Second Order Sufficient Conditions for Optimal Control Problems with Non-Unique Minimizers: An Abstract Framework

C. Gavriel* and R. B. Vinter[†]

March 3, 2014

Abstract

Standard second order sufficient conditions in optimal control theory provide not only the information that an extremum is a weak local minimizer, but also tell us that the extremum is locally unique. It follows that such conditions will never cover problems in which the extremum is continuously embedded in a family of constant cost extrema. Such problems arise in periodic control, when the cost is invariant under time translations, in shape optimization, where the cost is invariant under Euclidean transformations (translations and rotations of the extremal shape), and other areas where the domain of the optimization problem does not really comprise elements in a linear space, but rather an equivalence class of such elements. We supply a set of sufficient conditions for minimizers that are not locally unique, tailored to problems of this nature. The sufficient conditions are in the spirit of earlier conditions for ‘non-isolated’ minima, in the context of general infinite dimensional nonlinear programming problems provided by Bonnans, Ioffe and Shapiro, and require coercivity of the second variation in directions orthogonal to the constant cost set. The emphasis in this paper is on the derivation of directly verifiable sufficient conditions for a narrower class of infinite dimensional optimization problems of special interest. The role of the conditions in providing easy-to-use tests of local optimality of a non-isolated minimum, obtained by numerical methods, is illustrated by an example in optimal control.

Keywords: Optimal Control, Second Order Conditions, Second Variation.

**Department of Electrical and Electronic Engineering, Imperial College London, Exhibition Road, London, SW7 2BT, UK, e-mail: christos.gavriel@imperial.ac.uk*

[†]*Department of Electrical and Electronic Engineering, Imperial College London, Exhibition Road, London, SW7 2BT, UK, e-mail: r.vinter@imperial.ac.uk*

1 Introduction

Techniques employed to solve optimal control problems typically generate extremals, that is to say state trajectories satisfying the Maximum Principle or related first order conditions of optimality. Either the relations of the first order conditions are solved to yield an optimal state trajectory by analysis alone, or an optimal state trajectory is computed by means of an algorithm, for which the stopping condition is that first order necessary conditions are satisfied to within a specified tolerance. While it is reasonable to expect in many cases that control functions obtained by solving the first order conditions will be local optimal controls, this is by no means guaranteed. Second order sufficient conditions have an important role then, providing tests of optimality, to be more precise local optimality, for extremals obtained by means of analysis or computation.

This paper is concerned with sufficient conditions of local optimality for optimal control problems characterized as follows: a state trajectory satisfying the constraints belongs to a continuously parameterized family of state trajectories, each satisfying the constraints and all having the same cost. Autonomous optimal control problems with periodic boundary conditions are of this nature, in consequence of the fact that, if we consider the periodic extension of a state trajectory satisfying periodic boundary conditions, then any time translation of the state trajectory is also periodic and has the same cost. But this is just one subclass of such problems. Other examples are provided by shape optimization problems formulated as optimal control problems, where the cost (such as the moment of inertia of the structure) is invariant under Euclidean transformations (translations and rotations). More broadly, examples are to be found among problems where the domain of the optimization problem comprises in effect not single elements in some linear space, but equivalence classes of constant cost elements.

Standard second order sufficient conditions in optimal control can be thought of as elaborations of the requirement, in nonlinear programming, that the Hessian of the Lagrangian is positive definite on the tangent space of the constraint set. A significant feature of such conditions, from the point of view of this paper, is this: when they are satisfied they imply not only local optimality but also local uniqueness. *If follows that tests of local optimality based on classical second order sufficient conditions cannot be used to assess the local optimality of optimal controls that belong to a continuously parameterized family of constant cost controls, because the optimal controls are not locally unique. This is the key issue addressed in this paper.*

There is an extensive literature on second order sufficient conditions for local optimality of control functions, with potential application in situations where optimal control functions are locally unique. See [4], [8], [18], [19] and further references therein. For second order sufficient conditions, suitable for use as tests for local optimality, in numerical optimal control see [6]. By contrast, the theory of second order sufficient conditions for general classes of problems with non-unique optimal controls has received less attention. A notable exception is the work of Speyer and his collaborators [12], [16] who have employed methods of analytical mechanics [9] to derive sufficient conditions covering periodic optimal control problems. These methods, providing criteria for orbit stability expressed in terms of the spectral properties of the monodromy matrix, are specific to periodic problems.

We present a general framework for deriving second order sufficient conditions of local optimality in circumstances when extremals can be continuously embedded in a family of extremals. It is shown how it can be used to generate sufficient conditions in the context of optimal control problems possessing this ‘embedding’ property. The optimal control problems covered include autonomous problems with periodic boundary conditions, but also more general classes of problems where the transformations generating the embeddings are not just time shifts as in periodic optimal control. While the focus is on sufficient conditions in optimal control theory for controlled differential equations, the fact that the techniques are developed in a general, linear space framework points to potential applications involving different kinds of optimization problems, including static problems and those where the underlying dynamic constraints involve functional differential or partial differential equations. Preliminary versions of the conditions derived in this paper were reported in [10], [5].

The abstract framework of Section 5 provides sufficient conditions for local optimality, for a subset of infinite dimensional optimization problems formulated as follows:

$$\underset{x \in Q}{\text{Minimize}} \ f(x) \quad \text{subject to} \quad G(x) \in K, \quad (1)$$

where $f(\cdot) : X \rightarrow \mathbb{R}$ and $G(\cdot) : X \rightarrow Y$ are given functions (for some Banach spaces X and Y) and Q and K are given convex sets. Define $\Phi := \{x \in Q \mid G(x) \in K\}$. Take a closed set $\mathcal{A} \subset \Phi$, on which $f(\cdot)$ takes constant value, and a point $\bar{x} \in \mathcal{A}$. Earlier research has been conducted into ‘non-isolated minima’ in the context of problem (1), the principal findings of which are brought together in Bonnans’ and Shapiro’s monograph [2]. The approach adopted is to seek to characterize situations in which the point $\bar{x}$ has the ‘quadratic growth property’ w.r.t. $\mathcal{A}$, namely: there exists $\epsilon > 0$ and $K > 0$ such that

$$f(x) \geq f(\bar{x}) + K d_{\mathcal{A}}(x)^2 \quad \text{for all } x \in \bar{x} + \epsilon \mathbb{B},$$

in which $d_{\mathcal{A}}(x)$ denotes the distance of the point x from $\mathcal{A}$ w.r.t. the X norm. This literature provides, in particular, sufficient conditions for $\bar{x}$ to have the quadratic growth property w.r.t. $\mathcal{A}$. (See [2], Thm. 3.63). These conditions require coercivity of the second variation on a neighborhood of $\bar{x}$, in directions lying in the ‘approximate cone of critical directions’ which are simultaneously proximal normal vectors to the set S . When X is finite dimensional it is possible to provide verifiable criteria for satisfaction of these sufficient conditions, in which the above cone is replaced by the simpler-to-characterize ‘cone of critical directions’. The task of providing verifiable criteria for infinite dimensional problems is, on the other hand, a challenging one, which is probably best studied in relation to special classes of problems of interest in applications. The sufficient conditions provided in this paper should be viewed in this light: directly verifiable sufficient conditions for a special but important class of infinite dimensional problems arising in optimal control.

The sufficient conditions of [2] relate to a broad class of optimization problems and are included in a general theory combining necessary and sufficient conditions. We provide merely sufficient conditions for a narrower class of problems. The two sets of sufficient conditions are similar in character, since both require coercivity of the second variation in directions having an orthogonality property in relation to $\mathcal{A}$. But detailed comparisons are difficult to make. In this paper, we exploit a special structure of the set $\mathcal{A}$ (motivated by problems in, for example, periodic optimal control) as the image of a continuous function $\alpha \rightarrow x^\alpha$, with $\bar{x} = x^\alpha|_{\alpha=0}$, assumed to have a gradient at the origin, in order to provide sufficient conditions for local optimality, in which ‘orthogonal vectors’ are interpreted as vectors in the nullspace of this gradient mapping. Higher order differentiability of the functions is required only in a restricted sense, relating to increments in a subset of the underlying Hilbert space. We do not demonstrate ‘quadratic growth’; it is, indeed, unlikely that our conditions, in general, imply this stronger property. The sufficient conditions for quadratic growth in [2], which place no restrictions on the closed set $\mathcal{A}$, require higher order differentiability on the whole space and interpret ‘orthogonal vectors’ as proximal normal vectors at points in $\mathcal{A}$ near $\bar{x}$. The application of such sufficient conditions requires checking the coercivity property for every neighbouring proximal normal, many of which may exist when, according to our hypotheses, $\alpha \rightarrow x^\alpha$ is assumed merely to be continuous and differentiable at the origin.

Notation. In k -dimensional Euclidean space, the Euclidean length of a vector x is written $|x|$. The open ball, with center x_0 and radius r , is written $\mathbb{B}^k(x_0, r)$. $\mathbb{B}_Z(x_0, r)$ denotes the open norm ball in a Hilbert space Z , with center $x_0 \in Z$ and radius r . Given Hilbert spaces X and Y , the set of continuous linear maps from X to Y is written $\mathcal{L}(X, Y)$. $I_X \in \mathcal{L}(X, X)$ denotes the identity operator on X . For $T > 0$ and $p = 1$ or 2 , $W^{1,p}([0, T]; \mathbb{R}^n)$ denotes the space of absolutely continuous $\mathbb{R}^n$ valued functions on $[0, T]$ whose derivatives are integrable (if $p = 1$) and square integrable (if $p = 2$).

2 Problem Formulation and Preliminary Concepts

Consider the optimal control problem,

$$(P) \begin{cases} \text{Minimize } J(x(\cdot), u(\cdot)) := c(x(0), x(T)) \\ \text{over } x(\cdot) \in W^{1,1}([0, T]; \mathbb{R}^n) \text{ and measurable functions } u(\cdot) : [0, T] \rightarrow \mathbb{R}^m \\ \text{subject to} \\ \dot{x}(t) = f(t, x(t), u(t)) \text{ a.e. } t \in [0, T] \\ m(x(0), x(T)) = 0. \end{cases}$$

The data for (P) are as follows: a number $T > 0$ and the functions $c : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$, $f : [0, T] \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$, $m : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^k$.

