WHEN ARE MINIMIZING CONTROLS ALSO MINIMIZING RELAXED CONTROLS?

Michele Palladino
Imperial College London
Electrical and Electronical Engineering Department
South Kensington Campus, London SW7 2AZ, UK

Richard B. Vinter
Imperial College London
Electrical and Electronical Engineering Department
South Kensington Campus, London SW7 2AZ, UK

(Communicated by the associate editor name)

Abstract. Relaxation refers to the procedure of enlarging the domain of a variational problem or the search space for the solution of a set of equations, to guarantee the existence of solutions. In optimal control theory relaxation involves replacing the set of permissible velocities in the dynamic constraint by its convex hull. Usually the infimum cost is the same for the original optimal control problem and its relaxation. But it is possible that the relaxed infimum cost is strictly less than the infimum cost. It is important to identify such situations, because then we can no longer study the infimum cost by solving the relaxed problem and evaluating the cost of the relaxed minimizer. Following on from earlier work by Warga, we explore the relation between the existence of an infimum gap and abnormality of necessary conditions (i.e. they are valid with the cost multiplier set to zero). Two kinds of theorems are proved. One asserts that a local minimizer, which is not also a relaxed minimizer, satisfies an abnormal form of the Pontryagin Maximum Principle. The other asserts that a local relaxed minimizer that is not also a minimizer satisfies an abnormal form of the relaxed Pontryagin Maximum Principle.

1. Introduction. Consider the optimal control problem

\[
\begin{align*}
(P) \quad \min & \quad g(x(0), x(1)) \\
\text{subject to} & \quad \dot{x}(t) = f(t, x(t), u(t)) \text{ a.e.,} \\
& \quad u(t) \in U(t) \text{ a.e.,} \\
& \quad (x(0), x(1)) \in C,
\end{align*}
\]

2010 Mathematics Subject Classification. Primary: 49N, 49K.

Key words and phrases. Necessary Conditions, Optimal Control, Ordinary Differential Equations, State Constraints.

The authors are supported by 7th Framework Programme FP7-PEOPLE-2010-ITN, grant agreement number 264735-SADCO.
the data for which comprise: functions \( g : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \) and \( f : [0, 1] \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \), a closed set \( C \subset \mathbb{R}^n \times \mathbb{R}^n \) and a multifunction \( U(\cdot) : [0, 1] \to \mathbb{R}^m \).

A process \((x(\cdot), u(\cdot))\) is a pair of functions, of which the first \(x(\cdot) : [0, 1] \to \mathbb{R}^n\) is an absolutely continuous function, and the second \(u(\cdot) : [0, 1] \to \mathbb{R}^m\) is a measurable function satisfying
\[
\dot{x}(t) = f(t, x(t), u(t)) \text{ a.e. and } u(t) \in U(t) \text{ a.e.}
\]
A process \((x(\cdot), u(\cdot))\) is said to be admissible if \((x(0), x(1)) \in C\). The first component \(x(\cdot)\) of an (admissible) process \((x(\cdot), u(\cdot))\) is called an (admissible) state trajectory and the second an (admissible) control function.

We say that a process \((\tilde{x}(\cdot), \tilde{u}(\cdot))\) is a minimizer if it achieves the minimum of \(g(x(0), x(1))\) over all admissible processes \((x(\cdot), u(\cdot))\). It is called a strong local minimizer if, for some \(\epsilon > 0\),
\[
g(x(0), x(1)) \geq g(\tilde{x}(0), \tilde{x}(1))
\]
for all processes \((x(\cdot), u(\cdot))\) such that \(||x(\cdot) - \tilde{x}(\cdot)||_L^\infty \leq \epsilon\).

We write \(\inf(P)\), ‘the infimum cost’, for the infimum value of the cost function \(g(x(0), x(1))\) over the set of admissible processes \((x(\cdot), u(\cdot))\).

On first acquaintance with optimal control, one might expect that a minimizer for problem \((P)\) exists, under hypotheses ensuring the existence of a unique state trajectory for every control function and initial state and the continuous dependence of this state trajectory on these quantities, the continuity of the cost function and closedness of \(C\) and the values of \(U(\cdot)\), and the non-emptiness and boundedness of the set of admissible processes. But, as is well-known, such hypotheses are not enough, and an extra hypothesis is required. The most commonly invoked additional hypothesis to guarantee existence of minimizers is ‘convexity of the velocity set’, namely
\[
(C): f(t, x, U(t)) \text{ is convex for all } t \in [0, 1], x \in \mathbb{R}^n.
\]

Yet situations arise when the convexity hypothesis is violated and no minimizers exist. Here, it is still of interest to calculate \(\inf(P)\), since this number provides a tight lower bound on all possible values of the cost. The concept of ‘relaxation’ was introduced in the 1960’s to deal with this eventuality. (See [11].) It takes inspiration from Hilbert’s 20th problem

‘Has not every regular variation problem a solution, provided certain assumptions regarding the given boundary conditions are satisfied . . ., and provided also if need be that the notion of a solution shall be suitably extended?’

In the optimal control context, relaxation involves adding additional admissible processes (‘relaxed admissible processes’) to guarantee existence of minimizers, leading to the relaxed problem, which we write \((R)\). Of course \(\inf(R) \leq \inf(P)\), because the relaxed problem involves minimizing the same cost over a larger domain. The set of relaxed processes is chosen to be close to the set of admissible processes, in
some sense, and so, usually, we have

$$\inf(R) = \inf(P).$$

In this case the relaxed problem serves the following purpose: using the theory of necessary conditions, or numerical techniques, we solve the relaxed problem (which has a solution). The infimum cost is obtained as the cost of this solution to the relaxed problem. Also, even though $(P)$ does not have a minimizer, we can often obtain an admissible process with cost arbitrarily close to the infimum cost, by approximating the relaxed minimizer by a ‘neighboring’ admissible process. The relaxed problem is taken to be

$$\begin{align*}
(R) \quad \left\{ \begin{array}{l}
\text{Minimize } g(x(0), x(1)) \\
on \text{ over absolutely continuous functions } x(.) : [0, 1] \to \mathbb{R}^n \\
\text{and measurable functions } (u_0(.), \ldots, u_n(.)), (\lambda_0(.), \ldots, \lambda_n(.)) \\
\text{satisfying} \\
x(t) = \sum_{j=0}^n \lambda_j(t) f(t, x(t), u_j(t)) \text{ a.e.} \\
(u_0(t), \ldots, u_n(t)) \in U(t) \times \ldots \times U(t) \text{ a.e.} \\
(\lambda_0(t), \ldots, \lambda_n(t)) \in \Sigma \text{ a.e.} \\
(x(0), x(1)) \in C.
\end{array} \right.
\end{align*}$$

Here, $\Sigma$ is the set of simplicial index values in $n$-dimensional space:

$$\Sigma := \{ (\lambda_0 \geq 0, \ldots, \lambda_n \geq 0) \mid \sum_{j=0}^n \lambda_j = 1 \}.$$

Processes, state trajectories and strong local minimizers for $(R)$ are referred to as ‘relaxed processes’, ‘relaxed state trajectories’ and ‘relaxed strong local minimizers’, respectively. A process $(x(.), u(.))$ is interpreted as the relaxed process $(x(.), \{(\lambda^k(.), u^k(.))\}_{k=0}^n)$, in which $\lambda^0 = 1$, $\lambda^1 = 0$, $\ldots$, $\lambda^n = 0$, and $u^0(.) = u^1(.) = \ldots = u^n(.) = u(.)$.

Notice that the velocity sets for the relaxed problem are

$$\left\{ \sum_{j=0}^n \lambda_j f(t, x, u_j) \mid (u_0, \ldots, u_n) \in U(t) \times \ldots \times U(t), (\lambda_0, \ldots, \lambda_n) \in \Sigma \right\} =$$

$$= \text{co}\{ f(t, x, U(t)) \} \quad \text{for } t \in [0, 1], x \in \mathbb{R}^n.$$

So the convexity hypothesis $(C)$ is satisfied and the relaxed problem has a minimizer.

It is possible however that there is an ‘infimum gap’, i.e.

$$\inf(R) < \inf(P).$$

It is important to identify such situations because, then, the above justification for studying the relaxed problem no longer applies.

The aim of this paper is to derive new conditions for an infimum gap to occur. These conditions require necessary conditions of optimality, expressed in terms of the Maximum Principle, to apply in abnormal form.

The link between the occurrence of an infimum gap and abnormality is a natural one. It is well known that the minimum cost of a nonlinear programming problem can fail
to be stable under perturbations of the constraints when there exists an abnormal set of Lagrange multipliers [1]. Since an optimal control problem is a special kind of infinite dimensional nonlinear programming problem, the existence of an infimum gap is a manifestation of instability of the infimum cost under perturbations to the endpoint and pathwise state constraints and the Maximum Principle is a kind of Lagrange multiplier rule, we would expect the occurrence of an infimum gap to be revealed by the abnormality of the Maximum Principle conditions.

Two consequences of an infimum gap are explored. In the first we focus attention on a strong local minimizer which cannot also be interpreted as a strong relaxed minimizer; in the second, on a relaxed minimizer, whose cost is strictly less than the infimum cost over admissible (non-relaxed) processes.

Type A: A strong local minimizer satisfies the Pontryagin Maximum Principle in abnormal form (i.e. with cost multiplier zero) if, when regarded as a relaxed admissible process, it is not also a relaxed strong local minimizer.

Type B: A relaxed strong local minimizer satisfies the relaxed Pontryagin Maximum Principle in abnormal form if its cost is strictly less than the infimum cost over all admissible processes, whose state trajectories are close (in the $L^\infty$ sense) to that of the relaxed strong local minimizer.

Warga was the first to investigate the relation between the existence of an infimum gap and validity of the Maximum Principle in abnormal form. Warga announced a Type A relation for state constraint-free optimal control problems with smooth data in his early paper [10]. In his monograph [11] he proved a Type B relation for optimal control problems with state constraints. In a subsequent paper [12], Warga generalized his earlier Type B results to allow for nonsmooth data, making use of local approximations based on `derivative containers', developed in [13].

We prove both Type A and Type B relations for a class of non-smooth state-constrained optimal control problems which subsume those considered by Warga, and under less restrictive hypotheses on the data. (We allow, for example, the endpoint constraint set $C$ to be a general closed set, whereas Warga requires $C$ to have a functional representation.) Our relations also differ because they are based on the, by now, standard form of the non-smooth Maximum Principle, originally derived by Clarke [4] (and generalized to allow for state constraints in [5]), expressed in terms of subdifferentials; our results are therefore better suited as analytical tools for future developments in optimal control theory. The proofs in this paper make use of perturbation techniques. They are very different from those of Warga, which are based on the construction of approximating cones to reachable sets.

