theta(s) = 0$. These hypotheses employ the notation $\mathcal{L}(X, Y)$ which is used to denote the set of continuous linear operators from X to Y. Unless stated otherwise, the default norm $\|\cdot\|$ is understood to mean $\|\cdot\|_Z$. Finally, the shorter symbols Dh^α, D^2h^α are used to denote, briefly, the derivatives evaluated at z^α i.e. $Dh(z^\alpha)$, $D^2h(z^\alpha)$ and so on.

Hypotheses

(H_1^A) (Continuity): For every $\epsilon > 0$ there exists $\delta > 0$ such that $z^\alpha \in \bar{z} + \epsilon\mathcal{D}$ for all $\alpha \in \delta\mathbb{B}^k$.

(H_2^A) (End-point differentiability): The mapping $\alpha \to \Pi(z^\alpha) : \mathcal{A} \to \mathbb{R}^r$ is continuous and differentiable at $\alpha = 0$. To this end define

$$\Gamma := \big(\nabla_\alpha \Pi(z^\alpha)|_{\alpha=0}\big)^T. \tag{2.10}$$

For all $\alpha \in \mathcal{A}$ there exist $Dh^\alpha \in \mathcal{L}(Z, \mathbb{R})$, $D\Pi^\alpha \in \mathcal{L}(Z, \mathbb{R}^r)$ and $D^2h^\alpha \in \mathcal{L}(Z, Z)$ such that, for any $\epsilon' > 0$ and $z \in z^\alpha + \epsilon'\mathcal{D}$

$$|h(z) - h(z^\alpha) - Dh^\alpha(z - z^\alpha) - \frac{1}{2}\langle (z - z^\alpha), D^2h^\alpha(z - z^\alpha)\rangle| \;\le\; \theta(\epsilon')\|z - z^\alpha\|^2$$

and

$$|\Pi(z) - \Pi(z^\alpha) - D\Pi^\alpha(z - z^\alpha)| \;\le\; \theta(\epsilon')\|z - z^\alpha\| \,.$$

(H_3^A) (Controllability w.r.t. linearised constraints): There exists $k > 0$ such that, for each $y \in Z$ and $\alpha \in \mathcal{A}$, an element $y' \in Z$ can be found such that

$$\Gamma\, D\Pi^\alpha y' \;=\; 0 \text{ and } \nabla G(\Pi(z^\alpha))\, D\Pi^\alpha y' \;=\; 0$$

and

$$\|y' - y\| \;\le\; k\Big(|\Gamma\, D\Pi^\alpha y| + |\nabla G(\Pi(z^\alpha)) D\Pi^\alpha y|\Big) \,.$$

Second Order Sufficient Conditions. The theorem statement below makes reference to the Lagrangian h^α in (2.9), its first and second order derivatives the existence of which is asserted in Hypothesis (H_2^A) and the $k \times r$ matrix Γ as in (2.10).

Theorem 2.7. *Consider the optimization problem (P_2). Take an element $(\bar z, \bar\nu) \in Z \times \mathbb{R}^l$ belonging to an α-parameterised family of extremals and a set $\mathcal{D} \subset \mathbb{B}_Z$. Assume ($H_1^A$) – ($H_3^A$) and, furthermore, that:*

(E) *the rows of Γ and $\nabla G\big(\Pi(\bar z)\big)$ are linearly independent,*

(C) *there exist $\bar\epsilon, \gamma, \bar\alpha > 0$ such that $\bar\alpha\mathbb{B}^k \subset \mathcal{A}$, and an $r \times r$ matrix valued function $(\alpha, \epsilon) \to W^{\alpha,\epsilon}$ on $\bar\alpha\mathbb{B}^k \times [0, \bar\epsilon]$, which is continuous at $(0, 0)$, such that for every $\epsilon \in [0, \bar\epsilon]$*

(i) $\inf\big\{\langle y, (D^2h^\alpha - \epsilon I)y\rangle \,|\, y \in Z\,, D\Pi^\alpha y = \xi\big\} \ge \xi^T W^{\alpha,\epsilon}\xi$

(ii) $\xi^T W^{0,0}\xi \ge \gamma|\xi|^2$ *for all ξ such that*

$$\Gamma\,\xi \;=\; 0 \quad \textit{and} \quad \nabla G(\Pi(\bar z))\xi \;=\; 0. \tag{2.11}$$

Then $\bar z$ is a $\mathcal{D}-$local minimizer.

Remarks.

1. A sufficient condition for local optimality of an extremal $(\bar{z}, \bar{\nu})$ for problem (P_2), more akin to classical sufficient conditions in optimal control, would (among other standard conditions) place the following requirements on the second derivative $D^2h(\bar{z})$ of the Lagrangian function: there exists $\gamma > 0$ such that

$$\xi^T W \xi \geq \gamma |\xi|^2$$

for all $\xi \in \mathbb{R}^r$ such that

$$\nabla G(\Pi(\bar{z}))\xi \;=\; 0\,,$$

where W is the unique symmetric matrix such that

$$\xi^T W \xi \;=\; \inf\left\{\langle y, D^2h(\bar{z})y\rangle \mid D\Pi(\bar{z})y = \xi\right\}.$$

2. Hypothesis (H_3^A) concerns the constraints

$$\Gamma\, D\Pi^\alpha y \;=\; 0 \text{ and } \nabla G(\Pi(z^\alpha))\, D\Pi^\alpha y \;=\; 0$$

on the elements $y \in Z$. This 'metric regularity' hypothesis requires that, given any $y \in Z$ which violates these constraints, an element $y' \in Z$ can be found which satisfies the constraints and for which the distance from y' is linearly estimated in terms of the constraint violation of y:

$$\Big(|\Gamma\, D\Pi^\alpha y| + |\nabla G(\Pi(z^\alpha)) D\Pi^\alpha y|\Big).$$

Proposition 2.8. *Consider a normal extremal $(\bar{z}, \bar{\nu}) \in Z \times \mathbb{R}^l$ belonging to an α-parameterised family of extremals and assume (H_2^A) and (E). Then, hypothesis (H_3^A) is satisfied for $\mathcal{A}$ replaced by some neighbourhood $\mathcal{A}' \subset \mathcal{A}$ of the origin under the following condition.*
$(\tilde{H}_3^A)$ For each $\alpha \in \mathcal{A}'$ the linear map $D\Pi^\alpha$ has a right inverse $(D\Pi^\alpha)^\dagger : \mathbb{R}^r \to Z$, i.e.

$$(D\Pi^\alpha)(D\Pi^\alpha)^\dagger = I_{r\times r},$$

$\alpha \to (D\Pi^\alpha)^\dagger$ is continuous and there exists $k > 0$ such that

$$\|(D\Pi^\alpha)^\dagger\| \leq k \quad \textit{for all } \alpha \in \mathcal{A}'.$$

The proposition provides directly verifiable sufficient conditions for satisfaction of the "controllability" hypothesis (H_3^A). These sufficient conditions combine the requirements that,

following linearization, arbitrary boundary values can be achieved and a non-degeneracy condition holds on the representation of the constraints and boundary values of the equal cost extremals.

Proof. Define

$$V^\alpha := \begin{pmatrix} \Gamma \\ \nabla G\big(\Pi(z^\alpha)\big) \end{pmatrix}.$$

Since $\alpha \to \Pi(z^\alpha)$ is continuous at $\alpha = 0$ and V^0 has full row rank, the matrix inverse $(V^\alpha V^{\alpha T})^{-1}$ exists and is bounded on some neighbourhood $\mathcal{A}' \subset \mathcal{A}$ of 0. Take any $y \in Z$ and $\alpha \in \mathcal{A}'$. Define

$$y' := y - (D\Pi^\alpha)^\dagger V^{\alpha T}(V^\alpha V^{\alpha T})^{-1} V^\alpha D\Pi^\alpha y.$$

Then

$$\begin{aligned} V^\alpha D\Pi^\alpha y' &= V^\alpha D\Pi^\alpha y - V^\alpha D\Pi^\alpha y \\ &= 0. \end{aligned}$$

It follows that

$$\Gamma D\Pi^\alpha y' = 0\,, \nabla G\big(\Pi(z^\alpha)\big) D\Pi^\alpha y' = 0.$$

Furthermore,

$$\|y' - y\| \le k\Big(|\Gamma D\Pi^\alpha y| + |\nabla G\big(\Pi(z^\alpha)\big) D\Pi^\alpha y|\Big)$$

for k a scalar uniform bound on the norm of

$$(D\Pi^\alpha)^\dagger V^{\alpha T}(V^\alpha V^{\alpha T})^{-1},$$

on $\alpha \in \mathcal{A}'$. ■

The following lemma makes reference to Figure 2.3. It asserts the existence of the ϵ and δ neighbourhoods around $\bar{z}$ (top figure) such that for any z in the $\epsilon-$neighbourhood there exists suitable α such that z^α is in a $\delta-$neighbourhood and, in addition, an orthogonality condition (bottom figure) is satisfied.

Lemma 2.9. *Suppose that* $(H_1^A) - (H_3^A)$ *and* (E) *hold. Define*

$$M(z, \alpha) := \Gamma\,(\Pi(z) - \Pi(z^\alpha)).$$

Take any $\delta > 0$ *and* $\bar{\alpha} > 0$ *such that* $\bar{\alpha}\mathbb{B}^k \subset \mathcal{A}$*. Then, there exists* $\epsilon > 0$ *with the property:*

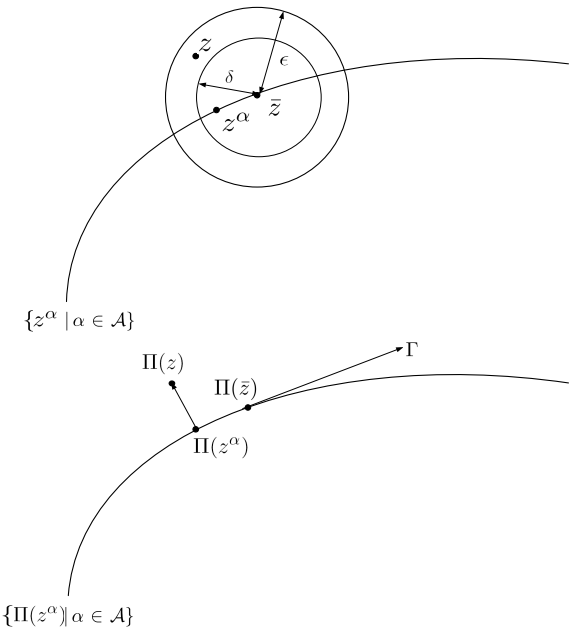


Figure 2.3: An intuitive interpretation of Lemma 2.9.

given any $z \in \bar{z} + \epsilon\mathcal{D}$, $\alpha \in \bar{\alpha}\mathbb{B}^k$ *can be found such that* $z^\alpha \in \bar{z} + \delta\mathcal{D}$ *and*

$$M(z, \alpha) = 0.$$

Proof. Fix $\delta > 0$ (and $\bar{\alpha} > 0$). Notice that, in view of (H_2^A),

$$\begin{aligned} \nabla_\alpha M(\bar{z}, 0) &= -\Gamma \nabla_\alpha \Pi(z^\alpha)|_{\alpha=0} \\ &= -\Gamma \Gamma^T. \end{aligned}$$

Moreover, in view of (E), $\nabla_\alpha M(\bar{z}, 0)$ is a nonsingular $k \times k$ matrix. Notice that

$$M(\bar{z}, 0) = \Gamma \big(\Pi(\bar{z}) - \Pi(z^0)\big) = 0$$

and hence

$$M(z, \alpha) = D_z M(\bar{z}, 0)(z - \bar{z}) + D_\alpha M(\bar{z}, 0)\alpha + \zeta$$

where $\zeta \in \mathbb{R}^k$ contains the higher order 'error' terms. Now define the mapping $W : Z \times \mathbb{R}^k \to \mathbb{R}^k$ by

$$W(z, \zeta) := M(z, \alpha(z, \zeta)) + \zeta,$$

in which

$$\alpha(z, \zeta) := -\big(D_\alpha M(\bar{z}, 0)\big)^{-1}[D_z M(\bar{z}, 0)(z - \bar{z}) + \zeta].$$

Note that

$$W(\bar{z},0) = M(\bar{z},0) = 0$$

and that

$$\begin{aligned} D_z W(\bar{z},0) &= D_z M(\bar{z},0) - \Gamma\,\Gamma^T(\Gamma\,\Gamma^T)^{-1}[D_z M(\bar{z},0)] \\ &= 0. \end{aligned}$$

Furthermore,

$$\begin{aligned} D_\zeta W(\bar{z},0) &= D_\alpha M(\bar{z},0) D_\zeta \alpha(\bar{z},0) + I_{k\times k} \\ &= -I_{k\times k} + I_{k\times k} = 0_{k\times k}. \end{aligned}$$

Now consider a first order expansion, about $(\bar{z},0)$, to obtain

$$\begin{aligned} |W(z,\zeta)| &= |W(z,\zeta) - W(\bar{z},0)| \\ &\leq |D_z W(\bar{z},0)(z-\bar{z})| + |D_\zeta W(\bar{z},0)\zeta| + \theta(\epsilon)(\|z-\bar{z}\| + |\zeta|) \\ &= 0 + \theta(\epsilon)(\|z-\bar{z}\| + |\zeta|) \end{aligned}$$

for some positive function $\theta(.)$ satisfying $\lim_{s\downarrow 0}\theta(s) = 0$. Now choose $\epsilon > 0$ such that

$$\theta(\epsilon)(\|z-\bar{z}\| + |\zeta|) \leq \epsilon.$$

for all $z \in \bar{z} + \epsilon\mathcal{D}$ and $|\zeta| \leq \epsilon$. This can be achieved by selecting ϵ such that

$$\theta(\epsilon) < \frac{1}{2}.$$

Since $\alpha(\bar{z},0) = 0$ and since it depends continuously on z and ζ it can be arranged, by further reducing the size of ϵ if necessary, that

$$\max_{\{(z,\zeta)|z\in\bar{z}+\epsilon\mathcal{D}, \text{ and } |\zeta|\leq\epsilon\}} \|z^{\alpha(z,\zeta)} - \bar{z}\| < \delta.$$

For any such z, it can be seen that $W(z,.)$ maps $\epsilon\mathbb{B}^k$ into itself and that the mapping is continuous. By the Schauder fixed point theorem (see for e.g. [78]) the mapping has a fixed point $\bar{\zeta}$. For such ζ write $\alpha = \alpha(z,\bar{\zeta})$. But then,

$$M(z,\alpha) = W(z,\bar{\zeta}) - \bar{\zeta} = 0$$

for any $z \in \bar{z} + \epsilon\mathcal{D}$ where $\alpha \in \bar{\alpha}\mathbb{B}^k$ and $\|z^\alpha - \bar{z}\| < \delta$. This confirms the assertions of the lemma. ■

Lemma 2.10. *Let $\mathcal{N}$ be a neighbourhood of the origin in $\mathbb{R}^r$. Consider $n \times n$ and $m \times n$ matrix valued functions $\beta \mapsto Q^\beta$, $\beta \mapsto D^\beta$ which are continuous at 0. Assume that there exists $\bar{\gamma} > 0$ such that*

$$x^T Q^0 x \geq \bar{\gamma}|x|^2$$

for all $x \in \mathbb{R}^n$ such that $D^0 x = 0$.

Then for any $\gamma' \in (0, \bar{\gamma})$ there exists a neighbourhood $\mathcal{N}' \subset \mathcal{N}$ of the origin such that, for each $\beta \in \mathcal{N}'$,

$$x^T Q^\beta x \geq \gamma'|x|^2$$

for all $x \in \mathbb{R}^n$ s.t. $D^\beta x = 0$.

Proof. Take $\bar{\gamma}$ as in the lemma statement and any $\gamma' \in (0, \bar{\gamma})$. Consider neighbourhoods $\mathcal{N}_i = \{\beta \mid |\beta| < \frac{1}{i}\}$. Then, if the assertions of the lemma are false, for each i sufficiently large there exists x_i such that

$$x_i^T Q^{\beta_i} x_i < \gamma'|x_i|^2 \quad \text{and } D^{\beta_i} x_i = 0. \tag{2.12}$$

Now, define $s_i := \frac{x_i}{|x_i|}$; note that $|s_i| = 1$ and extract a subsequence $s_i \to \bar{s}$. Then,

$$s_i^T Q^{\beta_i} s_i < \gamma' \quad \text{and } D^{\beta_i} s_i = 0.$$

Taking the limit $i \to \infty$ yields $\bar{s}^T Q^0 \bar{s} \leq \gamma'$ and $D^0 \bar{s} = 0$. Since $|\bar{s}| = 1$, this contradicts the hypotheses and confirms the lemma. ■

With these preliminary results in place, the proof of Theorem 2.7 can now be presented. This hinges upon the fact that given a nominal $\bar{z}$ and an arbitrary z such that $\|z - \bar{z}\| \leq \epsilon$, a z' can be found that approximates $(z - \bar{z})$ but at the same time satisfies the linear equality and orthogonality constraints. A quadratic lower estimate of the difference in the 'cost' (from $\bar{z}$ to z) which is expressed in terms of z' can then be found. Thinking ahead in terms of optimal control, this would be equivalent to an initial state and control that when applied to the linearised system result in an arc that satisfies the linearised endpoint constraints and the orthogonality condition.

Proof (of Theorem 2.7). Take $\delta > 0$, $\epsilon > 0$, $\bar{\alpha} > 0$ such that $\bar{\alpha}\mathbb{B}^k \subset \mathcal{A}$, the values of which will be chosen in the sequel. Furthermore, take $h^\alpha(.)$ as defined in (2.9) and consider an extremal $\bar{z}$ and feasible $z \in \bar{z} + \epsilon\mathcal{D}$ such that $G \circ \Pi(z) = 0$. For any $\alpha \in \bar{\alpha}\mathbb{B}^k$ such that $z^\alpha \in \bar{z} + \delta\mathcal{D}$,

$$\begin{aligned} F \circ \Pi(z) - F \circ \Pi(\bar{z}) &= F \circ \Pi(z) - F \circ \Pi(z^\alpha) \\ G \circ \Pi(\bar{z}) &= G \circ \Pi(z^\alpha) = 0. \end{aligned}$$

Moreover, taking into account (H_2^A),

$$\begin{aligned} F\circ\Pi(z)-F\circ\Pi(\bar{z}) &= F\circ\Pi(z)-F\circ\Pi(z^\alpha)\\ &= F\circ\Pi(z)-F\circ\Pi(z^\alpha)+\nu^\alpha\cdot[G\circ\Pi(z)-G\circ\Pi(z^\alpha)]\\ &\geq Dh^\alpha(z-z^\alpha)+\frac{1}{2}\langle(z-z^\alpha),D^2h^\alpha(z-z^\alpha)\rangle-\theta(\epsilon+\delta)\|z-z^\alpha\|^2\\ &\geq 0+\frac{1}{2}\langle(z-z^\alpha),D^2h^\alpha(z-z^\alpha)\rangle-\theta(\epsilon+\delta)\|z-z^\alpha\|^2. \end{aligned}$$

In view of Lemma 2.9, for any $\delta>0$ it can be arranged, by reducing the magnitude of ϵ if necessary, that there exists $\alpha\in\bar{\alpha}\mathbb{B}^k$ such that $\|z^\alpha-\bar{z}\|\leq\delta$

$$\Gamma\big(\Pi(z)-\Pi(z^\alpha)\big)=0 \tag{2.13}$$

and

$$G\circ\Pi(z)-G\circ\Pi(z^\alpha)=0. \tag{2.14}$$

Combining the differentiability hypotheses with (2.13) and (2.14), it follows that

$$|\Gamma D\Pi^\alpha(z-z^\alpha)|+|\nabla G(\Pi(z^\alpha))D\Pi^\alpha(z-z^\alpha)|\leq\theta_2(\epsilon+\delta)\|z-z^\alpha\|,$$

for some continuity modulus $\theta_2(.)$. By the controllability hypothesis, there exists z' such that

$$\Gamma D\Pi^\alpha z'=0 \text{ and } \nabla G\big(\Pi(z^\alpha)\big)D\Pi^\alpha z'=0.$$

Furthermore,

$$\|z'-(z-z^\alpha)\|\leq\theta_3(\epsilon+\delta)\|z-z^\alpha\|$$

for some new continuity modulus $\theta_3(.)$. By reducing the magnitude of ϵ and δ it can be arranged that $\theta_3(\epsilon+\delta)<1$. Then,

$$\|z'-(z-z^\alpha)\|\leq\theta_3(\epsilon+\delta)(\|z'-(z-z^\alpha)\|+\|z'\|)$$

and so

$$\|z'-(z-z^\alpha)\|\leq\left(\frac{\theta_3(\epsilon+\delta)}{1-\theta_3(\epsilon+\delta)}\right)\|z'\|. \tag{2.15}$$

In view of the properties of an $\alpha-$parameterised normal extremal in Definition 2.4 and writing $\Delta z:=z-z^\alpha$,

$$F\circ\Pi(z)-F\circ\Pi(\bar{z})\geq\frac{1}{2}\langle\Delta z,D^2h^\alpha\Delta z\rangle-\theta(\epsilon+\delta)\|\Delta z\|^2. \tag{2.16}$$

Noting the identity

$$\langle \Delta z, D^2h^\alpha \Delta z\rangle = \langle z', D^2h^\alpha z'\rangle + 2\langle z', D^2h^\alpha(\Delta z - z')\rangle + \langle (\Delta z - z'), D^2h^\alpha(\Delta z - z')\rangle$$

it follows that

$$\begin{aligned}
\langle \Delta z, D^2h^\alpha \Delta z\rangle &\geq \langle z', D^2h^\alpha z'\rangle - 2\|D^2h^\alpha\|\|z'\|\|\Delta z - z'\| - \|D^2h^\alpha\|\|\Delta z - z'\|^2 \\
&\geq \langle z', D^2h^\alpha z'\rangle - \|D^2h^\alpha\| \left(2\frac{\theta_3(\epsilon+\delta)}{1-\theta_3(\epsilon+\delta)} + \left(\frac{\theta_3(\epsilon+\delta)}{1-\theta_3(\epsilon+\delta)}\right)^2 \right) \|z'\|^2 \\
&= \langle \Delta z, D^2h^\alpha \Delta z\rangle \geq \langle z', D^2h^\alpha z'\rangle - \theta_4(\epsilon+\delta)\|D^2h^\alpha\|\|z'\|^2 \qquad (2.17)
\end{aligned}$$

where $\theta_4(.)$ is given by

$$\theta_4(\epsilon+\delta) = \frac{2\theta_3(\epsilon+\delta)}{1-\theta_3(\epsilon+\delta)} + (\frac{\theta_3(\epsilon+\delta)}{1-\theta_3(\epsilon+\delta)})^2.$$

By inequality (2.15), it can now be deduced that

$$\begin{aligned}
\|\Delta z\| \leq (\|\Delta z - z'\| + \|z'\|) &\leq \left(\frac{\theta_3(\epsilon+\delta)}{1-\theta_3(\epsilon+\delta)} + 1\right)\|z'\| \\
&= \left(\frac{1}{1-\theta_3(\epsilon+\delta)}\right)\|z'\|. \qquad (2.18)
\end{aligned}$$

Substituting (2.17) and (2.18) into (2.16) gives

$$F\circ\Pi(z) - F\circ\Pi(\bar{z}) \geq \frac{1}{2}\langle z', D^2h^\alpha z'\rangle - q(\epsilon,\delta)\|z'\|^2 \qquad (2.19)$$

where

$$q(\epsilon,\delta) = \frac{1}{2}\theta_4(\epsilon+\delta)\|D^2h^\alpha\|_{\mathcal{L}(Z,Z)} + \theta(\epsilon+\delta)(\frac{1}{1-\theta_3(\epsilon+\delta)})^2.$$

(Notice that $q(\epsilon,\delta)$ can be made arbitrarily small by reducing the size of ϵ and δ.) But

$$\begin{aligned}
\langle z', D^2h^\alpha z'\rangle - q(\epsilon,\delta)\|z'\|^2 &= \langle z', (D^2h^\alpha - q(\epsilon,\delta)I)z'\rangle \\
&\geq (D\Pi^\alpha z')^T W^{\alpha,q(\epsilon,\delta)}(D\Pi^\alpha z'), \qquad (2.20)
\end{aligned}$$

where $W^{\alpha,q(\epsilon,\alpha)}$ is the matrix of condition $(C)(i)$ in the theorem statement. Take any $\gamma' \in (0,\gamma)$, where γ is the constant in condition $(C)(ii)$. Consequently, $|\alpha| \leq \bar{\alpha}$ and $D\Pi^\alpha z'$ satisfies the constraint

$$\Gamma\, D\Pi^\alpha z' = 0 \text{ and } \nabla G\big(\Pi(z^\alpha)\big) D\Pi^\alpha z' = 0.$$

By (H_1^A) and (H_2^A), $G\big(\Pi(z^{\alpha'})\big) \to G\big(\Pi(\bar{z})\big)$ as $\alpha' \to 0$. It follows now from Lemma 2.10

that, following a further reduction in the size of δ, ϵ and $\bar{\alpha}$ if required,

$$(D\Pi^{\alpha}z')^T W^{\alpha,q(\epsilon,\delta)}(D\Pi^{\alpha}z') \;\geq\; \gamma'|(D\Pi^{\alpha}z')|^2.$$

Then, from (2.19) and (2.20),

$$F\circ\Pi(z) - F\circ\Pi(\bar{z}) \;\geq\; \frac{1}{2}\gamma'|(D\Pi^{\alpha}z')|^2 \;\geq\; 0.$$

Since z was an arbitrary feasible point in $\bar{z}+\epsilon\mathcal{D}$, it has been proved that $\bar{z}$ is a $\mathcal{D}$-local minimizer for (P_2). ■

What remains is to examine how this framework links to problems from optimal control. A key aspect of this analysis, which is the subject of Chapter 3, is the verification of the hypotheses of this abstract framework in an optimal control setting.

2.5 Conclusion

This chapter has given a brief account of the well-known first order necessary and standard second order sufficient conditions of optimality. For the standard sufficient conditions in nonlinear programming a numerical example demonstrates that the strict positivity condition on the Hessian of the Lagrangian may not always be satisfied on the required tangent space. Consequently, local minimizers cannot be identified using standard second order conditions when the solution is locally non-unique. Refinements are then proposed in Section 2.3 for problems of finite dimensions with equality constraints. Specifically, the new conditions explore the positive definiteness of the Hessian over a smaller subspace which is determined by an orthogonality condition which ensures that the 'perturbation direction' is normal to the 'direction of non-uniqueness'. The proof techniques employed here are based on the compactness of the unit sphere, which is of finite dimensionality.