A *control function* is a measurable $\mathbb{R}^m$ -valued function on $[0, T]$. Given a control function $u(\cdot)$, a *state trajectory* $x(\cdot)$ corresponding to $u(\cdot)$ is an a.e. solution to the differential equation $\dot{x}(t) = f(t, x(t), u(t))$ on $[0, T]$, in the space $W^{1,1}([0, T]; \mathbb{R}^n)$. A pair $(x(\cdot), u(\cdot))$ comprising a state trajectory $x(\cdot)$ and an associated control function $u(\cdot)$ is referred to as a *process*. If $x(\cdot)$ satisfies the endpoint constraints $m(x(0), x(T)) = 0$ then the process (x, u) is said to be *admissible*.

An admissible process $(\bar{x}(\cdot), \bar{u}(\cdot))$ that achieves the minimum value of $c(x(0), x(T))$ over all admissible processes is called a *minimizer*, and $\bar{u}(\cdot)$ is called an *optimal control*. A more restrictive notion of optimality is as follows: an admissible process $(\bar{x}(\cdot), \bar{u}(\cdot))$ is said to be a *weak minimizer* if there exists $\epsilon > 0$ such that

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

for all admissible processes $(x(\cdot), u(\cdot))$ satisfying

$$\|x(\cdot) - \bar{x}(\cdot)\|_{L^\infty} \text{ and } \|u(\cdot) - \bar{u}(\cdot)\|_{L^\infty} \leq \epsilon. \quad (2)$$

We say that $(\bar{x}(\cdot), \bar{u}(\cdot))$ is a *weak, locally unique minimizer* if there exists $\epsilon > 0$ such that

$$c(x(0), x(T)) > c(\bar{x}(0), \bar{x}(T))$$

for all admissible processes $(x(\cdot), u(\cdot))$ that are distinct from $(\bar{x}(\cdot), \bar{u}(\cdot))$ and satisfy (2).

We define the Hamiltonian function $H : [0, T] \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m$

$$H(t, x, p, u) := p^T f(t, x, u) \quad (3)$$

and the ‘end-point’ Lagrangian $h : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^k \rightarrow \mathbb{R}$

$$h(x_0, x_1; \nu) = c(x_0, x_1) + \nu^T m(x_0, x_1). \quad (4)$$

An admissible process $(\bar{x}(\cdot), \bar{u}(\cdot))$ is said to be a *weak normal extremal* if it satisfies the conditions of the Pontryagin Maximum Principle in the following restricted form: there exists $p(\cdot) \in W^{1,1}([0, T]; \mathbb{R}^n)$ and $\nu \in \mathbb{R}^k$ such that

$$\begin{aligned} -\dot{p}(t) &= f_x(t, \bar{x}(t), \bar{u}(t))^T p(t) & \text{a.e. } t \in [0, T] \\ H_u(t, \bar{x}(t), p(t), \bar{u}(t)) &= 0 & \text{a.e. } t \in [0, T] \\ [p^T(0), p^T(T)] &= \nabla_{x_0, x_1} h(\bar{x}(0), \bar{x}(T); \nu). \end{aligned}$$

$p(\cdot)$ and ν are the Lagrange multipliers for the extremal $(\bar{x}(0), \bar{u}(\cdot))$. It will be convenient (if slightly ambiguous) also to speak of the weak, normal extremal $((\bar{x}(\cdot), \bar{u}(\cdot)), p(\cdot), \nu)$ in these circumstances.

The descriptors ‘weak’ and ‘normal’ indicate, respectively, that the customary maximization property ‘ $u \rightarrow -H(t, \bar{x}(t), p(t), u)$ is maximized at $\bar{u}(t)$ ’ in the Pontryagin Maximum Principle has been replaced by the weaker, turning point condition $H_u(t, \bar{x}(t), p(t), u = \bar{u}(t)) = 0$, and that the cost multiplier is

set to 1.

It is well-known that any weak, local minimizer $(\bar{x}(\cdot), \bar{u}(\cdot))$ for (P) is a weak normal extremal, under the hypotheses we shall impose below, $(H1) - (H4)$. This fact can be deduced, for example, from [14, Thm. 6.2.3]. Notice that the controllability hypothesis $(H3)$ ensures validity of the conditions in normal form.

3 Second Order Conditions

Take a weak normal extremal $(\bar{x}(\cdot), \bar{u}(\cdot))$, for which the Lagrange multipliers are $p(\cdot)$ and ν . For purposes of comparison, we briefly review standard second order optimality conditions for problem (P) . (See, for example, [4].) These center on the analysis of the ‘accessory problem’, defined as follows. For each $t \in [0, T]$, let

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

$$\begin{bmatrix} Q(t) & D(t) \\ D(t)^T & R(t) \end{bmatrix} := \begin{bmatrix} H_{xx} & H_{xu} \\ H_{xu}^T & H_{uu} \end{bmatrix} ,$$

in which H is as defined in (3). The entries of the block matrix on the right side are evaluated at $(t, \bar{x}(t), p(t), \bar{u}(t))$. Let

$$C := \begin{bmatrix} h_{x_0 x_0} & h_{x_0 x_1} \\ h_{x_0 x_1} & h_{x_1 x_1} \end{bmatrix} ,$$

in which h is as defined in (4) and where the entries of the block matrix on the right side are evaluated at $(\bar{x}(0), \bar{x}(T))$ and ν . Define the function $J_A : W^{1,1}([0, T]; \mathbb{R}^n) \times L^2([0, T]; \mathbb{R}^m) \rightarrow \mathbb{R}$, referred to as the *second variation*, to be

$$J_A(y(\cdot), v(\cdot)) := \frac{1}{2} [y(0)^T y(T)^T] C \begin{bmatrix} y(0) \\ y(T) \end{bmatrix} + \frac{1}{2} \int_0^T y(t)^T Q(t) y(t) + 2y(t)^T D(t) v(t) + v(t)^T R(t) v(t) dt .$$

The accessory problem is

$$(A) \begin{cases} \text{Minimize } J_A(y(\cdot), v(\cdot)) \text{ over } y \in W^{1,1}([0, T]; \mathbb{R}^n) \text{ and } v \in L^2([0, T]; \mathbb{R}^m) \\ \text{subject to} \\ \dot{y} = A(t)y(t) + B(t)v(t) \quad \text{a.e. } t \in [0, T] \\ \nabla_{x_0} m(\bar{x}(0), \bar{x}(T))y(0) + \nabla_{x_1} m(\bar{x}(0), \bar{x}(T))y(T) = 0 . \end{cases}$$

Couples $(y(\cdot), v(\cdot))$ satisfying the differential equation constraint of (A) are called processes for (A) . They are referred to as admissible processes for (A) if, additionally, they satisfy the endpoint constraints of (A) .

Consider the following hypotheses, in which $\theta(\cdot) : [0, \infty) \rightarrow [0, \infty)$ is a modulus of continuity, i.e. a function such that $\lim_{s \downarrow 0} \theta(s) = 0$. (The term ‘modulus of continuity’ will always be understood in this sense below.)

$(H1)$ $c(\cdot, \cdot)$ and $m(\cdot, \cdot)$ are of class C^2 . $f(t, \cdot, \cdot)$ is of class C^2 for all $t \in [0, T]$, and $f(\cdot, \cdot, \cdot)$ and its derivatives with respect to the (x, u) variables, up to order 2, are continuous.

$(H2)$ There exists $\epsilon > 0$ such that

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

$(H3)$ $(A(\cdot), B(\cdot))$ is controllable on $[0, T]$.

$(H4)$ $\bar{u}(\cdot)$ is essentially bounded.

Standard second order conditions, expressed in terms of the second variation, are as follows:

Theorem 3.1. *Take a weak normal extremal for $(\bar{x}(\cdot), \bar{u}(\cdot))$. Assume hypotheses (H1) - (H4) are satisfied.*

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

$$J_A(y(\cdot), v(\cdot)) \geq 0 \quad (5)$$

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

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

$$J_A(y(\cdot), v(\cdot)) \geq \gamma \left(|y(0)|^2 + \int_0^T |v(t)|^2 dt \right) \quad (6)$$

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

Condition (5) and its strengthened form (6) are referred to as positivity and coercivity of the second variation respectively. Thus, positivity of the second variation is a necessary condition and coercivity of the second variation is a sufficient condition, for weak local optimality.

Coercivity of the second variation is often difficult to confirm directly. There is therefore interest in more easily verifiable tests, such as those expressed in terms the Riccati equation associated with the accessory problem

$$\dot{P}(t) + P(t)A(t) + A(t)^T P(t) + Q(t) - (B(t)^T P(t) + D(t)^T)^T R^{-1}(t) (B(t)^T P(t) + D(t)^T) = 0 \quad (7)$$

and Hamilton's system of equations

$$(H) \begin{cases} \dot{y}(t) = A(t)y - B(t)R^{-1}(t) (B(t)^T p(t) + D(t)^T y(t)) \\ -\dot{p}(t) = A(t)^T p(t) + Q(t)y(t) - D(t)R^{-1}(t) (B(t)^T p(t) + D(t)^T y(t)) \end{cases}$$

Let $\{\Phi(s, t) | 0 \leq s, t \leq T\}$ be the transition matrix associated with Hamilton's system of equations (H) for (A) , i.e. for each $s \in [0, T]$ $\Phi(\cdot, s)$ is the unique absolutely continuous, $n \times n$ matrix valued function satisfying:

$$\begin{cases} \frac{d}{dt}\Phi(t, s) = \mathcal{A}\Phi(t, s) \quad \text{a.e. } t \in [0, T] \\ \Phi(s, s) = I, \end{cases}$$

where

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

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

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

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

$$\mathcal{W} := C + \begin{bmatrix} \phi_{22}\phi_{12}^{-1} & \phi_{21} - \phi_{22}\phi_{12}^{-1}\phi_{11} \\ -\phi_{12}^{-1} & \phi_{12}^{-1}\phi_{11} \end{bmatrix}. \quad (8)$$

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

Notice that defining $\mathcal{W}$ requires the invertibility of $\phi_{12}(0, T)$, a sufficient condition for which is given by the following lemma.

Lemma 3.2. *Take a weak normal extremal $(\bar{x}(\cdot), \bar{u}(\cdot))$. Assume (H1) - (H4) are satisfied and there exists a symmetric solution $P(\cdot)$ to (7) on $[0, T]$. Then,*

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

See [3, Section 24]. For completeness, a concise proof is provided in the Appendix.