The main results of this paper can be summarized as

(A): ‘$(\bar{x}(.), \bar{u}(.) )$ is a strong local minimizer but not a relaxed strong local minimizer’ implies ‘$(\bar{x}(.), \bar{u}(.) )$ satisfies an `averaged’ version of the Maximum Principle in abnormal form’, and

(B): ‘$(\bar{x}(.), \bar{u}(.) )$ is a relaxed strong local minimizer with cost strictly less than that of any admissible (non-relaxed) process, whose state trajectory is close (in the
$L^\infty$ sense) to that of the relaxed strong local minimizer’ implies ‘($\bar{x}(.), \bar{u}(.)$) satisfies the relaxed Maximum Principle in abnormal form’.

The second statement links the occurrence of an infimum gap and abnormality of the Maximum Principle precisely (in relation to relaxed minimizers). But the first statement, we notice, does a little bit less, because it invokes a weaker, *averaged*, version of the Maximum Principle, which, for smooth data, involves the adjoint inclusion

$$-\dot{p}(t) \in \text{co} \{p(t)f^T_x(t, \bar{x}(t), u) \mid u \in U(t)\}$$

in place of the expected

$$-\dot{p}(t) = p(t)f^T_x(t, \bar{x}(t), \bar{u}(t)) \text{ a.e.}$$

It remains an open question whether a sharper Type (A) relation is valid, involving the adjoint equation (a some related non-smooth generalization) in place of its averaged version. In the final section, however, we show that the two adjoint relations are the same when the dynamics are affine with respect to the control variable.

The conditions for existence of an infimum gap are given for optimal control problems with and without pathwise state constraints. While the state constraint-free problem is a special case of the state constraint problem, we state the state constraint-free conditions separately, to bring out the underlying relationships more clearly, without the distraction of ‘measure multipliers’ and other complications in the statement of the state constrained Maximum Principle.

Type (A) and Type (B) relations have been previously derived for optimal control problems in which the dynamic constraint takes the form of a differential inclusion. See [7], [2], [3]. Here a link is established between the existence of an infimum gap and satisfaction of necessary conditions in abnormal form, when the necessary condition involved is Clarke’s Hamiltonian inclusion. The theory developed in these papers for optimal control problems, in which the dynamic constraint is taken to be a differential inclusion, employs the Hamiltonian inclusion in both Type (A) and (B) relations, not some averaged version, and so provides rather more precise relations than those can currently be obtained when the dynamic constraint takes the form of a controlled differential equation. Nonetheless, the relations of this paper are of interest because most applications of optimal control theory are based on formulations involving a controlled differential equation, not a differential inclusion.

The following notation will be used throughout the paper: for vectors $x \in \mathbb{R}^n$, $|x|$ denotes the Euclidean length. $B$ denotes the closed unit ball in $\mathbb{R}^n$. Given a multifunction $\Gamma(.) : \mathbb{R}^n \rightrightarrows \mathbb{R}^k$, $\text{Gr} \Gamma(.)$ is the set $\{(x, v) \in \mathbb{R}^n \times \mathbb{R}^k \mid v \in \Gamma(x)\}$. Given a set $A \subset \mathbb{R}^n$ and a point $x \in \mathbb{R}^n$, we denote by $d_A(x)$ the Euclidean distance of a point $x \in \mathbb{R}^n$ from $A$:

$$d_A(x) := \inf\{|x - y| \mid y \in A\}.$$  

$W^{1,1}([0, 1]; \mathbb{R}^n)$ is the space of absolutely continuous $\mathbb{R}^n$-valued functions $x(.)$ with norm $|x_0| + ||\dot{x}(.)||_{L^1}$.

We denote by $NBV^+[0, 1]$ the space of increasing, real-valued functions $\mu(.)$ on $[0, 1]$ of bounded variation, vanishing at the point 0 and right continuous on $(0, 1)$. The total variation of a function $\mu(.) \in NBV^+[0, 1]$ is written $||\mu||_{TV}$. As is well
known, each point $\mu(\cdot) \in NBV^+[0,1]$ defines a Borel measure on $[0,1]$. The associated measure is also denoted $\mu$. We write $W^{1,1}$ in place of $W^{1,1}([0,1],\mathbb{R}^n)$, $NBV^+$ in place of $NBV^+[0,1]$, etc. when the meaning is clear.

We shall use several constructs of nonsmooth analysis. Given a closed set $D \subset \mathbb{R}^k$ and a point $\bar{x} \in D$, the normal cone $N_D(\bar{x})$ of $D$ at $\bar{x}$ is defined to be

$$N_D(\bar{x}) := \left\{ p \mid \exists x_i \xrightarrow{D} \bar{x}, \quad p_i \longrightarrow p \quad \text{s.t.} \quad \limsup_{x \xrightarrow{D} x_i} \frac{p_i \cdot (x - x_i)}{|x - x_i|} \leq 0 \quad \text{for each } i \in \mathbb{N} \right\}.$$ 

Here, the notation $y_i \xrightarrow{D} y$ is employed to indicate that all points in the convergent sequence $\{y_i\}$ lie in $D$.

Given a lower semicontinuous function $f : \mathbb{R}^k \to \mathbb{R} \cup \{+\infty\}$ and a point $\bar{x} \in \text{dom } f := \{x \in \mathbb{R}^k \mid f(x) < +\infty\}$, the subdifferential of $f$ at $\bar{x}$ (termed the ‘limiting subdifferential’ in [6]) is denoted $\partial f(\bar{x})$:

$$\partial f(\bar{x}) := \left\{ \xi \mid \exists \xi_i \to \xi \text{ and } x_i \xrightarrow{\text{dom } f} \bar{x} \text{ such that} \right.$$

$$\left. \limsup_{x \rightarrow x_i} \frac{\xi_i \cdot (x - x_i) - \varphi(x) + \varphi(x_i)}{|x - x_i|} \leq 0 \text{ for all } i \in \mathbb{N} \right\}.$$ 

For details of definition and properties of these objects, we refer the reader to [6], [8] and [9].

2. Conditions for Non-Coincidence of Infima. In this section we state two theorems relating the existence of a gap between the infimum costs for the optimal control problem (P) and its relaxed counterpart (R), and the validity of a Maximum Principle in abnormal form. The following hypotheses, in which $\bar{x}(\cdot)$ is a given absolutely continuous function, will be invoked.

(H1) : $f(\cdot, x, u)$ is $\mathcal{L}$-measurable and $f(t, \cdot, \cdot)$ is continuous. $U(\cdot)$ is a Borel measurable multifunction taking values compact sets.

(H2) : There exist $\varepsilon > 0$, $k(\cdot) \in L^1$ and $c(\cdot) \in L^1$ such that

$$|f(t, x, u) - f(t, x', u)| \leq k(t)|x - x'| \quad \text{and} \quad |f(t, x, u)| \leq c(t)$$

for all $x, x' \in \bar{x}(t) + \varepsilon B$, $u \in U(t)$, a.e. $t \in [0,1]$.

Define the Hamiltonian function $H(t, x, u, p) := p \cdot f(t, x, u)$.

The first theorem is a Type A relation.

**Theorem 2.1.** Let $(\bar{x}(\cdot), \bar{u}(\cdot))$ be a strong local minimizer for problem (P). Assume that hypotheses (H1) and (H2) are satisfied and that $g(\cdot, \cdot)$ is Lipschitz continuous on a neighborhood of $(\bar{x}(0), \bar{x}(1))$.

(a) : Then there exist an arc $p(\cdot) \in W^{1,1}([0,1]; \mathbb{R}^n)$, and $\lambda \geq 0$ such that

(i) $\|p\|_{L^\infty} + \lambda \neq 0$,

(ii) $-\dot{p}(t) \in co \partial_x H(t, \bar{x}(t), p(t), \bar{u}(t))$ a.e.,

(iii) $(p(0), -p(0)) \in \lambda \partial g(\bar{x}(0), \bar{x}(1)) + N_{C}(\bar{x}(0), \bar{x}(1))$,

(iv) $H(t, \bar{x}(t), p(t), \bar{u}(t)) \geq H(t, \bar{x}(t), p(t), u)$,

for all $u \in U(t)$, a.e. $t \in [0,1]$,

in which $\partial_x H(t, x, p, u)$ denotes the limiting subdifferential of $H(t, \cdot, p, u)$. 


(b): Suppose that, for every $\epsilon > 0$, there exists an admissible relaxed process $(\bar{x}(.), \{(\bar{\lambda}^k(., \bar{u}^k(.))\}_{k=0}^n)$ such that

$$g(\bar{x}(0), \bar{x}(1)) > g(x(0), x(1))$$

and $\|x(.)-\bar{x}(.))\|_{L^\infty} \leq \epsilon$, i.e. $(\bar{x}(.), \bar{u}(.)$ is not also a local relaxed minimizer).

Then conditions (i)-(iv) above are satisfied for some choice of multipliers $(p(.), \lambda)$ such that $\lambda = 0$ and in which (ii) is replaced by

$$(ii)' \quad -\dot{p}(t) \in \co \bigcup_{u \in U(t)} \partial_x H(t, \bar{x}(t), p(t), u) \quad \text{a.e. } t \in [0,1].$$

Comments:

(1) Part (a) is a standard version of the state constrained Maximum Principle (See, e.g. ([9], Ch. 9)). Interest focuses on part (b), which is a Type A relation.

(2) The contrapositive statement of part (b) is a sufficient condition for the absence of an infimum gap (in an appropriate ‘local’ sense): if $(\bar{x}(.), \bar{u}(.)$ is a strong local minimizer such that, given any multipliers $p(.) \in W^{1,1}$, and $\lambda \geq 0$ satisfying conditions (i), (ii)', (iii) and (iv), we have $\lambda \neq 0$, then $(\bar{x}(.), \bar{u}(.)$ is also a strong local relaxed minimizer.

The second theorem is a Type B relation:

**Theorem 2.2.** Let $(\bar{x}(.), \{(\bar{\lambda}^k(., \bar{u}^k(.))\}_{k=0}^n)$ be a relaxed admissible process related to problem (R). Assume that hypotheses (H1) and (H2) are satisfied.