The second part of the chapter focuses on developing second order sufficient conditions for an optimization problem posed over an abstract space. In this set-up an intermediate function, Π, maps elements from a Hilbert space to a finite dimensional space. This function has analogs in optimal control, where it can be thought of as the evaluation of a state trajectory at its boundaries, given an initial state x_0 and control u. In this sense, the framework developed here plays a central role in developing refined second order sufficient conditions for optimal control problems, which is the topic of Chapter 3.

Chapter 3

Non-unique local minimizers in optimal control

Analogous results to the second order sufficient conditions for optimality in nonlinear programming can be derived for optimal control problems where a quadratic approximation to the cost functional, known as the accessory cost, is required to be strictly positive to guarantee optimality of a solution. Conditions of this nature can be thought to correspond to the strict positivity test of the Hessian for problems in nonlinear programming. Not surprisingly then, they also convey the information that a minimizer is locally unique. For a large class of optimal control problems, however, which includes periodic optimal control problems, this is not the case and these tests fail to identify optimal solutions. It is of interest, then, to derive more general conditions that can also be applied for problems in this class.

This chapter contains a treatment of non-autonomous optimal control problems with endpoint constraints that admit non-unique local solutions. For these problems it is shown that the requirements of the standard tests cannot be met and a new framework is presented which can accommodate them. This is based on the abstract framework of Chapter 2, in which an optimal control problem is posed over pairs comprising initial states and controls.

The assertions of Theorem 2.7, in Chapter 2, carry over to optimal control provided that the hypotheses of the abstract optimization problem and the hypotheses of the optimal control problem are compatible. It is shown that the 'typical' hypotheses, supplemented with mild continuity assumptions, are sufficient to provide this common ground.

Finally, two applications, featuring periodic and mixed endpoint constraints, serve to illustrate the superiority of these new tests in terms of applicability and the computations involved.

3.1 Introduction

The optimal control problem studied in this chapter is the following:

$$(P)\begin{cases} \text{minimize } g(x(0),x(T)) \\ \text{over } x(.)\in W^{1,1}([0,T];\mathbb{R}^n) \text{ and measurable functions } u(.):[0,T]\to\mathbb{R}^m \\ \text{subject to} \\ \dot{x}(t)=f(t,x(t),u(t)) \ \text{ a.e. } t\in[0,T] \qquad (3.1) \\ d(x(0),x(T))=0. \qquad (3.2) \end{cases}$$

Data for (P) comprise: a real number $T>0$, the cost function $g:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$, the nonlinear dynamical constraints $f:\mathbb{R}\times\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}^n$ and the mixed boundary constraint function $d:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}^l$. The cost function here generalises the special cases where the cost comprises an integral and possibly an additional cost term imposed on the final state.

Definition. A measurable m-valued function $u(.)$ is called the *control function*. The $W^{1,1}$ n-valued function $x(.)$ that satisfies (3.1) is known as the *state*. A pair $(x(.),u(.))$, in which the state $x(.)$ is the solution corresponding to the control $u(.)$ is referred to as a *process* for (P). If $x(.)$ (corresponding to control $u(.)$) satisfies (3.2) then the process $(x(.),u(.))$ is called an *admissible process*.

Definition. An admissible process $(\bar{x}(.),\bar{u}(.))$ which achieves the minimum cost over all admissible processes is called a *minimizer* for (P) and $\bar{u}(.)$ is called an optimal control. A more restrictive notion of optimality is given below.

Definition. An admissible process $(\bar{x}(.),\bar{u}(.))$ is called a *weak local minimizer* for (P) if it is a minimizer in the following sense: there exists $\epsilon>0$ such that

$$g(x(0),x(T))\geq g(\bar{x}(0),\bar{x}(T))$$

for all admissible processes satisfying

$$\|x(.)-\bar{x}(.)\|_{L^\infty}\leq\epsilon \text{ and } \|u(.)-\bar{u}(.)\|_{L^\infty}\leq\epsilon.$$

First order necessary conditions in optimal control can be thought of as analogues to the requirement that the gradient of the Lagrangian is zero in nonlinear programming. As such, they can be used to provide candidate solutions for the minimization problem. A widely known set of first order necessary conditions in optimal control is the *Pontryagin Maximum Principle* or PMP for short [13]. The conditions of the PMP have, naturally, been weakened over the years to accommodate a wider range of problems. As an example, one could refer to [65] for a Maximum Principle for discontinuous state trajectories or

to [64] for a Maximum Principle with general end point constraints, "nonsmooth" dynamic constraints and pathwise state constraints. A simpler form of the Maximum Principle is used in the analysis of the second order sufficient conditions developed in this thesis. This involves the Hamiltonian function $H : [0,T] \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m$

$$H(t,x,p,u) := p^T f(t,x,u)\,, \tag{3.3}$$

in which $p(.) \in W^{1,1}([0,T];\mathbb{R}^n)$ is called an *adjoint arc*. The conditions also make reference to a multiplier $\nu \in \mathbb{R}^l$ (accompanying the endpoint constraints) and a multiplier $\lambda \in \mathbb{R}$ (accompanying the cost), not both zero, such that

$$\begin{aligned} -\dot{p}(t) &= f_x^T(t,\bar{x}(t),p(t),\bar{u}(t))\,p(t) && (3.4)\\ -p^T(0) &= \lambda g_{x_0}(\bar{x}(0),\bar{x}(T)) + \nu^T d_{x_0}(\bar{x}(0),\bar{x}(T)) && (3.5)\\ p^T(T) &= \lambda g_{x_T}(\bar{x}(0),\bar{x}(T)) + \nu^T d_{x_T}(\bar{x}(0),\bar{x}(T)) && (3.6) \end{aligned}$$

and the minimization condition

$$H(t,\bar{x}(t),p(t),\bar{u}(t)) = \min_u H(t,\bar{x}(t),p(t),u(t)). \tag{3.7}$$

These conditions are collectively referred to as the Maximum Principle because, when initially formulated, the adjoint arc $p(.)$ was replaced by $p'(.) = -p(.)$. When (3.4)-(3.7) are expressed in terms of $p'(.)$ the minimization property (3.7) is replaced by a maximization property.

An admissible process $(\bar{x}(.),\bar{u}(.))$ which satisfies the conditions of the PMP is referred to as an *extremal process*. Furthermore, an extremal is called a *normal extremal* if $\lambda > 0$, in which case by scaling the multipliers it can be arranged that $\lambda = 1$.

An admissible process $(\bar{x}(.),\bar{u}(.))$ is referred to as a *weak normal extremal* when the weaker condition

$$H_u(t,\bar{x}(t),p(t),u(t))|_{u=\bar{u}} = 0 \quad \text{a.e. } t \in [0,T] \tag{3.8}$$

replaces (3.7). Notice that the minimization of the Hamiltonian also implies

$$H_{uu}(t,\bar{x}(t),p(t),u(t))|_{u=\bar{u}} \geq 0 \quad \text{a.e. } t \in [0,T] \tag{3.9}$$

known as the Legendre condition.

3.2 Second Order Conditions in Optimal Control: A Review

Second order conditions of optimality for optimal control problems focus on the analysis of the so-called *accessory problem* developed in, for example, [70], [33, Theorem 3.1] or [76, Theorem 5.1]. The problem is defined as follows.

For each $t \in [0,T]$, let

$$(A(t), B(t)) := (f_x(t, \bar{x}(t), \bar{u}(t)), f_u(t, \bar{x}(t), \bar{u}(t)))$$

and

$$\begin{pmatrix} Q(t) & D(t) \\ D^T(t) & R(t) \end{pmatrix} = \begin{pmatrix} H_{xx}(t, \bar{x}(t), p(t), \bar{u}(t)) & H_{xu}(t, \bar{x}(t), p(t), \bar{u}(t)) \\ H_{xu}^T(t, \bar{x}(t), p(t), \bar{u}(t)) & H_{uu}(t, \bar{x}(t), p(t), \bar{u}(t)) \end{pmatrix}$$

in which $H(t, \bar{x}(t), p(t), \bar{u}(t))$ is defined as in (3.3). Also define the 'end-point' Lagrangian $\tilde{h} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$

$$\tilde{h}(x_0, x_T; \nu) := g(x_0, x_T) + \nu^T d(x_0, x_T)$$

and

$$G := \begin{pmatrix} \tilde{h}_{x_0 x_0} & \tilde{h}_{x_0 x_T} \\ \tilde{h}_{x_0 x_T}^T & \tilde{h}_{x_T x_T} \end{pmatrix},$$

in which the entries of the block matrix on the right are evaluated at $(\bar{x}(0), \bar{x}(T))$ and ν. Finally define the accessory cost, or *second variation*, to be

$$\begin{aligned} J_A(y(.), v(.)) \quad := \quad & \frac{1}{2} \int_0^T y^T(t) H_{xx}(t) y(t) + 2 y^T(t) H_{xu}(t) v(t) + v^T(t) H_{uu}(t) v(t)\, dt + \\ & \frac{1}{2} \begin{pmatrix} y^T(0) & y^T(T) \end{pmatrix} G \begin{pmatrix} y(0) \\ y(T) \end{pmatrix}, \end{aligned}$$

which is a quadratic estimate of the change in the cost between $J(\bar{x}(.), \bar{u}(.))$ and $J(x(.), u(.))$. The accessory problem associated with a weak normal extremal $(\bar{x}(.), \bar{u}(.))$ is

$$(A) \begin{cases} \text{minimize } J_A(y(.), v(.)) \\ \text{over } y(.) \in W^{1,1}([0,T]; \mathbb{R}^n) \text{ and } v(.) \in L^2([0,T], \mathbb{R}^m) \\ \text{subject to} \\ \dot{y}(t) = A(t) y(t) + B(t) v(t) \text{ a.e. } t \in [0,T] & (3.10) \\ d_{x_0} y(0) + d_{x_T} y(T) = 0 & (3.11) \end{cases}$$

in which $(y(.), v(.))$ is a process satisfying the linear dynamics (3.10) and linear endpoint constraints (3.11). The d_{x_i} are understood to be evaluated at $(\bar{x}(0), \bar{x}(T))$.

Hypotheses

(H_1') $\bar{u}(.)$ is essentially bounded,

(H_2') $f(t,.,.)$ is of class C^2 for all $t \in [0,T]$, and $f(.,.,.)$ and its derivatives with respect to the (x,u) variables, up to order 2, are continuous. The functions g and d are also of class C^2.

(H_3') There exists $\epsilon > 0$ such that

$$R(t) > \epsilon I \quad \text{for all } t \in [0,T].$$

(H_4') The pair $(A(.),B(.))$ is controllable on $[0,T]$.

The PMP supplemented by conditions on the accessory cost provides second order optimality conditions. These well-known results are summarised by the following theorem.

Theorem 3.1. *Consider a weak normal extremal for $(\bar{x}(.),\bar{u}(.))$. Assume hypotheses (H_1') - (H_4') are satisfied.*

(a) *(Necessary Condition) Suppose that $(\bar{x}(.),\bar{u}(.))$ is a weak local minimizer. Then*

$$J_A(y(.),v(.)) \geq 0 \tag{3.12}$$

for all admissible processes $(y(.),v(.))$ of (A).

(b) *(Sufficient Condition) Suppose that there exists $\gamma > 0$ such that*

$$J_A(y(.),v(.)) \geq \gamma \left(|y(0)|^2 + \int_0^T |v(t)|^2 dt \right) \tag{3.13}$$

for all admissible processes $(y(.),v(.))$ of (A). Then $(\bar{x}(.),\bar{u}(.))$ is a weak locally unique minimizer.

The coercivity condition (3.13) cannot be easily verified directly. There is, therefore, interest in having sufficiency criteria that are more susceptible to numerical evaluation. Such conditions are typically expressed in terms of the Riccati equation associated with (A) (omitting the argument (t))

$$\begin{cases} \dot{P} + PA + A^T P + Q - (B^T P + D^T)^T R^{-1} (B^T P + D^T) = 0 \\ P(.) = P^T(.) \end{cases} \tag{3.14}$$

and the Hamiltonian system of equations

$$(HAM)\begin{cases} \dot{y} = Ay - BR^{-1}(B^T q + D^T y) \\ -\dot{q} = A^T q + Qy - DR^{-1}(B^T q + D^T y). \end{cases}$$

Let $\{\Phi(s,t)\,|\,0 \le s,t \le T\}$ be the transition matrix associated with Hamilton's system of equations (HAM), i.e. for each $s \in [0,T]$, $\Phi(.,s)$ satisfies:

$$\begin{cases} \frac{d}{dt}\Phi(t,s) \;=\; \mathcal{A}\Phi(t,s) \quad \text{for all } t \in [0,T] \\ \Phi(s,s) = I \end{cases} \tag{3.15}$$

where

$$\mathcal{A} = \begin{pmatrix} A - BR^{-1}D^T & -BR^{-1}B^T \\ -Q + DR^{-1}D^T & -A^T + DR^{-1}B^T \end{pmatrix}.$$

Partition $\Phi(0,T)$ into $n \times n$ blocks

$$\Phi(0,T) \;=\; \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}.$$

Now define the $2n \times 2n$ matrix W to be

$$W \;:=\; G \;+\; \begin{pmatrix} \phi_{22}\phi_{12}^{-1} & \phi_{21} - \phi_{22}\phi_{12}^{-1}\phi_{11} \\ -\phi_{12}^{-1} & \phi_{12}^{-1}\phi_{11} \end{pmatrix}. \tag{3.16}$$

Conditions for coercivity of the second variation can be expressed in terms of solutions to these equations.

Notice that defining W requires the invertibility of $\phi_{12}(0,T)$ which, according to Lemma 3.2, is guaranteed under $(H_1^{'}) - (H_4^{'})$.

Lemma 3.2. *Consider a normal extremal $(\bar{x}(.), \bar{u}(.))$. Assume $(H_1^{'})$ - $(H_4^{'})$ are satisfied and there exists a (symmetric) solution $P(.)$ to (3.14) on $[0,T]$. Then,*

$$\det[\phi_{12}(0,T)] \neq 0.$$

Proof. [15, Section 24]. ■

Proposition 3.3. *Consider a weak normal extremal $(\bar{x}(.), \bar{u}(.))$. Assume that $(H_1^{'})$ - $(H_4^{'})$ are satisfied. Select a positive number γ. Then the following conditions are equivalent.*

(a) (i) *The Riccati equation* (3.14) *has a (symmetric) solution on* $[0,T]$,

(*ii*) $\begin{pmatrix} x_0^T & x_1^T \end{pmatrix} W \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} > \gamma |(x_0, x_1)|^2$

for all non-zero $x_0, x_1 \in \mathbb{R}^n$ *such that*

$$d_{x_0} x_0 + d_{x_T} x_1 = 0.$$

(*b*) $J_A(y(.), v(.)) \geq \gamma \left(|y(0)|^2 + \int_0^T |v(t)|^2 \, dt \right)$

for all admissible processes $(y(.), v(.))$ *for* (A).

Proof. [1, Section 7]. ■

A sufficient condition for (*b*), called the Riccati sufficient condition, is presented in [49,61] and is expressed in terms of an inequality involving the solution to (3.14). It takes the form

$$\begin{pmatrix} y_0^T & y_1^T \end{pmatrix} \left[G + \begin{pmatrix} P(0) & 0 \\ 0 & P(T) \end{pmatrix} \right] \begin{pmatrix} y_0 \\ y_1 \end{pmatrix} > 0 \tag{3.17}$$

for all $y_0, y_1 \in \mathbb{R}^n$ satisfying $d_{x_0} y_0 + d_{x_T} y_1 = 0$. It is in fact reported in [61] that this condition can be extended by considering three differential equations in place of the single Riccati equation (3.14). This fact is used in [1] to develop weaker sufficient conditions, for periodic optimal control problems. Omitting (t) for brevity, the three equations referred to are:

$$\begin{cases} \dot{P} + PA + A^T P + Q - (PB + D) R^{-1} (B^T P + D^T) = 0 \\ P(.) = P^T(.) \end{cases} \tag{3.18}$$

$$\dot{S} + (A^T - (PB + D) R^{-1} B^T) S = 0 \tag{3.19}$$

$$\dot{M} - S^T B R^{-1} B^T S = 0. \tag{3.20}$$

These equations can be obtained when the mixed boundary constraints of the accessory problem are separated by restating the original accessory problem in $2n$ variables. This is achieved by considering an extended system (y, z) in which z satisfies $\dot{z}(t) = 0$ and $y(0) = z(0)$. The original cost is posed in terms of the final values of the extended state while the new boundary conditions are expressed in terms of the initial states. The solution to the new Riccati equation at each $t \in [0, T]$, denoted $\tilde{P}(t)$, comprises the block matrices

$$\tilde{P}(t) = \begin{pmatrix} P(t) & S(t) \\ S^T(t) & M(t) \end{pmatrix}$$

in which $P(t), S(t), M(t) \in \mathbb{R}^{n \times n}$. The improved condition is obtained by replacing P, G and the endpoints in (3.17) with the corresponding ones of the extended system.

3.2.1 Uniqueness of Local Minimizers

A fact often not stressed enough is that the local minimizer satisfying by second order sufficient conditions such as those of the preceding section is locally unique. This fact follows from the coercivity condition and [1, Lemma 6.1] which states that given $\delta \in (0, \gamma)$, a small enough $\epsilon > 0$ can be found such that for any admissible process $(x(.), u(.))$ of (P) satisfying

$$\|x(.) - \bar{x}(.)\|_{L^\infty} \leq \epsilon \quad \text{and} \quad \|u(.) - \bar{u}(.)\|_{L^\infty} \leq \epsilon \tag{3.21}$$

there exists an admissible process $(y(.), v(.))$ for (A) such that

$$J(x(.), u(.)) - J(\bar{x}(.), \bar{u}(.)) \geq J_A(y(.), v(.)) - \delta \left(|y(0)|^2 + \|v\|_{L^2}^2 \right)$$

and

$$|x(0) - \bar{x}(0)|^2 + \|u - \bar{u}\|_{L^2}^2 \leq K[|y(0)|^2 + \|v\|_{L^2}^2],$$

where

$$\|v\|_{L^2}^2 := \int_0^T |v(t)|^2 \, dt$$

and $K > 0$. It follows that

$$J(x(.), u(.)) - J(\bar{x}(.), \bar{u}(.)) \quad \geq \quad \left(\frac{\gamma - \delta}{K} \right) \left(|x(0) - \bar{x}(0)|^2 + \|u - \bar{u}\|_{L^2}^2 \right) > 0$$

for all $(x(.), u(.))$ different from $(\bar{x}(.), \bar{u}(.))$, which implies local uniqueness.

3.3 Second Order Sufficient Conditions for Non-unique Local Minimizers

It is clear from the argument of the preceding section that the standard second order sufficient conditions are never satisfied unless minimizers are locally unique. This section develops refined second order sufficient conditions of optimality that are satisfied even by weak, local extremals that are not locally unique minimizers, in certain cases of interest. These are cases in which a weak, local extremal can be continuously embedded in an $\alpha-$*parameterised family of extremals*, as follows.

Definition. The weak normal extremal $((\bar{x}(.), \bar{u}(.)), (p(.), \nu))$ is said to be continuously embedded in an $\alpha-$parameterised family of weak extremals

$$\{((x^\alpha(.), u^\alpha(.)), (p^\alpha(.), \nu^\alpha)) \mid \alpha \in \mathcal{A}\}$$

if $\mathcal{A}$ is an open ball in some finite dimensional vector space $\mathbb{R}^k$, centre the origin, and the

following conditions are satisfied.

(F_1) For each $\alpha \in \mathcal{A}$ $(x^\alpha(.), u^\alpha(.)), (p^\alpha(.), \nu^\alpha)$ is a weak normal extremal such that

$$\begin{aligned} g(x^\alpha(0), x^\alpha(T)) &= g(\bar{x}(0), \bar{x}(T)) \text{ and} \\ d(x^\alpha(0), x^\alpha(T)) &= d(\bar{x}(0), \bar{x}(T)) = 0, \end{aligned}$$

(F_2) $\alpha \to (x^\alpha(.), u^\alpha(.), p^\alpha(.), \nu^\alpha) : \mathcal{A} \to L^\infty \times L^\infty \times L^\infty \times \mathbb{R}^l$ is strongly continuous,

(F_3) $\alpha \to (x^\alpha(0), x^\alpha(T)) : \mathcal{A} \to \mathbb{R}^n \times \mathbb{R}^n$ is of class C^1,

(F_4) the rows of the $p \times 2n$ matrix

$$\Gamma := \left(\nabla_\alpha (x^\alpha(0), x^\alpha(T))|_{\alpha=0}\right)^T \tag{3.22}$$

are linearly independent.

In fact, Chapter 2 provides a convenient platform for studying such problems provided that, firstly, problem (P) can be reformulated in a way that fits the abstract optimization framework and, secondly, the hypotheses of the abstract optimization framework are satisfied under the 'usual' hypotheses of the optimal control problem.

3.3.1 Reformulation of the Optimal Control Problem

To match the optimal control problem (P) to the optimization problem of Section 2.4, the following identifications are made. Define the Hilbert space

$$Z := \mathbb{R}^n \times L^2([0,T]; \mathbb{R}^m) \tag{3.23}$$

which is associated with pairs $(x_0, u(.))$ and the induced norm

$$\|(x_0, u)\|_Z^2 = |x(0)|^2 + \|u\|_{L^2}^2 .$$

Also define

$$\mathcal{D} = \{(x_0, u(.)) \in Z \mid |x_0| \leq 1 \text{ and } \|u\|_{L^\infty} \leq 1\}.$$

Furthermore, $x(t)$ is understood to be an abbreviated notation for $x(t; x_0, u(.))$ and solves the ordinary differential equation

$$\begin{cases} \dot{x}(t) = f(t, x(t), u(t)) \\ x(0) = x_0. \end{cases}$$

In addition, interpret $\Pi : \mathbb{R}^n \times L^2([0,T];\mathbb{R}^m) \to \mathbb{R}^n \times \mathbb{R}^n$ as

$$\Pi(x_0, u(.)) := (x_0, x(T)), \tag{3.24}$$

which is the evaluation of the state trajectory $x(.)$ (obtained from $(x_0, u(.))$) at its boundary points.

In this respect, (P) can equivalently be stated as

$$\text{minimize } g \circ \Pi(x_0, u(.))$$

over pairs $(x_0, u(.))$ subject to the constraint

$$d \circ \Pi(x_0, u(.)) = 0\,,$$

which resembles the optimization problem of the abstract framework. Here it is noted that an admissible process achieving the minimum cost over processes satisfying

$$|x_0 - \bar{x}_0| \leq \epsilon \quad \text{and} \quad \|u - \bar{u}\|_{L^\infty} \leq \epsilon$$

for some $\epsilon > 0$ (i.e. is a $\mathcal{D}-$local minimizer) is also minimizing according do the definition of weak local minimizers, i.e. over process $(x(.), u(.))$ such that

$$\|x - \bar{x}\|_{L^\infty} \leq \epsilon \quad \text{and} \quad \|u - \bar{u}\|_{L^\infty} \leq \epsilon.$$

This follows from the fact that for any $\epsilon > 0$

$$\{ \text{ admissible processes } (x(.), u(.)) \mid \|x(.) - \bar{x}(.)\|_{L^\infty} \text{ and } \|u(.) - \bar{u}(.)\|_{L^\infty} \leq \epsilon\} \subset$$
$$\{ \text{ admissible processes } (x(.), u(.)) \mid |x(0) - \bar{x}(0)| \text{ and } \|u - \bar{u}\|_{L^\infty} \leq \epsilon\}\,.$$

Finally, for a given pair (x_0, u) and $\alpha \in \mathcal{A}$ introduce the Lagrangian function

$$h^\alpha(x_0, u) := h(x_0, u; v^\alpha) = (g + (v^\alpha)^T d) \circ \Pi(x_0, u)$$

and recall (with a slight change of notation) the 'end-point' Lagrangian

$$\tilde{h}^\alpha(x_0, x_T) := \tilde{h}(x_0, x_T; v^\alpha) = g(x_0, x_T) + (v^\alpha)^T d(x_0, x_T).$$

3.3.2 Statement of the New Sufficient Conditions

In the sequel the following hypotheses are imposed on the optimal control problem (P).

Hypotheses

(H_1^C) $\bar{u}(.)$ is essentially bounded and

$$\|x^\alpha(.) - \bar{x}(.)\|_{L^\infty} \to 0, \|u^\alpha(.) - \bar{u}(.)\|_{L^\infty} \to 0$$

as $\alpha \to 0$.

(H_2^C) (Differentiability): $f(., x, u)$ is Lebesgue measurable for each (x, u) and $f(t, ., .)$ is of class C^2 for each $t \in [0, T]$; g and d are also of class C^2.

(H_3^C) (Uniform boundedness): Given any bounded subset $Q \subset \mathbb{R}^n \times \mathbb{R}^m$ there exists a function $\theta : \mathbb{R}^+ \to \mathbb{R}^+$ with the property that $\lim_{\epsilon \downarrow 0} \theta(\epsilon)/\epsilon = 0$ such that

$$|f(t, z') - f(t, z) - \nabla f(t, z)(z' - z) - \frac{1}{2}(z' - z)^T \nabla^2 f(t, z)(z' - z)| \leq \theta(|z' - z|^2)$$

for all $t \in [0, T]$ and $z', z \in Q$.

(H_4^C) There exists $\gamma > 0$ such that $H_{uu}(t, \bar{x}(t), p(t), \bar{u}(t)) \geq \gamma I$ for all $t \in [0, T]$.

(H_5^C) $A(.) \in L^\infty([0, T]; \mathbb{R}^{n \times n})$, $B(.) \in L^\infty([0, T]; \mathbb{R}^{n \times m})$.

(H_6^C) The linear system $(A(.), B(.))$ is controllable.

The new, refined second order sufficient conditions which are a direct implication of Theorem 2.7 are summarised in the following theorem.

Theorem 3.4. *Consider a weak normal extremal for $(\bar{x}(.), \bar{u}(.))$. Assume hypotheses (H_1^C) - (H_6^C) are satisfied and that $(\bar{x}(.), \bar{u}(.))$ can be continuously embedded in a family of extremals satisfying the conditions (F_1) - (F_4). Suppose that*

(i) the Riccati equation (3.18) *has a symmetric solution on* $[0, T]$,

(ii) there exists $\gamma > 0$ such that

$$\begin{pmatrix} y_0^T & y_1^T \end{pmatrix} W \begin{pmatrix} y_0 \\ y_1 \end{pmatrix} > \gamma |(y_0, y_1)|^2$$

for all non-zero $(y_0, y_1) \in \mathbb{R}^n \times \mathbb{R}^n$ *such that*

$$d_{x_0} y_0 + d_{x_T} y_1 = 0 \text{ and } \Gamma \begin{pmatrix} y_0 \\ y_1 \end{pmatrix} = 0.$$

Then $(\bar{x}(.), \bar{u}(.))$ *is a weak local minimizer.*

Remark. In the absence of non-uniqueness $\Gamma = 0$ and the orthogonality condition is automatically satisfied. The above sufficient conditions are then reduced to checking strict positivity of W solely over the subspace defined by the null space of the constraints. In this case, the standard sufficient conditions are recovered.