Proposition 3.3. *Take a weak normal extremal $(\bar{x}(\cdot), \bar{u}(\cdot))$. Assume that (H1) - (H4) are satisfied. Take a positive number γ . Then the following conditions are equivalent:*

(a) (i) *the Riccati equation (7) has a symmetric solution on $[0, T]$,*

$$(ii) \begin{bmatrix} y_0^T & y_1^T \end{bmatrix} \mathcal{W} \begin{bmatrix} y_0 \\ y_1 \end{bmatrix} > \gamma \left\| \begin{bmatrix} y_0 \\ y_1 \end{bmatrix} \right\|^2$$

for all non-zero $y_0, y_1 \in \mathbb{R}^n$ such that $m_{x_0}y_0 + m_{x_1}y_1 = 0$.

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

We deduce from Thm. 3.1 and Prop. 3.3:

Theorem 3.4. *Take a weak normal extremal for $(\bar{x}(\cdot), \bar{u}(\cdot))$. Assume hypotheses (H1) - (H4) are satisfied. Suppose that*

(i) *the Riccati equation (7) has a symmetric solution on $[0, T]$,*

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

$$\begin{bmatrix} y_0^T & y_1^T \end{bmatrix} \mathcal{W} \begin{bmatrix} y_0 \\ y_1 \end{bmatrix} > \gamma \left\| \begin{bmatrix} y_0 \\ y_1 \end{bmatrix} \right\|^2$$

for all $y_0, y_1 \in \mathbb{R}^n$ such that $m_{x_0}y_0 + m_{x_1}y_1 = 0$. (Note that matrix $\mathcal{W}$ is well defined, in view of Lemma 3.2). Then $(\bar{x}(\cdot), \bar{u}(\cdot))$ is a weak locally unique minimizer.

The sufficient conditions of Thm. 3.1(b) and Thm. 3.4 are well-known, though the ‘extra’ assertions that the weak local minimizer is actually locally unique is not usually emphasized. We note that the extra assertion follows from (b) of Prop. 3.3 and the fact that, given $\delta \in (0, \gamma)$ sufficiently small, $\epsilon > 0$ may be found such that for any admissible process of $(x(\cdot), u(\cdot))$ for (P) satisfying

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

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

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

and

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

It follows that

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

which implies local uniqueness. (See the analysis in the Appendix of [1]).

4 Sufficient Conditions for Non-Unique Minimizers

In this section we supply refined second order sufficient conditions, which are satisfied even by weak, local extremals that are not locally unique minimizers in certain cases of of interest. These are cases in which the weak local extremal can be continuously embedded in a family of weak local extremals in the following sense:

Definition 4.1. The weak normal extremal $((\bar{x}(\cdot), \bar{u}(\cdot)), (p(\cdot), \nu))$ is said to be continuously embedded in a family of weak normal extremals

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

in which $\mathcal{A}$ is an open ball, center the origin, in some finite dimensional vector space $\mathbb{R}^p$, such that $(x^0(\cdot), u^0(\cdot)) = (\bar{x}(\cdot), \bar{u}(\cdot))$ and the following conditions are satisfied:

(F1) For each $\alpha \in \mathcal{A}$ $(x^\alpha(\cdot), u^\alpha(\cdot)), (p^\alpha(\cdot), \nu^\alpha)$ is a weak normal extremal such that

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

(F2) $\alpha \rightarrow (x^\alpha(\cdot), u^\alpha(\cdot), p^\alpha(\cdot), \nu^\alpha) : \mathcal{A} \rightarrow L^\infty \times L^\infty \times L^\infty \times \mathbb{R}^l$ is strongly continuous,

(F3) $\alpha \rightarrow (x^\alpha(0), x^\alpha(T)) : \mathcal{A} \rightarrow \mathbb{R}^{2n}$ is of class C^1 .

(F4) Define the $2n \times p$ matrix

$$\Gamma := \begin{bmatrix} \nabla_\alpha x^\alpha(0) \\ \nabla_\alpha x^\alpha(T) \end{bmatrix}_{\alpha=0}. \quad (9)$$

Then the $(p+k) \times 2n$ matrix

$$\begin{bmatrix} \Gamma^T \\ \nabla_{x_0, x_1} m(\bar{x}(0), \bar{x}(T)) \end{bmatrix}$$

has full row rank.

The refined second order sufficient condition, in which matrix $\mathcal{W}$ is as defined in (8), is as follows:

Theorem 4.2. Take a weak normal extremal for $(\bar{x}(\cdot), \bar{u}(\cdot))$. Assume hypotheses (H1) - (H4) are satisfied and that $(\bar{x}(\cdot), \bar{u}(\cdot))$ can be continuously embedded in a family of extremals satisfying the conditions (F1) - (F4) above. Suppose that

(i) the Riccati equation (7) has a symmetric solution on $[0, T]$,

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

$$\begin{bmatrix} y_0^T & y_1^T \end{bmatrix} \mathcal{W} \begin{bmatrix} y_0 \\ y_1 \end{bmatrix} > \gamma \left\| \begin{bmatrix} y_0 \\ y_1 \end{bmatrix} \right\|^2$$

for all non-zero $y_0, y_1 \in \mathbb{R}^n$ satisfying

$$m_{x_0} y_0 + m_{x_1} y_1 = 0 \text{ and } \Gamma^T \begin{bmatrix} y_0 \\ y_1 \end{bmatrix} = 0.$$

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

Notice that the the conclusions of the theorem are that $(\bar{x}(\cdot), \bar{u}(\cdot))$ is a weak local minimizer, not a weak, *locally unique* minimizer. These new sufficient conditions differ from the standard sufficient conditions involving the Riccati equation and Hamilton's system of equations by requiring positivity of the $\mathcal{W}$ matrix *only on the strict linear subspace of points in the linearized endpoint constraint set, satisfying the condition*

$$\Gamma^T \begin{bmatrix} y_0 \\ y_1 \end{bmatrix} = 0.$$

We return to the proof of this theorem after developing related second order conditions in an abstract setting.

5 An Abstract Framework

Consider the optimization problem

$$(AF) \begin{cases} \text{Minimize } F \circ \Pi(z) \text{ over elements } z \in Z \\ \text{subject to} \\ G \circ \Pi(z) = 0. \end{cases}$$

The data comprise: a Hilbert space Z (for which the inner product is written $\langle \cdot, \cdot \rangle_Z$ and norm is written $\|\cdot\|_Z$), functions $\Pi : Z \rightarrow \mathbb{R}^r$, $F : \mathbb{R}^r \rightarrow \mathbb{R}$ and $G : \mathbb{R}^r \rightarrow \mathbb{R}^l$. Here “ $\circ$ ” denotes composition of functions. Points $z \in Z$ are said to be *feasible* if they satisfy the constraint $G \circ \Pi(z) = 0$.

Definition 5.1. Take a convex set $\mathcal{D} \subset \mathbb{B}_Z(0, 1)$, containing the origin. Then a point $\bar{z}$ is said to be a $\mathcal{D}$ -local minimizer if it is feasible for (AF) and, for some $\epsilon > 0$, we have

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

for all feasible points $z \in \bar{z} + \epsilon\mathcal{D}$.

An element $\bar{z} \in Z$ is said to be a *normal extremal* if $\bar{z}$ is feasible and there exists an element $\bar{\nu} \in \mathbb{R}^l$ such that, writing

$$h(z) = (F + \bar{\nu}^T G) \circ \Pi(z),$$

we have

$$Dh(\bar{z}) = 0.$$

The element $\bar{\nu}$ is referred to as the multiplier (associated with $\bar{z}$). We shall write, briefly, ‘ $(\bar{z}, \bar{\nu})$ is a normal extremal’ to signify that $\bar{z}$ is a normal extremal with associated multiplier $\bar{\nu}$.

Definition 5.2. Given a normal extremal $(\bar{z}, \bar{\nu})$, we shall say $(\bar{z}, \bar{\nu})$ is embedded in a family of extremals $\{(z^\alpha, \nu^\alpha) \mid \alpha \in \mathbb{B}^p(0, \bar{\alpha})\}$, in which $\bar{\alpha} > 0$, if $\alpha \rightarrow (z^\alpha, \nu^\alpha)$ is a mapping from $\mathbb{B}^p(0, \bar{\alpha})$ to $Z \times \mathbb{R}^l$ such that $(\bar{z}, \bar{\nu}) = (z^0, \nu^0)$ and for, each $\alpha \in \mathbb{B}^p(0, \bar{\alpha})$,

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

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

$$h^\alpha(z) := (F + (\nu^\alpha)^T G) \circ \Pi(z).$$

Given a convex set $\mathcal{D} \subset \mathbb{B}_Z(0, 1)$ containing the origin and an extremal $(\bar{z}, \bar{\nu})$ embedded in a family of extremals $\{(z^\alpha, \nu^\alpha) \mid \alpha \in \mathbb{B}^p(0, \bar{\alpha})\}$, we shall impose the following hypotheses which give, in particular, precise meaning to the derivatives $Dh^\alpha(z^\alpha)$ and $Dh(\bar{z}) := Dh^0(z^0)$ in the definitions above.

(AH1) The functions $F(\cdot)$ and $G(\cdot)$ are of class C^2 .

(AH2) (Continuity of the embedding) There exists a modulus of continuity $\eta(\cdot) : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ such that $z^\alpha \in \bar{z} + \eta(\alpha')\mathcal{D}$ for all $\alpha' \in [0, \bar{\alpha}]$, $\alpha \in \mathbb{B}^p(0, \alpha')$.

(AH3) (Endpoint differentiability) The mapping $\alpha \rightarrow \Pi(z^\alpha) : \mathbb{B}^p(0, \bar{\alpha}) \rightarrow \mathbb{R}^r$ is continuous and differentiable at $\alpha = 0$. Define the $r \times p$ matrix

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

There exists a modulus of continuity $\theta(\cdot)$ and $K > 0$ with the following property:

for each $\alpha \in \mathbb{B}^p(0, \bar{\alpha})$ there exist $Dh^\alpha(z^\alpha) \in \mathcal{L}(Z, \mathbb{R})$, $D\Pi^\alpha(z^\alpha) \in \mathcal{L}(Z, \mathbb{R}^r)$ and a self-adjoint $D^2h^\alpha(z^\alpha) \in \mathcal{L}(Z, Z)$ such that

$$\begin{aligned} \|Dh^\alpha(z^\alpha)\|_{\mathcal{L}(X, X)} &\leq K, \\ |h(z) - h(z^\alpha) - Dh^\alpha(z^\alpha)(z - z^\alpha) - \frac{1}{2}\langle z - z^\alpha, D^2h^\alpha(z^\alpha)(z - z^\alpha) \rangle| &\leq \theta(\epsilon)\|z - z^\alpha\|_Z^2, \\ |\Pi(z) - \Pi(z^\alpha) - D\Pi^\alpha(z^\alpha)(z - z^\alpha)| &\leq \theta(\epsilon)\|z - z^\alpha\|_Z \end{aligned}$$

for all $\epsilon > 0$ and $z \in z^\alpha + \epsilon \mathcal{D}$.