(a): Assume that $g(.,.)$ is Lipschitz continuous on a neighborhood of $(\bar{x}(0), \bar{x}(1))$. Now suppose that $(\bar{x}(.), \{(\bar{\lambda}^k(., \bar{u}^k(.))\}_{k=0}^n)$ is a strong local relaxed minimizer. Then there exist an arc $p(.) \in W^{1,1}([0,1]; \mathbb{R}^n)$ and $\lambda \geq 0$ such that

(i) $\|p\|_{L^\infty} + \lambda \neq 0$,

(ii) $-\dot{p}(t) = \sum_{k=0}^n \bar{\lambda}^k(t) \co \{\partial_x H(t, \bar{x}(t), p(t), \bar{u}^k(t))\}$ a.e. ,

(iii) $(p(0), -p(1)) \in \lambda \partial g(\bar{x}(0), \bar{x}(1)) + N_C(\bar{x}(0), \bar{x}(1))$,

(iv) For $k = 0, \ldots, n$

$$H(t, \bar{x}(t), p(t), \bar{u}^k(t)) \geq H(t, \bar{x}(t), p(t), u),$$

for all $u \in U(t)$, a.e. $t \in [0,1]$.

(b): Assume also that the following condition is satisfied:

(S): $g(.,.)$ is continuous on a neighborhood of $(\bar{x}(0), \bar{x}(1))$, and numbers $\epsilon > 0$ and $\delta > 0$ can be chosen such that

$$g(x(0), x(1)) \geq g(\bar{x}(0), \bar{x}(1)) + \delta$$

for all admissible processes $(x(.), u(.))$ such that $\|x(.)-\bar{x}(.)\|_{L^\infty} \leq \epsilon$.

Then relations (i)-(iv) above are satisfied for some set of multipliers $p(.) \in W^{1,1}$ and $\lambda \geq 0$, such that $\lambda = 0$.

Comments:

(1): Part (a) of Thm. 2.2 is a version of the well known state constrained Maximum Principle, applied to the relaxed problem. Interest resides rather in part (b), which is a Type (B) relation.

(2): By focusing attention on the case when $(\bar{x}(.), \{(\bar{\lambda}^k(., \bar{u}^k(.))\}_{k=0}^n)$ is a strong local relaxed minimizer, we deduce from the contrapositive statement of Thm. 2.2 part (b) another sufficient condition for the non-existence of an infimum gap: if $(\bar{x}(.), \{(\bar{\lambda}^k(., \bar{u}^k(.))\}_{k=0}^n)$ is a relaxed strong local minimizer and, given any multipliers $p(.) \in W^{1,1}$ and $\lambda \geq 0$ satisfying conditions (i) - (iv), we
have \( \lambda \neq 0 \), then the cost of \((\bar{x}(\cdot), (\bar{\lambda}^k(\cdot), \bar{u}^k(\cdot))_{k=0}^n)\) is the infimum cost of admissible processes \((x(\cdot), u(\cdot))\) satisfying \(\|x(\cdot) - \bar{x}(\cdot)\|_{L^\infty} \leq \epsilon\), for some \(\epsilon > 0\).

(3): Part (a) of the theorem is valid in a stronger form than in the theorem statement, in which the costate differential inclusion is replaced by

\[
-\dot{p}(t) \in \text{co} \{ \sum_{k=0}^n \bar{\lambda}^k(t) H(t, \bar{x}(t), p(t), \bar{u}^k(t)) \} \text{ a.e. (1)}
\]

We mention that this is a distinction that applies only to nonsmooth data; if \(f(t,\cdot, \bar{u}^k(t))\) is continuously differentiable at \(\bar{x}(t)\), for all \(k\) and a.e. \(t \in [0,1]\), then (ii) and (ii)' both reduce to the relation:

\[
(ii)' \ -\dot{p}(t) = \sum_{k=0}^n \bar{\lambda}^k(t) \nabla_x H(t, \bar{x}(t), p(t), \bar{u}^k(t)) \text{ a.e. (1)}
\]

It remains an open question, whether part (b) is also valid for the more precise version of the costate differential inclusion (1).

3. State Constraints. Now consider a refinement of \((P)\), which includes a pathwise state constraint:

\[
(S) \begin{cases}
\text{Minimize } g(x(0), x(1)) \\
\text{over absolutely continuous functions } x(\cdot) : [0,1] \to \mathbb{R}^n \\
\text{and measurable functions } u(\cdot) \text{ satisfying} \\
\dot{x}(t) = f(t, x(t), u(t)) \text{ a.e.,} \\
u(t) \in U(t) \text{ a.e.,} \\
h(t, x(t)) \leq 0 \text{ for all } t \in [0,1], \\
(x(0), x(1)) \in C',
\end{cases}
\]

for which the new data is an upper semi-continuous function \(h : [0,1] \times \mathbb{R}^n \to \mathbb{R}\).

Problem \((S)\) has the relaxed counterpart:

\[
(RS) \begin{cases}
\text{Minimize } g(x(0), x(1)) \\
\text{over absolutely continuous functions } x(\cdot) : [0,1] \to \mathbb{R}^n \\
\text{and measurable functions } (u_0(\cdot), \ldots, u_n(\cdot)), (\lambda_0(\cdot), \ldots, \lambda_n(\cdot)) \text{ satisfying} \\
\dot{x}(t) = \sum_{j=0}^n \lambda_j(t) f(t, x(t), u_j(t)) \text{ a.e.} \\
(u_0(t), \ldots, u_n(t)) \in U(t) \times \ldots \times U(t) \text{ a.e.} \\
(\lambda_0(t), \ldots, \lambda_n(t)) \in \Sigma \text{ a.e.} \\
h(t, x(t)) \leq 0 \quad \forall t \in [0,1] \\
(x(0), x(1)) \in C'.
\end{cases}
\]

‘Admissible process’, ‘relaxed admissible process’ ‘strong local minimizer’, ‘relaxed strong local minimizer’, etc., have their previous meanings, except an admissible process \((x(\cdot), u(\cdot))\) is required, also, to satisfy the condition \(h(t, x(t)) \leq 0\) for all \(t \in [0,1]\). Concerning the pathwise constraint, we shall assume that (for the state trajectory \(\bar{x}(\cdot)\) of interest and some number \(\epsilon > 0\)) the following hypothesis is satisfied:

\[
(H3): \ h(\cdot,\cdot) \text{ is upper semicontinuous on } \{(t,x) \in [0,1] \times \mathbb{R}^n | x \in \bar{x}(t) + \epsilon B \} \text{ and there exists a constant } k_h \text{ such that} \\
\left| h(t, x) - h(t, x') \right| \leq k_h |x - x'|
\]
Theorem 2.1 generalizes to allow for the presence of the pathwise state constraint as follows:

**Theorem 3.1.** Let \((\bar{x}(\cdot), \bar{u}(\cdot))\) be a strong local minimizer for problem (S). Assume that hypotheses (H1), (H2) and (H3) are satisfied. Assume also that \(g(\cdot, \cdot)\) is Lipschitz continuous on a neighborhood of \((\bar{x}(0), \bar{x}(1))\).

(a): Then there exist an arc \(p(\cdot) \in W^{1,1}([0,1]; \mathbb{R}^n)\), a function \(\mu(\cdot) \in NBV^+ [0,1]\), a \(\mu\)-integrable function \(m(\cdot)\) and \(\lambda \geq 0\) such that

(i) \(\|p\|_{L^\infty} + \|\mu\|_{T,V} + \lambda \neq 0\),
(ii) \(-\dot{p}(t) \in \text{co} \{\partial_x H(t, \bar{x}(t), p(t) + \int_{[0,t]} m(s)\mu(ds)), \bar{u}(t)\}\) a.e.
(iii) \(\left[p(0) - \left[p(1) + \int_{[0,1]} m(s)\mu(ds)\right]\right] \in \lambda \partial g(\bar{x}(0), \bar{x}(1)) + NC(\bar{x}(0), \bar{x}(1))\),
(iv) \(H(t, \bar{x}(t), p(t) + \int_{[0,t]} m(s)\mu(ds)), \bar{u}(t)) \geq H(t, \bar{x}(t), p(t) + \int_{[0,t]} m(s)\mu(ds), u)\)
for all \(u \in \text{U}(t)\), a.e. \(t \in [0,1]\),
(v) \(m(t) \in \partial^2_x h(t, \bar{x}(t)) \mu\)-a.e. and \(\text{supp}(\mu) \subset \{t : h(t, \bar{x}(t)) = 0\}\).

(b): Suppose that, for every \(\epsilon > 0\), there exists a feasible relaxed process \((x(\cdot), \{(\bar{x}^k(\cdot), \bar{u}^k(\cdot))\}_{k=0}^n)\) such that

\[g(\bar{x}(0), \bar{x}(1)) > g(x(0), x(1))\]
and \(\|x(\cdot) - \bar{x}(\cdot)\|_{L^\infty} \leq \epsilon\), (i.e. \((\bar{x}(\cdot), \bar{u}(\cdot))\) is not also a relaxed strong local minimizer).

Then conditions (i)-(v) above are satisfied for some choice of multipliers \(p(\cdot) \in W^{1,1}([0,1]; \mathbb{R}^n)\), \(\mu(\cdot) \in NBV^+ [0,1]\) and a \(\mu\)-integrable function \(m(\cdot)\), such that \(\lambda = 0\) and in which (ii) is replaced by

\[(ii)' \quad -\dot{p}(t) \in \bigcup_{u \in \text{U}(t)} \partial_x H(t, \bar{x}(t), p(t) + \int_{0}^{t} m(s)\mu(ds), u) \quad \text{a.e. } t \in [0,1].\]

Here, \(\partial^*_x h(t, x)\) is the set

\[\partial^*_x h(t, x) := \text{co} \{\xi | \text{there exist } x_i \to x, t_i \to t, \xi_i \to \xi \text{ s.t.}, \text{for each } i, \nabla_x h(t_i, x_i) \text{ exists, } \xi_i = \nabla_x h(t_i, x_i) \text{ and } h(t_i, x_i) > 0\} .\]

There follows now a generalization of Thm. 2.2 to allow for the state constraint:

**Theorem 3.2.** Let \((\bar{x}(\cdot), \{(\bar{x}^k(\cdot), \bar{u}^k(\cdot))\}_{k=0}^n)\) be a relaxed feasible process related to problem (RS). Assume that hypotheses (H1), (H2) and (H3) are satisfied.

(a): Assume that \(g(\cdot, \cdot)\) is Lipschitz continuous on a neighborhood of \((\bar{x}(0), \bar{x}(1))\).