3.3.3 Proof of Theorem 3.4

Hypotheses $(H_1^C) - (H_6^C)$ ensure that the hypotheses $(H_1^A) - (H_3^A)$ of the abstract framework are satisfied. The detailed verification of the hypotheses, which follows, makes use of the following fact: for each $\alpha \in \mathcal{A}$ such that $(x^\alpha(.), u^\alpha(.))$ belongs to the $\alpha-$parameterised family of weak normal extremals, there exist $p^\alpha(.) \in W^{1,1}([0,T]; \mathbb{R}^n)$ and $\nu^\alpha \in \mathbb{R}^l$ such that

$$\begin{aligned}
-\dot{p}^\alpha(t) &= p^\alpha(t)^T f_x(t, x^\alpha(t), u^\alpha(t)) && (3.25)\\
p^\alpha(t)^T f_u(t, x^\alpha(t), u^\alpha(t)) &= 0 \quad \text{a.e. } t \in [0,T] && (3.26)\\
-p^\alpha(0) &= g_{x_0}(x^\alpha(0), x^\alpha(T)) + v^{\alpha T} d_{x_0}(x^\alpha(0), x^\alpha(T)) && (3.27)\\
p^\alpha(T) &= g_{x_T}(x^\alpha(0), x^\alpha(T)) + v^{\alpha T} d_{x_T}(x^\alpha(0), x^\alpha(T)) && (3.28)
\end{aligned}$$

Furthermore, write

$$A^\alpha(t) := f_x(t, x^\alpha(t), u^\alpha(t)) \quad \text{and} \quad B^\alpha(t) := f_u(t, x^\alpha(t), u^\alpha(t)).$$

The case $\alpha = 0$ is denoted by the absence of a superscript. Similarly, the superscript α indicates an underlining α-parameterised base point when taking derivatives. For example

$$g^\alpha_{x_0,x_T} \text{ means } \nabla_{(x_0,x_T)} g(x_0^\alpha, x_T^\alpha).$$

For the rest of this section assume that $(H_1^C) - (H_6^C)$ are satisfied.

Proposition 3.5. *There exists a continuity modulus* $\theta(.)$ *such that for any* $\epsilon' > 0$ *sufficiently small, all* $\alpha \in \mathcal{A}$ *and* $z \in z^\alpha + \epsilon' \mathcal{D}$,

(i) $|\Pi(z) - \Pi(z^\alpha) - D\Pi^\alpha(z^\alpha)(z - z^\alpha)| \leq \theta(\epsilon')\|z - z^\alpha\|$,

(ii) $|h(z) - h(z^\alpha) - Dh^\alpha(z^\alpha)(z - z^\alpha) - \frac{1}{2}\langle z - z^\alpha, D^2 h^\alpha(z^\alpha)(z - z^\alpha)\rangle| \leq \theta(\epsilon')\|z - z^\alpha\|^2$,

where Z , h, $\mathcal{D}$ and Π are as identified for the optimal control problem.

Proof. (i) Take $\bar{\alpha} > 0$, $\epsilon > 0$, the values of which will be determined in the sequel, $\alpha \in \bar{\alpha}\mathbb{B}^k$ and consider an extremal $(x^\alpha(0), u^\alpha(.)) = (x_0^\alpha, u^\alpha)$. Also recall that

$$\Pi(z) = \Pi\big(x_0, u\big) = \big(x_0, x(T)\big).$$

To compute the differential $D\Pi^\alpha(y_0^\alpha, v^\alpha)$ $(= D\Pi\left((x_0^\alpha, u^\alpha); (y_0^\alpha, v^\alpha)\right))$ consider

$$\lim_{\epsilon\to 0} \frac{1}{\epsilon}\Big(\Pi(x_0^\alpha + \epsilon y_0^\alpha, u^\alpha + \epsilon v^\alpha) - \Pi(x_0^\alpha, u^\alpha)\Big).$$

By continuity of $\nabla^2 f$ and Taylor's theorem [3]

$$\begin{aligned} D\Pi^\alpha(y_0^\alpha, v^\alpha) &= \lim_{\epsilon\to 0} \frac{1}{\epsilon}\Big(\epsilon y_0^\alpha, \epsilon y_0^\alpha + \int_0^T f(t, x^\alpha(t) + \epsilon y^\alpha(t), u^\alpha(t) + \epsilon v^\alpha(t)) - f(t, x^\alpha(t), u^\alpha(t))\, dt\Big) \\ &= \Big(y_0^\alpha, y_0^\alpha + \int_0^T A^\alpha(t) y^\alpha(t) + B^\alpha(t) v^\alpha(t)\, dt\Big) \\ &= \big(y_0^\alpha, y^\alpha(T)\big) \end{aligned}$$

where $y^\alpha(t)$ satisfies

$$\begin{cases} \dot{y}^\alpha(t) \;=\; A^\alpha(t) y^\alpha(t) + B^\alpha(t) v^\alpha(t) \\ y^\alpha(0) \;=\; y_0^\alpha \;=\; (x_0 - x_0^\alpha)\,. \end{cases}$$

But then,

$$\begin{aligned} |\Pi(x_0, u) - \Pi(x_0^\alpha, u^\alpha) - D\Pi^\alpha(x_0 - x_0^\alpha, u - u^\alpha)| &= |x(T) - x^\alpha(T) - y^\alpha(T)| \\ &\le \|\Delta x^\alpha - y^\alpha\|_{L^\infty}. \end{aligned} \tag{3.29}$$

By employing Filippov's existence theorem [64, Section 2.4] and in view of the uniform boundedness hypothesis (H_3^C) one can assert that given any $\Delta x^\alpha(.) \in W^{1,1}([0,T]; \mathbb{R}^n)$, there exists a solution $y'(.)$ to the linear system of equations with initial condition $y(0) = x_0 - x_0^\alpha$ such that for each $\alpha \in \bar{\alpha}\mathbb{B}^k$

$$\begin{aligned} &\|\Delta x^\alpha - y'\|_{L^\infty} \\ &\quad \le k'(\alpha) \int_0^T |f(t, x(t), u(t)) - f(t, x^\alpha(t), u^\alpha(t)) - A^\alpha(t)\Delta x^\alpha(t) - B^\alpha(t)\Delta u^\alpha(t)|\, dt \\ &\quad \le k(\alpha) \int_0^T |\Delta x^\alpha(t)|^2 + |\Delta u^\alpha(t)|^2\, dt \end{aligned}$$

for some finite bounds k', k that depend on α. However, in view of the continuous dependence of $\Delta x^\alpha(.)$ and $\Delta u^\alpha(.)$ on α, and the continuous dependence of solutions to differential

equations on α, one can instead find a uniform bound $k > 0$ which is independent of α. Finally, labelling the solution $y'(.)$ at each α as $y^\alpha(.)$ yields

$$\|\Delta x^\alpha - y^\alpha\|_{L^\infty} \leq 2kT\epsilon^2. \tag{3.30}$$

It is also known that the L^2 norm satisfies the parallelogram identity and hence,

$$\|X+Y\|^2 + \|X-Y\|^2 = 2\|X\|^2 + 2\|Y\|^2.$$

Writing v^α for Δu^α and taking $X = (y^\alpha, v^\alpha)$, $Y = (y^\alpha - \Delta x^\alpha, 0)$ gives

$$\begin{aligned}\int_0^T |\Delta x^\alpha(t)|^2 + |v^\alpha(t)|^2\,dt &\leq 2\int_0^T |y^\alpha(t)|^2 + |v^\alpha(t)|^2\,dt + 2\int_0^T |\Delta x^\alpha(t) - y^\alpha(t)|^2\,dt \\ &\leq 2\int_0^T |y^\alpha(t)|^2 + |v^\alpha(t)|^2\,dt + \\ &\quad 4k^2T^2\epsilon^2\int_0^T |\Delta x^\alpha(t)|^2 + |v^\alpha(t)|^2\,dt.\end{aligned}$$

It follows that

$$\int_0^T |\Delta x^\alpha(t)|^2 + |v^\alpha(t)|^2\,dt \leq 2(1-4k^2T^2\epsilon^2)^{-1}\left(\int_0^T |y^\alpha(t)|^2 + |v^\alpha(t)|^2\,dt\right). \tag{3.31}$$

It can also be deduced (see for example [1]) that for all $\alpha \in \bar{\alpha}\mathbb{B}^k$ and any process of the accessory problem a constant $c > 0$ can be found such that

$$\int_0^T |y^\alpha(t)|^2 dt \leq c\Big(|y_0^\alpha|^2 + \int_0^T |v^\alpha(t)|^2\,dt\Big). \tag{3.32}$$

Choosing ϵ to satisfy $4k^2T^2\epsilon^2 < 1$ and substituting (3.32) and (3.31) into (3.30) yields

$$\|\Delta x^\alpha - y^\alpha\|_{L^\infty} \leq \theta(\epsilon)\Big(|y_0^\alpha|^2 + \int_0^T |v^\alpha(t)|^2\,dt\Big)^{1/2}$$

for all $\alpha \in \bar{\alpha}\mathbb{B}^k$, $\epsilon' \in (0,\epsilon)$ and a function $\theta(.) : \mathbb{R}^+ \to \mathbb{R}^+$, which is a uniform bound, with the property $\lim_{s\downarrow 0}\theta(s) = 0$. Further substitution into the RHS of (3.29) confirms the required inequality.

(ii) Take $\bar{\alpha} > 0$, $\epsilon > 0$, the values of which will be adjusted presently, $\alpha \in \bar{\alpha}\mathbb{B}^k$ and consider an extremal $(x^\alpha(0), u^\alpha(.)) = (x_0^\alpha, u^\alpha)$. For any $(x_0, u) \in (x_0^\alpha, u^\alpha) + \epsilon\mathcal{D}$,

$$h^\alpha(x_0,u) - h^\alpha(x_0^\alpha, u^\alpha) = \tilde{h}^\alpha(x_0, x_T) - \tilde{h}^\alpha(x_0^\alpha, x_T^\alpha)\,. \tag{3.33}$$

Now add to (3.33) the zero term

$$\int_0^T p^\alpha(t)\cdot[(\dot{x}(t)-\dot{x}^\alpha(t))-(f(t,x(t),u(t))-f(t,x^\alpha(t),u^\alpha(t)))]\,dt$$

and integrate $\int_0^T p^\alpha(t)\cdot[(\dot{x}(t)-\dot{x}^\alpha(t))]\,dt$ by parts to get

$$\begin{aligned}
&h^\alpha(x_0,u)-h^\alpha(x_0^\alpha,u^\alpha)\\
&\quad=\tilde{h}^\alpha(x_0,x_T)-\tilde{h}^\alpha(x_0^\alpha,x_T^\alpha)-\\
&\qquad\qquad\int_0^T p^\alpha(t)\cdot[(\dot{x}(t)-\dot{x}^\alpha(t))-(f(t,x(t),u(t))-f(t,x^\alpha(t),u^\alpha(t)))]\,dt\\
&\quad=\tilde{h}^\alpha(x_0,x_T)-\tilde{h}^\alpha(x_0^\alpha,x_T^\alpha)-p^\alpha(T)\cdot\Delta x^\alpha(T)+p^\alpha(0)\cdot\Delta x_0^\alpha+\\
&\qquad\qquad\int_0^T \dot{p}^\alpha(t)\cdot\Delta x^\alpha(t)\,dt+\int_0^T p^\alpha(t)\cdot[f(t,x(t),u(t))-f(t,x^\alpha(t),u^\alpha(t))]\,dt.
\end{aligned}$$

Expanding the RHS of the equation to second order (in view of the uniform boundedness hypothesis (H_3^C)) and substituting the conditions of the Maximum Principle (since $(x^\alpha(.),u^\alpha(.))$ is an extremal process) gives

$$\begin{aligned}
&h^\alpha(x_0,u)-h^\alpha(x_0^\alpha,u^\alpha)\geq\\
&\quad\frac{1}{2}(\Delta x^\alpha(0),\Delta x^\alpha(T))^T\nabla^2\tilde{h}^\alpha(x_0^\alpha,x_T^\alpha)(\Delta x^\alpha(0),\Delta x^\alpha(T))+\\
&\quad\frac{1}{2}\int_0^T(\Delta x^\alpha(t),\Delta u^\alpha(t))^T\nabla^2H^\alpha(t)(\Delta x^\alpha(t),\Delta u^\alpha(t))\,dt-\\
&\quad\theta_1(\epsilon)\Big(|\Delta x^\alpha(0)|^2+|\Delta x^\alpha(T)|^2+\int_0^T|\Delta x^\alpha(t)|^2+|\Delta u^\alpha(t)|^2\,dt\Big)
\end{aligned}\tag{3.34}$$

for some function $\theta_1(.):\mathbb{R}^+\to\mathbb{R}^+$ with the property that $\theta_1(s)\downarrow 0$ as $s\downarrow 0$. Noting the identity,

$$x^TAx=y^TAy+(x-y)^TA(x+y)$$

for some square matrix A and vectors x and y of appropriate dimensions, (3.34) can be expressed as

$$\begin{aligned}
&h^\alpha(x_0,u)-h^\alpha(x_0^\alpha,u^\alpha)\geq\\
&\quad\frac{1}{2}(y^\alpha(0),y^\alpha(T))^T\nabla^2\tilde{h}^\alpha(x_0^\alpha,x_T^\alpha)(y^\alpha(0),y^\alpha(T))+\\
&\quad\frac{1}{2}\int_0^T(y^\alpha(t),v^\alpha(t))^T\nabla^2H^\alpha(t)(y^\alpha(t),v^\alpha(t))\,dt-\\
&\quad\theta_1(\epsilon)\Big(|\Delta x^\alpha(0)|^2+|\Delta x^\alpha(T)|^2+\int_0^T|\Delta x^\alpha(t)|^2+|\Delta u^\alpha(t)|^2\,dt\Big)-k_1e,
\end{aligned}\tag{3.35}$$

where

$$e = \|(\Delta x^\alpha, \Delta u^\alpha) - (y^\alpha, v^\alpha)\|_{L^\infty} \cdot \|(\Delta x^\alpha, \Delta u^\alpha) + (y^\alpha, v^\alpha)\|_{L^\infty}$$

and k_1 a positive constant. It can be deduced from (3.30)-(3.32) that by appropriately choosing the numbers ϵ and $\bar{\alpha}$ it can be arranged that for all $\alpha \in \bar{\alpha}\mathbb{B}^k$ there exists a function $\theta_2(.) : \mathbb{R}^+ \to \mathbb{R}^+$ such that

$$\|(\Delta x^\alpha, \Delta u^\alpha) - (y^\alpha, v^\alpha)\|_{L^\infty} = \|\Delta x^\alpha - y^\alpha\|_{L^\infty} \leq \theta_2(\epsilon)\Big(|y_0^\alpha|^2 + \int_0^T |v(t)|^2\, dt\Big) \tag{3.36}$$

and $\theta_2(\epsilon) \downarrow 0$ as $\epsilon \downarrow 0$. Furthermore, in view of (3.36), the triangle inequality and uniqueness of solutions to differential equations

$$\begin{aligned} \|(\Delta x^\alpha, \Delta u^\alpha) + (y^\alpha, v^\alpha)\|_{L^\infty} &\leq \|(\Delta x^\alpha, \Delta u^\alpha) - (y^\alpha, v^\alpha)\|_{L^\infty} + 2\|(\Delta x^\alpha, \Delta u^\alpha)\|_{L^\infty} \\ &= \|\Delta x^\alpha - y^\alpha\|_{L^\infty} + 2\|(\Delta x^\alpha, \Delta u^\alpha)\|_{L^\infty} \\ &\leq \theta_3(\epsilon) \end{aligned}$$

for some function $\theta_3(.) : \mathbb{R}^+ \to \mathbb{R}^+$ such that $\theta_3(s) \downarrow 0$ as $s \downarrow 0$. It can, therefore, be concluded that

$$e \leq \theta_3(\epsilon)\Big(|y_0^\alpha|^2 + \int_0^T |v^\alpha(t)|^2 dt\Big). \tag{3.37}$$

It is also found that (using the triangle inequality),

$$\begin{aligned} |\Delta x^\alpha(T)|^2 &\leq |y^\alpha(T)|^2 + |y^\alpha(T) + \Delta x^\alpha(T)| \cdot |y^\alpha(T) - \Delta x^\alpha(T)| \\ &\leq |y^\alpha(T)|^2 + (\|\Delta x^\alpha - y^\alpha\|_{L^\infty} + 2|\Delta x^\alpha(T)|) \cdot \|\Delta x^\alpha - y^\alpha\|_{L^\infty} \\ &\leq |y^\alpha(T)|^2 + \theta_4(\epsilon)\Big(|y_0^\alpha|^2 + \int_0^T |v^\alpha(t)|^2 dt\Big) \end{aligned}$$

for some new function $\theta_4(.)$ with the same properties as $\theta_3(.)$. Moreover, since $y(.)$ is the solution of the linearized system (A^α, B^α) with initial condition y_0^α and control v^α, and in view of (H_5^C) it follows that

$$|y^\alpha(T)|^2 \leq c_1\Big(|y_0^\alpha|^2 + \int_0^T |v^\alpha(t)|^2\, dt\Big)$$

for some constant $c_1 > 0$. Then,

$$|\Delta x^\alpha(T)|^2 \leq \theta_5(\epsilon)\Big(|y_0^\alpha|^2 + \int_0^T |v^\alpha(t)|^2\, dt\Big) \tag{3.38}$$

for some new function $\theta_5(.)$ with the same properties as before. Substituting (3.31), (3.37)

and (3.38) into (3.36) gives

$$\begin{aligned} h^\alpha(x_0,u) - h^\alpha(x_0^\alpha,u^\alpha) \geq \\ \frac{1}{2}(y^\alpha(0),y^\alpha(T))^T\nabla^2\tilde{h}^\alpha(x_0^\alpha,x_T^\alpha)(y^\alpha(0),y^\alpha(T)) + \\ \frac{1}{2}\int_0^T (y^\alpha(t),v^\alpha(t))^T\nabla^2 H^\alpha(t)(y^\alpha(t),v^\alpha(t))\,dt - \theta_6(\epsilon)\Big(|y_0^\alpha|^2 + \int_0^T |v^\alpha(t)|^2\,dt\Big) \end{aligned}$$

for all $\alpha \in \bar{\alpha}\mathbb{B}^k$ and $\epsilon' \in (0,\epsilon)$, where $\theta_6(.) : \mathbb{R}^+ \to \mathbb{R}^+$ and $\lim_{s\downarrow 0}\theta_6(s) = 0$. Identifying

$$Dh^\alpha(z - z^\alpha) = 0$$

and

$$\begin{aligned} \langle (z-z^\alpha), D^2h^\alpha(z-z^\alpha)\rangle &= \frac{1}{2}\Big((y^\alpha(0),y^\alpha(T))^T\nabla^2\tilde{h}^\alpha(x_0^\alpha,x_T^\alpha)(y^\alpha(0),y^\alpha(T)) + \\ &\quad \int_0^T (y^\alpha(t),v^\alpha(t))^T\nabla^2 H^\alpha(t)(y^\alpha(t),v^\alpha(t))\,dt\Big) \end{aligned}$$

completes the proof. ■

To verify the controllability hypothesis of the abstract framework it suffices to verify the auxiliary controllability condition $(\tilde{H}_3^A)$ of Proposition 2.8.

Proposition 3.6. *The conditions:*

(*i*) $(D\Pi^\alpha)(D\Pi^\alpha)^\dagger = I_{r\times r}$, *for each* $\alpha \in \mathcal{A}$,

(*ii*) *there exists* $k > 0$ *such that*

$$\|(D\Pi^\alpha)^\dagger\| \leq k \quad \textit{for all } \alpha \in \mathcal{A},$$

are satisfied for Z *and* Π *as identified for the optimal control problem.*

Proof. Take $\alpha \in \mathcal{A}$ and consider the linear map $L^\alpha \in \mathcal{L}(L^2,\mathbb{R}^n)$ such that for any $v \in L^2([0,T];\mathbb{R}^m)$

$$L^\alpha v := \int_0^T \Phi^\alpha(T,t)^T B^\alpha(t)^T v(t)\,dt,$$

where $\Phi^\alpha(t_2,t_1)$ is the state transition matrix associated with the linear system $(A^\alpha(.),B^\alpha(.))$ on some interval $[t_1,t_2]$. Define the controllability Grammian on $[0,T]$

$$\mathcal{W}^\alpha(T,0) := \int_0^T \Phi^\alpha(T,t)B^\alpha(t)B^\alpha(t)^T\Phi^\alpha(T,t)^T\,dt.$$

In the optimal control case, the linear map $D\Pi^\alpha$ is given by

$$D\Pi^\alpha := \begin{pmatrix} I_{n\times n} & 0_{n\times n} \\ \Phi(T,0) & L^\alpha \end{pmatrix}$$

By (H_6^C), the linear system $(A(.), B(.))$ is controllable and $\mathcal{W}^0(T,0) > 0$. By (H_1^C) the mapping $\alpha \to \mathcal{W}^\alpha(T,0)$ is continuous and by reducing the size of ϵ, if necessary, (and hence the $\mathcal{A}$ neighbourhood), it can be arranged that $\mathcal{W}^\alpha(T,0) > 0$ for any $\alpha \in \mathcal{A}$. For any $(y_0, y_1) \in \mathbb{R}^n \times \mathbb{R}^n$, there exists $\eta \in \mathbb{R}^n$ such that

$$y_1 - \Phi^\alpha(T,0)y_0 = \mathcal{W}^\alpha(T,0)\eta.$$

Moreover, a control v^α that achieves the required transfer is

$$v(.) = B^\alpha(.)^T\Phi^\alpha(T,.)^T\mathcal{W}^\alpha(T,0)^{-1}\begin{pmatrix} -\Phi^\alpha(T,0) & I_{n\times n}\end{pmatrix}\begin{pmatrix} y_0 \\ y_1\end{pmatrix}.$$

Let the right inverse $(D\Pi^\alpha)^\dagger$ be given by

$$(D\Pi^\alpha)^\dagger := \begin{pmatrix} I_{n\times n} & 0_{n\times n} \\ -B^\alpha(.)^T\Phi^\alpha(T,.)^T\mathcal{W}^\alpha(T,0)^{-1}\Phi^\alpha(T,0) & B^\alpha(.)^T\Phi^\alpha(T,.)^T\mathcal{W}^\alpha(T,0)^{-1}\end{pmatrix}$$

Then, $\big(D\Pi^\alpha\big)\big(D\Pi^\alpha\big)^\dagger = I_{2n\times 2n}$ as required. Boundedness of $\|(D\Pi^\alpha)^\dagger\|$ follows automatically from the fact that $B^\alpha(.)$ and $\Phi^\alpha(T,.)$ are essentially bounded. ■

Finally, what remains to be confirmed is the existence of the matrix valued function $W^{\alpha,\gamma}$ as proposed by condition $(C)(i)$ of Theorem 2.7. For this purpose take $\bar\alpha, \bar\gamma > 0$ and consider for each $\alpha \in \bar\alpha\mathbb{B}^k$ and $\gamma \in [0, \bar\gamma]$, the accessory problem:

$$(A_p)\begin{cases} \text{minimize } J_A^{\alpha,\gamma}(y(.), v(.)) \\ \text{over } y(.) \in W^{1,1}([0,T];\mathbb{R}^n) \text{ and } v(.) \in L^2([0,T],\mathbb{R}^m) \\ \text{subject to} \\ \dot y(t) = A^\alpha(t)y(t) + B^\alpha(t)v(t) \text{ a.e. } t \in [0,T] \\ y(0) = y_0,\ y(T) = y_1 \end{cases}$$

where $y_0, y_1 \in \mathbb{R}^n$ and the cost function is defined by

$$\begin{aligned} 2J_A^{\alpha,\gamma}(y(.), v(.)) \ &:= \ \begin{pmatrix} y^T(0) & y^T(T)\end{pmatrix} G^{\alpha,\gamma}\begin{pmatrix} y(0) \\ y(T)\end{pmatrix} + \\ &\int_0^T y(t)^TQ^\alpha(t)y(t) + 2y(t)^TD^\alpha(t)v(t) + v(t)^TR^{\alpha,\gamma}(t)v(t)\,dt, \end{aligned}$$

where

$$G^{\alpha,\gamma} \;:=\; \nabla^2(g+\nu^\alpha d)(x_0^\alpha, x^\alpha(T)) - \begin{pmatrix} \gamma I_{n\times n} & 0_{n\times n} \\ 0_{n\times n} & 0_{n\times n} \end{pmatrix}.$$

The $\alpha-$parameterised, time-dependent matrices Q, D and R present in the accessory cost are, in turn, defined as follows:

$$\begin{pmatrix} Q^\alpha(t) & D^\alpha(t) \\ D^\alpha(t)^T & R^{\alpha,\gamma}(t) \end{pmatrix} := \begin{pmatrix} H^\alpha_{xx}(t) & H^\alpha_{xu}(t) \\ H^\alpha_{xu}(t)^T & H^\alpha_{uu}(t) - \gamma I \end{pmatrix}.$$

$G^{\alpha,\gamma}$ is continuous in (α,γ), and $R^{\alpha,\gamma}(t)$ is continuous in (α,γ) uniformly w.r.t. $t\in[0,T]$. By reducing the size of $\bar\gamma$ if required, it can be arranged that, for some $K>0$, which does not depend on the choice of (α,γ) or $t\in[0,T]$,

$$(R^{\alpha,\gamma})^{-1}(t) \;\le\; K\;.$$

Consider now the Riccati equation associated with (A_p)

$$\begin{aligned} &\dot P(t) + P(t)A^\alpha(t) + A^\alpha(t)^T P(t) + Q^\alpha(t) - \\ &\quad \left(B^\alpha(t)^T P(t) + D^\alpha(t)^T\right)^T (R^{\alpha,\gamma})^{-1}(t)\left(B^\alpha(t)^T P(t) + D^\alpha(t)^T\right) = 0 \end{aligned} \tag{3.39}$$

and the Hamiltonian involving $L_A^{\alpha,\gamma}(t,y(t),v(t))$, which is the integrand in the accessory cost formulation. The Hamiltonian system associated with the accessory problem is the following:

$$\begin{cases} \dot y(t) = A^\alpha(t)y(t) + B^\alpha(t)v_{min}(y(t),q(t)) \\ -\dot q(t) = (A^\alpha)^T(t)q(t) + \dfrac{\partial}{\partial y}L_A^{\alpha,\gamma}(t,y(t),v_{min}(y(t),q(t))), \end{cases}$$

where

$$v_{min}(y(t),q(t)) = \arg\min_v \Big\{ q(t)\cdot[A^\alpha(t)y(t) + B^\alpha(t)v(t)] + L_A^{\alpha,\gamma}(t,y(t),v(t)) \Big\},$$

and is given by $v_{min} = -(R^{\alpha,\gamma})^{-1}\left((D^\alpha)^T y + (B^\alpha)^T q\right)$. Substituting for v_{min} and $L_A^{\alpha,\gamma}$ in the Hamiltonian system gives

$$(HAM_A)\begin{cases} \dot y = A^\alpha y - B^\alpha(R^{\alpha,\gamma})^{-1}\left((D^\alpha)^T y + (B^\alpha)^T q\right) \\ -\dot q = (A^\alpha)^T q + Q^\alpha y - D^\alpha(R^{\alpha,\gamma})^{-1}\left((D^\alpha)^T y + (B^\alpha)^T q\right). \end{cases}$$

The $2n\times 2n$ transition matrix associated with (HAM_A) is the block matrix

$$\Phi^{\alpha,\gamma}(t,s) = \begin{pmatrix} \phi_{11}^{\alpha,\gamma}(t,s) & \phi_{12}^{\alpha,\gamma}(t,s) \\ \phi_{21}^{\alpha,\gamma}(t,s) & \phi_{22}^{\alpha,\gamma}(t,s) \end{pmatrix} \tag{3.40}$$

with $0 \leq t, s \leq T$ and $\phi_{ij}(t,s)$ $n \times n$ matrices.