(AH4) (Constraint regularity)

- (a) The set of r -vectors comprising the rows of the matrices Γ^T and $\nabla G(\Pi(\bar{z}))$ are linearly independent.
- (b) There exists $k > 0$ such that, for each $y \in Z$ and $\alpha \in \mathbb{B}^p(0, \bar{\alpha})$, an element $y' \in Z$ can be found such that

$$\Gamma^T D\Pi^\alpha(z^\alpha)y' = 0 \text{ and } \nabla G(\Pi(z^\alpha)) D\Pi^\alpha(z^\alpha)y' = 0$$

and

$$\|y' - y\|_Z \leq k \left(|\Gamma^T D\Pi^\alpha(z^\alpha)y| + |\nabla G(\Pi(z^\alpha)) D\Pi^\alpha(z^\alpha)y| \right).$$

(AH5) There exists $\bar{\delta} > 0$, $\gamma > 0$ and an $(r \times r)$ -matrix valued function $(\alpha, \delta) \rightarrow \mathcal{W}^{\alpha, \delta}$ on $\mathbb{B}^p(0, \bar{\alpha}) \times [0, \bar{\delta}]$, which is continuous at $(0, 0)$ and satisfies the conditions

- (i) for every $\alpha \in \mathbb{B}^p(0, \bar{\alpha})$, $\delta \in [0, \bar{\delta}]$ and $\xi \in \mathbb{R}^r$ we have

$$\inf \{ \langle z, (D^2 h^\alpha(z^\alpha) - \delta I_Z)z \rangle \mid z \in Z, D\Pi^\alpha(z^\alpha)z = \xi \} \geq \xi^T \mathcal{W}^{\alpha, \delta} \xi$$

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

$$\Gamma^T \xi = 0 \quad \text{and} \quad \nabla G(\Pi(\bar{z})) \xi = 0.$$

Theorem 5.3. *Consider the optimization problem (AF). Take a normal extremal $(\bar{z}, \bar{\nu}) \in Z \times \mathbb{R}^l$ that is embedded in a family of extremals $\{(z^\alpha, \nu^\alpha) \mid \alpha \in \mathbb{B}^p(0, \bar{\alpha})\}$, and a set $\mathcal{D} \subset \mathbb{B}_Z(0, 1)$. Assume (AH1) – (AH5). Then $\bar{z}$ is a $\mathcal{D}$ -local minimizer.*

Comments.

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

$$\xi^T \mathcal{W} \xi \geq \gamma |\xi|^2 \quad \text{for all } \xi \in \mathbb{R}^r \text{ such that } \nabla G(\Pi(\bar{z})) \xi = 0.$$

where $\mathcal{W}$ is the unique symmetric matrix such that

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

Hypothesis (AH5) of Thm. 5.3 imposes ‘stability’ requirements on the matrix $\mathcal{W}$ relative to perturbations of the extremal (condition (i)) and, more significantly, condition (ii) that requires the positivity of $\mathcal{W}$ merely on the subset of $\{\xi \mid \nabla G(\Pi(\bar{z})) \xi = 0\}$ comprising elements orthogonal to the columns of the matrix Γ . Condition (AH5)(ii) is satisfied in some cases of interest, when $\mathcal{W}$ fails to be positive definite on the whole of the subspace $\{\xi \mid \nabla G(\Pi(\bar{z})) \xi = 0\}$ itself.

- (ii) Consider Hypothesis (AH4)(b) concerning the constraints

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

on elements $y \in Z$. This can be interpreted as a metric regularity-type hypothesis, requiring 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 , as measured by:

$$\left(|\Gamma^T D\Pi^\alpha y| + |\nabla G(\Pi(z^\alpha)) D\Pi^\alpha y| \right).$$

The following sufficient condition for satisfaction of Hypothesis (AH4)(b), which involves the right inverse of the linear map $D\Pi^\alpha(z^\alpha)$, is often convenient for optimal control applications, because it can be more directly linked to verifiable, controllability-type condition on the data for the underlying control system.

Proposition 5.4. *Consider a normal extremal $(\bar{z}, \bar{\nu}) \in Z \times \mathbb{R}^l$ for the optimization problem (AF) that is embedded in a family of extremals $\{(z^\alpha, \nu^\alpha) \mid \alpha \in \mathbb{B}^p(0, \bar{\alpha})\}$. Assume (AH3) and (AH4)(a). Then (AH4)(b) is satisfied (following, if necessary, a reduction of the size of $\bar{\alpha}$), under the following conditions:*

(AH3)'(b) : *For each $\alpha \in \mathbb{B}^p(0, \bar{\alpha})$ the linear map $D\Pi^\alpha(z^\alpha)$ has a right inverse $(D\Pi^\alpha(z^\alpha))^\dagger \in \mathcal{L}(\mathbb{R}^r; Z)$, i.e.*

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

$\alpha \rightarrow D\Pi^\alpha(z^\alpha)^\dagger$ is continuous w.r.t. the uniform operator topology on $\mathcal{L}(\mathbb{R}^r; Z)$ and there exists $k > 0$ such that

$$\|D\Pi^\alpha(z^\alpha)^\dagger\| \leq k \quad \text{for all } \alpha \in \mathbb{B}^p(0, \bar{\alpha}) .$$

Proof. Define

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

Since $\alpha \rightarrow \Pi(z^\alpha)$ is continuous and V^0 has full row rank, the matrix inverse $(V^\alpha V^{\alpha T})^{-1}$ exists and is bounded on $\mathbb{B}^p(0, \alpha_1)$, for some $\alpha_1 \in (0, \bar{\alpha})$. Take any $y \in Z$ and $\alpha \in \mathbb{B}^p(0, \alpha_1)$. Define

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

Then

$$V^\alpha D\Pi^\alpha(z^\alpha) y' = V^\alpha D\Pi^\alpha(z^\alpha) y - V^\alpha D\Pi^\alpha(z^\alpha) y = 0 .$$

It follows that

$$\Gamma^T D\Pi^\alpha(z^\alpha) y' = 0, \nabla G(\Pi(z^\alpha)) D\Pi^\alpha(z^\alpha) y' = 0$$

and

$$\|y' - y\|_Z \leq k \left(|\Gamma^T D\Pi^\alpha(z^\alpha) y| + |\nabla G(\Pi(z^\alpha)) D\Pi^\alpha(z^\alpha) y| \right) ,$$

for k a uniform bound on $\|(D\Pi^\alpha)^\dagger V^{\alpha T} (V^\alpha V^{\alpha T})^{-1}\|_Z$ over $\alpha \in \mathbb{B}^p(0, \alpha_1)$. We have shown that condition (AH4)(b) is satisfied, for $\bar{\alpha}$ replaced by α_1 . $\square$

Proof of Thm. 5.3 In the sequel we shall omit the argument z^α in expressions involving Π^α and h^α , writing $D\Pi^\alpha(z^\alpha)$ briefly as $D\Pi^\alpha$, etc. Take a normal extremal $(\bar{z}, \bar{\nu}) \in Z \times \mathbb{R}^l$ that is embedded in a family of extremals $\{(z^\alpha, \nu^\alpha) \mid \alpha \in \mathbb{B}^p(0, \bar{\alpha})\}$, and a set $\mathcal{D} \subset \mathbb{B}_Z(0, 1)$. Assume (AH1)–(AH5).

The following lemma summarizes a key step in the proof of Thm. 5.3. The lemma asserts that, given a point z close to $\bar{z}$, we can adjust α so that the deviation of the boundary values of z from those of z^α , namely $\Pi(z) - \Pi(z^\alpha)$, is normal to the tangent space of $\{\Pi(z^\alpha) \mid \alpha \in \mathbb{B}^p(0, \bar{\alpha})\}$ at $\alpha = 0$.

Lemma 5.5. *For each $\alpha' \in (0, \bar{\alpha})$, $\epsilon' > 0$ can be found with the property: given any $z \in \bar{z} + \epsilon' \mathcal{D}$, there exists $\alpha \in \mathbb{B}^p(0, \alpha')$ such that*

$$\Gamma^T(\Pi(z) - \Pi(z^\alpha)) = 0 .$$

Proof. Fix $\alpha' \in (0, \bar{\alpha})$. Take any $\alpha \in \mathbb{B}^p(0, \alpha')$. We know from (AH3) that

$$|\Pi(z^\alpha) - \Pi(\bar{z}) - \Gamma \alpha| \leq \eta_1(\alpha') |\alpha| , \tag{10}$$

for some modulus of continuity $\eta_1(\cdot)$. Define

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

Notice that, $M(\bar{z}, \cdot)$ is differentiable at $\alpha = 0$ and

$$\frac{d}{d\alpha} M(\bar{z}, 0) = -\Gamma^T \frac{d}{d\alpha} \Pi(z^\alpha)|_{\alpha=0} = -\Gamma^T \Gamma$$

is a nonsingular matrix, since Γ has linearly independent columns. Define also the mapping $\mathcal{M} : Z \times \mathbb{R}^p \rightarrow \mathbb{R}^p$, where

$$\mathcal{M}(z, \xi) := M(z, \alpha(z, \xi)) + \xi$$

in which

$$\alpha(z, \xi) := -(\Gamma^T \Gamma)^{-1} [\Gamma^T D\Pi^0(z - \bar{z}) + \xi] .$$

Note that

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

Take a number $\epsilon' > 0$ (we shall set its value presently), and consider an arbitrary point $z \in \bar{z} + \epsilon' \mathcal{D}$. Since $\mathcal{D} \subset \mathbb{B}_Z(0, 1)$ we have that $\|z - \bar{z}\|_Z \leq \epsilon'$. It follows from (10) and (AH3) that

$$\begin{aligned} |\mathcal{M}(z, \xi)| &= |\mathcal{M}(z, \xi) - \mathcal{M}(\bar{z}, 0)| \\ &= |\Gamma^T \left((\Pi(z) - \Pi(\bar{z})) - (\Pi(z^{\alpha(z, \xi)}) - \Pi(\bar{z})) \right) + \xi| \\ &\leq |\Gamma^T (D\Pi^0(z - \bar{z}) - \Gamma(\Gamma^T \Gamma)^{-1} (\Gamma^T D\Pi^0(z - \bar{z}) + \xi)) + \xi| + e(z, \alpha) \\ &= 0 + e(z, \alpha) \end{aligned}$$

where

$$e(z, \xi) = |\Gamma| \left[\theta(\|z - \bar{z}\|_Z) \|z - \bar{z}\|_Z + \eta_1(|\alpha(z, \xi)|) |\alpha(z, \xi)| \right] .$$