Now suppose that \((\bar{x}(\cdot), \{(\bar{x}^k(\cdot), \bar{u}^k(\cdot))\}_{k=0}^n)\) is a relaxed strong local minimizer. Then there exist an arc \(p(\cdot) \in W^{1,1}([0,1]; \mathbb{R}^n)\), a function \(\mu(\cdot) \in NBV^+ [0,1]\), a \(\mu\)-integrable function \(m(\cdot)\) and \(\lambda \geq 0\) such that

(i) \(\|p\|_{L^\infty} + \|\mu\|_{T,V} + \lambda \neq 0\),
(ii) \(-\dot{p}(t) \in \sum_{k=0}^{n} \bar{x}^k(\cdot) \text{co} \{\partial_x H(t, \bar{x}(t), p(t) + \int_{[0,t]} m(s)\mu(ds)), \bar{u}^k(t)\}\) a.e.
(iii) \(\left[p(0) - \left[p(1) + \int_{[0,1]} m(s)\mu(ds)\right]\right] \in \lambda \partial g(\bar{x}(0), \bar{x}(1)) + NC(\bar{x}(0), \bar{x}(1))\),
(iv) \(H(t, \bar{x}(t), p(t) + \int_{[0,t]} m(s)\mu(ds), \bar{u}^k(t)) \geq H(t, \bar{x}(t), p(t) + \int_{[0,t]} m(s)\mu(ds), u)\)
for all \(u \in \text{U}(t)\), a.e. \(t \in [0,1]\),
(v) $m(t) \in \partial^\infty_c h(t, \bar{x}(t)) \mu$-a.e. and $\text{supp}\{\mu\} \subset \{t : h(t, \bar{x}(t)) = 0\}$.

(b): Assume also that the following condition is satisfied:

(S): $g(\cdot, \cdot)$ is continuous on a neighborhood of $(\bar{x}(0), \bar{x}(1))$, and numbers $\epsilon > 0$ and $\delta > 0$ can be chosen such that

$$g(x(0), x(1)) \geq g(\bar{x}(0), \bar{x}(1)) + \delta$$

for all feasible processes $(x(\cdot), u(\cdot))$ such that $\|x(\cdot) - \bar{x}(\cdot)\|_{L^\infty} \leq \epsilon$.

Then relations (i)-(iv) above are satisfied for some set of multipliers $p(\cdot) \in W^{1,1}$, $\mu(\cdot) \in NBV^+$ and $\lambda \geq 0$, and some $\mu$-integrable function $m(\cdot)$, such that $\lambda = 0$.

In Section 5 we provide a proof, not of Thm. 3.2, but of Thm. 3.3 below. Thm. 3.3 is more general but, perhaps, of lesser interest in the context of this paper, because it does not make explicit the link between the non-coincidence of the infimum costs (over admissible processes and admissible relaxed processes) and the existence of an abnormal multiplier set. Notice that Thm. 3.3 part (b)' makes no reference at all to the cost function $g(\cdot, \cdot)$.

**Theorem 3.3.** The assertions of Thm. 3.2 remain valid when part (b) is replaced by:

(b)': Suppose that there exists $\epsilon > 0$ such that, for any process $(x(\cdot), u(\cdot))$,

$$\|x(\cdot) - \bar{x}(\cdot)\|_{L^\infty} \leq \epsilon \quad \text{implies} \quad (x(0), x(1)) \notin C \quad \text{or} \quad \max_{t\in[0,1]} h(t, x(t)) > 0'.$$

Then conditions (i)-(iv) above are satisfied for some choice of multipliers $p(\cdot) \in W^{1,1}$, $\mu(\cdot) \in NBV^+$ and $\lambda \geq 0$, and some $\mu$-integrable function $m(\cdot)$, such that $\lambda = 0$.

To show that Thm. 3.2 part (b) follows from Thm. 3.3, we must show that, under condition (S) in the Thm. 3.2, the hypotheses of Thm. 3.3 part (b)' are satisfied. Since $g(\cdot, \cdot)$ is continuous on a neighborhood of $(\bar{x}(0), \bar{x}(1))$ we can arrange, by reducing the size of $\epsilon$ if necessary, that

$$g(x(0), x(1)) - g(\bar{x}(0), \bar{x}(1)) < \delta$$

for any process $(x(\cdot), u(\cdot))$ for which $\|x(\cdot) - \bar{x}(\cdot)\|_{L^\infty} \leq \epsilon$. It follows from (S) that, for any such process $(x(\cdot), u(\cdot))$, either $(x(0), x(1)) \notin C$ or $\max_{t\in[0,1]} h(t, x(t)) > 0$.

We have confirmed the hypotheses of Thm. 3.3 part (b)' as required.

4. **An Example.** In this section we present an example of an optimal control problem which illustrates the assertions of Thms. 3.1 and 3.2. Earlier examples of optimal control problems in which the infimum costs over admissible processes and over relaxed admissible processes do not coincide are to be found, for example, in ([11], p. 246)

$$\begin{cases}
\text{Minimize } -x_1(1) \\
\text{over } (x(\cdot) = (x_1(\cdot), x_2(\cdot), x_3(\cdot)), u(\cdot)) \text{ satisfying } \\
(x_1(t), x_2(t), x_3(t)) = (0, x_1(t)u(t), x_2(t)^2) \\
u(t) \in \{-1\} \cup \{+1\} \\
x_2(0) = x_3(0) = x_3(1) = 0.
\end{cases}$$

This is an example of $(P)$, in which $n = 3$, $m = 1$, $f(t, x_1, x_2, x_3, u) = (0, x_1 u, x_2^2)$, $h(t, x) \equiv -1$, $g((x_1^0, x_2^0, x_3^0)), (x_1^1, x_2^1, x_3^1)) = -x_1^1$ and $C = (\mathbb{R} \times \{0\} \times \{0\}) \times (\mathbb{R} \times \mathbb{R} \times \{0\})$. 
Claim: \( (\bar{x}(\cdot) \equiv (0, 0, 0), \bar{u}(\cdot) \equiv 1) \) is a minimizer for \((E)\).

To validate the claim, suppose there exists an admissible process \((x(\cdot), u(\cdot))\) with lower cost than that of \((\bar{x}(\cdot), \bar{u}(\cdot))\). Since \(\dot{x}_1(t) = 0\) and the cost is \(-x_1(1)\), we must have \(x_1 \equiv k\) for some \(k > 0\). We have

\[
x_3(1) - x_3(0) = \int_0^1 |x_2(t)|^2 dt
\]

It follows from this relation and the fact that \(x_3(0) = x_3(1) = 0\), that \(x_2(\cdot) \equiv 0\). But then \(\dot{x}_2(t) \equiv 0\) a.e. Since \(\dot{x}_2(t) = ku(t)\) (for some \(k \neq 0\) and \(u(t) \in \{-1\} \cup \{+1\}\) a.e. we deduce that \(\dot{x}_2(t) \neq 0\) on a set of full measure. From this contradiction it follows that no admissible process exists with cost less than that of \((\bar{x}(\cdot) \equiv (0, 0, 0), \bar{u}(\cdot) \equiv 1)\), as claimed.

Notice however that \((\bar{x}(\cdot) \equiv (0, 0, 0), \bar{u}(\cdot) \equiv 1)\) cannot be interpreted as a relaxed local minimizer. This is because, for any \(\alpha > 0\), the relaxed admissible process

\[
(x^\alpha(\cdot) \equiv (\alpha, 0, 0),
\quad (\lambda_0^\alpha, \lambda_1^\alpha, \lambda_2^\alpha, \lambda_3^\alpha) \equiv (1/2, 1/2, 0, 0),
\quad (u_0^\alpha, u_1^\alpha, u_2^\alpha, u_3^\alpha) \equiv (+1, -1, 0, 0))
\]

(3)

has state trajectory that (by adjustment of \(\alpha\)) can be made to approximate \(\bar{x}(\cdot)\) arbitrarily closely (w.r.t. the \(L^\infty\) norm), yet has cost \(-\alpha\) which is strictly less than that of \((\bar{x}(\cdot) \equiv (0, 0, 0), \bar{u}(\cdot) \equiv 1)\).

Illustration of Thm. 3.1: The Maximum Principle conditions at \((\bar{x}(\cdot) \equiv (0, 0, 0), \bar{u}(\cdot) \equiv 1)\), in which the adjoint inclusion takes the weaker ‘averaged’ form (ii’), are as follows: there exist \(p(\cdot) = (p_1(\cdot), p_2(\cdot), p_3(\cdot))\) and \(\lambda \geq 0\), not both zero, such that

\[
-\dot{p}_1(t) = c^1_0 \{ -p_2(t) \} \cup \{ p_2(t) \},
\quad -\dot{p}_2(t) = 2p_3(t)\dot{x}_2(t),
\quad \dot{p}_3(t) \equiv 0.
\]

Because the state constraint is inactive, the state constraint multiplier \(\mu\) drops out of the condition. The Weierstrass, or maximization of Hamiltonian, condition conveys no information.

Taking note of the specified \((\bar{x}(\cdot), \bar{u}(\cdot))\), we see that \(p(\cdot)\) and \(\lambda\) must satisfy:

\[
p_1(\cdot) \equiv 0,\quad p_2(\cdot) \equiv 0,\quad p_3(\cdot) \equiv k\quad \text{and}\quad \lambda = 0.
\]

for some \(k \neq 0\). Thus there exists a set of multipliers (it is, in fact, a unique set modulo scaling of the \(k\) parameter) which is abnormal, as predicted by Thm. 3.1 part (b).

Illustration of Thm. 3.2: Now fix \(\alpha > 0\) and consider the admissible relaxed process (3). The Maximum Principle conditions, with reference to this relaxed admissible process, are as follows: there exist \(p(\cdot) = (p_1(\cdot), p_2(\cdot), p_3(\cdot))\) and \(\lambda \geq 0\), not both zero, such that

\[
-\dot{p}_1(t) = p_2(t) (\lambda_0^\alpha u_0^\alpha(t) + \lambda_1^\alpha u_1^\alpha(t)),
\quad -\dot{p}_2(t) = 2p_3(t),
\quad \dot{p}_3(t) \equiv 0.
\]

Noting the specified \((\bar{x}(\cdot), \bar{u}(\cdot))\), we conclude that, one again, \(p(\cdot)\) and \(\lambda\) must satisfy:

\[
p_1(\cdot) \equiv 0,\quad p_2(\cdot) \equiv 0,\quad p_3(\cdot) \equiv k\quad \text{and}\quad \lambda = 0.
\]

for some \(k \neq 0\). We see that there exists a set of abnormal multipliers (it is the unique multiplier set modulo scaling) which is abnormal, as predicted by Th. 3.2.
5. **Proofs.** This section provides proofs of the theorems in Sections 2 and 3. In fact it suffices to proof Thms. 3.1 and 3.3. This is because, as we have already shown, Thm. 3.2 follows from Thm. 3.3. Furthermore, Thms. 2.1 and 2.2 are special cases of Thms. 3.1 and 3.2 in which the state constraint functional \( h(.,.) \equiv -1 \).