It is known, by assumption, that (3.39) has a (symmetric) solution $P(.)$ on $[0,T]$ and that $\phi_{12}^{\alpha,\gamma}(0,T)$ is invertible, in the case $(\alpha,\gamma) = (0,0)$. It is known, furthermore, that the Riccati equation continues to have a solution on $[0,T]$ if the right boundary value is replaced by $P(T) + kI_{n\times n}$, for any $k \geq 0$ by continuous dependence. By choosing k sufficiently large, it can be arranged that both

(i) there exists a solution to $P^{\alpha,\gamma}(t)$ to (3.39) on $[0,T]$ and

(ii) $\phi_{12}^{\alpha,\gamma}(0,T)$ and $\phi_{11}^{\alpha,\gamma}(0,T) + \phi_{12}^{\alpha,\gamma}(0,T)P^{\alpha,\gamma}(T)$ are invertible

for $(\alpha,\gamma) = (0,0)$. But then, by 'continuous dependence', (i) and (ii) remain true for all choices of α and γ, provided $\bar{\alpha}$ and $\bar{\gamma}$ are sufficiently small. Now define,

$$W^{\alpha,\gamma} := G^{\alpha,\gamma} + \begin{pmatrix} \phi_{22}^{\alpha,\gamma}(\phi_{12}^{\alpha,\gamma})^{-1} & \phi_{21}^{\alpha,\gamma} - \phi_{22}^{\alpha,\gamma}(\phi_{12}^{\alpha,\gamma})^{-1}\phi_{11}^{\alpha,\gamma} \\ -(\phi_{12}^{\alpha,\gamma})^{-1} & (\phi_{12}^{\alpha,\gamma})^{-1}\phi_{11}^{\alpha,\gamma} \end{pmatrix}.$$

Lemma 3.7. *For $\bar{\alpha} > 0$ and $\bar{\gamma} > 0$ sufficiently small*

$$\inf\left\{J^{\alpha,\gamma}(y(0),v) \,|\, y(0) = y_0, y(T) = y_1\right\} = \begin{pmatrix} y_0^T & y_1^T \end{pmatrix} W^{\alpha,\gamma} \begin{pmatrix} y_0 \\ y_1 \end{pmatrix}$$

for all $\alpha \in \bar{\alpha}\mathbb{B}^k$ and γ s.t. $|\gamma| \leq \bar{\gamma}$.

Proof. The proof is the same for all (α,γ) sufficiently small in magnitude. So, assuming that $(\alpha,\gamma) = (0,0)$ the superscripts α and (α,γ) can be omitted throughout. Take any $y_0, y_1 \in \mathbb{R}^n$. Let $P(.)$ be the given solution to the Riccati equation (3.14) on $[0,T]$ and let $\Phi(.,.)$ be the transition matrix for the Hamiltonian system of equations (HAM_A). To simplify notation, the time argument t shall be dropped. Similarly, write Φ for $\Phi(0,T)$. Let $s(.)$ and $m(.)$ be solutions on $[0,T]$ of

$$\begin{cases} \dot{s}(t) + (A^T - \left(B^T P + D^T\right)^T R^{-1} B^T)\, s(t) = 0 \\ s(T) = \xi \end{cases}$$

and

$$\begin{cases} \dot{m}(t) - s^T(t) B R^{-1} B^T s(t) = 0 \\ m(T) = 0 \end{cases}$$

where

$$\xi = \phi_{12}^{-1}(y_0 - [\phi_{11} + \phi_{12}P(T)]y_1). \tag{3.41}$$

Define

$$V(t,y) = \frac{1}{2}y^T P(t)y + s^T(t)y + m(t).$$

It can be shown by direct calculation that

$$V_t + \min_v\{V_y \cdot (Ay + Bv) + \frac{1}{2}y^T Qy + y^T Dv + \frac{1}{2}v^T Rv\} = 0 \tag{3.42}$$

and

$$\begin{aligned} X(t,y) &:= \arg\min_v\{V_y \cdot (Ay + Bv) + \frac{1}{2}y^T Qy + y^T Dv + \frac{1}{2}v^T Rv\} \\ &= -R^{-1}((PB + D)^T y + B^T s) \end{aligned}$$

for all $t \in [0,T]$ and $y \in \mathbb{R}^n$.

Let $(\bar{y}, \bar{v})$ be the process on $[0,T]$ satisfying

$$\begin{cases} \dot{\bar{y}} = A\bar{y} + B\bar{v} & (3.43) \\ \bar{v} = X(t,y) & (3.44) \\ \bar{y}(0) = y_0. & (3.45) \end{cases}$$

It follows, by integrating across (3.42), that for any (y,v) satisfying

$$\dot{y} = Ay + Bv,$$

$y(0) = y_0$ and $y(T) = \bar{y}(T)$,

$$\begin{aligned} \frac{1}{2}\int_0^T y^T Qy + 2y^T Dv + v^T Rv\, dt &\geq \frac{1}{2}\int_0^T \bar{y}^T Q\bar{y} + 2\bar{y}^T D\bar{v} + \bar{v}^T R\bar{v}\, dt \\ &= V(0,y_0) - V(T,\bar{y}(T)). \end{aligned}$$

The assertions of the lemma will have been confirmed, then, if it can be shown that

(a) $\bar{y}(T) = y_1$,

(b) $V(0,y_0) - V(T,y_1) = \begin{pmatrix} y_0^T & y_1^T \end{pmatrix} \begin{pmatrix} \phi_{22}\phi_{12}^{-1} & \phi_{21} - \phi_{22}\phi_{12}^{-1}\phi_{11} \\ -\phi_{12}^{-1} & \phi_{12}^{-1}\phi_{11} \end{pmatrix} \begin{pmatrix} y_0 \\ y_1 \end{pmatrix}$.

To show (a) observe that $(\bar{y}, p)$ solves Hamilton's system of equations with

$$p(t) = P(t)\bar{y}(t) + s(t). \tag{3.46}$$

It follows that

$$\begin{aligned} y_0 &= \phi_{11}\bar{y}(T) + \phi_{12}p(T) \\ &= \phi_{11}\bar{y}(T) + \phi_{12}[P(T)\bar{y}(T) + \xi]. \end{aligned}$$

It can be deduced from (3.41) and the assumed invertibility of $(\phi_{11} + \phi_{12}P(T))$ that $y(T) = y_1$. (a) is, thus, proved.

Attention now focuses on the verification of (b). From (3.46)

$$V(0, y_0) - V(T, y_1) = \frac{1}{2}(p^T(0)y_0 - p^T(T)y_1) + \frac{1}{2}(s(0)y_0 - s(T)y_1) + (m(0) - m(T)).$$

But,

$$\begin{aligned} \frac{d}{dt}(s^T(t)y(t) + m(t)) &= \dot{s}^T(t)y(t) + s^T(t)\dot{y}(t) + \dot{m}(t) \\ &= \left[A^T - (B^TP + D^T)^TR^{-1}B^Ts(t)\right]^Ty(t) + \\ &\quad s^T(t)\left[A - BR^{-1}(PB + D)^Ty(t) + B^Ts(t)\right] + s^T(t)BR^{-1}B^Ts(t) \\ &= 0. \end{aligned}$$

It follows that

$$V(0, y_0) - V(T, y_1) = \frac{1}{2}(p(T)^Ty_1 - p(0)^Ty_0). \tag{3.47}$$

Furthermore, it is known that

$$\begin{pmatrix} y(0) \\ p(0) \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} y(T) \\ p(T) \end{pmatrix}.$$

Rearranging to obtain expressions for $p(0), p(T)$ in terms of $y(0), y(T)$ and substituting into (3.47) yields

$$V(0, y_0) - V(T, y_1) = \begin{pmatrix} y_0^T & y_1^T \end{pmatrix} \begin{pmatrix} \phi_{22}\phi_{12}^{-1} & \phi_{21} - \phi_{22}\phi_{12}^{-1}\phi_{11} \\ -\phi_{12}^{-1} & \phi_{12}^{-1}\phi_{11} \end{pmatrix} \begin{pmatrix} y_0 \\ y_1 \end{pmatrix}$$

as required to confirm (b). The proof of Theorem 3.4 is thus complete. ■

The following sections will examine how Theorem 3.4 can be used to generate numerical tests for local optimality, in specific problems. Two applications will be examined: the first concerns periodic optimal control, where initial and final states are required to match while the second combines both periodic and fixed endpoint constraints.

3.4 Application I: Periodic Optimal Control

Problems in periodic optimal control have a long history and originate from the work of Horn and Lin [37] in the field of chemical processing during the 1960s. The question arising at the time was "Can periodic operation achieve lower average cost than steady state?". Their work showed that steady state operation, commonly applied in the chemical industry, could be improved by cyclic operation. Using variational methods they derived a second order sufficient condition for the existence of a periodic control achieving improved performance over steady state operation. This condition was subsequently strengthened in [6].

Further sufficient conditions for the existence of periodic solutions are also reported in the work of Bittanti, Fronza and Guardabassi [26] in terms of a frequency criterion known as the $\pi-$test. Improvements on this test are reported by Bernstein and Gilbert in [9] who have also considered a somewhat more general problem with regards to the cost and constraint formulation. Necessary conditions, for this matter, are rather scarce and appear in [25] in terms a non-concavity requirement of the Hamiltonian in (x, u).

More interesting, from the point of view of this thesis, are sufficient conditions for optimality of periodic processes such as those presented in [44]. Specifically, they require the solution to the Hamilton-Jacobi equation with modified boundary conditions to suit the periodic constraints of the problem. Periodic optimal control problems with fixed period T have also been treated in [11, 12] in which the authors have developed second order sufficient conditions for optimality requiring the existence of a periodic solution to the Riccati equation (3.14) as well as the nonexistence of unity eigenvalues in the state transition matrix associated with the linearised dynamics.

More recent results concerning the periodic optimal control problem have been obtained by Speyer and his collaborators in [12] and [17]. In these papers they have actually demonstrated that the requirement concerning the non-existence of unity eigenvalues in the transition matrix is contradicted by the fact that the matrix in question has in fact two unit eigenvalues. Refinements to the results of [12] are then proposed in [60, 67, 68].

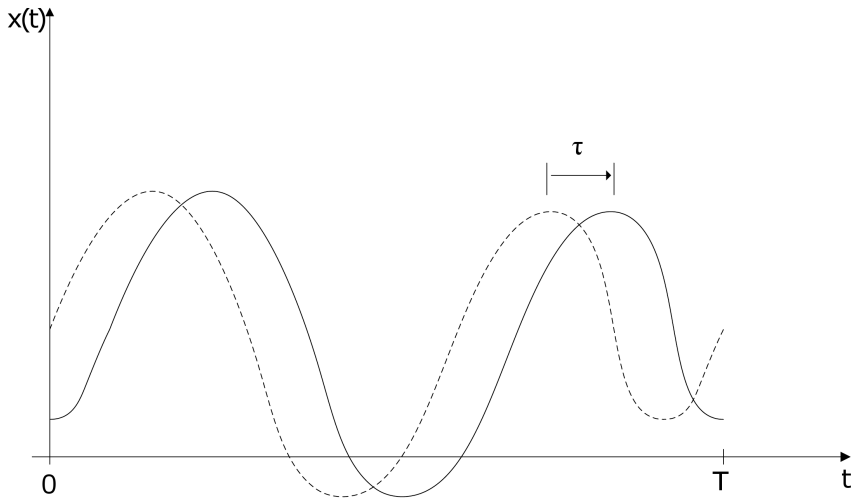


Figure 3.1: Time invariance of periodic solution.

3.4.1 Problem Statement

The periodic optimal control problem studied here takes the form

$$(P_1)\begin{cases} \text{minimize } \displaystyle\int_0^T L(x(t),u(t))\,dt \\ \text{subject to} \\ \dot{x}(t) = f(x(t),u(t)) \text{ a.e. } t \in [0,T] \\ x(0) = x(T) \end{cases}$$

where $T > 0$ and is fixed, while $f, L, g \in C^2$.

Take $\big(\bar{x}(.), \bar{u}(.)\big)$ to be an admissible process for (P_1) obtained either numerically or by analytical solution of optimality conditions. Suppose that the control $\bar{u}(.)$ is continuous and non-constant. A new process can be obtained by introducing a time delay τ in both the control and state functions and 'wrapping around' as in Figure 3.1. Time translation yields a new process $(x_\tau(.), u_\tau(.))$ for which $x_\tau(0) = \bar{x}(\tau)$ and $u_\tau(.) = \bar{u}_{ext}(. + \tau)$ on the interval $[0, T]$. Here $\bar{u}_{ext}(.)$ denotes the periodic extension of $\bar{u}(.)$ over $(-\infty, \infty)$. The integral cost, however, involving $(x_\tau(.), u_\tau(.))$ is identical to that of $(\bar{x}(.), \bar{u}(.))$ because the integral of a periodic function, over its period, is translationally invariant. Since the control function has been assumed to be continuous, any new control obtained by time translation can be arranged to be arbitrarily close to the original one by adjusting

the magnitude of the delay τ. Furthermore, as the control function is assumed to be non-constant, any time translation (however small) produces a *distinct* control function. But the classical sufficient conditions, if satisfied, provide the information that the process under consideration is locally unique. So, in fact, such conditions can never be satisfied for such problems.

Hypotheses

(H_1^P) The control $\bar{u}(.)$ is continuous and $\bar{u}(0) = \bar{u}(T)$,

(H_2^P) f, L are of class C^2 and their derivatives with respect to the (x, u) variables, up to order 2, are continuous,

(H_3^P) there exists $\epsilon > 0$ such that

$$R(t) > \epsilon I \quad \text{for all } t \in [0, T],$$

(H_4^P) $(A(.), B(.))$ is controllable on $[0, T]$.

In view of (H_1^P) and the translational invariance of the cost, it can be easily deduced that the conditions $(F_1) - (F_4)$ are automatically satisfied. Moreover,

$$\frac{d}{d\alpha}\begin{pmatrix} x^\alpha(0) \\ x^\alpha(T) \end{pmatrix}_{\alpha=0} = \lim_{\alpha\to 0}\begin{pmatrix} \frac{\bar{x}(\alpha)-\bar{x}(0)}{\alpha} \\ \frac{\bar{x}(T+\alpha)-\bar{x}(T)}{\alpha} \end{pmatrix} = \frac{d}{dt}\begin{pmatrix} \bar{x}(0) \\ \bar{x}(T) \end{pmatrix}$$

and hence,

$$\begin{pmatrix} \Gamma \\ d_{x_0}(\bar{x}(0), \bar{x}(T)) \quad d_{x_1}(\bar{x}(0), \bar{x}(T)) \end{pmatrix} = \begin{pmatrix} \dot{\bar{x}}^T(0) & \dot{\bar{x}}^T(T) \\ I_{n\times n} & -I_{n\times n} \end{pmatrix}.$$

The set $\mathcal{A}$ is the open interval $(-T,\, T)$. In view of the periodicity of the constraints define the matrix

$$W := \phi_{21}(0, T) + [I_{n\times n} - \phi_{22}(0, T)]\, \phi_{12}^{-1}(0, T)\, [\phi_{11}(0, T) - I_{n\times n}]\,.$$

Furthermore, $\dot{\bar{x}}(0) = \dot{\bar{x}}(T)$. The refined sufficient conditions for optimality then simplify as follows.

Theorem 3.8. *Consider a weak normal extremal for $(\bar{x}(.), \bar{u}(.))$. Assume hypotheses (H_1^P) - (H_4^P) are satisfied. Also suppose that*

(i) the Riccati equation (3.14) *has a (symmetric) solution on* $[0, T]$,

(*ii*) *there exists* $\gamma > 0$ *such that*

$$\xi^T W \xi > \gamma|\xi|^2$$

for all non-zero $\xi \in \mathbb{R}^n$ *such that*

$$\xi^T \dot{\bar{x}}(0) = 0.$$

Then $(\bar{x}(.), \bar{u}(.))$ *is a weak local minimizer.*

3.4.2 Speyer's Sailboat Problem

A simple example which illustrate the applicability of the new sufficient conditions is the so-called Speyer sailboat problem, presented by Speyer and his collaborators in [59] and also used in [1]. The integral cost in the formulation contains negative quadratic terms which provides the nonconvexity required to induce periodic minimizing paths. For this problem, the π test of [26], [9] has been employed in [60] to show that the coefficient of the control term determines whether the solution will be static or periodic (reducing the value of the coefficient yields periodic solutions). The interpretation given to this problem by Speyer is that of a sailboat trying to maximize its average speed into the wind.

$$(S)\begin{cases} \text{minimize } \displaystyle\int_0^1 x_1^2(t) - x_2^2(t) + x_2^4(t) + 0.01u^2(t)\,dt \\ \text{over } x \in W^{1,1}([0,1];\mathbb{R}^2) \text{ and continuous functions } u : [0,1] \to \mathbb{R} \\ \text{subject to} \\ \dfrac{d}{dt}\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = \begin{pmatrix} x_2(t) \\ u(t) \end{pmatrix} \text{ for all } t \in [0,1] \\ x(0) = x(1), \end{cases}$$

where x_1 and x_2 represent the lateral displacement and velocity of the boat, respectively, while the term in the Lagrangian $-x_2^2(t) + x_2^4(t)$ represents the velocity component in the direction of the wind. The accompanying accessory problem is

$$(S_A)\begin{cases} \text{minimize } \dfrac{1}{2}\displaystyle\int_0^1 \begin{pmatrix} y_1(t) & y_2(t) \end{pmatrix} Q(t) \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} + v(t)Rv(t)\,dt \\ \text{over } y \in W^{1,1}([0,1];\mathbb{R}^2) \text{ and } v \in L^2([0,1];\mathbb{R}) \\ \text{subject to} \\ \dot{y}(t) = Ay(t) + bv(t) \text{ for all } t \in [0,1] \\ y(0) = y(1) \end{cases}$$

in which, the matrices relating to the Riccati equation (3.18) are given by

$$Q(t) = \begin{pmatrix} 1 & 0 \\ 0 & 6\bar{x_2}^2(t) - 1 \end{pmatrix}, \; A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \; b = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \; R = 0.01, \; D = 0.$$

A solution to problem (S) has been obtained by using the numerical solver *IPOPT* [66]. Plots of the resulting state trajectories and control function are shown in Figures 3.2 and 3.3 respectively. Initially, it needs to be established that there exists a bounded

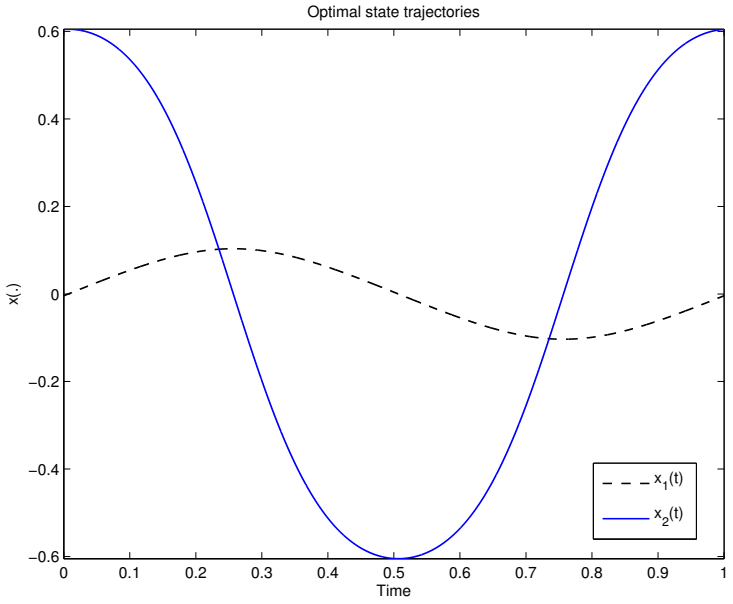


Figure 3.2: State plot

solution to the Riccati equation (3.14) on the interval $[0, 1]$. Indeed this is the case, as illustrated by Figure 3.4. The solution $P(.)$ has been computed using a terminal condition of $P(1) = 0.5I_{2\times 2}$ and integrating backwards in time. It should be noted that the solutions $P_{12}(.)$ and $P_{21}(.)$ obviously overlap due to the symmetric nature of $P(.)$. Furthermore, since the minimizing control for the accessory problem takes the form $\bar{v} = -R^{-1}(B^T P y + B^T s)$ (see Lemma 3.7 and the reference to the solutions $P(.), S(.), M(.)$ of the 3 matrix equations (3.18)-(3.18)) it is not a requirement that the boundary points $P(0)$ and $P(1)$ match in order to preserve continuity of the control.

To verify the conditions of Theorem 3.8 the transition matrix $\Phi(0, 1)$ has to be computed. To this end, consider the Hamiltonian system linked to the accessory problem,

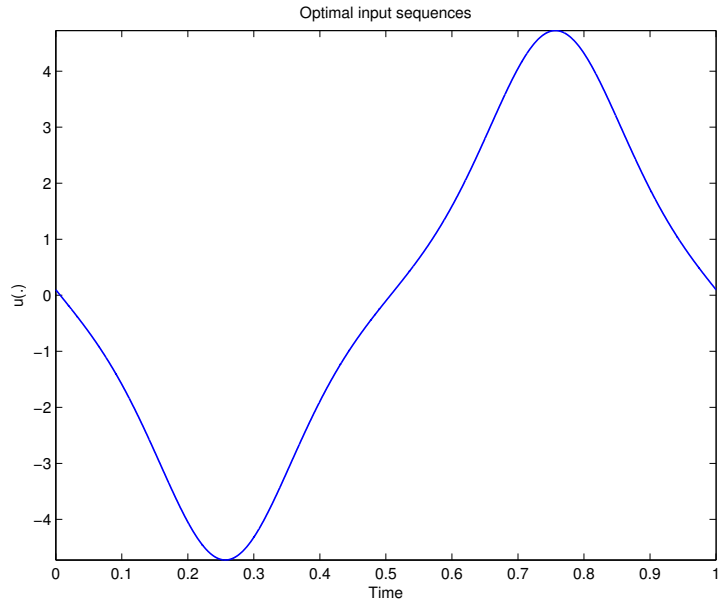


Figure 3.3: Control plot

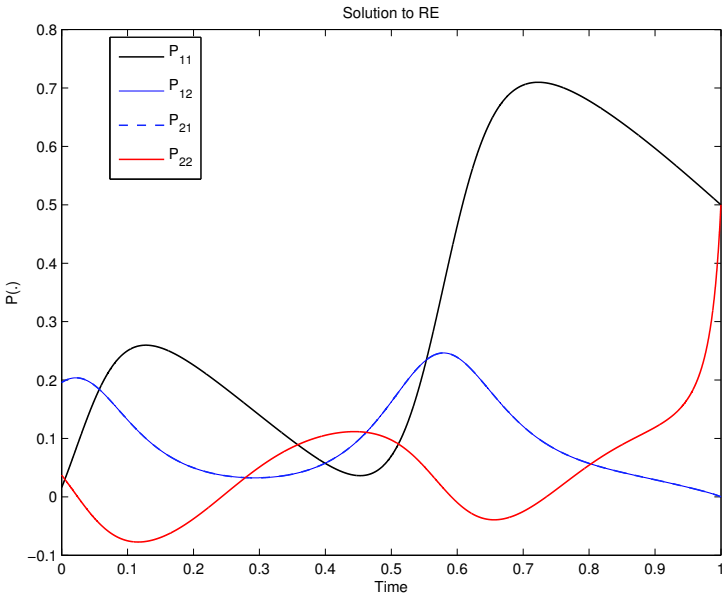


Figure 3.4: Solution to Riccati equation

$$\frac{d}{dt}\begin{pmatrix} y_1(t) \\ y_2(t) \\ q_1(t) \\ q_2(t) \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -50 \\ -2 & 0 & 0 & 0 \\ 0 & (2-12\bar{x}_2^2(t)) & -1 & 0 \end{pmatrix} \begin{pmatrix} y_1(t) \\ y_2(t) \\ q_1(t) \\ q_2(t) \end{pmatrix}, \tag{3.48}$$

for which $\bar{v}(t) = -50q_2(t)$. It is known from the properties of transition matrices that Φ itself satisfies the Hamiltonian system and that $\Phi(1,1) = I_{4\times4}$. $\Phi(0,1)$ can then be obtained by inverting the signs of the elements in the matrix of (3.48) starting from $\Phi(1,1) = I_{4\times4}$ and integrating backwards in time. One way to do that, see for example [69] and [52], is to independently obtain solutions with each column of the identity matrix as an initial condition. The solution, in this case, will be the corresponding column of the transition matrix. For the computations the following second order Runge-Kutta algorithm has been used:

$$\begin{aligned} z_{k+1} &= z_k + \frac{\Delta t}{2}(k_1 + k_2) \\ \text{where} & \\ k_1 &= A_k z_k \\ k_2 &= A_{k+1}(z_k + \Delta t A_k z_k) \end{aligned}$$

where k is the step number, Δt the step size, z the column vector solution and $-A$ the matrix in (3.48). The solution was found to be

$$\Phi(0,1) = \begin{pmatrix} -3.0581 & 3.2828 & -6.4777 & 7.4065 \\ -0.8315 & 3.7613 & -9.7855 & 1.0110 \\ -0.2928 & 1.0988 & -3.2433 & 0.3400 \\ -1.0277 & 1.7632 & -3.6490 & 2.6068 \end{pmatrix}.$$

From this it can be deduced that $\phi_{12}(0,1)$ is invertible and that

$$W = \begin{pmatrix} 0.0166 & -0.0773 \\ -0.0773 & 0.4680 \end{pmatrix}.$$

The orthogonality condition requires,

$$\eta^T \begin{pmatrix} 0.6044 \\ 0.0992 \end{pmatrix} = 0$$

and so $\eta = (-0.0992, 0.6044)\rho$ for any nonzero real number ρ. The check is

$$\rho(-0.0992, 0.6044)^T W(-0.0992, 0.6044)\rho = 0.1806\rho^2 > 0,$$

for all nonzero ρ. It can now be concluded that the solution obtained numerically is indeed a minimizer.