But, for constants k_1 and k_2

$$|\alpha(z, \xi)| \leq k_1 \|z - \bar{z}\|_Z + k_2 |\xi| .$$

It follows that

$$|\mathcal{M}(z, \xi)| \leq |\Gamma| (\theta(\epsilon') \epsilon' + \eta_1(k_1 \epsilon' + k_2 |\xi|)) (k_1 \epsilon' + k_2 |\xi|) .$$

It follows that we can choose $\sigma > 0$ and $\epsilon' > 0$ such that, for all $z \in \bar{z} + \epsilon' \mathcal{D}$ and $\xi \in \mathbb{B}(0, \sigma)$,

$$|\mathcal{M}(z, \xi)| \leq \sigma .$$

We see that, for fixed $z \in \bar{z} + \epsilon' \mathcal{D}$, $\mathcal{M}(z, \cdot)$ maps the closed ball $\mathbb{B}^p(0, \sigma)$ into itself. Furthermore, for every $\xi \in \mathbb{B}^p(0, \sigma)$ we have that $\alpha(z, \xi) \in \mathbb{B}^p(0, \alpha')$. In view of (AH3), this map is continuous. Invoking the Schauder Fixed Point Theorem [11] we deduce that the map has a fixed point $\bar{\xi}$ in $\mathbb{B}^p(0, \sigma)$. But $\bar{\xi} = \mathcal{M}(z, \bar{\xi})$ implies $M(z, \bar{\xi}) = 0$, i.e.

$$\Gamma^T (\Pi(z) - \Pi(z^\alpha)) = 0 ,$$

where $\alpha = \alpha(z, \bar{\xi}) \in \mathbb{B}^p(0, \alpha')$. This confirms the assertions of the lemma. $\square$

The following lemma, the proof of which is based on a simple sequential compactness analysis and which we therefore omit, is also required.

Lemma 5.6. *Take $\mathcal{N}$ a neighborhood of the origin $\mathbb{R}^r$ and $n \times n$ and $m \times n$ matrix valued functions $\beta \mapsto Q^\beta$, $\beta \mapsto D^\beta$ on $\mathcal{N}$, 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 neighborhood $\mathcal{N}' \subset \mathcal{N}$ 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. $D^\beta x = 0$.

We are now ready to complete the proof of the theorem. Take any $\alpha' \in (0, \bar{\alpha}]$. Let ϵ' be as in Lemma 5.5. (ϵ' will depend on α'). Take $\epsilon \in (0, \epsilon']$ and consider an arbitrary feasible point $z \in \bar{z} + \epsilon\mathcal{D}$. (Notice that, because $\mathcal{D} \subset \mathbb{B}_Z(0, 1)$, $\|z^\alpha - \bar{z}\|_Z \leq \epsilon$). Then $F \circ \Pi(\bar{z}) = F \circ \Pi(z^\alpha)$ and $G \circ \Pi(\bar{z}) = G \circ \Pi(z^\alpha) = 0$. Taking account also of (AH2) and (AH3), we conclude that

$$\begin{aligned}
F \circ \Pi(z) - F \circ \Pi(\bar{z}) &= F \circ \Pi(z) - F \circ \Pi(z^\alpha) \\
&= F \circ \Pi(z) - F \circ \Pi(z^\alpha) + \nu^\alpha \cdot [G \circ \Pi(z) - G \circ \Pi(z^\alpha)] \\
&= h(z) - h(z^\alpha) \\
&\geq Dh^\alpha(z - z^\alpha) + \frac{1}{2} \langle (z - z^\alpha), D^2h^\alpha(z - z^\alpha) \rangle - \theta(\epsilon + \eta(\alpha')) \|z - z^\alpha\|_Z^2 \\
&= 0 + \frac{1}{2} \langle (z - z^\alpha), D^2h^\alpha(z - z^\alpha) \rangle - \theta(\epsilon + \eta(\alpha')) \|z - z^\alpha\|_Z^2.
\end{aligned} \tag{11}$$

By Lemma 5.5 there exists $\alpha \in \mathbb{B}^p(0, \alpha')$ such that

$$\Gamma^T(\Pi(z) - \Pi(z^\alpha)) = 0. \tag{12}$$

Since z and z^α are both feasible,

$$G \circ \Pi(z) - G \circ \Pi(z^\alpha) = 0. \tag{13}$$

It follows from (12), (13) and (AH3) that

$$|\Gamma^T D\Pi^\alpha(z - z^\alpha)| + |\nabla G(\Pi(z^\alpha)) D\Pi^\alpha(z - z^\alpha)| \leq \theta_2(\epsilon + \eta(\alpha')) \|z - z^\alpha\|,$$

for some continuity modulus $\theta_2(\cdot)$. By (AH4) there exists $z' \in Z$ such that

$$\Gamma^T D\Pi^\alpha z' = 0 \text{ and } \nabla G(\Pi(z^\alpha)) D\Pi^\alpha z' = 0.$$

Furthermore,

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

for some continuity modulus $\theta_3(\cdot)$. By reducing the magnitude of α' and ϵ , while continuing to ensure that $\epsilon \in (0, \epsilon']$, it can be arranged that $\theta_3(\epsilon + \eta(\alpha')) < 1$. Then, by the triangle inequality,

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

and so

$$\|z' - (z - z^\alpha)\|_Z \leq \frac{\theta_3(\epsilon + \eta(\alpha'))}{1 - \theta_3(\epsilon + \eta(\alpha'))} \|z'\|_Z. \tag{14}$$

Writing $\Delta z := z - z^\alpha$, we can express (11) as

$$F \circ \Pi(z) - F \circ \Pi(\bar{z}) \geq \frac{1}{2} \langle \Delta z, D^2h^\alpha \Delta z \rangle - \theta(\epsilon + \eta(\alpha')) \|\Delta z\|_Z^2. \tag{15}$$

In view of the identity

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

we have

$$\langle \Delta z, D^2h^\alpha \Delta z \rangle \geq \langle z', D^2h^\alpha z' \rangle - 2 \|D^2h^\alpha\|_{\mathcal{L}(Z, Z)} \|z'\|_Z \|\Delta z - z'\|_Z - \|D^2h^\alpha\|_Z \|\Delta z - z'\|_Z^2.$$

But then, by (14),

$$\langle \Delta z, D^2h^\alpha \Delta z \rangle \geq \langle z', D^2h^\alpha z' \rangle - \theta_4(\epsilon + \eta(\alpha')) K \|z'\|_Z^2, \tag{16}$$

where $\theta_4(\cdot)$ is the continuity modulus

$$\theta_4(\epsilon + \eta(\alpha')) = \frac{2\theta_3(\epsilon + \eta(\alpha'))}{1 - \theta_3(\epsilon + \eta(\alpha'))} + \left(\frac{\theta_3(\epsilon + \eta(\alpha'))}{1 - \theta_3(\epsilon + \eta(\alpha'))} \right)^2$$

and K is the uniform bound on $\|D^2h^\alpha\|_{\mathcal{L}(Z,Z)}$ of hypothesis (AH3). From (14) we also deduce that

$$\|\Delta z\|_Z^2 \leq (\|\Delta z - z'\|_Z + \|z'\|_Z)^2 \leq \left(\frac{\theta_3(\epsilon + \eta(\alpha'))}{1 - \theta_3(\epsilon + \eta(\alpha'))} + 1 \right)^2 \|z'\|_Z^2.$$

Combining this relation with (15) and (16) yields

$$F \circ \Pi(z) - F \circ \Pi(\bar{z}) \geq \langle z', D^2h^\alpha z' \rangle - \frac{1}{2}q(\epsilon, \alpha')\|z'\|_Z^2 \quad (17)$$

in which

$$q(\epsilon, \alpha') = \frac{1}{2}\theta_4(\epsilon + \eta(\alpha'))K + \theta(\epsilon + \eta(\alpha')) \left(\frac{1}{1 - \theta_3(\epsilon + \eta(\alpha'))} \right)^2.$$

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

$$\begin{aligned} \langle z', D^2h^\alpha z' \rangle - q(\epsilon, \alpha')\|z'\|_Z^2 &= \langle z', (D^2h^\alpha - q(\epsilon, \alpha')I)z' \rangle \\ &\geq (D\Pi^\alpha z')^T \mathcal{W}^{\alpha, q(\epsilon, \alpha')} (D\Pi^\alpha z'), \end{aligned} \quad (18)$$

where $\mathcal{W}^{\alpha, q(\epsilon, \alpha')}$ is as in hypothesis (AH5). Take any $\gamma' \in (0, \gamma)$, where γ is the constant in condition (AH5)(ii). We know that $\alpha' \in (0, \bar{\alpha}]$, $\alpha \in \mathbb{B}^p(0, \alpha')$ and $D\Pi^\alpha z'$ satisfies the constraint

$$\Gamma^T D\Pi^\alpha z' = 0 \text{ and } \nabla G(\Pi(z^\alpha)) D\Pi^\alpha z' = 0.$$

By (AH2) and (AH3), $G(\Pi(z^\alpha)) \rightarrow G(\Pi(\bar{z}))$ as $\alpha' \rightarrow 0$. We conclude now from Lemma 5.6 that, following a further reduction in the size of ϵ and α' (if required), we have

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

Then, from (17) and (18),

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

Since z was an arbitrary feasible point in $\bar{z} + \epsilon\mathcal{D}$, we have proved that $\bar{z}$ is a $\mathcal{D}$ -local minimizer for (AF). $\square$

6 Proof of Theorem 4.2

We now make use of the abstract framework above to prove the sufficient conditions for weak local optimality for the optimal control problem, stated as Thm. 4.2. Take a weak normal extremal $(\bar{x}(\cdot), \bar{u}(\cdot))$ for (P). Assume that $(\bar{x}(\cdot), \bar{u}(\cdot))$ is continuously embedded in a family of weak normal extremals

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

satisfying conditions (F1) - (F4) and that hypotheses (H1) - (H4) are valid.