5.1. **Proof of Thm. 3.3:** We need only proof part (b). Define the arc \( \tilde{\xi}(.) : [0,1] \rightarrow \mathbb{R}^n \) as

\[
\tilde{\xi}(t) = \int_0^t \sum_{j=0}^{n} \tilde{\lambda}^j(s)e^j ds,
\]

where \( e^0 = (0, \ldots, 0), e^1 = (1, 0, \ldots, 0), \ldots, e^n = (0, \ldots, 0, 1) \). We observe that \((\tilde{x}(.),\tilde{\xi}(.))\) is a solution of the differential inclusion

\[
(\tilde{x}(t), \tilde{\xi}(t)) \in coF(t,x(t)),
\]

in which

\[
F(t,x) := \bigcup_{k=0}^{n} \{(f(t,x,\tilde{u}^k(t)), e^k)\}.
\]

Take a sequence \( \{\rho'_i\} \subset (0,1) \) such that \( \rho'_i \downarrow 0 \). Invoking the Relaxation Theorem (see, e.g. [9, Thm. 2.7.2]), we can find, for each \( i \), a pair of arcs \((x_i(.),\xi_i(.))\) such that \( \xi_i(0) = 0 \),

\[
(\tilde{x}_i(t), \tilde{\xi}_i(t)) \in F(t,x_i(t)) \quad \text{a.e.}
\]

and

\[
||(x_i(.),\xi_i(.)) - (\tilde{x}(.),\tilde{\xi}(.))||_{L^\infty} \leq \rho'_i. \tag{4}
\]

Consider now the set \( \mathcal{S} \) of couples \((x(.),\xi(.)),(u(.),\omega(.))\) comprising absolutely continuous functions \( x(.) \) and \( \xi(.) \) and measurable functions \( u(.) \) and \( \omega(.) \):

\[
\mathcal{S} := \{((x(.),\xi(.)),(u(.),\omega(.))) | (\tilde{x}(t),\tilde{\xi}(t)) = (f(t,x(t),u(t)),\omega(t)) \}
\]

\[
(\rho(u(t),\omega(t)) \in U(t) \times V \quad \text{a.e.,} \quad \xi(0) = 0, \quad ||x - \tilde{x}||_{L^\infty} \leq \varepsilon \},
\]

where \( V := \bigcup_{k=0}^{n} \{e^k\} \).

For sufficiently small \( \varepsilon > 0 \), \((\mathcal{S},d_\mathcal{S}(.,.))\) is a complete metric space, with distance

\[
d_\mathcal{S}(((x(.),\xi(.)),(u(.),\omega(.))),((x'(.),\xi'(.),(u'(.),\omega'(.)))) :=

|\rho(0) - \rho'(0)| + m\{\rho : (u(t),\omega(t)) \neq (u'(t),\omega'(t))\}.
\]

Define the function \( J(.) : \mathcal{S} \rightarrow R \)

\[
J((x(.),\xi(.)),(u(.),\omega(.))) := \max\{d_\mathcal{S}(x(0),x(1)), \max_{t \in [0,1]} h(t,x(t))\}.
\]

\( J(.) \) is continuous on \((\mathcal{S},d_\mathcal{S}(.,.))\). For each \( i \), \((x_i(.),\xi_i(.))\) can be interpreted as the state trajectory component of a process \((x_i(.),\xi_i(.)),(u_i(.),\omega_i(.))\) for the dynamical system:

\[
\begin{align*}
(\dot{x}(t),\dot{\xi}(t)) &= (f(t,x(t),u(t)),\omega(t)) \\
(u(t),\omega(t)) &\in U(t) \times V,
\end{align*}
\]

in which, for each index value \( i \) and a.e. \( t \),

\[
(u_i(t),\omega_i(t)) = (\tilde{u}^{j(i,t)}(t),e^{j(i,t)}).
\]

Here, \( j(i,t) \) is the unique index value such that

\[
\frac{d}{dt}\xi_i(t) = e^{j(i,t)}.
\]
For some constant $K$ independent of $i$,

$$0 \leq \max_{t \in [0,1]} \{d_C(x_i(0), x_i(1)), h(t, x_i(t))\} \leq K \rho_i^\prime.$$ 

Writing $\rho_i = K \rho_i^\prime$ we see that, for each $i$, $((x_i(.), \xi_i(.)), (v_i(.), \omega_i(.)))$ is a $\rho_i$-minimizer for the optimization problem:

$$\begin{cases}
\text{Minimize } J((x(.), \xi(.)), (u(.), \omega(.))) \\
\text{over } ((x(.), \xi(.)), (u(.), \omega(.))) \in S
\end{cases}$$

In consequence of Ekeland’s Theorem, there exists, for each $i$, a minimizer $((y_i(.), \eta_i(.)), (v_i(.), \beta_i(.))) \in S$ for the perturbed optimization problem:

$$\begin{cases}
\text{Minimize } \max_{t \in [0,1]} \{d_C(x(0), x(1)), h(t, x(t))\} + \\
\rho_i^{1/2} \left( |x(0) - y_i(0)| + \int_0^1 m_i(t, (u(t), \omega(t))) dt \right) \\
\text{over } ((x(.), \xi(.)), (u(.), \omega(.))) \in S
\end{cases}$$

Furthermore,

$$|x_i(0) - y_i(0)| + \int_0^1 m_i(t, (u_i(t), \omega_i(t))) dt \leq \rho_i^{1/2}. \quad (5)$$

Here $m_i(., ., .)$ is the function

$$m_i(t, u, \omega) := \begin{cases}
1 & \text{if } (u, \omega) \neq (v_i(t), \beta_i(t)) \\
0 & \text{otherwise}
\end{cases}.$$

Under the stated hypotheses, (4) and (5) imply

$$||(y_i(.), \eta_i(.)) - (\bar{x}(.), \bar{\xi}(.))||_{L^\infty} \to 0 \text{ as } i \to \infty. \quad (6)$$

Also, along some subsequence (we do not relabel),

$$(\dot{y}_i(.), \dot{\eta}_i(.)) \to (\dot{\bar{x}}(.), \dot{\bar{\xi}}(.)) \text{ weakly in } L^1. \quad (7)$$

Define $c_i := \max_{t \in [0,1]} h(t, y_i(t))$. Then $((y_i(.), \xi_i(.), c_i(.)) = c_i), (v_i(.), \beta_i(.)))$ can be interpreted as a strong local minimizer for the state constrained optimal control problem:

$$\begin{cases}
\text{Minimize } d_C(x(0), x(1)) + c(1) + \\
\rho_i^{1/2} \left( |x(0) - y_i(0)| + \int_0^1 m_i(t, (u(t), \omega(t))) dt \right) \\
\text{over } (x(.), \xi(.), c(.)) \text{ satisfying} \\
(\dot{x}(t), \dot{\xi}(t), \dot{c}(t)) = (f(t, x(t), u(t)), \omega(t), 0) \text{ a.e.} \\
(u(t), \omega(t)) \in U(t) \times V \text{ a.e.} \\
h(t, x(t)) - c(t) \leq 0 \text{ for } t \in [0,1] \\
(\xi(0)) = 0
\end{cases}$$

(\text{Q}_i)

Here ‘$a \lor b$’ denotes ‘$\max\{a, b\}$‘. In view (6), (7) and (2),

$$\max_{t \in [0,1]} \{d_C(y_i(0), y_i(1)), h(t, y_i(t))\} > 0 \quad (8)$$

for $i$ sufficiently large.

Presently, we shall apply the state constrained Maximum Principle to (Q$_i$) (with reference to the strong local minimizer), the hypotheses for the validity of which are
here satisfied. Before doing so, we note that, following extraction of an appropriate subsequence, we may restrict attention to one of the following two cases.

Case (1): \( c_i > 0 \) for all \( i \),

Case (2): \( c_i \leq 0 \) for all \( i \).

Consider first case (1). Application of the state constrained Maximum Principle will require estimation of the ‘hybrid’ subdifferential \( \partial_{\nu_2} \hat{h}(t, x, c) \) of the state constraint functional \( \hat{h}(t, x, c) := h(t, x) - c \). In view of (9) we may express it in terms of \( \partial_{\nu} h(t, x) \), thus:

\[
\partial_{\nu_2} \hat{h}(t, y_i(t), c_i) = \partial_{\nu} h(t, x) \times \{-1\},
\]

because, in this case, ‘\( h(t, y_i) - c_i > 0 \)’ implies ‘\( h(t, y_i) > 0 \)’.

We also require the following estimate of \( \partial_\{dx(x_0, x_1) \cap c\} \):

\[
(\alpha', \beta', \gamma') \in \partial_\{d_C(x_0, x_1) \cap c\}((x_0, x_1, c) = (y_i(0), y_i(1), c_i))
\]

implies

\[
(\alpha', \beta') = (1 - \sigma)(p', q') \text{ and } \gamma' = \sigma
\]

for some \( \sigma \in [0, 1] \) and \( (p', q') \) such that

\[
\langle p', q' \rangle \in \partial d_C(y_i(0), y_i(1)) \text{ and } \|p', q'\| = 1.
\]

Properties (12)-(14) follow from the ‘max rule’ of subdifferential calculus, the fact that elements in \( \partial d_C(y_i(0), y_i(1)) \) have unit norm if \( (y_i(0), y_i(1)) \notin C \) and condition (9), which assures that, if \( d_C(y_i(0), y_i(1)) \geq c_i \), then \( d_C(y_i(0), y_i(1)) > 0 \).