Remark. The second order sufficiency test developed here compares favourably with other existing second order sufficiency tests that in many cases are not even applicable. The

Riccati condition, for example, presented in [61], [75] specialised to problem (S) requires that the symmetric solution to (3.18) satisfies $P(0) - P(T) > 0$. As anticipated this test was found not to be satisfied over a wide range of terminal values $P(T)$. In fact, the extended Riccati condition (see Section 3.2) is also not satisfied due to the presence of the zero eigenvalue in the W matrix.

Sufficient conditions specific to the periodic optimal control problems exist in literature. Speyer and his collaborators, for example, have developed conditions that are more general than the ones of this section, since they treat the free period problem but they are more difficult to evaluate numerically. The second order sufficiency condition found in [60, Theorem 2] requires that there exists a solution to the Riccati equation that satisfies a periodic condition $P(0) = P(T)$. This is more restrictive than what is hypothesised here and inevitably this requires searching for such a solution over terminal conditions and is by no means a quick task. Similar requirements in terms of periodic solution to the Riccati equation are also reported by Bittanti et. al. in [12] and Chang [17]. Weaker hypotheses that do not require a global solution of the Riccati equation on $[0, T]$, are actually reported in [67]. It should be noted that even though the conditions there are more general to the ones of Theorem 3.8 they are not generally applicable to other problems exhibiting non-uniqueness characteristics beyond the special case of periodic optimal control.

3.5 Application II: Other Endpoint Constraints

A variant of Speyer's sailboat problem [59] is presented in this section. This problem features an isometric constraint on the average displacement of the first state. In this case, the assertions of sufficient conditions specialised to periodic optimal control problem are not applicable due to the different problem formulation and the standard second order sufficient conditions fail due to non-uniqueness of the solution. Consider,

$$(SM') \begin{cases} \text{minimize } \displaystyle\int_0^1 x_1^2(t) - x_2^2(t) + x_2^4(t) + 0.01u^2(t)\, dt \\ \text{subject to} \\ \dfrac{d}{dt}\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = \begin{pmatrix} x_2(t) \\ u(t) \end{pmatrix} \text{ for all } t \in [0,1] \\ \displaystyle\int_0^1 x_1(t)dt = 1 \\ x(0) = x(1), \end{cases}$$

which following state augmentation can be written as

$$(SM)\begin{cases} \text{minimize } \displaystyle\int_0^1 x_1^2(t) - x_2^2(t) + x_2^4(t) + 0.01u^2(t)\,dt \\ \text{subject to} \\ \dfrac{d}{dt}\begin{pmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{pmatrix} = \begin{pmatrix} x_2(t) \\ u(t) \\ x_1(t) \end{pmatrix} \text{ for all } t \in [0,1] \\ x_{1,2}(0)x_{1,2}(1) \\ x_3(0) = 0 \\ x_3(1) = 1. \end{cases} \tag{3.49}$$

A numerical solution obtained by *IPOPT* [66] is shown in Figures 3.5 and 3.6. In line with previous analysis, the corresponding accessory problem is

$$(SM_A)\begin{cases} \text{minimize } \displaystyle\int_0^1 \begin{pmatrix} y_1(t) & y_2(t) \end{pmatrix} Q(t) \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} + v(t)Rv(t)\,dt \\ \text{over } y \in W^{1,1}([0,1];\mathbb{R}^3) \text{ and } v \in L^2([0,1];\mathbb{R}) \\ \text{subject to} \\ \dot{y}(t) = Ay(t) + bv(t) \text{ for all } t \in [0,1] \\ y_{1,2}(0) = y_{1,2}(1) \\ y_3(0) = y_3(1) = 0, \end{cases}$$

where

$$Q(t) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 6\bar{x_2}^2(t) - 1 & 0 \\ 0 & 0 & 0 \end{pmatrix},\ A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},\ b = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix},\ R = 0.01, D = 0.$$

Similarly, the Hamiltonian system linked to (SM_A) is,

$$\frac{d}{dt}\begin{pmatrix} y(t) \\ q(t) \end{pmatrix} = \begin{pmatrix} A & -50bb^T \\ -2Q & -A^T \end{pmatrix}\begin{pmatrix} y(t) \\ q(t) \end{pmatrix}. \tag{3.50}$$

By solving the linear system (3.50) as before, the partition $\phi_{12}(0,T)$ is found to be

$$\phi_{12}(0,T) = \begin{pmatrix} 0.7068 & 37.8438 & -1.8036 \\ -34.4995 & -437.2376 & 8.4988 \\ 1.7343 & 8.2528 & 0.3752 \end{pmatrix}$$

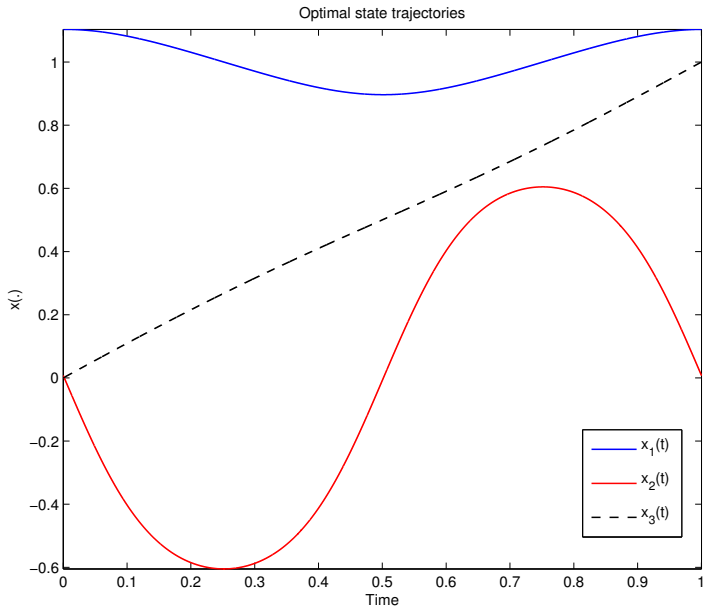


Figure 3.5: State plot

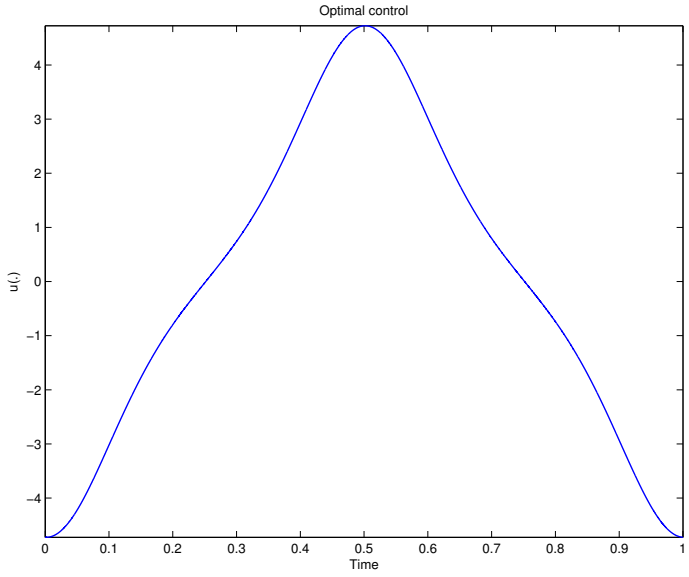


Figure 3.6: Control plot

which is invertible (indeed this is a consequence of the system being controllable). Condition (i) of Theorem 3.4 is confirmed by solving the Riccati equation with terminal condition $P(T) = 0.5I_{9\times 9}$. Figure 3.7 shows the 6 solutions (due to symmetry). It now remains

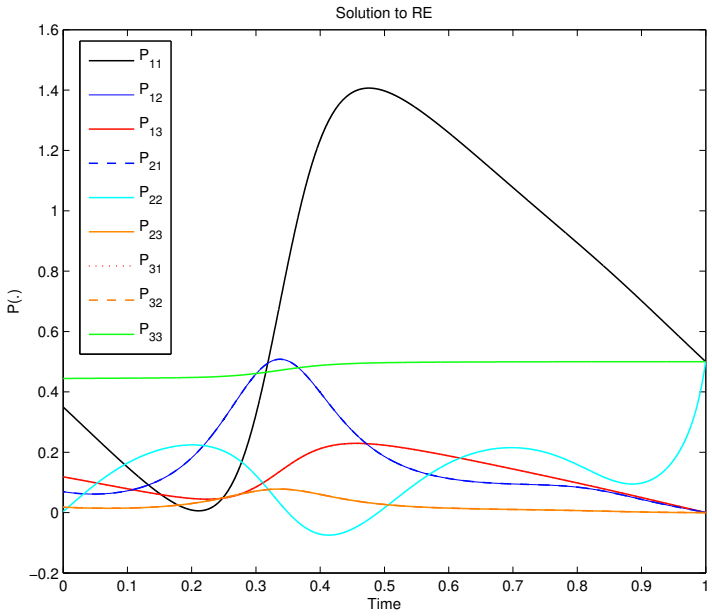


Figure 3.7: Solution to Riccati equation

to confirm condition (ii). The 6×6 W matrix calculated by solving the system (3.50) is calculated to be

$$W = \begin{pmatrix} 8.5491 & 1.0969 & 16.9130 & 8.3637 & -0.9887 & -16.9130 \\ 1.0969 & 0.1219 & 2.1356 & 1.0387 & -0.1212 & -2.1356 \\ 16.9130 & 2.1356 & 35.5898 & 16.6764 & -2.0076 & -35.5898 \\ 8.3637 & 1.0387 & 16.6764 & 8.3125 & -1.0189 & -16.6764 \\ -0.9887 & -0.1212 & -2.0076 & -1.0189 & 0.1210 & 2.0076 \\ -16.9130 & -2.1356 & -35.5898 & -16.6764 & 2.0076 & 35.5898 \end{pmatrix}.$$

The 1×6 matrix $\Gamma = \gamma^T$ of Theorem 3.4 is

$$\begin{aligned} \gamma^T &= \Big(\dot{\tilde{x}}_1(0),\, \dot{\tilde{x}}_2(0),\, 0,\, \dot{\tilde{x}}_1(0),\, \dot{\tilde{x}}_2(0),\, 0\Big) \\ &= \Big(0.0180, -4.7216, 0, 0.0180, -4.7216, 0\Big). \end{aligned}$$

It can be seen that γ and the rows of the matrix

$$\begin{pmatrix} d_{x_0} & d_{x_1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

are linearly independent. Thus (F_4) is satisfied. To apply the sufficiency test it remains only to check positivity of W on appropriate subspaces.

Application of standard sufficient conditions (such as the extended Riccati condition for example given in [1]) requires W to be positive definite on the null-space of $[d_{x_0}\, d_{x_1}]$. But this is not the case, because γ is a non-zero vector satisfying $[d_{x_0}\, d_{x_1}]\gamma = 0$, for which $W\gamma = 0$. Thus the standard sufficient condition fails to establish the weak local optimality of $(\bar{x}(.), \bar{u}(.))$.

On the other hand, the refined sufficiency test of Theorem 3.4 requires positive definiteness of W on the one-dimensional subspace of $\mathbb{R}^6$ comprising vectors $(\xi 0, \xi_1)$ that satisfy the condition

$$d_{x_0}\xi_0 + d_{x_1}\xi_1 = 0 \quad \text{and} \quad \Gamma \begin{pmatrix} \xi_0 \\ \xi_1 \end{pmatrix} = 0\,. \tag{3.51}$$

This subspace is generated by the vector

$$\begin{aligned} \eta &= \Big(-\dot{\bar{x}}_2(0), \dot{\bar{x}}_1(0), 0, -\dot{\bar{x}}_2(0), \dot{\bar{x}}_1(0), 0 \Big) \\ &= \Big(4.7216, 0.0180, 0, 4.7216, 0.0180, 0 \Big). \end{aligned}$$

It is found that

$$\eta^T W \eta \;=\; 748.8 \;>\; 0\,.$$

So W restricted to the subspace defined by (3.51) is positive definite. Theorem 3.4, therefore, confirms that the weak local extremal is a weak local minimizer.

3.6 Conclusion

Chapter 3 has presented second order sufficient conditions of local optimality for control problems having non-unique local solutions. The new, refined sufficiency criteria have been derived by reformulating the optimal control problem as a special case of the abstract framework of Chapter 2. The novelty of the new sufficient conditions is that they can be applied to a wide range of problems. Moreover, they comprise simple numerical tests which can be easily implemented computationally.

Finally, applications from the field of periodic optimal control as well as other problems with mixed boundary constraints have been employed to illustrate the applicability of the new conditions in cases where existing conditions cannot be used.

Chapter 4

Regularity of Minimizers for Second Order Problems

This chapter examines regularity properties of minimizers for problems in the calculus of variations in which the integrand involves derivatives up to second order. In particular a new set of hypotheses is proposed under which a minimizer for the problem in question has bounded highest derivative. The new results are obtained by employing a proof technique which allows for earlier, known hypotheses to be relaxed.

The structure of the chapter is as follows. Section 4.1 introduces the problem under consideration and summarizes known techniques from regularity theory. Section 4.2 summarizes results from regularity theory for first order problems, because they are suggestive of the kind of regularity properties that might be expected in a second order context. An example, however, illustrates that merely extrapolating from the hypotheses for the first order problem does not yield the required results; instead, the hypotheses for the second order problem need to be strengthened. Whilst such hypotheses have been developed in the literature, the aim of this chapter is to show that they can actually be weakened and therefore be applicable to a wider class of problems. The new results are presented in Section 4.3 and section 4.4 exposes the details of the proof technique.

4.1 Introduction

The problem to be considered here, labelled as (SOP), is a second order, autonomous problem in one independent variable. Consider,

$$(SOP)\begin{cases}\text{minimize } \displaystyle\int_0^1 L(x(t),\dot{x}(t),\ddot{x}(t))\,dt\\ \text{over arcs } x(.)\in W^{2,1}([0,1];\mathbb{R}^n)\\ \text{subject to}\\ (x(0),\dot{x}(0))=(x_0,v_0)\\ (x(1),\dot{x}(1))=(x_1,v_1).\end{cases}$$

'Autonomous' refers to the fact that the integrand does not depend explicitly on the independent variable t. The cost is expressed in terms of the function $L:\mathbb{R}^n\times\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$ and the vectors $x_0,v_0,x_1,v_1\in\mathbb{R}^n$ represent boundary restrictions. The space $W^{2,1}$ is as defined in Chapter 1, Section 1.2.

Problems of this nature can be found, for example, in rod theory [56, Section 4.9]. Apparently, the work done in deforming an elastic rod of fixed length that is clamped at both ends can be expressed as an integral involving the first and second derivatives of a curve $y(x)$ (the rod in Cartesian coordinates) and is stored as potential energy of strain in the rod. It turns out that a curve that minimizes this energy describes the shape of a bent elastic rod in stable equilibrium.

Unrestrictive conditions for existence of solutions to (SOP) ('$W^{2,1}$ minimizers') are well-known (see for example [54]). A simple set of such conditions, referred to as the 'basic existence hypotheses', are:

$$(H_E)\begin{cases}L(.,.,.) \text{ is lower semi-continuous,}\\ L(x,v,w) \text{ is convex in } w \text{ for fixed } (x,v) \text{ and}\\ L(x,v,.) \text{ is uniformly coercive,}\end{cases}\tag{4.1}$$

where '$L(z,.)$ is uniformly coercive' means that there exist a number $\beta>0$ and a positive, scalar valued, monotone function, $\theta(.):(0,\infty)\to(0,\infty)$ such that $\theta(r)/r\to\infty$ as $r\to\infty$ and

$$L(z,w)\geq-\beta|w|+\theta(|w|)\quad\text{for all } z \text{ and } w.\tag{4.2}$$

Conditions for existence of minimizers have an important role in a classical approach to the solution of (SOP) (the 'Direct Method' [71], [19]), championed by Caratheodory:

(i) verify conditions for existence of minimizers,

(ii) identify extremals (arcs satisfying first order necessary conditions akin to the Euler-

Lagrange equation),

(*iii*) seek an extremal with least cost.

A defect of the approach, regarded as a general methodology for solving (SOP), is that the hypotheses for existence of solutions (H_E) on the one hand, and the hypotheses under which necessary conditions can be rigorously derived on the other, are not wholly compatible. Indeed, the data for a particular problem may satisfy (H_E), but not the hypotheses that guarantee minimizers satisfy extremality conditions (such as the higher order Euler-Lagrange condition); the approach does not identify minimizers in these circumstances.

The discrepancy between the two sets of hypotheses is in most evidence when a minimizer $\bar{x}$ has unbounded second derivatives. The fact that the cost is finite ensures that $t \to L(\bar{x}(t), \dot{\bar{x}}(t), \ddot{\bar{x}}(t))$ is integrable. But it cannot be directly deduced from (H_E) that the unbounded functions $\partial_x L$, $d/dt\, \partial_{\dot{x}} L$, $d^2/dt^2\, \partial_{\ddot{x}} L$ have the integrability properties when evaluated along $\bar{x}(t)$, required even for the statement of the second order Euler-Lagrange condition to make sense.

The difficulty can be overcome in particular cases, by establishing a common set of hypotheses covering the problems at hand, for existence and validity of necessary conditions. Since the discrepancy between the two types of hypotheses typically arises only when minimizers have unbounded second derivatives, the goal will is often achieved if unrestrictive supplementary conditions can be found, to add to (H_E), guaranteeing that minimizers are $W^{2,\infty}$ functions.

The objective of this chapter is to weaken known conditions, supplementing the existence hypotheses (H_E) for $W^{2,\infty}$ regularity, i.e. for minimizers to have bounded second derivatives. It has so far been discussed how dealing with the logical inconsistencies implicit in Caratheodory method is one form of justification for this line of research. Another is the need of computational optimization: prior knowledge of the regularity properties of minimizers influences the choice of numerical schemes because of their effect on rates of convergence.

An obvious technique for establishing regularity properties of a minimizer $\bar{x}$ is to apply necessary conditions and examine the regularity properties of arcs satisfying these conditions. But this technique is not directly applicable here, in establishing $W^{2,\infty}$ regularity, because the standard necessary conditions are not valid. Instead, necessary conditions can be used to establish regularity, but only indirectly; the idea is to exploit the fact that minimizers for (SOP) are also minimizers for certain 'auxiliary problems' to which necessary conditions can be applied instead.

The issues addressed in this chapter, centered on the interplay between hypotheses for existence of minimizers, hypotheses for the rigorous derivation of extremality conditions

and minimizer regularity, have been extensively addressed, in the context of first order problems in one independent variable (the so-called 'Basic Problem of the Calculus of Variations') in [20] and also [18] for example. Higher order problems (and in particular second order problems) have also been addressed in [21], but have received much less attention. It will be shown here, how existing conditions for $W^{2,\infty}$ regularity of minimizers for second order problems can be significantly improved.

Apart from the 'compactification' techniques employed in [55], there have been two main approaches to regularity analysis for problems to which the optimality conditions akin to the Euler-Lagrange condition are not directly applicable. One is based on analyzing the Tonelli set (the complement of the set in $[0,1]$ on which the highest derivative of the minimizer is locally essentially bounded) and showing this must be empty [63], [20]; the other on time parameterization [4].

The novelty of the technical approach employed in this chapter is that it is based on the simultaneous application of both of these approaches, in a way that produces stronger results than would have been achievable using either approach alone.

4.2 Conditions for Minimizer Regularity

4.2.1 Minimizer Regularity for First Order Problems

A helpful insight to minimizer regularity is obtained by examining the simple, first order variational problem

$$(FOP)\begin{cases} \text{minimize } \displaystyle\int_0^1 L(x(t),\dot{x}(t))\,dt \\ \text{over } x \in W^{1,1}([0,1],\mathbb{R}^n) \\ \text{subject to} \\ x(0)=x_0,\ x(1)=x_1. \end{cases}$$

It would be fair to say that first order problems have been examined exhaustively over the years. It is still remarkable, however, that minimizers for these problems are $W^{1,\infty}$ functions under (H_E) and a very mild supplementary 'boundedness' hypothesis. This fact was proved in [4]. An earlier proof, based on a generalization of Tonelli regularity theory, and valid under the additional assumption that L is locally Lipschitz continuous, is to be found in [20]. The following theorem summarizes regularity results for first order problems.

Theorem 4.1. [64, Section 11.5] *Let $\bar{x}$ be a minimizer for (FOP). Assume that*

(*i*) $L(.,.)$ *is Borel measurable and bounded on bounded sets,*

(*ii*) $L(x,v)$ *is convex in* v *for fixed* x*,*

(*iii*) $L(x,.)$ *is uniform coercive.*

Then $\bar{x} \in W^{1,\infty}([0,1];\mathbb{R}^n)$.

4.2.2 Minimizer Regularity for Second Order Problems

Analogous hypotheses to those of Theorem 4.1, for the second order problem (SOP) are:

(H_1^F) $L(.,.,.)$ is Borel measurable and bounded on bounded sets.

(H_2^F) $L(x,v,w)$ is convex in w for fixed (x,v).

(H_3^F) $L(x,v,.)$ is uniformly coercive.

Extrapolating from the first order case, covered by Theorem 4.1, leads one to conjecture that hypotheses (H_1^F)-(H_3^F) alone suffice to guarantee that minimizers for (SOP) have bounded second derivatives. An example from [21], involving the Lagrangian

$$f_\epsilon(x,v,w) \;=\; |x^2 - v^5|^2|w|^{22} + \epsilon|w|^2\,,$$

salient properties of which are summarized in the proposition below, indicates however, that this conjecture is false.

Proposition 4.2. [21] *There exists a number* $\bar{\epsilon} > 0$ *and a* C^∞ *function* $k : (-\bar{\epsilon},\bar{\epsilon}) \to \mathbb{R}$ *with the properties: for any* $\epsilon \in (0,\bar{\epsilon})$ *consider the problem* $P(\epsilon)$*:*

$$P(\epsilon)\begin{cases} \textit{minimize } \displaystyle\int_0^1 f_\epsilon(x(t),\dot{x}(t),\ddot{x}(t))\,dt \\ \textit{over arcs } x(.) \in W^{2,1}([0,1];\mathbb{R}) \\ \textit{subject to} \\ x(0)=0,\ \dot{x}(0)=0,\ x(1)=k(\epsilon),\ \dot{x}(1)=(5/3)k(\epsilon). \end{cases}$$

Then for each $\epsilon \in (0,\bar{\epsilon})$*,* $P(\epsilon)$ *satisfies* (H_1^F)-(H_3^F) *above, and the function*

$$x_\epsilon(t) \;=\; k(\epsilon)t^{5/3},$$

which is a $W^{2,1}$ *function but not a* $W^{2,\infty}$ *function, satisfies*

$$\left(L_x - \frac{d}{dt}L_v + \frac{d^2}{dt^2}L_w\right)(x_\epsilon(t),\dot{x}_\epsilon(t),\ddot{x}_\epsilon(t)) \;=\; 0 \quad a.e. \tag{4.3}$$

where $L = f_\epsilon$.

The proposition offers evidence that, under hypotheses (H_1^F)-(H_3^F), minimizers may fail to be $W^{2,\infty}$ functions, because, as shown in [21], (4.3) is a necessary condition for $x_\epsilon(.)$ to be a minimizer. Note that condition (4.3) is weaker than the standard second order extremality conditions, that also require $t \to L_v$ and $t \to L_w$, evaluated along $\bar{x}(.)$, to be integrable. The arc $x_\epsilon(.)$ in the proposition fails to satisfy these integrability conditions.

The example suggests that, to achieve $W^{2,\infty}$ regularity, (H_1^F)-(H_3^F) need to be supplemented with additional hypotheses. One such set of supplementary hypothesis achieving this purpose was earlier identified in [21].

Theorem 4.3. *Let* $\bar{x}$ *be a minimizer for* (SOP). *Assume*

$(\tilde{H}_1^F)$ $L(.,.,.)$ *is locally Lipschitz continuous.*

(H_2^F) $L(x,v,w)$ *is convex in* w *for fixed* (x,v).

(H_3^F) $L(x,v,.)$ *is uniformly coercive.*

Assume furthermore that

(S) *there exists some non-negative* $\alpha(.) \in L^1([0,1];\mathbb{R})$ *such that*

$$|\partial_{x,v} L(\bar{x}(t), \dot{\bar{x}}(t), \ddot{\bar{x}}(t))| \leq \alpha(t), \quad \text{a.e. } t \in [0,1].$$

Then $\bar{x}(.) \in W^{2,\infty}([0,1];\mathbb{R}^n)$.

Remark. In the case that L is continuously differentiable, (S) can alternatively be expressed as:

$$t \to |\nabla_x L(\bar{x}(t), \dot{\bar{x}}(t), \ddot{\bar{x}}(t))| + |\nabla_v L(\bar{x}(t), \dot{\bar{x}}(t), \ddot{\bar{x}}(t))| \text{ is integrable.} \tag{4.4}$$

4.3 New Conditions for $W^{2,\infty}$ Regularity

The new conditions for $W^{2,\infty}$ regularity require a weaker form of the supplementary condition (S). These are the subject of two theorems, which differ slightly according to the nature of the hypotheses imposed.

Theorem 4.4. *Let* $\bar{x}$ *be a minimizer for* (SOP). *Assume*

(H_1^S) $L(.,.,.)$ *is Borel measurable and bounded on bounded sets.*

(H_2^S) $L(x,v,w)$ *is convex in* w *for each* (x,v).

(H_3^S) *L(x,v,.) is uniformly coercive.*

(H_4^S) *There exist $\epsilon > 0$ and $\bar{\sigma} > 0$ such that, for each $\sigma \in [-\bar{\sigma}, \bar{\sigma}]$, $x \to L(x, \dot{\bar{x}}(t), (1+\sigma)\ddot{\bar{x}}(t))$ is Lipschitz continuous on $\bar{x}(t) + \epsilon\mathbb{B}^n$ with Lipschitz constant $k(t,\sigma)$, where $k(.,.)$ is Borel measurable and $t \to k(t,0)$ is an integrable function.*

Then $\bar{x}(.) \in W^{2,\infty}([0,1];\mathbb{R}^n)$.

The next theorem asserts that if (H_1^S) is modified to assume that the Lagrangian is locally Lipschitz continuous, then (H_4^S) can also be replaced by a less restrictive integrability hypothesis as follows.