For $(x_0, u(\cdot)) \in \mathbb{R}^n \times L^\infty([0, T]; \mathbb{R}^m)$ let $\xi(t; x_0, u(\cdot))$ denote the solution on $[0, T]$ of

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

As before, we write

$$(A^\alpha(t), B^\alpha(t)) = (f_x(t, x^\alpha(t), u^\alpha(t)), f_u(t, x^\alpha(t), u^\alpha(t))).$$

(Here, and elsewhere, the absence of the superscript α indicates that $\alpha = 0$. Thus $(A(t), B(t)) = (A^\alpha(t), B^\alpha(t))|_{\alpha=0}$.)

For given $(y_0, v(\cdot)) \in \mathbb{R}^n \times L^2$ let $\xi_L^\alpha(t; y_0, v(\cdot))$ denote the solution on $[0, T]$ of

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

We now make the appropriate identifications of functions and sets to match the optimal control problem (P) to the abstract framework of Section 5. Accordingly, set

$$\begin{aligned} Z &= \mathbb{R}^n \times L^2([0, T]; \mathbb{R}^m) & \mathcal{D} &= \{(x_0, u(\cdot)) \in Z \mid |x_0| + \|u\|_{L^\infty} \leq 1\} \\ \Pi(x_0, u(\cdot)) &= (x_0, \xi(T; x_0, u(\cdot))) & F(x_0, x_1) &= c(x_0, x_1) \\ G(x_0, x_1) &= m(x_0, x_1) & h^\alpha(x_0, u(\cdot)) &= \left(g + (\nu^\alpha)^T m\right)(x_0, \xi(T; x_0, u(\cdot))). \end{aligned}$$

Somewhat lengthy but routine calculations, taking into account the conditions of the Maximum Principle, yield the following derivative formulae

$$\begin{aligned} \Gamma &= \frac{d}{d\alpha} \left[x^\alpha(0) \right]_{\alpha=0}, \quad D\Pi^\alpha(x_0^\alpha, u^\alpha(\cdot))(y_0, v(\cdot)) = \begin{bmatrix} y_0 \\ y^\alpha(T) \end{bmatrix}, \\ Dh^\alpha(x_0^\alpha, u^\alpha(\cdot))(y_0, v(\cdot)) &= \\ &= (\nabla(g + \nu^\alpha) + (p^\alpha(T), -p^\alpha(0))) \cdot (y(0), y^\alpha(T)) + \int_0^T \nabla_u H(t, x^\alpha(t), p^\alpha(t), u^\alpha(t))v(t) dt, \end{aligned}$$

and the bilinear form defining the self-adjoint operator $D^2h^\alpha(x_0^\alpha, u^\alpha(\cdot))$ is

$$\begin{aligned} \langle (y_0, v(\cdot)), D^2h^\alpha(x_0^\alpha, u^\alpha(\cdot))(y_0, v(\cdot)) \rangle &= \begin{bmatrix} y_0^T & y^\alpha(T)^T \end{bmatrix} C^\alpha \begin{bmatrix} y_0 \\ y^\alpha(T) \end{bmatrix} + \\ &+ \int_0^T \begin{bmatrix} y^\alpha(t)^T & v(t)^T \end{bmatrix} \nabla_{x,u}^2 H(t, x^\alpha(t), p^\alpha(t), u^\alpha(t)) \begin{bmatrix} y^\alpha(t) \\ v(t) \end{bmatrix} dt \end{aligned}$$

in which $C^\alpha := \nabla^2(g + (\nu^\alpha)^T m)$ and $y^\alpha(T)$ is the vector $y^\alpha(T) = \xi_L^\alpha(T; y_0, v(\cdot))$. Derivatives of $(g + (\nu^\alpha)^T m)$ are evaluated at $(y_0, y^\alpha(T))$. Notice that $Dh^\alpha(x_0^\alpha, u^\alpha(\cdot))(y_0, v(\cdot)) = 0$ by the extremality property.

Hypotheses $(H1)$ - $(H4)$ ensure that, following conversion to the abstract framework, hypotheses $(AH1)$ - $(AH3)$ are satisfied. We note, in particular, that the validity of $(AH4)$ ('constraint regularity') can be deduced from the controllability hypothesis $(H4)$ with the help of the characterization of the constraint regularity property provided by Prop. 5.5. Here we make use of the following formula for the right inverse $D\Pi^\alpha(x^\alpha(0), u^\alpha(\cdot))^\dagger$ of $D\Pi^\alpha(x^\alpha(0), u^\alpha(\cdot))$:

$$D\Pi^\alpha(x^\alpha(0), u^\alpha(\cdot))^\dagger(y_0, y_1) = (y_0, v(t) \equiv B^\alpha(t)^T \Phi_{A^\alpha}(T, t)^T W^\alpha(0, T)^{-1} (y_1 - \Phi_{A^\alpha}(T, 0)) y_0),$$

in which $\Phi_{A^\alpha}(\cdot, \cdot)$ is the fundamental matrix for $\dot{y}(t) = A^\alpha(t)y(t)$ and $W^\alpha(0, T)$ is the controllability Grammian

$$W^\alpha(0, T) = \int_0^T \Phi_{A^\alpha}(T, s) B^\alpha(s) B^\alpha(s)^T \Phi_{A^\alpha}(T, s)^T ds.$$

(Notice that $W^\alpha(0, T)$ is invertible at $\alpha = 0$, by the controllability assumption, and therefore invertible

for all α in some ball about the origin by continuous dependence.)

With the above identifications we shall show that, for all (α, ϵ) confined to a sufficiently small ball about the origin, there exists a matrix $\mathcal{W}^{\alpha, \epsilon}$ meeting the requirements of the final hypothesis (AH5) of Thm. 5.3. This will permit us to conclude that $(\bar{x}_0, \bar{u}(\cdot))$ is a $\mathcal{D}$ -local minimizer for (P) formulated as a special case of (AF). It will then follow that $(\bar{x}(\cdot), \bar{u}(\cdot))$ is a weak local minimizer and

$$\begin{aligned} \{ \text{feasible processes } (x(\cdot), u(\cdot)) \mid \|x(\cdot) - \bar{x}(\cdot)\|_{L^\infty}, \|u(\cdot) - \bar{u}(\cdot)\|_{L^\infty} \leq \delta \} \subset \\ \{ (x(\cdot), u(\cdot)) \mid |x(0) - \bar{x}(0)|, \|u(\cdot) - \bar{u}(\cdot)\|_{L^\infty} \leq \delta \}, \end{aligned}$$

for $\delta > 0$ sufficiently small. This is what we set out to prove.

Consider parameters $\bar{\alpha} > 0$ and $\bar{\epsilon} > 0$ (the values of which will be chosen presently) and take any $\alpha \in \mathbb{B}^p(0, \bar{\alpha}) \cap \mathcal{A}$ and $\epsilon \in [0, \bar{\epsilon}]$. For arbitrary $\xi = \begin{bmatrix} y_0 \\ y_1 \end{bmatrix} \in \mathbb{R}^{2n}$ we have

$$\inf \left\{ \langle z, (D^2 h^\alpha(z^\alpha) - \epsilon I)z \rangle \mid z \in Z, D\Pi^\alpha(z^\alpha)z = \xi \right\} = \inf \left\{ J_A^{\alpha, \epsilon}(y(\cdot), v(\cdot)) \mid (y(0), y(T)) = (y_0, y_1) \right\}, \quad (19)$$

where

$$\begin{aligned} J_A^{\alpha, \epsilon}(y(\cdot), v(\cdot)) := & \frac{1}{2} \begin{bmatrix} y_0^T & y^\alpha(T)^T \end{bmatrix} C^{\alpha, \epsilon} \begin{bmatrix} y_0 \\ y^\alpha(T) \end{bmatrix} + \\ & \frac{1}{2} \int_0^T y(t)^T Q(t)y(t) + 2y(t)^T D(t)v(t) + v(t)^T R^{\alpha, \epsilon}(t)v(t) dt, \end{aligned}$$

in which

$$\begin{aligned} C^{\alpha, \epsilon} &:= C^\alpha - \begin{bmatrix} \epsilon I_{n \times n} & 0_{n \times m} \\ 0_{m \times n} & 0_{m \times m} \end{bmatrix} \\ R^{\alpha, \epsilon}(t) &:= \nabla_u^2 H(t, x^\alpha(t), p^\alpha(t), u^\alpha(t)) - \epsilon I_{m \times m}, \end{aligned}$$

and $\inf \{ J_A^{\alpha, \epsilon}(y(\cdot), v(\cdot)) \mid (y(0), y(T)) = (y_0, y_1) \}$ denotes the infimum cost of the optimal control problem

$$(A^{\alpha, \epsilon}) \begin{cases} \text{Minimize } J_A^{\alpha, \epsilon}(y(\cdot), v(\cdot)) \text{ over } (y(\cdot), v(\cdot)) \in W^{1,1} \times L^2 \\ \text{subject to} \\ \dot{y}(t) = A^\alpha(t)y(t) + B^\alpha(t)v(t) \quad \text{a.e.} \\ (y(0), y(T)) = (y_0, y_1). \end{cases}$$

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

$$(R^{\alpha, \epsilon})^{-1}(t) \leq K.$$

Consider now the modified Riccati equation and Hamiltonian system of equations

$$\begin{aligned} \dot{P}(t) + P(t)A^\alpha(t) + A^\alpha(t)^T P(t) + Q^\alpha(t) - \\ (B^\alpha(t)^T P(t) + D^\alpha(t)^T)^T (R^{\alpha, \epsilon})^{-1}(t) (B^\alpha(t)^T P(t) + D^\alpha(t)^T) = 0 \end{aligned} \quad (20)$$

and the modified Hamilton's system of equations

$$\begin{cases} \dot{y}(t) = A^\alpha(t)y - B^\alpha(t)(R^{\alpha, \epsilon})^{-1}(t) (B^\alpha(t)^T p(t) + D^\alpha(t)^T y(t)) \\ -\dot{p}(t) = A^\alpha(t)^T p(t) + Q^\alpha(t)y(t) - D^\alpha(t)(R^{\alpha, \epsilon})^{-1}(t) (B^\alpha(t)^T p(t) + D^\alpha(t)^T y(t)). \end{cases} \quad (21)$$