The state constrained Maximum Principle [9], applied to a version of \( (Q_i) \) in which the integral cost term has been eliminated by state augmentation, together with the above subdifferential estimates, yield the following information: for each \( i \) there exist an arc \( p_i(.) \in W^{1,1} \), \( q_i(.) \in W^{1,1} \) (the value of the costate arc associated with the state component \( c \)), a function \( \mu_i(.) \in NBV^+[0, 1] \), a \( \mu_i \)-integrable function \( m_i(.) \), \( \sigma_i \in [0, 1] \) and \( \lambda_i \geq 0 \) such that

(i): \( ||p_i(.)||_{L^\infty} + ||c_i(.)||_{L^\infty} + ||\mu_i||_{T.V.} + \lambda_i = 1 \),

(ii): \( -\hat{p}_i(t) \in \co \{\partial_x H(t, y_i(t), (p_i(t) + \int_{[0,t]} m_i(s)\mu_i(ds)), v_i(t))\} \) a.e. ,

(iii): \( (p_i(0), -\left[p_i(1) + \int_{[0,1]} m_i(s)\mu_i(ds)\right]) \in\)

\[
(1 - \sigma_i)\lambda_i\{(\alpha, \beta) \in \partial d_C(y_i(0), y_i(1)) \mid |(\alpha, \beta)| = 1\} + \lambda_i\rho_i^{1/2} B,
\]

(iv): \( (p_i(t) + \int_{[0,t]} m_i(s)\mu_i(ds)) \cdot \hat{y}_i(t) \geq \max_{u \in U(t)} (p_i(t) + \int_{[0,t]} m_i(s)\mu_i(ds)) \cdot f(t, y_i(t), u) - \lambda_i\rho_i^{1/2} a.e.
\]

(v): \( m_i(t) \in \partial_{\nu_2} h(t, y_i(t)) \mu_i a.e. \text{ and } \text{supp}\{\mu_i\} \subset \{ t : h(t, y_i(t)) < c_i = 0 \} \).

(vi): \( \hat{q}_i(t) = 0 \text{ a.e.}, q_i(0) = 0, q_i(1) + \int_{[0,1]} m_i(ds) = \lambda_i\sigma_i \).

(Notice that the costate arc associated with the state component \( \eta_i(.) \) is the zero function, and so does not feature in the above conditions.)

We deduce from condition (iii) that \( ||p_i(.)||_{L^\infty} + ||\mu_i||_{T.V.} \geq (1 - \sigma_i)\lambda_i + \rho_i^{1/2} \lambda_i \). It then follows from (i) and (vi) that

\[
2||p_i(.)||_{L^\infty} + 3||\mu_i||_{T.V.} + \lambda_i \geq 1 + (1 - \sigma_i)\lambda_i - \rho_i^{1/2} \lambda_i + \sigma_i\lambda_i.
\]
Bearing in mind that $\lambda \in [0,1] \text{ and } \rho_i \leq 1/2$, we deduce that

$$2\|p_i(.)\|_{L^\infty} + 3\|\mu_i\|_{T.V.} \geq 1 - \sqrt{\frac{1}{2}}. \quad (15)$$

Next consider case (2). Conditions (8) and (10) imply that $d_C(y_i(0), y_i(1)) > c_i$. It follows that $((y_i(.) , \xi_i(.) , c_i(.)) = c_i)$, $(v_i(.) , \beta_i(.) )$ remains a strong local minimizer when $c_i$ is replaced by $c_i' := c_i + \delta$, for some suitably small $\delta > 0$. Following this adjustment, we find that state constraint $h(t, y_i(t)) - c_i' \leq 0$ is inactive on $[0,1]$. But then $((y_i(.) , \xi_i(.) , (v_i(.) , \beta_i(.) ))$ is a strong local minimizer for the state constraint-free version of $(Q_i)$. Applying the (state constraint-free) Maximum Principle, we deduce the existence of $p_i(.) \in W^{1,1}$, such that conditions (i)-(v) are satisfied with $\mu_i = 0$, $\sigma_i = 0$ and $\lambda_i = 1$.

Extracting salient information from (ii)-(vi) and (15), and scaling the multipliers appropriately, we arrive at the following properties (for either case (1) or case (2)): there exist an arc $p_i(.) \in W^{1,1}$, a function $\mu_i(.) \in NBV^+[0,1]$, a $\mu_i$-integrable function $m_i(.)$ and $\lambda_i \geq 0$ such that, for some $K \geq 0$ that does not depend on $i$,

(i)': $\|p_i(.)\|_{L^\infty} + \|\mu_i\|_{T.V.} = 1$.

(ii)' : $-\dot{p}_i(t) \in \partial H(t, y_i(t), (p_i(t) + \int_{[0,t]} m_i(s)\mu_i(ds)), v_i(t))$ a.e.,

(iii)' : $\left(p_i(0) , -\left[p_i(1) + \int_{[0,1]} m_i(s)\mu_i(ds)\right]\right) \in K\partial d_C(y_i(0), y_i(1)) + \rho_i^{1/2} KB$.

(iv)' : $\left(p_i(t) + \int_{[0,t]} m_i(s)\mu_i(ds)\right) \cdot \dot{y}_i(t) \geq 0$.

$\max_{u \in U(t)} \left[p_i(t) + \int_{[0,t]} m_i(s)\mu_i(ds)\right] \cdot f(t, y_i(t), u) - K\rho_i^{1/2} a.e.$

(v)' : $m_i(t) \in \partial^2 h(t, y_i(t)) \mu_i - a.e.$ and $\text{supp}\{\mu_i\} \subset \{t : h(t, y_i(t)) - c_i = 0\}$.

Now the functions $p_i(.)$ are uniformly bounded and have uniformly integrably bounded derivatives. It follows that, along some subsequence (we do not relabel), $p_i(.) \rightharpoonup p(.)$ strongly in $L^\infty$ and $\dot{p}_i(.) \rightharpoonup \dot{p}(.)$ weakly in $L^1$ for some $p(.) \in W^{1,1}$. We can also arrange (by subsequence extraction) that $\mu_i(.) \rightharpoonup \mu(.)$ weakly * in $C^*([0,1])$ and that $m_i(s)\mu_i(ds) \rightharpoonup m(s)\mu(ds)$ for some Borel measurable function $m(.)$ such that $m(t) \in \partial^2 h(t, \bar{x}(t)) \mu$-a.e. (The details of this standard convergence analysis are to be found in ([9], Ch. 9.)

Now define the functions $q_i(.)$:

$$r_i(t) := p_i(t) + \int_{[0,t]} m_i(s)\mu_i(ds).$$

The $r_i(.)$'s are uniformly integrably bounded and converge a.e. to the function

$$r(t) := p(t) + \int_{[0,t]} m(s)\mu(ds).$$

By the dominated convergence theorem then, $r_i(.) \rightharpoonup r(.)$ strongly in $L^1$. For any selector $u(t) \in U(t)$ we have, from (iv)',

$$\int_{[0,1]} r_i(t)\dot{y}_i(t)dt \geq \int_{[0,1]} r_i(t) \cdot f(t, y_i(t), u(t))dt - K\rho_i^{1/2}.$$

The convergence properties of the $r_i(.)$'s etc, ensure that, in the limit as $i \rightarrow \infty$,

$$\int_{[0,1]} r(t)\dot{x}(t)dt \geq \int_{[0,1]} r(t) \cdot f(t, \bar{x}(t), u(t))dt.$$
Since this relation is valid for arbitrary selectors \( u(.) \) we can conclude that
\[
(p(t)+\int_{[0,t]} m(s)\mu(ds)) \cdot \dot{x}(t) = \max_{u \in U(t)} (p(t)+\int_{[0,t]} m(s)\mu(ds)) \cdot f(t, \bar{x}(t), u) \text{ a.e. (16)}
\]

Conditions (i)', (iii)' and (v)' yield, in the limit as \( i \to \infty \),
\[
\left( p(0), -\left[ p(1) + \int_{[0,1]} m(s)\mu(ds) \right] \right) \in K\partial d_C(\bar{x}(0), \bar{x}(1)) \subset N_C(\bar{x}(0), \bar{x}(1)) \quad (17)
\]
and
\[
m(t) \in \partial^2_x h(t, \bar{x}(t)) \mu\text{-a.e. and } \supp \{ t : h(t, \bar{x}(t) = 0 \}
\]
and
\[
\|p(.)\|_{L^\infty} + \|\mu\|_{T.V.} = 1.
\] (19)

Now define
\[
\mathcal{A}_i := \{ t : (v_i(t), \beta_i(t)) \neq (u_i(t), \omega_i(t)) \}.
\]
By (5)
\[
(v_i(t), \beta_i(t)) \in \bigcup_{j=0}^{n} \{ (\bar{u}^j(t), e^j) \} \text{ a.e. } t \in [0,1] \setminus \{ \mathcal{A}_i \}.
\]
It follows from condition (ii)' that, for a.e. \( t \in [0,1] \setminus \mathcal{A}_i \),
\[
(-\dot{p}_i(t), \dot{y}_i(t), \dot{\eta}_i(t)) \in \bigcup_{j=0}^{n} \{ (\co \partial_x H(t, y_i(t), r_i(t), \bar{u}^j(t)) \times \{ (f(t, y_i(t), \bar{u}^j(t), e^j) \}) \}
\]
But (along some subsequence) \( (p_i(\cdot), y_i(\cdot), \eta_i(\cdot)) \to (p(\cdot), \bar{x}(\cdot), \bar{\xi}(\cdot)) \) uniformly, \( (\dot{p}_i(\cdot), \dot{y}_i(\cdot), \dot{\eta}_i(\cdot)) \to (\dot{p}(\cdot), \dot{x}(\cdot), \dot{\xi}(\cdot)) \) weakly in \( L^1 \). A convergence analysis similar to in ([9], Ch. 9) yields, in the limit as \( i \to \infty \),
\[
(-\dot{p}(t), \dot{x}(t), \dot{\xi}(t)) \in \co \left( \bigcup_{j=0}^{n} \{ (\partial_x H(t, \bar{x}(t), r(t), \bar{u}^j(t)) \times \{ (f(t, \bar{x}(t), \bar{u}^j(t), e^j) \}) \} \right) \text{ a.e.}
\]
By the Carathéodory Representation Theorem (see [11]), there exist measurable functions \( \lambda^j(\cdot), j = 0, 1, \ldots, n \) such that \( \{ \lambda^j(t) \} \in \Lambda \text{ a.e. and} \)
\[
(-\dot{p}(t), \dot{x}(t), \dot{\xi}(t)) \in \bigcup_{j=0}^{n} \lambda^j(t) \co \{ (\partial_x H(t, \bar{x}(t), \bar{u}^j(t)) \times \{ (f(t, \bar{x}(t), \bar{u}^j(t), e^j) \}) \} \text{ a.e.}
\]
But then
\[
\dot{\xi}(t) = \sum_{j=0}^{n} \lambda^j(t)e_j = \sum_{j=0}^{n} \bar{\lambda}^j(t)e_j, \text{ a.e.}
\]
Since the vectors \( e_0, \ldots, e_n \) are in ‘general position’, it follows that \( \lambda_j(t) = \bar{\lambda}_j(t) \) a.e., \( j = 0, \ldots, n \). But then
\[
(-\dot{p}(t), \dot{x}(t), \dot{\xi}(t)) \in \bigcup_{j=0}^{n} \bar{\lambda}^j(t) \co \{ (\partial_x H(t, \bar{x}(t), q(t), \bar{u}^j(t)) \times \{ (f(t, \bar{x}(t), \bar{u}^j(t), e^j) \}) \} \text{ a.e.}
\]
This relation implies that:
\[
-\dot{p}(t) \in \sum_{j=0}^{n} \bar{\lambda}^j(t) \co \partial_x H(t, \bar{x}(t), q(t), \bar{u}^j(t)) \text{ a.e. (20)}
\]
(Note that if, in the preceding analysis, we had omitted to augment the state trajectories \( x(.) \) by addition of the arcs \( \xi(.) \), we would have obtained a weaker form
of the adjoint differential inclusion (20), in which the \( \lambda'(.)'s \) were replaced by some other, possibly different, weight functions \( \lambda'(.)'s \). The reason for augmenting the state trajectories \( x(.) \) was precisely to ensure that they coincide.