Theorem 4.5. *Let $\bar{x}$ be a minimizer for (SOP). Assume*

$(\tilde{H}_1^S)$ *$L(.,.,.)$ is locally Lipschitz continuous.*

(H_2^S) *$L(x,v,w)$ is convex in w for each (x,v).*

(H_3^S) *$L(x,v,.)$ is uniformly coercive.*

$(\tilde{H}_4^S)$ *There exists some non-negative $\alpha(.) \in L^1([0,1];\mathbb{R})$ such that*

$$\partial_x L(\bar{x}(t), \dot{\bar{x}}(t), \ddot{\bar{x}}(t)) \ \in \ \alpha(t)\mathbb{B}^n, \quad \text{a.e. } t \in [0,1].$$

Then $\bar{x}(.) \in W^{2,\infty}([0,1];\mathbb{R}^n)$. Furthermore, there exists $p_1(.), p_2(.) \in W^{1,\infty}([0,1];\mathbb{R}^n)$ such that

$$(\dot{p}_1(t), \dot{p}_2(t) + p_1(t), p_2(t)) \in \operatorname{co} \partial L(\bar{x}(t), \dot{\bar{x}}(t), \ddot{\bar{x}}(t)). \tag{4.5}$$

Remarks

1. The condition (4.5) is a form of the Euler Lagrange inclusion which, in the smooth case, specializes to

   $$(\dot{p}_1(t), \dot{p}_2(t) + p_1(t), p_2(t)) = \nabla L(\bar{x}(t), \dot{\bar{x}}(t), \ddot{\bar{x}}(t)).$$

   The condition follows automatically from the Maximum Principle (for example [64]), the hypotheses for application of which are satisfied since $\bar{x}(.) \in W^{2,\infty}$.

2. If $L(x,v,w)$ is continuously differentiable in x then $(\tilde{H}_4^S)$ becomes

   $$t \to |\nabla_x L(\bar{x}(t), \dot{\bar{x}}(t), \ddot{\bar{x}}(t))| \text{ is integrable.} \tag{4.6}$$

   (H_4^S) is a slightly stronger 'uniform' version of this condition.

3. The key difference between the earlier, stronger, hypothesis (4.4) and (4.6) is that the former places integrability conditions on *both* the x and v derivatives of the

Lagrangian along the minimizing arc, whereas the new condition imposes only integrability conditions on the x derivatives. Consider, for example, Lagrangians of the form,

$$L(x, v, w) = L_0(x, v) + L_1(v, w)$$

in which L_0 and L_1 are locally Lipschitz continuous functions and $L_1(v, .)$ is uniformly coercive. These Lagrangians satisfy (4.6) but will not, in general, satisfy (4.4), because the partial subgradients of L_1 with respect to v may fail to have the required integrability properties along the optimal arc. It must be mentioned that, the regularity conditions in [55], based on compactification techniques, are applicable to optimal control problems of more general structure, but when specialized to the problems considered here, require at least integrability of both x and v derivatives of the Lagrangian.

4.4 Proof of Theorems 4.4 and 4.5

The proof technique for the theorems presented is based on an optimal control problem, which is obtained by parameterising the time variable of the calculus of variations problem. The underlying idea here is that first order necessary conditions of optimality will be applicable to this problem (and hence can provide useful information about minimizer regularity) and at the same time the minimizers of the optimal control problem and (SOP) coincide. The new problem, which will be referred to as the *auxiliary problem*, is the following.

$$(A)\begin{cases} \text{minimize } \displaystyle\int_0^1 l(s, z(s), u(s))\, ds \\ \text{over arcs } (z(.), y(.)) \in W^{1,1}([0,1]; \mathbb{R}^{n+1}) \text{ and measurable functions } u : [0,1] \to \mathbb{R} \\ \text{subject to} \\ \begin{pmatrix} \dot{z}(s) \\ \dot{y}(s) \end{pmatrix} = \begin{pmatrix} u(s)\dot{\tilde{x}}(s) \\ u(s) \end{pmatrix} \quad \text{a.e. } s \in [0,1] \\ u(s) \in [1-\alpha, 1+\alpha] \quad \text{a.e. } s \in [0,1] \\ (z(0), y(0)) = (x_0, 0),\ (z(1), y(1)) = (x_1, 1), \end{cases}$$

where $\alpha \in (0,1)$ and the function l is defined to be

$$l(s, z, u) \ = \ L(z, \dot{\tilde{x}}(s), \ddot{\tilde{x}}(s)/u)u.$$

In problem (A), (z, y) is regarded as a vector in $\mathbb{R}^{n+1}$ and u as the control variable. Observe that under the hypotheses, the cost function is well-defined for any $u(.)$ and $(z(.), y(.))$

satisfying the constraints of (A), since $l(.,.,.)$ is a Borel measurable function and $l(s,z,u)$ is bounded below by $-\beta' |u|$, for some constant $\beta' > 0$.

Proposition 4.6. $(\bar{z}(.) = \bar{x}(.), \bar{y}(s) \equiv s, \bar{u}(s) \equiv 1)$ *is a minimizer for* (A).

Proof. To prove the proposition, take $\alpha \in (0,1)$ and introduce, for an arbitrary measurable function $u(.) : [0,1] \to [1-\alpha, 1+\alpha]$, a change of independent variable $t = \tau(s)$, where $\tau : [0,1] \to [0,1]$ is the strictly increasing Lipschitz continuous function with Lipschitz continuous inverse:

$$t = \tau(s) \;=\; \int_0^s u(\sigma)\, d\sigma.$$

Choose any $(z(.), y(.))$ and $u(.)$ satisfying the constraints of (A). The relationship between t and s is therefore dictated by the ordinary differential equation

$$\frac{dt}{ds} = u(s).$$

Now, write $\dot{\eta}(t) := (\dot{\bar{x}} \circ \tau^{-1})(t)$. It follows from the solution to the dynamics equation that

$$z(s) = x_0 + \int_0^s \dot{\bar{x}}(\sigma) u(\sigma)\, d\sigma,$$

which after a change of variables becomes

$$\eta(t) = x_0 + \int_0^t \dot{\eta}(t')\, dt'.$$

Hence $\eta(.)$ is a $W^{1,1}$ function. Finally, it follows from the definition of $\dot{\eta}(t)$ that

$$\begin{aligned} \frac{1}{u(s)} \ddot{\bar{x}}(s) \;&=\; \frac{1}{u(s)} \frac{d}{ds} \dot{\bar{x}}(s) \\ &=\; \ddot{\eta}(t). \end{aligned}$$

Notice that $\tau(0) = 0$ and $\tau(1) = 1$, which follows from the constraints $y(0) = 0$ and $y(1) = 1$. Applying the change of independent variables to the Lagrangian gives

$$\begin{aligned} \int_0^1 l(s, z(s), u(s))\, ds \;&=\; \int_0^1 L(z(s), \dot{\bar{x}}(s), \frac{\ddot{\bar{x}}(s)}{u(s)}) u(s)\, ds \\ &=\; \int_0^1 L(\eta(t), \dot{\eta}(t), \ddot{\eta}(t))\, dt. \end{aligned}$$

It follows from the optimality of $\bar{x}(.)$ that

$$\begin{aligned}\int_0^1 l(s,z(s),u(s))\,ds &= \int_0^1 L(\eta(t),\dot{\eta}(t),\ddot{\eta}(t))\,dt \\ &\geq \int_0^1 L(\bar{x}(t),\dot{\bar{x}}(t),\ddot{\bar{x}}(t))\,dt \;=\; \int_0^1 l(s,\bar{z}(s),\bar{u}(s))\,ds\;.\end{aligned}$$

The proposition is thus confirmed. ■

Remark. Note that even though the boundary constraints $\dot{x}(0) = v_0$, $\dot{x}(1) = v_1$ are absent in the formulation of the auxiliary problem , the vectors $\dot{y}(0)$ and $\dot{y}(1)$ are 'forced' to the correct values. This follows from the definition of the new Lagrangian l and the fact that following time parameterisation the times $s = 0, 1$ in the 's' domain map to the times $t = 0, 1$ in the 't' domain.

Having confirmed the minimizer for the auxiliary problem, attention now focuses on applying first order necessary conditions required in the proof of Theorem 4.4.

Proof (**of Theorem 4.4**). In what follows, fix $\alpha \in (0,1)$ such that the control $u(s) \in [1-\alpha, 1+\alpha]$ a.e. $s \in [0,1]$. The first task is to show that $l(s,z,.)$ is convex on $(0,\infty)$. Take arbitrary points $u_1, u_2 \in (0,\infty)$. It must be shown that for any (s,z),

$$\epsilon l(s,z,u_1) + (1-\epsilon) l(s,z,u_2) \;\geq\; l(s,z,\epsilon\, u_1 + (1-\epsilon)\, u_2)\;.$$

It can, however, be deduced from the convexity of $L(x,v,.)$ that

$$\frac{\epsilon u_1}{\epsilon u_1 + (1-\epsilon)u_2} L(z,\dot{\bar{x}},\frac{\ddot{\bar{x}}}{u_1}) + \frac{(1-\epsilon)u_2}{\epsilon u_1 + (1-\epsilon)u_2} L(z,\dot{\bar{x}},\frac{\ddot{\bar{x}}}{u_2}) \geq L(z,\dot{\bar{x}},\frac{\ddot{\bar{x}}}{\epsilon u_1 + (1-\epsilon)u_2}).$$

Multiplying across by $\epsilon u_1 + (1-\epsilon)u_2$ and noting the definition of $l(s,z,u)$ gives the desired inequality.

The aim is to establish regularity properties of $\bar{x}(.)$ by applying the Maximum Principle to (A) and studying the implications of the conditions thereby obtained. This is not possible, however, unless it is guaranteed that the Maximum Principle can be applied in normal form (i.e. the cost multiplier can be taken to be non-zero). A standard criterion for normality is that the reachable set for the underlying control system, linearized about the optimal trajectory, has non-empty interior ('local reachability'). The control system for (A) does not in general possess the reachability property, and so this criterion is not directly applicable. But reachability can be induced, and therefore normality of the Maximum Principle conditions, by reducing the dimension of the state space. This is the next task.

Henceforth it is assumed that

$$x_0 \;=\; 0.$$

There is no loss of generality in making this assumption, because it can always be arranged that it holds, by translating the origin of the state space. Now consider a modification of the control system associated with (A), obtained by removing the control constraint $u(s) \in [1-\alpha, 1+\alpha]$:

$$(S)\begin{cases} \begin{pmatrix} \dot{z}(s) \\ \dot{y}(s) \end{pmatrix} = \begin{pmatrix} u(s)\dot{\bar{x}}(s) \\ u(s) \end{pmatrix} & \text{a.e. } s \in [0,1] \\ (z(0), y(0)) = (0,0). \end{cases}$$

For each $s \in [0,1]$, let $\mathcal{R}_{[0,s]}$ be the reachable set of (S) at time s:

$$\mathcal{R}_{[0,s]} = \{(z,y)(s) \mid ((z,y),u) \text{ is a process for which } u \text{ is continuous}\}.$$

Relevant properties of the family of sets $\{\mathcal{R}_{[0,s]} \mid \; 0 \le s \le 1\}$ are listed below.

- $\mathcal{R}_{[0,s]}$ is a subspace for each $s \in [0,1]$.
- $\mathcal{R}_{[0,s]} \subseteq \mathcal{R}_{[0,s']}$ for $0 \le s \le s' \le 1$.
- The multifunction $s \to \mathcal{R}_{[0,s]}$ is left continuous, $\lim_{s' \uparrow s} \mathcal{R}_{[0,s']} = \mathcal{R}_{[0,s]}$.
- For a.e. $s \in [0,1]$

$$(u\dot{\bar{x}}(s), u) \;\in\; \mathcal{R}_{[0,s]}, \quad \text{for all } u \in (-\infty, \infty).$$

Now let D_1 be a $(n+1)\times r$ matrix, with r columns that span $\mathcal{R}_{[0,1]}$. The notation $I_{r\times(n+1)}$ is taken to mean an $r\times(n+1)$ matrix with block columns the r-dimensional identity matrix $I_{r\times r}$ followed by $(n+1)-r$ zero columns. Moreover define the $(n+1)\times(n+1)$ matrix

$$D^{-1} := [\; (D_1^T D_1)^{-1} D_1^T,\; D_2 \;]$$

for some matrix D_2. It can be shown that given a measurable function u and absolutely continuous function d such that

$$(\tilde{S})\begin{cases} \dot{d}(s) = I_{r\times(n+1)}\, D^{-1}(u(s)\,\dot{\bar{x}}(s), u(s)) \\ u(s) \;\in [1-\alpha, 1+\alpha] \\ d(0) = 0, \end{cases}$$

then

$$(z(s), y(s)) = D_1\, d(s)$$

satisfies

$$\frac{d}{ds}(z(s), y(s)) = (u(s)\,\dot{\bar{x}}, u(s)).$$

Define

$$\bar{d}(s) = I_{r\times(n+1)} D^{-1}(\bar{x}(s), \bar{y}(s) = s).$$

Now write $\mathcal{R}^d_{[0,1]}$ for the reachable set of $(\tilde{S})$ at time $s = 1$

$$\mathcal{R}^d_{[0,1]} = \{d(1) \mid (d, u) \text{ is a process for } (\tilde{S})\}.$$

Since $\bar{u} \equiv 1$ is the nominal solution, it follows, by integrating the dynamics differential equation and in view of the properties of the reachable set, that $\bar{d}(1)$ lies in the interior of the reachable set. Consequently, the control system has the following local reachability property: there exists $\epsilon > 0$ such that

$$\bar{d}(1) + \epsilon\mathbb{B}^r \;\subset\; \mathcal{R}^d_{[0,1]}. \tag{4.7}$$

It is clear from the above properties linking (S) and $(\tilde{S})$ that $(\bar{d}, \bar{u})$ is a minimizer for the following problem

$$\begin{cases} \text{minimize } \displaystyle\int_0^1 l(s, I_{n\times(n+1)} D_1 d, u)\, ds \\ \text{over arcs } d(.) \in W^{1,1}([0,1]; \mathbb{R}^r) \text{ and measurable functions } u : [0,1] \to \mathbb{R} \\ \text{subject to} \\ \dot{d}(s) \;=\; I_{r\times(n+1)} D^{-1}(u\dot{\bar{x}}(s), u) \quad \text{a.e. } s \in [0,1] \\ u \;\in [1-\alpha, 1+\alpha] \quad \text{a.e. } s \in [0,1] \\ d(0) = 0,\; d(1) = I_{r\times(n+1)} D^{-1}(x_1, 1). \end{cases}$$

Under hypotheses $(H^S_1) - (H^S_4)$, the Maximum Principle applies to this problem. Furthermore, the reachability condition implies that the cost multiplier in the Maximum Principle conditions can be set to 1. Taking account of the definition of l in terms of L, one can deduce the existence of $p(.) \in W^{1,1}([0,1]; \mathbb{R}^r)$ such that

$$\dot{p}(s) \;\in\; (I_{n\times(n+1)} D_1)^T \circ \operatorname{co} \partial_x L(\bar{x}(s), \dot{\bar{x}}(s), \ddot{\bar{x}}(s)) \tag{4.8}$$

and

$$h(s,u)|_{u=1} \geq h(s,u') \text{ a.e.} \tag{4.9}$$

for any $u' \in [1-\alpha, 1+\alpha]$, where

$$h(s,u) := p(s) \cdot I_{r\times(n+1)} D^{-1}(\dot{\bar{x}}, 1)u - L(\bar{x}(s), \dot{\bar{x}}(s), \ddot{\bar{x}}(s)/u)u.$$

Since a local maximizer of a concave function is in fact a global maximizer, $u = 1$ is a maximizer in (4.9) over all $u' \in (0, \infty)$. In particular, taking $u = 1 + |\ddot{\bar{x}}(s)|$, it can be seen that

$$L(\bar{x}(s), \dot{\bar{x}}(s), \ddot{\bar{x}}(s)) \leq$$
$$L(\bar{x}(s), \dot{\bar{x}}(s), \frac{\ddot{\bar{x}}(s)}{1 + |\ddot{\bar{x}}(s)|})(1 + |\ddot{\bar{x}}(s)|) - p(s) \cdot I_{r\times(n+1)} D^{-1}(\dot{\bar{x}}(s), 1)|\ddot{\bar{x}}(s)|.$$

Since

$$\left| \frac{\ddot{\bar{x}}(s)}{1 + |\ddot{\bar{x}}(s)|} \right| < 1\,,$$

there are, by (H_1^S), numbers $k > 0$ and $k' > 0$ such that

$$\|L(\bar{x}(s), \dot{\bar{x}}(s), \frac{\ddot{\bar{x}}(s)}{1 + |\ddot{\bar{x}}(s)|})\|_{L^\infty} \;\leq\; k, \tag{4.10}$$
$$\|L(\bar{x}(s), \dot{\bar{x}}(s), \frac{\ddot{\bar{x}}(s)}{1 + |\ddot{\bar{x}}(s)|}) \;-\; p(s) \cdot I_{r\times(n+1)} D^{-1}(\dot{\bar{x}}(s), 1)\|_{L^\infty} \leq k'.$$

These inequalities imply that for a.e. $s \in [0, 1]$,

$$L(\bar{x}(s), \dot{\bar{x}}(s), \ddot{\bar{x}}(s)) \;\leq\; k + k'|\ddot{\bar{x}}(s)|. \tag{4.11}$$

Taking $\theta(.)$ and β as in (4.2) (the coercivity condition), $\bar{\alpha}$ can be chosen such that $\theta(\alpha) - \beta\alpha > k + k'\alpha$ if $\alpha > \bar{\alpha}$. Now, assert that for a.e. $s \in [0, 1]$

$$|\ddot{\bar{x}}(s)| \;\leq\; \bar{\alpha}. \tag{4.12}$$

Indeed, if this were not the case there would exist a set of times s' of positive measure such that

$$|\ddot{\bar{x}}(s')| > \bar{\alpha}$$

and (4.11) satisfied. This in turn would imply that

$$\theta(|\ddot{\bar{x}}(s')|) - \beta|\ddot{\bar{x}}(s')| > k + k'|\ddot{\bar{x}}(s')|$$

and by the coercivity hypothesis (H_3^S)

$$L(\bar{x}(s'), \dot{\bar{x}}(s'), \ddot{\bar{x}}(s')) > k + k'|\ddot{\bar{x}}(s')|.$$

However, this last inequality would clearly contradict (4.11). Theorem 4.4 is therefore proved. ■

Proof (**of Theorem 4.5**). Since now $(\tilde{H}_1^S)$, (H_2^S) and (H_3^S) are in force, consider the following regularity property of the minimizer $\bar{x}$, proved in [21], which asserts that 'bad behaviour' of the minimizer $\bar{x}$ is confined to a closed set of zero measure (the 'Tonelli set').

There exists a closed subset $\mathcal{S} \subset [0,1]$ of zero measure, with the following property: given any $s \in [0,1]\backslash\mathcal{S}$, $\ddot{\bar{x}}(.)$ is essentially bounded on a relative neighbourhood of s.

Take an arbitrary point t_0 in $(0,1)\backslash\mathcal{S}$ and define

$$t_1 \;=\; \sup\{s \in (t_0,1] \,|\, \ddot{\bar{x}}(.) \text{ is essentially bounded on } [t_0,s]\}\,.$$

It follows from the above property that either $\ddot{\bar{x}}(.)$ is essentially bounded on $[t_0,1]$ or

$$\liminf_{s\uparrow t_1} \|\ddot{\bar{x}}(.)\|_{L^\infty(t_0,s)} \;=\; \infty\,. \tag{4.13}$$

Now it will be shown that, actually, possibility (4.13) cannot arise. This will mean then that $\ddot{\bar{x}}(.)$ is essentially bounded on $[t_0,1]$. An analogous analysis 'backwards in time' will then say that $\ddot{\bar{x}}(.)$ is essentially bounded on $[0,t_0]$ and the proof of Theorem 4.5 will be complete.

Assume (4.13). It will be shown that this gives rise to a contradiction.

Take an increasing sequence $\{t_i\} \subset (t_0,t_1]$ such that $t_i \uparrow t_1$. Then, for each i, the restriction of $\bar{x}(.)$ to $[t_0,t_i]$ is a modified version of (SOP), posed over $[t_0,t_i]$ and with fixed endpoints, those of $\bar{x}(.)$.

The same analytical techniques as in the proof of Theorem 4.4 are employed for each i, but now when the underlying time interval is $[t_0,t_i]$. In view of the upper semi-continuity property of $s \to \mathcal{R}_{[t_0,s]}$, arrange, by discarding initial terms in the sequence, that

$$\mathcal{R}_{[t_0,t_i]} \;=\; \mathcal{R}_{[t_0,t_1]}\,.$$

This means that the auxiliary problem on $[t_0,t_i]$ (with reduced state space) is the same for each i.

For each i there is an adjoint arc $p_i(.) \in W^{1,1}([t_0,t_i];\mathbb{R}^r)$ such that

$$\dot{p}_i(s) \;\in\; (I_{n\times(n+1)}D_1)^T \circ \operatorname{co}\partial_x L(\bar{x}(s),\dot{\bar{x}}(s),\ddot{\bar{x}}(s)) \tag{4.14}$$

and

$$h_i(s,u)|_{u=1} \geq h_i(s,u') \text{ a.e.} \tag{4.15}$$

for any $u' \in [1-\alpha,1+\alpha]$, where

$$h_i(s,u) \;=\; p_i(s)\cdot I_{r\times(n+1)}D^{-1}(\dot{\bar{x}},1)u - L(\bar{x}(s),\dot{\bar{x}}(s),\ddot{\bar{x}}(s)/u)u\,.$$

From these relations and taking $u(s) = 1+|\ddot{\bar{x}}(s)|$, it can be deduced that for a.e. $s \in [t_0, t_i]$

$$L(\bar{x}(s), \dot{\bar{x}}(s), \ddot{\bar{x}}(s)) \;\leq\; k + k_i^{'}|\ddot{\bar{x}}(s)| \tag{4.16}$$

where $k > 0$ is a constant satisfying (4.10) and $k_i^{'} > 0$ satisfies

$$|L(\bar{x}(s), \dot{\bar{x}}(s), \frac{\ddot{\bar{x}}(s)}{1+|\ddot{\bar{x}}(s)|}) - p_i(s) \cdot I_{r\times(n+1)} D^{-1}(\dot{\bar{x}}(s), 1)| \;\leq\; k_i^{'}.$$

But it may be deduced from (4.15) that the $p_i(t_0)$ are confined to a bounded set. It follows then from (4.14) and $(\tilde{H}_4^S)$ that the $p_i(.)$ are confined to a bounded set in L^∞. This in turn implies that the $k_i^{'}$ have a uniform bound. Consequently, $\ddot{\bar{x}}(.)$ is essentially bounded on $[t_0, t_1]$ as in the proof of Theorem 4.4. It has, thus, been shown that (4.13) cannot arise, as required to complete the proof of Theorem 4.5. ■

4.5 Conclusion

This chapter has examined the conditions under which minimizers of the second order problem in the calculus of variations have bounded second derivatives. In particular, new regularity results have been derived by combining two different approaches; time parameterisation and Tonelli's regularity theorem. This approach has resulted in stronger results than either of the two techniques alone.

The characteristic feature of the new conditions is that they *only* impose integrability requirements on the partial derivative of the Lagrangian L (along the minimizing arc) with respect to x and *not* $\dot{x}$. The new regularity results can therefore be used to identify a larger class of problems for which the minimizing arcs have essentially bounded second derivatives.

Chapter 5

Regularity of Minimizers for Higher Order Problems

Chapter 4 has presented new regularity results for autonomous second order problems of the calculus of variations. Chapter 5 extends these results to problems the Lagrangian of which depends explicitly on the time variable t as well as derivatives of the state x of up to any order N, where $N \geq 1$.

State of art hypotheses guaranteeing boundedness of the N^{th} derivative of x have, so far, been obtained by supplementing known existence hypotheses with integrability requirements of derivatives of the Lagrangian, up to order $N-1$. In this chapter it is shown that these additional hypotheses can be significantly reduced. This is made possible by employing a proof technique which is based on the technique presented in the preceding chapter. Whilst the idea is similar and is based on a combination of time re-parameterisation and the Tonelli's regularity theorem, the extension to the general case of arbitrary highest derivative, N, is not straight forward and requires an intricate analysis.

The chapter is organised as follows. The problem statement as well as known regularity results are summarised in Sections 5.1 and 5.2. The new, improved conditions are then presented in Section 5.3 and an in-depth analysis which covers the proofs appears in Section 5.4.

5.1 Introduction

The problem considered is the following.

$$(HP)\begin{cases} \text{minimize } \displaystyle\int_0^1 L(t, x(t), D^{(1)}x(t), \ldots, D^{(N)}x(t))\, dt \\ \text{over arcs } x(.) \in W^{N,1} \\ \text{subject to} \\ (x, D^{(1)}x, \ldots, D^{(N-1)}x)(0) = (x_0^0, x_0^1, \ldots, x_0^{N-1}) \\ (x, D^{(1)}x, \ldots, D^{(N-1)}x)(1) = (x_1^0, x_1^1, \ldots, x_1^{N-1}), \end{cases}$$

the data for which comprises an integer $N > 0$, a function $L : [0,1] \times \mathbb{R}^{N+1} \to \mathbb{R}$, and N-vectors $(x_0^0, x_0^1, \ldots, x_0^{N-1})$ and $(x_1^0, x_1^1, \ldots, x_1^{N-1})$. The superscripts accompanying the elements of the endpoint vector refer to the corresponding derivative of x which is in turn denoted by $D^{(i)}x$.

Standard hypotheses for existence of solutions, referred to as $W^{N,1}$ minimizers, to (HP) are as follows [53, 54].

$$(H_E^H)\begin{cases} (t, (x_0, \ldots, x_{N-1}, x_N)) \to L(t, x_0, \ldots, x_{N-1}, x_N) \text{ is } \mathcal{L} \times \mathcal{B} \text{ measurable,} \\ (x_0, \ldots, x_{N-1}, x_N) \to L(t, x_0, \ldots, x_{N-1}, x_N) \text{ is lower semi-continuous for each } t, \\ L(t, x_0, \ldots, x_{N-1}, x_N) \text{ is convex in } x_N \text{ for fixed } (t, x_0, \ldots, x_{N-1}) \text{ and} \\ L(t, x_0, \ldots, x_{N-1}, .) \text{ is uniformly coercive.} \end{cases}$$

In the preceding conditions $\mathcal{L} \times \mathcal{B}$ denotes the product σ-algebra of the Lebesgue subsets of $[0,1]$ and the Borel subsets of $\mathbb{R}^{N+1}$. The statement '$L(t, x_0, \ldots, x_{N-1}, .)$ is uniformly coercive' is taken to mean: there exist a number $\beta > 0$ and a positive, scalar valued, monotone function, $\theta(.) : (0, \infty) \to (0, \infty)$ such that $\theta(r)/r \to \infty$ as $r \to \infty$ and

$$L(t, x_0, \ldots, x_{N-1}, x_N) \geq -\beta|x_N| + \theta(|x_N|) \quad \text{for all} \quad (t, x_0, \ldots, x_{N-1}) \text{ and } x_N. \tag{5.1}$$

Henceforth, these hypotheses are referred to as the Basic Existence Hypotheses and are labelled (H_E^H).

