Trajectories Satisfying a State Constraint: Improved Estimates and New Non-Degeneracy Conditions

Piernicola Bettiol and Richard Vinter

Abstract

For a state-constrained control system described by a differential inclusion and a single functional inequality state constraint, it is known that, under an ‘inward pointing condition’, the $W^{1,1}$ distance of an arbitrary state trajectory to the set of state trajectories which have the same left endpoint and which satisfy the state constraint, is linearly related to the state constraint violation. In this paper we show that, in situations where the state-constrained control system is described instead by a controlled differential equation, this estimate can be improved by replacing the $W^{1,1}$ distance on state trajectories by the Ekeland metric of the distance of the control functions. A counter-example reveals that a refinement of this nature is not in general valid for state constrained differential inclusions. Finally we show how the refined estimates may be used to establish new conditions for non-degeneracy of the state constrained Maximum Principle, in circumstances when the data depends discontinuously on the control variable.

Keywords: Optimal Control, Differential Inclusions, State Constraints, Sensitivity.
I. INTRODUCTION

Control systems subject to a state constraint have been widely studied, either in a controlled differential equations framework:

\[ \dot{x}(t) = f(t, x(t), u(t)) \ \text{a.e.} \ t \in [S, T] \]
\[ u(t) \in U(t) \]
\[ x(t) \in A(t) \quad \text{for all} \ t \in [S, T], \]

in which \([S, T]\) is a given interval, \(f(., ., .) : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n\) is a given function, and \(U(.) : \mathbb{R} \rightharpoonup \mathbb{R}^m\) is a given multifunction, or a differential inclusions framework

\[ \dot{x}(t) = F(t, x(t)) \ \text{a.e.} \ t \in [S, T] \]
\[ x(t) \in A(t) \quad \text{for all} \ t \in [S, T], \]

in which \(F(., .) : \mathbb{R} \times \mathbb{R}^n \rightharpoonup \mathbb{R}^n\) is a given multifunction.

We shall assume, in either case, that the time dependent state constraint set \(A(t)\) has the functional inequality representation

\[ A(t) = \{ x \in \mathbb{R}^n : h(t, x) \leq 0 \}, \quad S \leq t \leq T. \]

for some given function \(h(., .) : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}\). With regard to (1), we shall refer to a couple \((x(.), u(.))\), comprising a measurable function \(u(.) : [S, T] \rightarrow \mathbb{R}^m\) and an absolutely continuous function \(x(.) : [S, T] \rightarrow \mathbb{R}^n\) which satisfy \(\dot{x}(t) = f(t, x(t), u(t))\) and \(u(t) \in U(t)\) a.e., as a process. The function \(x(.)\) is called a state trajectory and the function \(u(.)\) is called a control function. With regard to (3), a state trajectory \(x(.)\) will be taken to mean an absolutely continuous function that satisfies \(\dot{x}(t) \in F(t, x(t))\). In either context, feasible state trajectories are state trajectories that satisfy the state constraint (2). We sometimes make explicit the underlying time interval \([S', T']\), which may be a subset of \([S, T]\), by writing ‘state trajectory on \([S', T']\)’, etc.

In the investigation of properties of state constrained control systems, an important role is played by estimates governing the distance of a given state trajectory \(\hat{x}(.)\) from the set of feasible state trajectories, of the type:

\[ \inf \{ m(x(.), \hat{x}(.)) \} \leq K \ h^+(\hat{x}(.)) \]
in which the infimum is taken over feasible state trajectories \( x(,) \) satisfying \( x(S) = \hat{x}(S) \). Here, \( m(, ,) \) is a metric on the set of state trajectories (with fixed left endpoint), the non-negative number

\[
h^+(\hat{x}(,)) = \max_{t \in [S,T]} (h(t, x(t)) \lor 0)
\]

quantifies the extent to which \( \hat{x} \) fails to satisfy the state constraint, and \( K \) is some number that does not depend on \( \hat{x}(,) \). Indeed, such estimates have been used to provide conditions under which the assertions of the state constrained Maximum Principle are non-degenerate, to validate characterizations of the minimum cost function in terms of the Hamilton Jacobi equation and in the derivation of sensitivity relations [7], [8], [3].

The stronger the metric \( m(, ,) \), the more information is conveyed by the estimate. First estimates of this nature [11], [6], obtained under an ‘inward pointer hypothesis, employed the \( L^\infty \) metric

\[
m(x(,), \hat{x}(,)) = ||x(,) - \hat{x}(,)||_{L^\infty}.
\]

These were subsequently improved to estimates involving the \( W^{1,1} \) metric

\[
m(x(,), \hat{x}(,)) = ||\dot{x}(,) - \dot{\hat{x}}(,)||_{L^1}.
\]

See [1]. (It was earlier claimed [6], [7], [8] and that this estimate was valid for multiple state constraints; [1] provides a counterexample illustrating that this is not in general the case and corrects earlier proofs relating the the one state constraint case.)