Reviewing (16), (17), (18), (19) and (20), we see that all the assertions of Thm. 3.3 have been proved.

5.2. Proof of Thm. 3.1. Again, we need to attend only to part (b). Take a sequence \( \epsilon_i \downarrow 0 \). Under the assumed conditions, there exists a sequence of admissible relaxed trajectories \( (x_i,\{\lambda^k_i, u^k_i\})_{k=0}^n \) such that

\[
g(x_i(0), x_i(1)) < g(\bar{x}(0), \bar{x}(1))
\]

and

\[
||x_i(.) - \bar{x}(.)||_{L^\infty} \leq \epsilon_i.
\]

This last relation implies that \( x_i(.) \to \bar{x}(.) \) uniformly. Now apply Thm. 2.1, with reference to the process \( (x_i,\{\lambda^k_i, u^k_i\})_{k=0}^n \), noting that condition (S) is satisfied for \( i \) sufficiently large. We deduce the existence of \( p_i(.) \in W^{1,1}([0,1];\mathbb{R}^n) \), \( \mu_i \in NBV^+[0,1] \) and a \( \mu_i \)-integrable function \( m_i(.) \) such that:

(i): \( ||\mu_i\|_{TV} + ||p_i||_{L^\infty} = 1 \),

(ii): \( \bar{p}_i(t) = \sum_{k=0}^n \lambda^k_i(t) \partial_x H(t, x_i(t), u^k_i(t), p_i(t) + \int_{[0,t]} m_i(s)\mu_i(ds)) \) a.e. \( t \in [0,1] \),

(iii): \( (p_i(0), -(p_i(1) + \int m_i(s)\mu_i(ds))) \in N_C(x_i(0), x_i(1)) \),

(iv): \( \forall k = 0, \ldots, n, \)

\[
\left(p_i(t) + \int_{[0,t]} m_i(s)\mu_i(ds)\right) \cdot f(t, x_i(t), u^k_i(t)) = \max_{u \in \mathcal{U}(t)} \left(p_i(t) + \int_{[0,t]} m_i(s)\mu_i(ds)\right) \cdot f(t, x_i(t), u)
\]

\( \mu_i \)-a.e. and \( \text{supp}\mu_i \subset \{ t : h(t, x_i(t)) = 0 \} \).

Define the sequence of arcs

\[
q_i(t) = p_i(t) + \int_{[0,t]} m_i(s)\mu_i(ds) \quad \text{a.e. } t \in [0,1].
\]

From (ii) it follows that

\[
-\bar{p}_i(t) = \sum_{k=0}^n \lambda^k_i(t) \partial_x H(t, x_i(t), u^k_i(t), q_i(t)) \subset \sum_{k=0}^n \lambda^k_i(t) \bigcup_{u \in \mathcal{U}(t)} \partial_x H(t, x_i(t), u, q_i(t)) = \bigcup_{u \in \mathcal{U}(t)} \partial_x H(t, x_i(t), u, q_i(t))
\]

a.e. \( t \in [0,1] \). It is easy to check that both the sequences \( \{x_i(.)\} \) and \( \{p_i(.)\} \) are uniformly bounded. Furthermore \( \{\dot{x}_i(.)\} \) and \( \{p_i(.)\} \) are equintegrable. From the Compactness of Trajectories Theorem [9, Thm. 2.5.3], it follows that \( x_i(.) \to \bar{x}(.) \) uniformly and \( \dot{x}_i(.) \to \dot{\bar{x}}(.) \) weakly in \( L^1 \), as well as \( p_i(.) \to \bar{p}(.) \) uniformly and \( \dot{p}_i(.) \to \dot{\bar{p}}(.) \) weakly in \( L^1 \). By further subsequence extraction, we have that \( \mu_i \to \mu \) in the weak* topology of \( NBV^+[0,1] \). A standard convergence analysis (see, e.g. [9, Chap. 9]) permits us to pass to the limit in (i)'-(v)', and thereby show:

(i): \( ||\mu\|_{TV} + ||\bar{p}\|_{L^\infty} \neq 0 \);

(ii): \( \bar{p}(t) \in \partial_x H(t, \bar{x}(t), \bar{p}(t) + \int_{[0,t]} m(s)\mu(ds), u) \) a.e. \( t \in [0,1] \);

(iii): \( (p(0), -(p(1) + \int_{[0,1]} m(s)\mu(ds))) \in N_C(\bar{x}(0), \bar{x}(1)) \);
Consider the optimal control problem

\[(iv): \left(p(t) + \int_{[0,t]} m(s)\mu(ds)\right) \cdot f(t, \bar{x}(t), \bar{u}(t)) = \max_{u \in U(t)} \left(p(t) + \int_{[0,t]} m(s)\mu(ds)\right) \cdot f(t, \bar{x}(t), u) \quad \text{a.e. } t \in [0,1];\]

\[(v): m(t) \in \partial^2 \mathcal{H}(t, \bar{x}(t)) \quad \mu - \text{a.e.} \quad \text{and} \quad \text{supp}\{\mu\} \subset \{t : h(t, \bar{x}(t)) = 0\}.\]

These are the desired necessary conditions, in which the cost multiplier is zero. The proof is complete.

6. A Special Case: Affine Control Dependence. We recall an ‘untidy’ feature of our earlier Type (A) Thm. 3.1 (and of Thm. 2.1) that the non-normal necessary conditions in part (b) are expressed not in terms of the standard adjoint relation, but a weaker, ‘averaged’ version. We show in this section that a cleaner version of Thm. 3.1 (and therefore of Thm. 2.1), involving the expected adjoint relation, is valid if we place restrictions on the nature of the dynamic constraint.

Consider the optimal control problem

\[
(A) \quad \begin{cases}
\text{Minimize } g(x(0), x(1)) \\
\text{over } x(\cdot) \in W^{1,1} \text{ and measurable functions } u(\cdot) \text{ satisfying}
\end{cases}
\]

\[
\dot{x}(t) = f_1(t, x(t)) + B(t, x(t))u(t) \quad u(t) \in U(t) \text{ a.e. } t \in [0,1] \quad (x(0), x(1)) \in C
\]

\[
h(t, x(t)) \leq 0 \quad \forall t \in [0,1]
\]

the data for which comprise functions \(g(\cdot, \cdot): \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}, f_1(\cdot, \cdot): [0,1] \times \mathbb{R}^n \rightarrow \mathbb{R}^n, B(\cdot, \cdot): [0,1] \times \mathbb{R}^n \rightarrow M_{n \times m}\) (where we denote with \(M_{n \times m}\) the set of matrices of dimension \(n \times m\)) and \(h(\cdot, \cdot): [0,1] \times \mathbb{R}^n \rightarrow \mathbb{R}\), a closed set \(C\) and a measurable multifunction \(U(\cdot)\). The main feature of problem (A) is that the right side of the underlying controlled differential equation

\[
\dot{x}(t) = f_1(t, x(t)) + B(t, x(t))u(t)
\]

is affine with respect to the control. Now the Hamiltonian takes the form

\[
H(t, x, p, u) = p \cdot (f_1(t, x) + B(t, x)u).
\]

The following theorem is a refinement of Thm. 3.1(b) (a Type A relation), involving the improved adjoint relation (ii), for Problem (A).

**Theorem 6.1.** Take a strong local minimizer \((\bar{x}(\cdot), \bar{u}(\cdot))\) for problem (A). Assume that the hypotheses of Thm. 3.1 are satisfied (with reference to \((\bar{x}(\cdot), \bar{u}(\cdot))\)) when we identify \(f(t, x, u) = f_1(t, x) + B(x)u\). Assume, furthermore, that \(B(\cdot, \cdot)\) is bounded on bounded sets, for some \(\bar{\epsilon} > 0\) and \(\alpha > 0\) \(B(t, \cdot)\) in continuously differentiable on \(\bar{x}(t) + \bar{\epsilon}B\) for a.e. \(t \in [0,1]\) and det \(\{B^T(t, x)B(t, x)\} \geq \alpha\) for all \(x \in \bar{x}(t) + \bar{\epsilon}B\), a.e. \(t\) in \([0,1]\).

Suppose that, for each \(\delta > 0\), there exists a relaxed admissible process \((x(\cdot), \{\lambda^k(\cdot), u^k(\cdot)\}_{k=0}^n)\) for which

\[
g(x(0), x(1)) < g(\bar{x}(0), \bar{x}(1))
\]

and

\[
\|x(\cdot) - \bar{x}(\cdot)\|_{L^\infty} \leq \delta
\]

(i.e. \((\bar{x}(\cdot), \bar{u}(\cdot))\) is not also a relaxed strong local minimizer for problem (A)).
Then there exist a \(p(.) \in W^{1,1} \), a function \(\mu(.) \in NBV^+[0,1] \) and a \(\mu\)-integrable function \(m(.)\) satisfying the following conditions (i.e. the necessary conditions of Thm. 3.1(a) with cost multiplier set to zero):

(i): \(||\mu||_{TV} + ||p(.)||_{L^\infty} = 1\);
(ii): \(-\tilde{\mu}(t) \in \co \partial_x H(t, \tilde{x}(t), p(t) + \int_{[0,t]} m(s)\mu(ds), \tilde{u}(t)) \) a.e. ,
(iii): \(\left(p(0), -\left(p(1) + \int_{[0,1]} m(s)\mu(ds)\right)\right) \in N_C(\tilde{x}(0), \tilde{x}(1))\),
(iv): \(H(t, \tilde{x}(t), p(t) + \int_{[0,t]} m(s)\mu(ds), \tilde{u}(t))\)

\[\begin{align*}
= \max\{H(t, \tilde{x}(t), p(t) + \int_{[0,t]} m(s)\mu(ds), u) \mid u \in U(t)\} \text{ a.e. },
\end{align*}\]
(v): \(m(t) \in \partial^*_\mu h(t, \tilde{x}(t))\) - a.e. and \(\text{supp}\{\mu\} \subset \{t : h(t, \tilde{x}(t)) = 0\}\).