A much studied question is the following: under what additional hypotheses do minimizers, whose existence is guaranteed under (H_E^H), have essentially bounded N^{th} order derivatives? The question is important because it underlies the construction of a framework for formulating variational problems and investigating their solutions by means of first order conditions of optimality (such as the Euler-Lagrange equation). Indeed, imposing the hypotheses of existence theory ensures that L, evaluated along an optimal trajectory $\bar{x}$, is integrable but that does not guarantee that *derivatives* of L along $\bar{x}$ are

integrable as required for the standard version of the Euler-Lagrange equation to hold.

If on the other hand it is known that the minimizers for the problem have essentially bounded N^{th} order derivatives, there is access under unrestrictive, regularity hypotheses on the data (local Lipschitz continuity or continuous differentiability, say) to first order optimality conditions for its solution.

Notation. It is convenient to write generic points in the domain of the Lagrangian as $(t, x_0, \ldots, x_N)$ for purposes of defining partial derivatives, subgradients, etc. In this way $\nabla_{t,x_0,\ldots,x_{N-2}} L$ denotes the partial derivative with respect to the first N arguments.

5.2 Standard Conditions for $W^{N,\infty}$ Regularity

A representative set of conditions from the literature for a solution $\bar{x}$ to (HP) to have an essentially bounded N^{th} derivative is summarised in the following theorem:

Theorem 5.1. *Let $\bar{x}$ be a minimizer for (HP). Assume (H_E^H) and also that*

(H_1^L) *L is continuously differentiable,*

(H_2^L) *there exists some $\alpha(.) \in L^1([0,1];\mathbb{R})$ such that*

$$|\nabla_{t,x_0,\ldots,x_{N-1}} L(t, \bar{x}(t), D^{(1)}\bar{x}(t), \ldots, D^{(N)}\bar{x}(t))| \leq \alpha(t), \quad \text{a.e. } t \in [0,1]. \tag{5.2}$$

Then $D^{(N)}\bar{x}(.)$ is essentially bounded.

A more general statement of this theorem is possible in which L is assumed merely to be locally Lipschitz continuous and hypothesis (H_2^L) is expressed in terms of partial subdifferentials. A proof of this theorem in cases when L does not depend on t and when $N = 1$ is given in [20, Theorem 2.1] and [21, Theorem 4.1] respectively. Validity of the assertions for general N is deduced from the Tonelli regularity theorem [21, Theorem 2.1] for higher order variational problems, by means of extension techniques employed, for example, in [21, Theorem 4.2].

The principal contribution of this chapter is to improve on the above conditions for essential boundedness of the N^{th} derivative of $\bar{x}$. In particular, it is shown that the assertions of Theorem 5.1 remain true when the condition (H_2^L) is replaced by the less restrictive condition

$(\tilde{H}_2^L)$ there exists some $\alpha(.) \in L^1([0,1];\mathbb{R})$ such that

$$|\nabla_{t,x_0,\ldots,x_{N-2}} L(t, \bar{x}(t), D^{(1)}\bar{x}(t), \ldots, D^{(N)}\bar{x}(t))| \leq \alpha(t), \quad \text{a.e. } t \in [0,1],$$

in which the integrability requirement is imposed on the partial derivatives of L only with respect to the $(t, x_0, \ldots, x_{N-2})$ variables, not the $(t, x_0, \ldots, x_{N-2})$ *and* x_{N-1} variables.

The validity of the assertions of Theorem 5.1 under the reduced hypotheses has been proved in the special case $N = 2$ (Second Order Problems) in Chapter 4. The findings reported in this chapter improve on those of Chapter 4 *because they place no restrictions on the value of the positive integer* N.

The approach of applying necessary conditions to an auxiliary problem was used in Chapter 4, to establish essential boundedness of the N^{th} derivative, in the case $N = 2$. Tonelli regularity theory permitted a further refinement of hypotheses. In broad terms, this approach is also followed here for arbitrary value of N. But the adaptation of the proof is by no means straightforward. A key step is to establish that the solution of the auxiliary problem is a normal extremal. This step, which is trivial in the $N = 2$ case because the dynamics of the auxiliary problem do not then depend on the state, is much more challenging for general N.

5.3 New Conditions for $W^{N,\infty}$ Regularity

This section develops new conditions for solutions to (HP) to have essentially bounded N^{th} order derivatives. These conditions appear, as in Chapter 4, in two theorems with differing hypotheses.

Theorem 5.2. *Let* $\bar{x}$ *be a minimizer for* (HP). *Assume*

(H_1^H) $(t, (x_0, \ldots, x_{N-1}, x_N)) \to L(t, x_0, \ldots, x_{N-1}, x_N)$ *is* $\mathcal{L} \times \mathcal{B}$ *measurable, and* L *bounded on bounded sets.*

(H_2^H) $L(t, x_0, \ldots, x_{N-1}, .)$ *is convex for each* $(t, x_0, \ldots, x_{N-1})$.

(H_3^H) $L(t, x_0, \ldots, x_{N-1}, .)$ *is uniformly coercive.*

(H_4^H) *there exist* $\epsilon > 0$, $\bar{\sigma} > 0$ *and a Borel measurable function* $k(.,.)$ *such that* $t \to k(t, 0)$ *is integrable and, for each* $\sigma \in [-\bar{\sigma}, +\bar{\sigma}]$

$$(t, x_0, \ldots, x_{N-2}) \to L(t, x_0, \ldots, x_{N-2}, D^{(N-1)}\bar{x}(t), (1 + \sigma)D^{(N)}\bar{x}(t))$$

is Lipschitz continuous on $\bar{x}(t) + \epsilon\mathbb{B}^n$ *with Lipschitz constant* $k(t, \sigma)$.

Then $D^{(N)}\bar{x}(.)$ *is essentially bounded.*

The next theorem asserts that if hypothesis (H_1^H) is changed to assume that the Lagrangian is locally Lipschitz continuous, then (H_4^H) can be replaced by a less restrictive integrability hypothesis.

Theorem 5.3. *Let $\bar{x}$ be a minimizer for (HP). Assume*

$(\tilde{H}_1^H)$ *L is locally Lipschitz continuous,*

(H_2^H) *$L(t, x_0, \ldots, x_{N-1}, .)$ is convex for each $(t, x_0, \ldots, x_{N-1})$,*

(H_3^H) *$L(t, x_0, \ldots, x_{N-1}, .)$ is uniformly coercive.*

$(\tilde{H}_4^H)$ *For almost everywhere $t \in [0, 1]$*

(i) $L(t, \bar{x}(t), D^{(1)}\bar{x}(t), \ldots, D^{(N-1)}\bar{x}(t), .)$ is differentiable at $D^{(N)}\bar{x}(t)$,

(ii) there exists some non-negative $\alpha(.) \in L^1([0, 1]; \mathbb{R})$ such that

$$\partial_{t,x_0,\ldots,x_{N-2}} L(t, \bar{x}(t), D^{(1)}\bar{x}(t), \ldots, D^{(N)}\bar{x}(t)) \ \in \alpha(t)\mathbb{B}^n. \tag{5.3}$$

Then $D^{(N)}\bar{x}(.)$ is essentially bounded. Furthermore, there exist

$$p_1(.), \ldots, p_N(.) \in W^{1,\infty}([0, 1]; \mathbb{R}^N)$$

such that

$$\begin{aligned}(\dot{p}_1(t), \dot{p}_2(t) + p_1(t), \ldots, \dot{p}_N(t) + p_{N-1}(t), p_N(t)) \\ \in \operatorname{co} \partial_{x_0,\ldots,x_N} L(t, \bar{x}(t), D^{(1)}\bar{x}(t), \ldots, D^{(N)}\bar{x}(t)).\end{aligned} \tag{5.4}$$

Remarks.

1. Condition (5.4) is recognized as the higher order Euler-Lagrange inclusion which, in the smooth case, specializes to

$$\begin{aligned}(\dot{p}_1(t), \dot{p}_2(t) + p_1(t), \ldots, \dot{p}_N(t) + p_{N-1}(t), p_N(t)) \\ = \nabla_{x_0,\ldots,x_N} L(t, \bar{x}(t), D^{(1)}\bar{x}(t), \ldots, D^{(N)}\bar{x}(t)).\end{aligned} \tag{5.5}$$

   The condition can be deduced from the Maximum Principle the hypotheses for application of which are satisfied since $\bar{x}(.) \in W^{N,\infty}$.

2. If $L(t, x, D^{(N-1)}\bar{x}(t), D^{(N)}\bar{x}(t))$ is continuously differentiable in x then $(\tilde{H}_4^H)$ becomes

$$t \to |\nabla_{t,x_0,\ldots,x_{N-2}} L(t, \bar{x}(t), D^{(1)}\bar{x}(t), \ldots, D^{(N)}\bar{x}(t))| \text{ is integrable.} \tag{5.6}$$

   (H_4^H) is a slightly stronger 'uniform' version of this condition.

3. The key difference between the earlier, stronger, hypothesis (5.2) and (5.6) is that the former places integrability conditions on all the $(t, x_0, x_1, \ldots, x_{N-1})$ derivatives of the Lagrangian along the minimizing arc, whereas the new condition imposes

only integrability conditions on the $(t, x_0, x_1, \ldots, x_{N-2})$ derivatives. Consider, for example, Lagrangians of the form,

$$L(t, x_0, \ldots, x_N) = L_0(t, x_0, \ldots, x_{N-2}) + L_1(x_{N-1}, x_N)$$

in which L_0 and L_1 are locally Lipschitz continuous functions and $L_1(x_{N-1}, .)$ is uniformly coercive. These Lagrangians satisfy (5.6) but do not, in general, satisfy (5.2), because the partial subgradients of L_1 with respect to x_{N-1} may fail to have the required integrability properties along the optimal arc.

4. Despite the fact that in this chapter N is arbitrary (and therefore the results in this regard are more general than those of the preceding chapter) it has to be mentioned that the regularity results established in Chapter 4 allow arcs to be vector valued. It still remains an open question whether the assertions of Theorems 5.2 and 5.3 remain valid for vector valued arcs, when $N > 2$.

5.4 Proof of Theorems 5.2 and 5.3

Central to the proofs is the auxiliary optimal control problem obtained by reformulation of (HP) to which first order necessary conditions of optimality can be applied. To this end, consider the problem

$$(A)\begin{cases} \text{minimize } \displaystyle\int_0^1 l(s, z(s), u(s))\, ds \\ \text{over arcs } z = (z_0, \ldots, z_{N-1})(.) \in W^{1,1}([0,1]; \mathbb{R}^N) \text{and meas. functions } u : [0,1] \to \mathbb{R} \\ \text{subject to} \\ \dot{z}(s) = u(s)\, Az(s) \,+\, b(s)u(s) \text{ a.e. } s \in [0,1] \\ u(s) \,\in\, [1-\alpha, 1+\alpha] \text{ a.e. } s \in [0,1] \\ z(0) = \big(0, x_0^0, x_0^1, \ldots, x_0^{N-2}\big) \\ z(1) = \big(1, x_1^0, x_1^1, \ldots, x_1^{N-2}\big). \end{cases}$$

In the above formulation the $N \times N$ matrix A and the $N-$valued function $b(.)$ are given, respectively, by

$$A = \begin{pmatrix} 0 & \ldots & 0 \\ 0_{(N-2)\times 1} & 0_{(N-2)\times 1} & I_{(N-2)\times(N-2)} \\ 0 & \ldots & 0 \end{pmatrix} \tag{5.7}$$

and

$$b(s) = \begin{pmatrix} 1 \\ 0_{(N-2)\times 1} \\ D^{(N-1)}\bar{x}(s) \end{pmatrix},$$

for all $s \in [0,1]$. Also, the function $l(.,.,.)$ is defined by

$$l(s,z,u) := L(z_0, \ldots, z_{N-1}, D^{(N-1)}\bar{x}(s),\ D^{(N)}\bar{x}(s)/u)u.$$

In problem (A), z is regarded as the state variable and u is the scalar control variable. Note that, under the hypotheses, the cost function is well-defined for any $u(.)$ and $z(.)$ satisfying the constraints of (A), since $l(.,.,.)$ is a Borel measurable function and $l(s,z,u)$ is bounded below by $-\beta' |u|$, for some constant $\beta' > 0$.

Proposition 5.4. $(\bar{z} = (s, \bar{x}, D^{(1)}\bar{x}, \ldots, D^{(N-2)}\bar{x}), \bar{u}(s) \equiv 1)$ *is a minimizer for* (A).

Proof. Let $\bar{x}(.)$ be a minimizing arc. Take any (z,u) satisfying the constraints of (A). Then

$$\begin{aligned} &\frac{d}{ds}z_0(s) = u(s), \\ &\frac{d}{ds}z_i(s) = u(s)z_{i+1}(s) \quad \text{for } i = 1, \ldots, N-2, \\ &\frac{d}{ds}z_{N-1}(s) = u(s)D^{(N-1)}\bar{x}(s). \end{aligned}$$

Once more, consider the change of independent variable $t = \tau(s)$, where $\tau : [0,1] \to [0,1]$ is the strictly increasing Lipschitz continuous function with Lipschitz continuous inverse:

$$\tau(s) \ = \ \int_0^s u(\sigma)\, d\sigma.$$

To transform the problem from the s time domain to the t time domain, define the vector valued function

$$y(t) = \begin{pmatrix} y_0(t) \\ \vdots \\ y_{N-1}(t) \end{pmatrix} = \begin{pmatrix} z_0 \circ \tau^{-1}(t) \\ \vdots \\ z_{N-1} \circ \tau^{-1}(t) \end{pmatrix} = z \circ \tau^{-1}(t).$$

Taking derivatives with respect to t gives

$$\frac{d}{dt}y_0(t) = \frac{d}{ds}x(s) \cdot \frac{1}{u(s)} = 1$$

and

$$\frac{d}{dt}y_i(t) = \frac{d}{ds}z_i(s) \cdot \frac{1}{u(s)} = y_{i+1}(t)$$

for all $t \in [0,1]$ and $i = 1, \ldots, N-2$. It also holds that

$$\frac{d}{dt}y_{N-1}(t) = D^{(N-1)}\bar{x}(s),$$

which follows from the dynamics, and furthermore that

$$\frac{d^2}{dt^2}y_{N-1}(t) = D^{(N)}\bar{x}(s) \cdot \frac{1}{u(s)}.$$

A change of variables, then, yields

$$\begin{aligned}
\int_0^1 l(s, z(s), u(s))\, ds &= \int_0^1 L(z(s), D^{(N-1)}\bar{x}(s), D^{(N)}\bar{x}(s)/u(s))u(s)\, ds \\
&= \int_0^1 L(z \circ \tau^{-1}(t), D^{(N-1)}\bar{x}(s), D^{(N)}\bar{x}(s)/u(s))\, dt \\
&= \int_0^1 L(y_0(t) = t, y_1(t), D^{(1)}y_1(t), \ldots, D^{(N-1)}y_1(t), D^{(N)}y_1(t))\, dt \\
&\geq \int_0^1 L(t, \bar{x}(t), D^{(1)}\bar{x}(t), \ldots, D^{(N-1)}\bar{x}(t), D^{(N)}\bar{x}(t))\, dt \\
&= \int_0^1 l(s, \bar{z}(s), \bar{u}(s))\, ds
\end{aligned}$$

which follows from the optimality of $\bar{x}$, the fact that

$$y_{N-1}(t) = D^{(N-2)}y_1(t)$$

and that $\tau(0) = 0$ and $\tau(1) = 1$. The proposition is therefore proved. ■

Remark. Note that even though the boundary constraints $D^{(N-1)}x(0) = x_0^{N-1}$, $D^{(N-1)}x(1) = x_1^{N-1}$ are absent in the formulation of the auxiliary problem (A) they are forced to the correct values. This follows from the definition of the new Lagrangian l and the fact that following time parameterisation the boundary values of $D^{N-1}\bar{x}(.)$ remain unchanged because the times $s = 0, 1$ in the 's' domain map to the times $t = 0, 1$ in the 't' domain.

In the sequel it is shown that $l(s, z, .)$ is convex on $(0, \infty)$. Take arbitrary points $u_1, u_2 \in (0, \infty)$. This implies that for any (s, z),

$$\epsilon l(s, z, u_1) + (1 - \epsilon)l(s, z, u_2) \;\geq\; l(s, z, \epsilon\, u_1 + (1 - \epsilon)\, u_2)\ .$$

Writing $q^N = D^{(N)}\bar{x}(s)$, it can be deduced from (H_2^H) that

$$\frac{\epsilon u_1}{\epsilon u_1 + (1-\epsilon)u_2} L(z, q^{N-1}, \frac{q^N}{u_1}) + \frac{(1-\epsilon)u_2}{\epsilon u_1 + (1-\epsilon)u_2} L(z, q^{N-1}, \frac{q^N}{u_2}) \geq L(z, q^{N-1}, \frac{q^N}{(\epsilon u_1 + (1-\epsilon)u_2)}) .$$

Now multiply across by $\epsilon u_1 + (1-\epsilon)u_2$. The desired inequality follows from the definition of $l(s, z, u)$.

5.4.1 Proof of Theorem 5.2

Regularity properties of $\bar{x}(.)$ will are established by applying the Maximum Principle to (A), with reference to the minimizer $(\bar{z}, \bar{u})$, and studying the implications of the conditions thereby obtained. It is readily checked that, under (H_1^H), (H_2^H) and (H_4^H), the hypotheses for validity of the Maximum Principle are satisfied. For the Maximum Principle to provide useful information in this case however, it needs to be shown that $(\bar{z}, \bar{u})$ is a normal extremal (i.e. the cost multiplier can be taken non-zero). A standard sufficient condition for normality is reachability of the control system (S), obtained by linearizing the underlying control system about the optimal trajectory and control, $\bar{z}$ and $\bar{u}$ respectively. The control system (S), in which the state and control variables are denoted x and v respectively, takes the form

$$(S) \begin{cases} \dot{x}(s) = Ax(s) + \tilde{b}(s)v(s) , & (5.8) \\ x(0) = 0, & (5.9) \end{cases}$$

in which A is as in (5.7) and

$$\begin{aligned} \tilde{b}(s) &= A\bar{z}(s) + b(s) \\ &= \begin{pmatrix} 1 \\ \dot{\bar{x}}(s) \\ \ddot{\bar{x}}(s) \\ \vdots \\ D^{(N-1)}\bar{x}(s) \end{pmatrix} . \end{aligned}$$

The following lemma asserts that, in establishing essential boundedness of $D^{(N)}\bar{x}(.)$, it is only essential to limit attention to the case when $(\bar{z}(.), \bar{u}(.))$ is a normal extremal for (A).

Lemma 5.5. *Suppose that (S) is not reachable (on some subinterval $[s_1, s_2] \subseteq [0, 1]$). Then $D^{(N)}\bar{x}(s) = 0$ a.e. on $[s_1, s_2]$ and so, in particular, $D^{(N)}\bar{x}(s)$ is essentially bounded on $[s_1, s_2]$.*

Proof. Define the reachability Grammian on the interval $[s_1, s_2]$

$$\mathcal{W}(s_2, s_1) = \int_{s_1}^{s_2} \Phi(s_2, t)\tilde{b}(t)\tilde{b}^T(t)\Phi^T(s_2, t)\, dt$$

where

$$\begin{aligned}\Phi(s_2, t) &= e^{A(s_2-t)} \\ &= I + A(s_2 - t) + \frac{1}{2}A^2(s_2 - t)^2 + \frac{1}{6}A^3(s_2 - t)^3 + .. + \frac{1}{(N-2)!}A^{N-2}(s_2 - t)^{N-2}.\end{aligned}$$

If the system is not reachable on $[s_1, s_2]$, there exists a non-zero $N-$vector $\alpha := (\alpha_0, \alpha_1, \ldots, \alpha_{N-1})$ in the null space of $\mathcal{W}(s_2, s_1)$ such that

$$\alpha^T \mathcal{W}(s_2, s_1)\alpha = 0.$$

This in turn implies that for all $s \in [s_1, s_2]$

$$\alpha^T \Phi(s_2, s)\tilde{b}(s) = 0.$$

It follows that

$$a_0 + \sum_{i=1}^{N-1} a_i l_i(s) \equiv 0, \tag{5.10}$$

where $(l_0(s), \ldots, l_{N-1}(s))^T = e^{A(s_2-s)}b(s)$. It can be easily verified that

$$\begin{pmatrix} l_0(s) \\ \vdots \\ l_{N-1}(s) \end{pmatrix} = \begin{pmatrix} 1 \\ \sum_{i=0}^{N-2} \frac{(s_2-s)^i}{i!} D^{(i+1)}\bar{x}(s) \\ \sum_{i=0}^{N-3} \frac{(s_2-s)^i}{i!} D^{(i+2)}\bar{x}(s) \\ \vdots \\ D^{(N-1)}\bar{x}(s)\ . \end{pmatrix}.$$

Observe that, by (5.10), $(a_1, \ldots, a_{N-1}) \neq 0$, since otherwise a_0 must also be zero, which is not allowed. Differentiating (5.10) and taking note of cancellations, yields

$$q(s)D^{(N)}\bar{x}(s) \ = \ 0 \quad \text{a.e. } s \in [s_1, s_2]\ , \tag{5.11}$$

where

$$q(s) := \Big(a_1 \frac{(s_2 - s)^{(N-2)}}{(N-2)!} + a_2 \frac{(s_2 - s)^{(N-3)}}{(N-3)!} + ... + a_{N-2}(s_2 - s) + a_{N-1}\Big). \tag{5.12}$$

But $q(.)$ is a non-zero weighted sum of linearly independent polynomials, and therefore has a finite number of roots in $[s_1, s_2]$. It follows from (5.12) that $D^{(N)}\bar{x}(s) = 0$ a.e. $s \in [s_1, s_2]$. So $D^{(N)}\bar{x}(s)$ is certainly essentially bounded. ■

It can thus be taken that $(\bar{z}, \bar{u})$ is a normal extremal for (A). Only the 'Maximization of the Hamiltonian' condition from the Maximum Principle is required. This asserts the existence of $p(.) \in W^{1,1}([0,1]; \mathbb{R}^N)$ such that, for a.e. $s \in [0,1]$,

$$H(s, \bar{z}(s), p(s), \bar{u}(s) = 1) \;=\; \max_u H(s, \bar{z}(s), p(s), u) \tag{5.13}$$

for all $u \in [1-\alpha, 1+\alpha]$, where

$$H(s, z, p(s), u) \;=\; p^T(s)\big[uAz + b(s)u\big] - l(s, z, u).$$

But $H(s, z, p(s), .)$ is concave on $(0, \infty)$ because $l(s, z, .)$ is convex on $(0, \infty)$. It follows that the local maximizer $\bar{u}(.) \equiv 1$ is in fact a maximizer over $(0, \infty)$. Inequality (5.13) is valid in particular when $u = 1 + |D^{(N)}\bar{x}(s)|$. Therefore,

$$\begin{aligned} L(s, \bar{z}(s), D^{(N-1)}\bar{x}(s), D^{(N)}\bar{x}(s)) & \\ &\leq L(s, \bar{z}(s), D^{(N-1)}\bar{x}(s), \frac{D^{(N)}\bar{x}(s)}{1 + |D^{(N)}\bar{x}(s)|})(1 + |D^{(N)}\bar{x}(s)|) \\ &\quad -p^T(s)\big[A\bar{z}(s) + b(s)\big]|D^{(N)}\bar{x}(s)| \\ &\leq k + k'|D^{(N)}\bar{x}(s)|, \end{aligned}$$

for some numbers $k > 0$ and $k' > 0$ which do not depend on s. By the coercivity hypothesis, however,

$$\begin{aligned} \theta(|D^{(N)}\bar{x}(s)|) \;&\leq\; L(s, \bar{z}(s), D^{(N-1)}\bar{x}(s), D^{(N)}\bar{x}(s)) + \beta|D^{(N)}\bar{x}(s)| \\ &\leq\; k + k_1|D^{(N)}\bar{x}(s)|, \end{aligned} \tag{5.14}$$

where $k_1 = k' + \beta$.

According to the properties of $\theta(\cdot)$, K and $\bar{\alpha} > 0$ can be chosen such that $\theta(\alpha) \geq K\alpha$ for all $\alpha \geq \bar{\alpha}$ and

$$(K - k_1)\alpha > k. \tag{5.15}$$

It can therefore be assumed that

$$|D^{(N)}\bar{x}(s)| \;\leq\; \bar{\alpha} \quad \text{a.e. } s \in [0,1].$$

For if this were not the case it would follow from (5.14) that, for all points in some set of

positive measure,

$$K|D^{(N)}\bar{x}(s)| \leq \theta(|D^{(N)}\bar{x}(s)|) \leq k + k_1|D^{(N)}\bar{x}(s)|.$$

For all points s in this set,

$$(K - k_1)\bar{\alpha} \;\leq\; (K - k_1)|D^{(N)}\bar{x}(s')| \leq k$$

which contradicts (5.15). Thus, Theorem 5.2 is proved. ■

5.4.2 Proof of Theorem 5.3

Since $(\tilde{H}_1^H)$, (H_2^H) and (H_3^H) are satisfied, the following regularity property of the minimizer $\bar{x}$, proved in [21] ('bad behaviour' of the minimizer $\bar{x}(.)$ is confined to a closed set of zero measure) holds.

There exists a closed subset $\mathcal{S} \subset [0,1]$ of zero measure, with the following property: $D^{(N)}\bar{x}(.)$ is locally essentially bounded on $[0,1]\backslash\mathcal{S}$, i.e. given $s \in [0,1]\backslash\mathcal{S}$, $D^{(N)}\bar{x}(.)$ is essentially bounded on a relative neighbourhood of s in $[0,1]$. Furthermore, if $D^{(N)}\bar{x}(.)$ is essentially bounded on some subinterval $I \subset [0,1]$, then I lies in the relative interior of $[0,1]\backslash\mathcal{S}$.

Take an arbitrary point in $(0,1)\backslash\mathcal{S}$ and define

$$\bar{t} \;=\; \sup\{t \in (t_0, 1] \,|\, D^{(N)}\bar{x}(.) \text{ is essentially bounded on } [t_0, t]\}\,.$$

As a consequence of the above stated property, there are two possibilities: either $D^{(N)}\bar{x}(.)$ is essentially bounded on $[t_0, 1]$ or

$$\liminf_{t \uparrow \bar{t}} \; ||D^{(N)}\bar{x}(.)||_{L^\infty(t_0,t)} \;=\; \infty\,. \tag{5.16}$$

Assuming that (5.16) is true gives rise to a contradiction. This means then that $D^{(N)}\bar{x}(.)$ is essentially bounded on $[t_0, 1]$. An analogous analysis 'backwards in time' then indicates that $D^{(N)}\bar{x}(.)$ is essentially bounded on $[0, t_0]$ and the proof of Theorem 5.3 is complete.