Let $\Phi^{\alpha, \epsilon}(\cdot, \cdot)$ be the transition matrix for (21). We know, by assumption, that (20) has a (symmetric) solution P on $[0, T]$ and that $\phi_{12}^{\alpha, \epsilon}(0, T)$ is invertible, in the case $(\alpha, \epsilon) = (0, 0)$. 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 $k \geq 0$ (see, e.g. [1]). By choosing k sufficiently large, we can arrange that both

(i) there exists a solution to $P^{\alpha, \epsilon}(t)$ to (20) on $[0, T]$ and

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

for $(\alpha, \epsilon) = (0, 0)$. But then, by ‘continuous dependence’, (i) and (ii) remain true for all choices of α and ϵ , provided $\bar{\alpha}$ and $\bar{\epsilon}$ are sufficiently small. Furthermore, we can choose $P^{\alpha, \epsilon}(t)$ to depend continuously on (α, ϵ) , uniformly w.r.t. $t \in [0, T]$. Now define

$$\tilde{\mathcal{W}}^{\alpha, \epsilon} := C^{\alpha, \epsilon} + \begin{bmatrix} \phi_{22}^{\alpha, \epsilon}(\phi_{12}^{\alpha, \epsilon})^{-1} & \phi_{21}^{\alpha, \epsilon} - \phi_{22}^{\alpha, \epsilon}(\phi_{12}^{\alpha, \epsilon})^{-1}\phi_{11}^{\alpha, \epsilon} \\ -(\phi_{12}^{\alpha, \epsilon})^{-1} & (\phi_{12}^{\alpha, \epsilon})^{-1}\phi_{11}^{\alpha, \epsilon} \end{bmatrix}.$$

Note that $(\alpha, \epsilon) \rightarrow \tilde{\mathcal{W}}^{\alpha, \epsilon}$ is continuous on a neighborhood of the origin. To prove Thm. 4.2, it remains to prove the lemma

Lemma 6.1. *For $\bar{\epsilon} > 0$ and $\rho > 0$ sufficiently small,*

$$\inf \{ J_A^{\alpha, \epsilon}(y(\cdot), v(\cdot)) \mid (y(0), y(T)) = (y_0, y_1) \} = \begin{bmatrix} y_0^T & y_1^T \end{bmatrix} \tilde{\mathcal{W}}^{\alpha, \epsilon} \begin{bmatrix} y_0 \\ y_1 \end{bmatrix}$$

for all ϵ such that $|\epsilon| \leq \bar{\epsilon}$ and $\alpha \in \mathbb{B}^p(0, \bar{\alpha})$.

Indeed, the assertions of Thm. 4.2 will then follow from Thm. 5.3, when we identify $\tilde{\mathcal{W}}^{\alpha, \epsilon}$ and $\mathcal{W}^{\alpha, \epsilon}$. We merely have to note that, in view of (19), the lemma tells us that, for arbitrary $(y_0, y_1) \in \mathbb{R}^n \times \mathbb{R}^n$,

$$\inf \{ \langle z, (D^2 h^\alpha(z^\alpha) - \epsilon I)z \rangle \mid z \in Z, D\Pi^\alpha(z^\alpha)z = \xi \} = \begin{bmatrix} y_0^T & y_1^T \end{bmatrix} \tilde{\mathcal{W}}^{\alpha, \epsilon} \begin{bmatrix} y_0 \\ y_1 \end{bmatrix}.$$

Thus condition (i) of Thm. 5.3 is satisfied. On the other hand, condition (ii) is also satisfied, in view of the hypotheses imposed on the matrix $\mathcal{W} = \mathcal{W}^{0,0}$ in Thm. 4.2.

Proof of Lemma 6.1. The proof is the same for all (α, ϵ) sufficiently small in magnitude. So we assume $(\alpha, \epsilon) = (0, 0)$ and omit superscripts α and (α, ϵ) throughout. Take any $y_0, y_1 \in \mathbb{R}^n$. Let $P(\cdot)$ be the given solution to the Riccati equation (7) on $[0, T]$ and let $\Phi(\cdot, \cdot)$ be the transition matrix for the Hamiltonian system of equations (satisfying conditions (i) and (ii) above). To simplify notation we shall drop the time argument in $A(t)$ etc, and write Φ for $\Phi(0, T)$. Let $s(\cdot)$ and $m(\cdot)$ be solutions on $[0, T]$ of

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

and

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

where

$$\xi = \phi_{12}^{-1}(y_0 - [\phi_{11} + \phi_{12}P(T)]y_1). \quad (22)$$

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

$$V_t + \min_v \left\{ V_x(Ay + Bv) + \frac{1}{2} y^T Q y + y^T D v + \frac{1}{2} v^T R v \right\} = 0 \quad (23)$$

and

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

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

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

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

It follows, by integrating across (23), that for any (y, v) satisfying

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

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

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

The assertions of the lemma will have been confirmed, then, if we can show

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

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

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

$$p(t) = P(t)\bar{y}(t) + s(t) . \tag{24}$$

It follows that

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

We deduce from (22) and the assumed invertibility of $(\phi_{11} + \phi_{12}P(T))$ that $y(T) = y_1$. (a) is, thus, proved.

We turn finally to the verification of (b). From (24)

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

But,

$$\begin{aligned} \frac{d}{dt}(s(t)^T y(t) + m(t)) &= \dot{s}(t)^T y(t) + s(t)^T \dot{y}(t) + \dot{m}(t) \\ &= (A^T - (B^T P + D)^T R^{-1} B^T s(t))^T y(t) \\ &\quad + s(t)^T (A - BR^{-1}(PB + D)^T y(t) + B^T s(t)) + s(t)^T BR^{-1} B^T s(t) \\ &= 0 . \end{aligned}$$

It follows

$$V(0, y_0) - V(T, y_1) = \frac{1}{2} (p(T)^T y_1 - p(0)^T y_0). \quad (25)$$

But we have that

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

Rearranging to obtain expressions for $p(0), p(T)$ in terms of $y(0), y(T)$ and substituting into (25) yields

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

as required to confirm (b). $\square$

7 Periodic Optimal Control

An important class of problems for which the minimizer is locally non-unique comprises optimal control problems with periodic endpoint constraints, such as

$$(P_p) \left\{ \begin{array}{l} \text{Minimize } \int_0^T L(x(t), u(t)) dt \\ \text{subject to} \\ \dot{x}(t) = f(x(t), u(t)) \text{ a.e. } t \in [0, T] \\ x(0) = x(T) \end{array} \right.$$

Take $(\bar{x}(\cdot), \bar{u}(\cdot))$ to be an admissible process for (P_p) , obtained by numerical or analytical solution of optimality conditions. We recover sufficient conditions of local optimality in this case, from the preceding theory. Let us suppose that the control associated with this process is continuous and non-constant. We may obtain a new process by introducing a time delay of τ in both state and control functions, and ‘wrapping around’. There results another admissible process $(\bar{x}(\cdot + \tau), \bar{u}(\cdot + \tau))$. The integral cost of $(\bar{x}(\cdot + \tau), \bar{u}(\cdot + \tau))$ is the same as that of $(\bar{x}(\cdot), \bar{u}(\cdot))$, because the integral of a periodic function, over the period, is *translationally invariant*. Since we have assumed that the control function is continuous, we can arrange that the new control function, obtained by time translation, is arbitrarily close to the original control function by adjusting the size of the time delay τ . Since the control function is assumed non-constant, the 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.

The hypotheses imposed on (P_p) are the following.

(PH1) The control $\bar{u}(\cdot)$ is continuous and $\bar{u}(0) = \bar{u}(T)$,

(PH2) f, L are of class C^2 and their derivatives with respect to the (x, u) variables, up to order 2, are continuous,

(PH3) there exists $\epsilon > 0$ such that

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

(PH4) $(A(\cdot), B(\cdot))$ is controllable on $[0, T]$.

In view of (PH1) and the translational invariance of the cost, it can easily be deduced that the conditions of the abstract framework (AF1) – (AF3) are automatically satisfied. In particular, we have

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

and hence,

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

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

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

In view of the periodicity of $\bar{x}(\cdot)$ we have that $\dot{\bar{x}}(0) = \dot{\bar{x}}(T)$. The refined sufficient conditions for optimality then simplify as follows:

Theorem 7.1. *Take a weak normal extremal for $(\bar{x}(\cdot), \bar{u}(\cdot))$. Assume hypotheses (PH1) - (PH4) are satisfied. Also suppose that*

(i) *the Riccati equation (7) has a symmetric solution on $[0, T]$,*

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

$$\xi^T \mathcal{W} \xi > \gamma |\xi|^2$$

for all non-zero $\xi \in \mathbb{R}^n$ satisfying

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

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

These sufficient conditions recover those early derived by Speyer [16] for a fixed period, when there exists a solution to the Riccati equation (7) on the closed interval $[0, T]$. The abstract methods of this paper can also be used to obtain sufficient conditions for free-time periodic problems, via state augmentation.

8 Example

To illustrate the application of the refined sufficient conditions, we consider the following optimal control problem with state dimension 3, which is a variant of a periodic optimal control problem investigated by Speyer [13]. Here, an additional isoperimetric constraint is imposed on the average displacement of the first state component over the time interval.

$$(SM) \begin{cases} \text{Minimize } \int_0^1 x_1^2(t) - x_2^2(t) + x_3^4(t) + 0.01u^2(t) dt \\ \text{subject to} \\ \dot{x}_1(t) = x_2(t), \dot{x}_2(t) = u(t), \dot{x}_3(t) = x_1(t) \\ (x_1(0) - x_1(1), x_2(0) - x_2(1), x_3(0), x_3(1)) = (0, 0, 0, 0). \end{cases}$$

Notice that, while the first two state variables are constrained to be periodic over the time interval $[0, 1]$, the third state variable is required to have fixed endpoints. (SM) is not then a periodic optimal control problem. Nonetheless, it is a straightforward exercise to show that any weak normal extremal $((\bar{x}(\cdot), \bar{u}(\cdot)), (p(\cdot), \nu))$ can be continuously embedded in a family of extremals, and that the hypotheses (PH1) - (PH4) and (F1) - (F3) of Section 4 are satisfied. These facts are all simply consequences of the controllability of the dynamic constraint and the following properties of weak normal extremals for this problem:

$(p_1(\cdot), p_2(\cdot))$ and $\bar{u}(\cdot)$ are periodic (on $[0, 1]$). $p_3(\cdot) \equiv c$, for some constant c . The four ‘end-point’ multipliers $(\nu_1, \nu_2, \nu_3, \nu_4)$ are related to the endpoints of the $p_i(\cdot)$ according to:

$$((\nu_1, \nu_2) = (-p_1(0), -p_2(0)), \nu_3 = -c, \nu_4 = +c). \quad (26)$$

The family of extremals, parameterized by $\alpha \in \mathcal{A} := (-1, +1)$, can be chosen to be

$$\left(x_1^\alpha(t) = \bar{x}_1(t + \alpha), x_2^\alpha(t) = \bar{x}_2(t + \alpha), x_3^\alpha(t) = \int_0^t x_1^\alpha(s) ds, u^\alpha(s) = \bar{u}(t + \alpha) \right)$$

with associated costate arcs $(p_1^\alpha(t) = p_1(t + \alpha), p_2^\alpha(t) = p_2(t + \alpha), c)$. (The ν^α are related to the endpoint values of the p_i^α , as in (26).) For evaluation of these formulae, $\bar{x}_1(\cdot)$, etc., are interpreted as their periodic extensions.