Proof. Take a sequence \(\epsilon_i \downarrow 0\). Under the assumptions of the theorem statement, there exists a sequence of relaxed processes \((x_i(.), \{\lambda^k_i(.), u^k_i(.)\}_{k=0}^n)\) such that

\[g(x_i(0), x_i(1)) < g(\tilde{x}(0), \tilde{x}(1))\]

and

\[||x_i(.) - \tilde{x}(.)||_{L^\infty} < \epsilon_i.\]

The hypotheses under which we may apply Thm. 3.2(b) are satisfied with reference to each of the relaxed processes \((x_i(.), \{\lambda^k_i(.), u^k_i(.)\}_{k=0}^n)\), in consequence there exist an absolutely continuous function \(p_i(.)\), a function \(\mu_i(.) \in NBV^+[0,1] \) and a \(\mu_i\)-integrable function \(m_i(.)\) such that:

(i): \(||\mu_i||_{TV} + ||p_i(.)||_{L^\infty} = 1\); 
(ii): \(-\tilde{\mu}_i(t) \in \co \partial_x((p_i(t) + \int_{[0,t]} m_i(s)\mu_i(ds) \cdot f_i(t, x_i(t))) + \nabla_x B(t, x_i(t)) \sum_{k=0}^n \lambda^k_i(t)u^k_i(t)) \) a.e. \(t \in [0,1]\);
(iii): \(\left(p_i(0), -(p_i(1) + \int_{[0,1]} m_i(s)\mu_i(ds))\right) \in N_C(x_i(0), x_i(1))\);
(iv): for \(k = 0, \ldots, n, \)

\[p_i(t) + \int_{[0,t]} m_i(s)\mu_i(ds) \cdot (f_i(t, x_i(t)) + B(t, x_i(t))u^k_i(t)) \geq \max_{u \in U(t)} \left(p_i(t) + \int_{[0,t]} m_i(s)\mu_i(ds) \cdot (f_i(t, x_i(t)) + B(t, x_i(t))u)\right), \text{ a.e. } t \in [0,1].\]
(v): \(m_i(t) \in \partial^*_\mu h(t, x_i(t))\) - a.e. and \(\text{supp}\{\mu_i\} \subset \{t : h(t, x_i(t)) = 0\}\).

Notice that we have made use of the continuous differentiability of \(B(.t)\) to write the adjoint inclusion as (ii)'.

From the Compactness of Trajectories Theorem, \(x_i(.) \to \tilde{x}(.)\) in \(L^\infty\) and \(\dot{x}_i(.) \to \dot{\tilde{x}}(.)\) weakly in \(L^1\), along a subsequence. We can arrange, by further subsequence extraction, that \(\mu_i \to \mu\) in the weak \(NBV^+[0,1]\) topology, \(p_i(.) \to p(.)\) uniformly and \(\dot{p}_i(.) \to \dot{p}(.)\) weakly in \(L^1\), for some \(\mu\) and \(p(.)\) and there exists a Borel measurable function \(m(.)\) such that \(m_i(.)d\mu(s) \to m(s)d\mu(s)\) weakly*.

Now write \(u_i(t) = \sum_{k=0}^n \lambda^k_i(t)u^k_i(t).\) Then, for each \(i\) and a.e. \(t \in [0,1],\)

\[u_i(t) = (B^T(t, x_i(t))B(t, x_i(t)))^{-1}B^T(t, x_i(t))\dot{x}_i(t) - f_i(t, x_i(t))\]

(The hypotheses ensure that the \((B^T(t, x)B(t, x))\) matrices are invertible and their inverses are uniformly bound on a tube about \(\tilde{x}(.)\).) We deduce from this equation and the weak \(L^1\) convergence of the \(x_i(.)\)'s to \(\tilde{x}(.)\), that

\[u_i(.) \to \bar{u}(.) \text{ weakly in } L^1,\]
in which \( \bar{u}(.) \) is the unique control such that

\[
\dot{x}(t) = f_1(t, \bar{x}(t)) + B(t, \bar{x}(t))\bar{u}(t) \quad \text{a.e.}
\]

A standard analysis (see \[9, \text{Chap. 9}\]) permits us to obtain from (i)', (iii)', (iv)' and (v)', in the limit as \( i \to \infty \),

(i): \( ||\mu||_{L^2(V)} + ||p(.)||_{L^\infty} = 1 \),

(iii): \( (p(0), -(p(1) + \int_{[0,1]} m(s)\mu(ds))) \in N_C(\bar{x}(0), \bar{x}(1)) \),

(iv): \( (p(t) + \int_{[0,t]} m(s)\mu(ds)) \cdot (f(t, \bar{x}(t)) + B(t, \bar{x}(t))\bar{u}(t)) = \max_{u \in U(t)} \left\{ (p(t) + \int_{[0,t]} m(s)\mu(ds)) \cdot (f(t, \bar{x}(t)) + B(t, \bar{x}(t))u) \right\} \text{a.e.,} \)

(v): \( m(t) \in \partial^*_x h(t, \bar{x}(t)) \mu - \text{a.e. and supp}\{\mu\} \subset \{t : h(t, \bar{x}(t)) = 0\} \).

Let us attend to the adjoint relation (ii)'). Write, for each \( i \),

\[
z_i(t) = p_i(t) + \int_{[0,t]} (p_i(t') + \int_{[0,t']} m_i(s)\mu_i(ds)) \cdot \nabla_x B(t', x_i(t')) h_i(t') dt'.
\]

Then \( \{z_i(.)\} \) is a sequence of absolutely continuous functions such that \( z_i(.) \to z(.) \) strongly in \( L^1 \) and \( \dot{z}_i(.) \to \dot{z}(.) \) weakly in \( L^1 \) for some absolutely continuous \( z(.) \) such that

\[
z(t) = p(t) + \int_{[0,t]} (p(t') + \int_{[0,t']} m(s)\mu(ds)) \cdot \nabla_x B(t', \bar{x}(t'))\bar{u}(t') dt'.
\]

Each \( z_i(.) \) satisfies the differential inclusion

\[
-\dot{z}_i(t) \in \text{co} \partial_x \{(p_i(t) + \int_{[0,t]} m_i(s)\mu_i(ds)) \cdot f_1(t, x_i(t))\}.
\]

With the help of the Compactness of Trajectories Theorem we obtain

\[
-\dot{z}(t) \in \text{co} \partial_x \{(p(t) + \int_{[0,t]} m(s)\mu(ds)) \cdot f_1(t, \bar{x}(t))\} \quad \text{a.e.}
\]

in the limit as \( i \to \infty \). Replacing \( \dot{z}(.) \) by its derivative

\[
\dot{z}(t) = \dot{p}(t) + (p(t) + \int_{[0,t]} m(s)\mu(ds)) \cdot \nabla_x B(t, \bar{x}(t))\bar{u}(t)
\]

we obtain

(ii): \( -\dot{p}(t) \in \text{co} \partial_x \left( \left( p(t) + \int_{[0,t]} m(s)\mu(ds) \right) f_1(t, \bar{x}(t)) \right) + \left( p(t) + \int_{[0,t]} m(s)\mu(ds) \right) \cdot \nabla_x B(t, \bar{x}(t))\bar{u}(t) \quad \text{a.e. } t \in [0,1]. \)

All the asserted relations have been verified. The proof is complete.

\[\square\]

REFERENCES

[1] A. L. Dontchev, R. T. Rockafellar, Implicit Functions and Solution Mappings: A View from Variational Analysis, Monographs in Mathematics, Springer, Berlin 2009.

[2] A. Ioffe, Euler Lagrange and Hamiltonian Formalisms in Dynamic Optimization, Trans. Amer. Math. Soc., 349, 1997, pp. 2871-2900.

[3] A. Ioffe, Optimal control of Differential Inclusions: New Developments and Open Problems, Proc. 41 IEEE Conference on Decision and Control, 349, 2002, pp. 3127-3132.

[4] F. H. Clarke, The Maximum Principle under Minimal Hypotheses, SIAM J. Control and Opt., 14, 1976, pp. 1078-1091.

[5] F. H. Clarke, Optimization and Nonsmooth Analysis, Wiley-Interscience, New York, 1983, reprinted as vol. 5 of Classics in Applied Mathematics, SIAM, Philadelphia, 1990.
WHEN ARE MINIMIZING CONTROLS ALSO MINIMIZING RELAXED CONTROLS?  

[6] F. H. Clarke, Y. S. Ledyaev, R. J. Stern and P. R. Wolenski, Nonsmooth Analysis and Control Theory, Graduate Texts in Mathematics Vol. 178, Springer Verlag, New York, 1998
[7] M. Palladino and R. B. Vinter, Minimizers that are not also Relaxed Minimizers, SIAM J. Control and Optim., 52, 4, 2014, pp. 2164 - 2179.
[8] T. T. Rockafellar and J.-B. Wets, Variational Analysis, Grundlehren der Mathematischen Wissenschaft, vol. 317, Springer Verlag, New York, 1998
[9] R. B. Vinter, Optimal Control, Birkhäuser, Boston, 2000.
[10] J. Warga, Normal Control Problems have no Minimizing Strictly Original Solutions, Bulletin of the Amer. Math. Soc., 77, 4, 1971, pp. 625-628.
[11] J. Warga, Optimal Control of Differential and Functional Equations, Academic Press, New York, 1972.
[12] J. Warga, Controllability, extremality, and abnormality in nonsmooth optimal control, J. Optim. Theory and Applic., 41, 1, 1983, pp. 239-260.
[13] J. Warga, Optimization and Controllability Without Differentiability Assumptions, SIAM J. Control and Optim., 21, 6, 1983, pp. 837-855.

Received xxxx 20xx; revised xxxx 20xx.

E-mail address: m.palladino@imperial.ac.uk
E-mail address: r.vinter@imperial.ac.uk