Suppose (5.16). Take an increasing sequence $\{t_i\} \subset (t_0, \bar{t})$ such that $t_i \uparrow \bar{t}$. Then, for each i, $\bar{x}_i(.) = \bar{x}(.)|_{[t_0,t_i]}$ is a solution to a modified version of (HP) (write it $(HP)_i$), posed over $[t_0, t_i]$ and with fixed endpoints, those of $\bar{x}(.)$ (and its higher derivatives) restricted to $[t_0, t_i]$.

For each i consider a variant on (A) in which the underlying time interval is $[t_0, t_i]$.

$$(A_i)\begin{cases} \text{minimize } \displaystyle\int_{t_0}^{t_i} l(s, z(s), u(s))\, ds \\ \text{over arcs } z(.) \in W^{1,1}([t_0, t_i]; \mathbb{R}^N) \text{ and measurable functions } u : [t_0, t_i] \to \mathbb{R} \\ \text{subject to} \\ \dot{z}(s) \;=\; u(s)\, Az(s) \;+\; b(s)u(s) \;\; \text{a.e. } s \in [t_0, t_i] \\ u(s) \;\in\; [1-\alpha, 1+\alpha]) \;\; \text{a.e. } s \in [t_0, t_i] \\ z(t_0) = (t_0, (\bar{x}, D^{(1)}\bar{x}, \ldots, D^{(N-2)}\bar{x})(t_0)) \\ z(t_i) = (t_i, (\bar{x}, D^{(1)}\bar{x}, \ldots, D^{(N-2)}\bar{x})(t_i)). \end{cases}$$

It can be shown, as before, that for each i, $(\bar{z}(.), \bar{u}(.))$ restricted to $[t_0, t_i]$ is a minimizer for (A_i). Linearizing the dynamic constraint about $(\bar{x}_i, \bar{u}_i)$ yields the linear system (S'):

$$(S')\begin{cases} \dot{x}(s) = Ax(s) + \tilde{b}(s)v(s) \\ x(t_0) = 0 \end{cases}$$

in which $\tilde{b}(s)$ is defined by $\tilde{b}(s) = (1, D^{(1)}\bar{x}(s), \ldots, D^{(N-1)}\bar{x}(s))$. By discarding early terms in the sequence $\{t_1, t_2, \ldots\}$ and relabelling it can be arranged that (S') is reachable on $[t_0, t_1]$ (and therefore on $[t_0, t_i]$ for all $i > 1$). This is possible by condition (5.16) and in view of Lemma 5.5.

As in the previous analysis it can be assumed, without loss of generality, that $(\bar{x}_i(.), \bar{u}_i(.))$ is a normal extremal. Noting that $\dot{\bar{z}}(s) = A\bar{z}(s) + b(s) = \tilde{b}(s)$, it can be deduced from the Maximum Principle (normal form) that for each i there exists an arc $p_i \in W^{1,1}([t_0, t_i]; \mathbb{R}^N)$ such that, for a.e. $s \in [t_0, t_i]$

$$-\dot{p}_i(s) \;=\; A^T p_i(s) - \gamma_i(s)\,, \tag{5.17}$$

$$p_i^T(s)\tilde{b}(s) - l(s, \bar{z}(s), 1) \;\geq\; p_i^T(s)\tilde{b}(s)u - l(s, \bar{z}(s), u) \tag{5.18}$$

for all $u \in [1-\alpha, 1+\alpha]$. Here $\gamma_i(.)$ is an integrable function satisfying

$$\gamma_i(s) \;\in\; \partial_{s, x_0, \ldots, x_{N-2}} L(s, \bar{x}(s), D^{(1)}\bar{x}(s), \ldots, D^{(N)}\bar{x}(s)) \quad \text{a.e. } s \in [t_0, t_i]$$

Since $L(s, \bar{x}(s), D^{(1)}\bar{x}(s), \ldots, D^{(N-1)}\bar{x}(s), .)$ is differentiable at $D^{(N)}\bar{x}(s)$ almost everywhere, (5.18) implies

$$\tilde{b}^T(s)p_i(s) = \nabla_u l(s, \bar{z}(s), 1). \tag{5.19}$$

The solution to (5.17) is given by the variation of constants formula:

$$p_i(s) \;=\; e^{-A^T(s-t_0)}p_i(t_0) + d_i(s) \quad \text{for } s \in [t_0, t_i],$$

where $d_i(.)$ is some continuous function the supremum norm of which admits a bound independent of i.

Lemma 5.6. *There exists* c_1 *such that*

$$\|p_i(.)\|_{L^\infty([t_0,t_i];\mathbb{R}^N)} \leq c_1 \quad \textit{for } i = 1, 2, \ldots$$

Proof. By (5.19),

$$\begin{aligned} \nabla_u l(s, \bar{z}(s), 1) &= \tilde{b}^T(s) p_i(s) \\ &= \tilde{b}^T(s) e^{A^T(t_0-s)} p_i(t_0) + \tilde{b}^T(s) d_i(s) \quad \text{a.e. } s \in [t_0, t_i]. \end{aligned}$$

Considering the case $i = 1$ gives

$$\begin{aligned} \nabla_u l(s, \bar{z}(s), 1) &= \tilde{b}^T(s) p_1(s) \\ &= \tilde{b}^T(s) e^{A^T(t_0-s)} p_1(t_0) + \tilde{b}^T(s) d_1(s) \text{ a.e. } s \in [t_0, t_1]. \end{aligned}$$

Now consider the change of notation $q_i(s) = p_i(s) - p_1(s)$. Subtracting the two previous equations yields

$$0 = \tilde{b}^T(s) e^{A^T(t_0-s)} q_i(t_0) + e_i(s) \quad \text{a.e. } s \in [t_0, t_1]$$

for some continuous function $e_i(.)$ the supremum norm of which admits a bound independent of i. It follows that

$$q_i^T(t_0) e^{A(t_0-t_1)} \mathcal{W}(t_1, t_0) e^{A^T(t_0-t_1)} q_i(t_0) \leq \sup_i \|e_i(.)\|_{L^2}^2,$$

where $\mathcal{W}(t_1, t_0)$ is the controllability Grammian associated to the linear system (S'), i.e.

$$\mathcal{W}(t_1, t_0) = \int_{t_0}^{t_1} e^{A(t_1-s)} \tilde{b}(s) \tilde{b}^T(s) e^{A^T(t_1-s)} \, ds.$$

But in view of Lemma 5.5, it can be assumed that (S') is controllable on $[t_0, t_1]$ and the matrix $\mathcal{W}(t_1, t_0)$ is positive definite; since $\|e_i(.)\|_{L^2}$ admits a bound independent of i, it follows that the $|p_i(t_0) - p_1(t_0)|$ and therefore also the $p_i(0)$, are uniformly bounded. Since the $p_i(.)$ are obtained by solving the linear differential equation (5.17) on a finite interval and since the $p_i(0)$ are uniformly bounded the existence of a constant c_i such that $\|p_i(.)\|_{L^\infty([t_0,t_i];\mathbb{R}^N)} \leq c_i$ for all i now follows from the fact that the γ_i in this equation are uniformly bounded in the L^1 norm. ■

To complete the proof of the theorem, consider (5.18). As before, this inequality remains valid when u is allowed to take arbitrary values in $(0, \infty)$. Set $u(s) = 1 + |D^{(N)} \bar{x}(s)|$. In

conclusion, for a.e. $s \in [t_0, t_i]$,

$$\begin{aligned} & L(s, \bar{z}(s), D^{(N-1)}\bar{x}(s), D^{(N)}\bar{x}(s)) \\ & \quad \leq \quad L(s, \bar{z}(s), D^{(N-1)}\bar{x}(s), \frac{D^{(N)}\bar{x}(s)}{1+|D^{(N)}\bar{x}(s)|})(1+|D^{(N)}\bar{x}(s)|) - p_i^T(s)\tilde{b}(s)|D^{(N)}\bar{x}(s)| \\ & \quad \leq \quad k + k'|D^{(N)}\bar{x}(s)|, \end{aligned}$$

for some numbers $k > 0$ and $k' > 0$. The preceding lemma has a crucial role here, ensuring that k and k' can be chosen independently of i.

Arguing as in the proof of Theorem 5.2, it can be deduced from this inequality and the coercivity hypothesis that $D^{(N)}\bar{x}(.)$ is essentially bounded and the numbers $\|D^{(N)}\bar{x}(.)\|_{L^\infty([t_0,t_i];\mathbb{R}^n)}$, $i = 1, 2, \ldots$ admit a uniform bound. But then $D^{(N)}\bar{x}(.)$ is essentially bounded on $[t_0, \bar{t}]$. The condition (5.16) has therefore been excluded and so the proof of the theorem is complete. ■

5.5 Conclusion

This chapter has extended the previous regularity results of Chapter 4 to problems with Lagrangians with no restriction on the order of the highest derivative, N.

While the techniques employed here are based on those of the preceding chapter, an intricate analysis is required to adapt them to this more general problem, particularly in establishing a key point in the analysis, namely normality of the extremal associated with the auxiliary control problem. The combination of Tonelli's existence theory and time reparemeterisation enable the development of new, stronger results that impose integrability requirements *only* on the first $N-2$ partial derivatives of the Lagrangian rather than the first $N-1$ partial derivatives which has so far been the trend in the field.

Chapter 6

Conclusions

New second order sufficient conditions of optimality have been derived for nonlinear optimal control problems with mixed endpoint constraints. These refined conditions have the advantage that they are also applicable to problems with non-unique local solutions.

A starting point for second order sufficient conditions is that first order necessary conditions are satisfied. In this thesis, it has been assumed that necessary conditions can be applied in 'normal' form, in other words that the multiplier associated with the cost is non-zero. This is the case, provided that the minimizer of the optimal control problem has certain regularity properties. The study of regularity properties of minimizers for the special class of problems in the calculus of variations has been the subject of the last two chapters. The novelty of the results in these chapters lies in the fact that a differentiability hypothesis has been significantly weakened and thus one can identify a larger class of problems for which the highest derivative is essentially bounded.

This concluding section of the thesis provides a brief summary of the main results and possible future research directions.

6.1 Summary

The main results of each chapter are summarised below.

- Chapter 2 has demonstrated, in a finite dimensional nonlinear programming setting at first, that the coercivity requirement of the standard second order sufficient conditions for optimality cannot be satisfied when problems have non-unique local solutions. Refined conditions have been developed for which the positivity criterion is associated with a smaller subspace. A similar approach has been followed

to derive sufficiency conditions for more abstract optimization problems posed over Hilbert spaces. In this setting, the smaller subspace is obtained by requiring that perturbations of the nominal solution are orthogonal to the 'constant cost family of extremals'.

- Chapter 3 centers on nonlinear optimal control problems with mixed boundary conditions. Such problems can be reformulated to match the abstract optimization problem of Chapter 2. Furthermore, it is seen that the optimal control problem hypotheses satisfy those of the abstract optimization framework. The assertions of the theorem for the abstract problem, then, remain valid for the optimal control problems examined here and specialize to yield simple numerical sufficiency tests. Problems with locally non-unique solutions arise naturally in fields like periodic optimal control which require that initial and final boundary conditions match ($x(0) = x(T)$). The refined conditions for these problems, simply involve checking the strict positivity of a certain matrix and the existence of a solution to a Ricatti equation. A well-known problem, called Speyer's sailboat problem, is used to highlight the superiority of the new conditions.

- Chapter 4 has focused on a special class of optimal control problems, namely second order problems in the calculus of variations and studied the conditions under which one can guarantee that the minimizer will have bounded second order time-derivative. Two different approaches have been combined here to give a stronger result that the latest state-of-art conditions in the literature: time parameterisation and Tonelli's regularity theory. The distinguishing feature of the new conditions is that they only require differentiability of the integrand $L(x(t), \dot{x}(t), \ddot{x}(t))$ in the x variable, contrasting earlier conditions in the literature requiring differentiability in both x and $\dot{x}$ variables.

- Chapter 5 has followed in line, applying the time parameterisation technique and Tonelli's existence theory to higher order problems in the calculus of variations. More precisely, the problems studied in this chapter are non-autonomous and the Lagrangian of which contains derivatives of arbitrary order N. While one could conjecture that this is a mere extrapolation of the results of Chapter 4 this is not the case because the extension of the earlier techniques to the higher order problems requires a more intricate analysis. This is based on the fact that normality of the Maximum Principle (applied to the auxiliary problem) does not follow automatically because the dynamics of the auxiliary problem also depend on the state. The new conditions only impose integrability requirements on the first $(N-2)$ derivatives of the Lagrangian as opposed to the first $(N-1)$ derivatives of existing conditions.

6.2 Future Research Directions

Finally, it is appropriate to present a plan of future research based on what has currently been achieved in this thesis. At first, it has been shown that the abstract problem of Chapter 2, which covers problems the solution of which might be non-unique in a local sense, can be specialised to optimal control problems with nonlinear dynamics and mixed endpoint constraints. A challenging extension would be to examine how this framework can be adapted to cover more general control problems with additional path-wise state constraints and also control constraints. In addition, the periodic optimal control problems that have been discussed in Chapter 3 feature an integral cost over a given, fixed time interval $[0, T]$. It would be interesting to examine how new conditions can be derived, under the same framework, for periodic problems with free final time since for many periodic optimal control problems, the period T of the control, is also an optimization variable.

Chapter 4 has addressed a minimization problem in the calculus of variations posed over $n-$valued arcs in $W^{2,1}$ by blending together two distinct approaches. Although a similar strategy has been employed in the study of the higher order calculus of variations problem in Chapter 5 the optimization domain only comprised scalar-valued arcs in $W^{N,1}$. Exactly as to how this methodology can be extended to derive similar results for problems with vector-valued arcs still remains an open question and it is certainly not a straightforward one to answer.

Publications

The material in this thesis provides the basis for the following publications:

Journal Papers

C. Gavriel and R. Vinter. Second Order Sufficient Conditions for Optimal Control Problems with Non-Unique Minimizers: An Abstract Framework, *SIAM J. of Control and Optimization* (submitted)

C. Gavriel and R. Vinter. Regularity of Minimizers for Second Order Variational Problems in One Independent Variable, *J. of Discrete and Cont. Dynamical Systems*, 29, 547 – 557, 2011

C. Gavriel, S. Lopes and R. Vinter. Regularity of Minimizers for Higher Order Variational Problems, *J. Annual Reviews in Control*, (submitted)

Conference Papers

C. Gavriel and R. Vinter. Second Order Sufficient Conditions for Optimal Control Problems with Non-Unique Minimizers, *Proceedings of the American Control Conference*, 2010

Bibliography

[1] J. Allwright and R. Vinter. Second order conditions for periodic optimal control problems. *Control and Cybernetics*, 34:618–643, 2005.

[2] W. Alt and K. Malanowski. The Lagrange-Newton method for nonlinear optimal control problems. *Computational Optimization and Applications*, (2):77–100, 1993.

[3] T. M. Apostol. *Mathematical Analysis : A Modern Approach to Advanced Calculus*. Addison-Wesley, 1965.

[4] O. Ascenti, L. Ambrosio, and G. Butazzo. Lipschitz regularity for minimizers of integral functionals with highly discontinuous integrands. *J. of Math. Anal. Applic.*, 142:301–316, 1989.

[5] A. Astolfi. Optimization - An Introduction. Lecture Notes, 2004.

[6] J. E. Bailey and F. J. M. Horn. Comparison between two sufficient conditions for improvement of an optimal steady-state process by periodic operation. *J. of Optim. Theory and Appl.*, 7:378–384, 1971.

[7] E. Barucci and P. Zezza. Does a life cycle exist for a hedonistic consumer? *Mathematical Social Sciences*, 32:57–69, 1996.

[8] R. Bellman. Dynamic programming. *Science*, 153:34–37, 1966.

[9] D. S. Bernstein and E. G. Gilbert. Optimal periodic control: The $\pi-$test revisited. *IEEE Trans. on Automatic Control*, 25:673–684, 1980.

[10] D. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, 2005.

[11] S. Bittanti and G. Guardabassi. Optimal periodic control and periodic systems analysis: An overview. *Proceedings of Conference on Decision and Control*, 1986.

[12] S. Bittanti, A. Locatelli, and C. Maffezzoni. Second variation methods in periodic optimization. *J. of Optim. Theory and Appl.*, 14(1), 1974.

[13] V. G. Boltyanskii, R. V. Gamkrelidze, E. F. Mishchenko, and L. S. Pontryagin. *The Mathematical Theory of Optimal Processes.* Interscience Publishers, 1962.

[14] J. M. Borwein. *Optimization With Respect to Partial Orderings.* PhD thesis, Oxford University, Jesus College, 1974.

[15] R. W. Brockett. *Finite Dimensional Linear Systems.* John Wiley & Sons, Inc., 1970.

[16] C. Buskens and H. Maurer. SQP-methods for solving optimal control problems with control and state constraints: adjoint variables, sensitivity analysis and real-time control. *J. of Computational and Applied Mathematics*, (120):85–108, 2000.

[17] K. S. Chang. *Second Variation for Periodic Optimization Problems*, volume 2 of *Periodic Optimization.* Springer Verlag, New York, 1972.

[18] F. H. Clarke. *Methods of Dynamic and Nonsmooth Optimization.* Regional Conference Series in Applied Mathematics. SIAM Publications, PA, 1989.

[19] F. H. Clarke. Necessary conditions in dynamic optimization. *Memoirs of the Amer. Math. Soc.*, 816:1–110, 2005.

[20] F. H. Clarke and R. B. Vinter. Regularity properties of solutions to the basic problem in the calculus of variations. *Trans. Amer. Math. Soc*, 289:73–98, 1985.

[21] F. H. Clarke and R. B. Vinter. A regularity theory for variational problems with higher order derivatives. *Trans. Amer. Math. Soc.*, 320:227–251, 1990.

[22] A. L. Dontchev, W. W. Hager, A. B. Poore, and B. Yang. Optimality, stability and convergence in nonlinear control. *Applied Mathematics and Optimization*, 31:297–326, 1995.

[23] J. C. Dunn. Second order optimality conditions in sets of L^∞ functions with range in a polyhedron. *SIAM J. Control and Optimization*, 33(5):1603–1635, 1995.

[24] J. C. Dunn. On L^2 sufficient conditions and the gradient projection method for optimal control problems. *SIAM J. Control and Optimization*, 34(4):1270–1290, 1996.

[25] G. Feichtinger, M. Han, and R. F. Hartl. Nonconcavity and proper optimal periodic control. *J. of Economic Dynamics and Control*, 18:975–990, 1994.

[26] G. Fronza, S. Bittanti, and G. Guardabassi. Periodic control: A frequency domain approach. *IEEE Trans. on Automatic Control*, 18:33–38, 1973.

[27] C. Gavriel, S. Lopes, and R. B. Vinter. Regularity of minimizers for higher order variational problems. *J. of Annual Reviews in Control*, (submitted).

[28] C. Gavriel and R. B. Vinter. Second order sufficient conditions for optimal control problems with non-unique minimizers. *Proceedings of American Control Conference*, pages 756–761, 2010.

[29] C. Gavriel and R. B. Vinter. Regularity of minimizers for second order variational problems in one independent variable. *Discrete and Cont. Dynamical Systems*, 29:547–557, 2011.

[30] C. Gavriel and R. B. Vinter. Second order sufficient conditions for optimal control problems with non-unique local minimizers: A general framework. *SIAM J. on Control and Optimization*, (submitted).

[31] I. M. Gelfand and S. V. Fomin. *Calculus of Variations*. Prentice Hall, Englewood Cliffs, N.J., 1963.

[32] E. G. Gilbert. Vehicle cruise: Improved fuel economy by periodic control. *Automatica*, 12:159–166, 1976.

[33] E. G. Gilbert and D. S. Bernstein. Second-order necessary conditions in optimal control: Accessory-problem results without normality conditions. *J. of Optim. Theory and Appl.*, 41(1):75–106, 1983.

[34] G. Guardabassi, A. Locatelli, and S. Rinaldi. Status of periodic optimization of dynamical systems. *J. of Optim. Theory and Appl.*, 14(1):1–20, 1974.

[35] M. Guignard. Generalized Kuhn-Tucker conditions for mathematical programming problems in a banach space. *SIAM J. Control*, 7(2):232–242, 1969.

[36] M. R. Hestenes. *Optimization Theory. The Finite-Dimensional Case*. John Wiley, New York, 1975.

[37] F. J. M Horn and R. C. Lin. Periodic processes: A variational approach. *Ind. Eng. Chem. Process Des. Dev.*, 6:21–30, 1967.

[38] A. D. Ioffe and V. M. Tikhomirov. *Theory of Extremal Problems*, volume 6. North - Holland, New York, 1979.

[39] H.Maurer K. Malanowski and S. Pickenhain. Second order sufficient conditions for state constrained optimal control problems. *J. of Optim. Theory and Appl.*, 123(3):595–617, 2004.

[40] R. E. Kalman. Contributions to the theory of optimal control. *Bol. Soc. Mat. Mexicana*, 5:102–119, 1960.

[41] H. W. Kuhn and A. W. Tucker. Nonlinear programming. In *Proceedings of 2nd Berkeley Symposium*, 1951.

[42] L. S. Ledyaev, R. J. Stern, F. H. Clarke, and P. R. Wolenski. *Nonsmooth Analysis and Control Theory*, volume 178. Springer Verlag, New York, 1998.

[43] D. G. Luenberger. *Introduction to Linear and Nonlinear Programming.* Addison-Wesley, Massachusettes, 1973.

[44] C. Maffezzoni. Hamilton-Jacobi theory for periodic control problems. *J. of Optim. Theory and Appl.*, 14:21–29, 1974.

[45] K. Malanowski and H. Maurer. Sensitivity analysis for parametric control problems with control-state constraints. *Computational Optimization and Applications*, (5):253–283, 1996.

[46] H. Maurer and H. J. Oberle. Second order sufficient conditions for optimal control prolems with free final time: The Riccati approach. *SIAM J. Control and Optimization*, 41(2):380–403, 2002.

[47] H. Maurer and N. P. Osmolovskii. Second order conditions for bang-bang control problems. *Control and Cybernetics*, 32(3):555–584, 2003.

[48] H. Maurer and N. P. Osmolovskii. Second order sufficient conditions for time-optimal bang-bang control. *SIAM J. Control and Optimization*, 42(6):2239–2263, 2004.

[49] H. Maurer and S. Pickenhain. Second order sufficient conditions for control problems with mixed control-state constraints. *J. of Optim. Theory and Appl.*, 86:649–667, 1995.

[50] H. Maurer and J. Zowe. First and second-order necessary and sufficient optimality conditions for infinite-dimensional programming problems. *Mathematical Programming*, 16:98–110, 1979.

[51] D. Q. Mayne. Sufficient conditions for a control to be strong minimum. *J. of Optim. Theory and Appl.*, (21):339–383, 1977.

[52] C. Moler and C. V. Loan. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *Industrial and Applied Mathematics*, 45:3–49, 2003.

[53] R. T. Rockafellar. Existence theorems for general control problems of Bolza and Lagrange. *Advances in Math.*, 15:312–333, 1975.

[54] R. T. Rockafellar and R. J. W. Wets. *Variational Analysis, Grundlehren der Mathematischen Wissenschaften*, volume 317. Springer Verlag, New York, 1998.

[55] A. V. Sarychev and D. F. Torres. Lipschitzian regularity of minimizers for optimal control problems with control-affine dynamics. *J. of Applied Math. Optim.*, 41:237–254, 2000.

[56] D. R. Smith. *Variational Methods in Optimization*. Prentice Hall, Englewood Cliffs, N.J., 1974.

[57] G. Sorger. Sufficient optimality conditions for nonconvex control problems with state constraints. *J. of Optim. Theory and Appl.*, 62(2):289–310, 1989.

[58] J. Speyer. Nonoptimality of the steady-state cruise for aircraft. *American Institute of Aeronautics and Astronautics*, 14:1604–1610, 1976.

[59] J. Speyer. Periodic optimal flight. *J. of Guidance, Control and Dynamics*, 19:745–754, 1996.

[60] J. Speyer and R. T. Evans. A second variational theory for optimal periodic processes. *IEEE Trans. on Automatic Control*, 2:138–148, 1984.

[61] G. Stefani and P. Zezza. Constrained regular LQ-control problems. *SIAM J. Control and Optimization*, 35(3), 1997.

[62] H. J. Sussmann and J. C. Willems. 300 years of optimal control: From the Brachystochrone to the Maximum Principle. *Proceedings of the Conference on Decision and Control*, 1996.

[63] L. Tonelli. *Fondamenti di Calcolo delle Variazioni*. Zanichelli, Bologna, 1921.

[64] R. B. Vinter. *Optimal Control*. Birkhauser, Boston, 2000.

[65] R. B. Vinter and F. M. Pereira. A Maximum Principle for optimal processes with discontinuous trajectories. *SIAM J. Control and Optimization*, 26:205–229, 1988.

[66] A. Wächter. *An Interior Point Algorithm for Large-Scale Nonlinear Optimization with Applications in Process Engineering*. PhD thesis, Carnegie Mellon University, 2002.

[67] Q. Wang and J. Speyer. Necessary and sufficient conditions for local optimality of a periodic process. *SIAM J. Control and Optimization*, 28(2):482–497, 1990.

[68] Q. Wang and J. Speyer. Regulators for optimal periodic processes. *IEEE Trans. on Automatic Control*, 40:1767–1776, 1995.

[69] X. Wang and J. K. Hale. On monodromy matrix computation. *Computer Methods in Applied Mechanics and Engineering*, 190:2263–2275, 2001.

[70] J. Warga. A second-order Lagrangian condition for restricted control problems. *J. of Optim. Theory and Appl.*, 24(3):475–483, 1978.

[71] L. C. Young. *Lectures on the Calculus of Variations and Optimal Control Theory.* Saunders, Philadelphia, 1969.

[72] V. Zeidan. Conjugate points criterion for optimal control. *Proceedings of Conference on Decision and Control*, pages 379–383, 1983.

[73] V. Zeidan. Sufficient conditions for the generalized problem of Bolza. *Trans. Amer. Math. Soc*, 275(2), 1983.

[74] V. Zeidan. Extended Jacobi sufficiency criterion for optimal control. *SIAM J. Control and Optimization*, 22(2):294–301, 1984.

[75] V. Zeidan. New second-order optimality conditions for variational problems with C^2-Hamiltonians. *SIAM J. Control and Optimization*, 40(2):577–609, 2001.

[76] V. Zeidan and P. Zezza. The conjugate point condition for smooth control sets. *J. of Math. Anal. Applic.*, 132:572–589, 1988.

[77] V. Zeidan and P. Zezza. Necessary conditions for optimal control problems: conjugate points. *SIAM J. Control and Optimization*, 26(3):592–608, 1988.

[78] E. Zeidler. *Nonlinear Functional Analysis and its Applications. Part 1: Fixed-Point Theorems.* Teubner Verlag, Leipzig, 1986.