Application of the numerical optimal control solver *IPOPT* [17] yields a state trajectory, control function and multipliers for (SM) , satisfying conditions for weak normal extremality. The computed state trajectory components and control function are shown in Figures 1 and 2 respectively.

Figure 1: State trajectory

Figure 2: Control function

Figure 3: Solution of Riccati equation

Let us now examine the implications of the sufficient conditions of Sections 3 and 4, regarding weak

local optimality of the extremal. This involves investigating the accessory problem:

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

in which

$$Q(t) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 6\bar{x}_2^2(t) - 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, A(t) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, b(t) = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, R(t) = 0.01, D(t) = 0.$$

The Hamiltonian system is as follows,

$$\frac{d}{dt} \begin{bmatrix} y \\ q \end{bmatrix} = \begin{bmatrix} A & -50bb^T \\ -2Q & -A^T \end{bmatrix} \begin{bmatrix} y \\ q \end{bmatrix}. \quad (27)$$

A second order Runge-Kutta algorithm [15] was used to compute the transition matrix $\Phi(0, 1)$, related to the system of equations (27).

The Riccati equation (7) has a solution $P(\cdot)$ in $[0, 1]$ when the right boundary condition is chosen to be $P(1) = \frac{1}{2}I_{9 \times 9}$. Figure 3 shows the 6 (symmetric) solutions.

It now remains to confirm condition (ii) of Theorem 4.2. The 6×6 matrix $\mathcal{W}$ is calculated to be

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

The 6×1 matrix Γ of (9) is the column vector

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

We see that Γ^T and the rows of the matrix

$$\begin{bmatrix} m_{x_0} & m_{x_1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

are linearly independent. Thus (F4) is satisfied. To apply the sufficiency tests of Sections 3 and 4, it remains only to check positivity of $\mathcal{W}$ on appropriate subspaces.

Application of Thm. 3.4 requires $\mathcal{W}$ to be positive definite on the null-space of $\begin{bmatrix} m_{x_0} & m_{x_1} \end{bmatrix}$. But this is not the case, because γ is a non-zero vector satisfying $\begin{bmatrix} m_{x_0} & m_{x_1} \end{bmatrix} \gamma = 0$, for which $\mathcal{W}\gamma = 0$. Thus the standard sufficient condition fails to establish the weak local optimality of $(\bar{x}(\cdot), \bar{u}(\cdot))$.

On the other hand, the refined sufficiency test of Thm. 4.2 requires examination of the positive

definiteness of $\mathcal{W}$ on the one dimensional subspace of $\mathbb{R}^6$ comprising vectors $[\xi_0^T \ \xi_1^T]^T$ that satisfy the condition

$$m_{x_0}\xi_0 + m_{x_1}\xi_1 = 0 \quad \text{and} \quad \gamma^T \begin{bmatrix} \xi_0 \\ \xi_1 \end{bmatrix} = 0. \quad (28)$$

This subspace is generated by the vector

$$\begin{aligned} \eta &= \begin{bmatrix} -\dot{x}_2(0) & \dot{x}_1(0) & 0 & -\dot{x}_2(0) & \dot{x}_1(0) & 0 \end{bmatrix}^T \\ &= \begin{bmatrix} 4.7216 & 0.0180 & 0 & 4.7216 & 0.0180 & 0 \end{bmatrix}^T. \end{aligned}$$

We find that

$$\eta^T \mathcal{W} \eta = 748.8 > 0.$$

So $\mathcal{W}$ restricted to the subspace defined by (28) is positive definite. Thm. 4.2 therefore confirms that the weak local extremal is a weak local minimizer.

Appendix

Proof of Lemma 3.2. Suppose there exists a solution to the Riccati equation (7) on $[0, T]$. We claim that for any $y_0 \in \mathbb{R}^n$, the problem

$$\min \{J_L(y(\cdot), v(\cdot)) \mid y(0) = y_0, y(T) = 0\}, \quad (29)$$

where

$$J_L(y(\cdot), v(\cdot)) = \frac{1}{2} \int_0^T y(t)^T Q(t) y(t) + 2y(t)^T D(t) v(t) + v(t)^T R(t) v(t) dt$$

has a solution $(\bar{y}, \bar{v})$. Because (A, B) is controllable, this will mean that the Maximum Principle is satisfied in normal form and, in consequence, $\bar{y}$ and the associated costate arc p satisfy Hamilton's system of equations. In particular,

$$y_0 = \phi_{12}(0, T)p(T).$$

But this is only possible (for arbitrary y_0) if $\phi_{12}(0, T)$ is invertible. To prove the claim, select $\epsilon > 0$ such that the modified Riccati equation, in which $R(t)$ is replaced by $R(t) - \epsilon I$, has a solution $P_\epsilon(\cdot)$. Then, for any (y, v) such that

$$\begin{cases} \dot{y}(t) = A(t)y(t) + B(t)v(t) \\ y(0) = y_0 \quad \text{and} \quad y(T) = 0, \end{cases}$$

(and one exists, by 'controllability') we have that

$$\begin{aligned} J_L(y, v) - \epsilon \|v\|_{L^2}^2 &\geq \inf \{J_L(y', v') - \epsilon \|v'\|_{L^2}^2 + y'(T)^T P_\epsilon(T) y'(T) \mid y'(0) = y_0\} \\ &= y_0^T P_\epsilon(0) y_0. \end{aligned}$$

It follows that, if $\{(y_i, v_i)\}$ is any minimizing sequence for (29), $\{\|v_i\|_{L^2}^2\}$ is uniformly bounded. We may arrange, by restricting attention to a subsequence, that $\{y_i(\cdot)\}$ converges uniformly to $\bar{y}(\cdot)$ and $\{v_i(\cdot)\}$ converges to $\bar{v}(\cdot)$, weakly in L^2 , for some process $(\bar{y}(\cdot), \bar{v}(\cdot))$ that satisfies the endpoint constraints. The cost $J_L(\cdot, \cdot)$, together with the linear dynamics and fixed initial condition, define a strongly continuous functional on the space of square integrable control functions. Since, however, $R(t) > 0$, this functional is convex. So it is also weakly lower semi-continuous. A standard argument permits us to deduce from this semi-continuity property that $(\bar{y}, \bar{v})$ is a minimizer. $\square$

Acknowledgment: This work was co-funded by the EPSRC grant "Control For Energy and Sustainability", grant reference EP/G066477/1, and by the European Union under the 7th Framework Programme "FP7-PEOPLE-2010-ITN", grant agreement number 264735-SADCO.

References

- [1] Allwright, J. and Vinter, R. “Second order conditions for periodic optimal control problems”, *Control and Cybernetics*, vol. 34, 2005, pp. 617–643.
- [2] Bonnans, J. F. and Shapiro, A. *Perturbation Analysis of Optimization Problems*, Springer, New York, 2000
- [3] Brockett, R. W. *Finite Dimensional Linear Systems*, John Wiley, New York, 1970.
- [4] Hestenes, M. R. “Applications of the theory of quadratic forms in Hilbert space to calculus of variations”, *Pacific J. Math*, vol. 1, 1951, pp. 525–581.
- [5] Liu, M. *First and Second Order Conditions for Optimal Control Problems*, PhD Thesis, Imperial College London, 2007.
- [6] Maurer, H. and Pickenhain, S. “Second order sufficient conditions for optimal control problems with mixed control-state constraints”, *J. Optim. Theory and Applic.*, vol. 86, 1995, pp. 649–667.
- [7] Maurer, H. and Oberle, S. “Second order sufficient conditions and sensitivity analysis for optimal control problems with free final time: the Riccati approach”, *SIAM J. Control Optim.*, vol. 41, 2002, pp.380–403.
- [8] Milyutin, A. A. and Osmolovskii, N. P. *Calculus of Variations and Optimal Control*, Translations of mathematical monographs 180, American Mathematical Society, Providence, Rhode Island, 1998.
- [9] Pars, L. *A Treatise on Analytical Mechanics*, Wiley, New York, 1965.
- [10] Gavriel, C. and Vinter, R. “Second order sufficient conditions for optimal control problems with non-unique minimizers”, *IEEE Proc. Amer. Control Conf.*, 2010, pp. 756–761.
- [11] Smart, D. R. *Fixed Point Theorems*, Cambridge University Press, Cambridge, England, 1974.
- [12] Speyer, J. L. and Evans, R. T. “A second variational theory for optimal periodic process”, *IEEE Trans. Aut. Control*, vol. 29, 1984, pp. 138–148.
- [13] Speyer, J. L. “Periodic optimal flight”, *J. Guidance, Control and Dynamics*, vol. 61, 1996, pp. 745–754.
- [14] Vinter, R. *Optimal Control*, Birkhäuser, Boston, 2000.
- [15] Wang, H. and Hale, J. K. “On monodromy matrix computation”, *Computer Methods in Applied Mechanics and Engineering*, vol. 190, 2001, pp. 2263–2275.
- [16] Wang, H. and Speyer J. L. “Necessary and sufficient conditions for local optimality of a periodic process”, *SIAM J. Control and Optim.*, vol. 28, 1990, pp. 2263–2275.
- [17] Wächter, A. and Biegler, L. T. “On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming”, *Mathematical Programming*, vol. 106, 2006, pp. 25–57.
- [18] Stefani, G. and Zezza, P. “Constrained regular LQ-control problems”, *SIAM J. Control and Optim.*, vol. 35, 1997, pp. 876–900.
- [19] Zeidan, V. “New second-order optimality conditions for variational problems with C^2 -Hamiltonians”, *SIAM J. Control and Optim.*, vol. 40, 2001, pp. 577–609.