It was recently shown (see Lemma 4.3 of [2]) that if \( F \) does not depend on \( (t, x) \) and \( h(t, x) = b \cdot x \) for some fixed vector \( b \), then the \( W^{1,1} \) metric can be replaced by the Ekeland metric on velocities of the state trajectories:

\[
m(x(,), \hat{x}(,)) = \text{meas} \{ t \mid \dot{x}(t), \dot{\hat{x}}(t) \}.
\]

The purpose of this paper is to explore whether estimates involving Ekeland type metrics are still valid when we allow \( F(, ,) \) and \( h(, ,) \) to be \( (t, x) \) dependent, and, if they are, to investigate the implications of such estimates. The outcome is perhaps not what one would expect:

(i): A counter-example illustrates that Ekeland type estimates are not in general valid within the differential inclusions framework, when \( F \) depends on \( x \).
(ii): A related estimate is valid in the controlled differential equations framework however, for 
\((t, x)\)-dependent \(f(., ., .)\’s\) and \(h(., .)\’s\), involving the Ekeland metric on control functions

\[
m((x(., u(./)), (\dot{x}(., \dot{u}(./)) = \text{meas}\{t | u(t), \dot{u}(t)\}.
\]

The first uses of the Ekeland metric in optimal control were to prove versions of the Maximum Principle for situations in which \(f(t, x, u)\) is discontinuous in \(u\), or there is in integral cost function discontinuous in the control or velocity variable. (See [5], [4]). Consistent with this early work, we show that that the Ekeland type estimates of this paper permit us to derive new conditions of non-degeneracy of the state constrained Maximum Principle covering situations in which \(f(t, x, u)\) is discontinuous in \(u\).

One final point: a backward look over the field of control theory would seem to indicate that the controlled differential equation and differential inclusion frameworks are virtually interchangeable, since parallel theories of necessary conditions and dynamic programming can be developed within either one. The findings of this paper are rather unusual then, because they reveal one area of state constrained control (dealing with discontinuities in control and velocity variables) where the choice of framework is highly significant.

Concerning the notation adopted in this paper, we denote by \(B\) the closed unit ball \(\{x \mid |x| \leq 1\}\) \((|x|\) is the length of a vector \(x\) in an Euclidean space). Take an extended-valued function \(\varphi : \mathbb{R}^k \to \mathbb{R} \cup \{+\infty\}\). Then the effective domain of \(\varphi\) is the set \(\text{dom} \varphi := \{x \in \mathbb{R}^k \mid \varphi(x) < +\infty\}\). The epigraph of \(\varphi\) is the set \(\text{epi} \varphi := \{(r, x) \in \mathbb{R}^{1+k} \mid r \geq \varphi(x)\}\). Now take a set \(D \subset \mathbb{R}^k\). The indicator function of the set \(D\), \(\chi_D\), is the function on \(\mathbb{R}^k\) taking value 0 on the set \(D\) and 1 on its complement. We write \(z_i \overset{D}{\to} z\) to describe a sequence \(\{z_i\}\) converging to \(z\) such that \(z_i \in D\) for all \(i\).

We also gather together some definitions that will be used in the application reported in Section IV. All the material is standard, and may be found in a number of texts, [10] and [12] for example.

**Definition 1.1:** Take 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

$$N_D(\bar{x}) := \{ p \mid \exists x_i \overset{D}{\rightarrow} \bar{x}, p_i \rightarrow p \text{ s.t.} \limsup_{x_i \rightarrow x_i} \frac{p_i \cdot (x - x_i)}{|x - x_i|} \leq 0 \text{ for all } i \in \mathbb{N} \}.$$  

**Definition 1.2:** Take an open set $\mathcal{O} \subset \mathbb{R}^k$, a lower semicontinuous function $\varphi : \mathcal{O} \rightarrow \mathbb{R} \cup \{+\infty\}$ and a point $\bar{x} \in \mathcal{O}$ such that $\bar{x} \in \text{dom } \varphi$. The subdifferential of $\varphi$ at $\bar{x} \in \text{dom } \varphi$ is

$$\partial \varphi(\bar{x}) = \{ \xi \mid \exists \xi_i \to \xi \text{ and } x_i \overset{\text{dom } \varphi}{\rightarrow} \bar{x} \text{ such that} \limsup_{x_i \rightarrow x_i} \frac{\xi_i \cdot (x - x_i) - \varphi(x) + \varphi(x_i)}{|x - x_i|} \leq 0 \text{ for all } i \in \mathbb{N} \}.$$  

Given a function $\Phi(.,.)$ of two vector variables $(x, y)$ and a point $(\bar{x}, \bar{y}) \in \text{dom } \Phi(.,.)$, we denote the subgradient of $\Phi(., \bar{y})$ at $(\bar{x}, \bar{y})$ by $\partial_x \Phi(\bar{x}, \bar{y})$.

We recall that the subdifferential of a lower semi-continuous extended valued function $\varphi$ at a point $\bar{x} \in \text{dom } \varphi$ can be expressed in terms of the normal cone of the epigraph of $\varphi$:

$$\partial \varphi(\bar{x}) = \{ \xi \mid (\xi, -1) \in N_{\text{epi } \varphi}(\bar{x}, \varphi(\bar{x})) \}. \quad (5)$$

**II. The Main Result: A Linear Distance Estimate**

Consider the state constrained control system (1)-(2) of the Section 1. This section provides an estimate on the distance of a control function $\hat{u}(.)$, associated with some process $\hat{x}(.)$, $\hat{u}(.)$, to the set of control functions $u(.)$ associated with a feasible process $(x(\cdot), u(\cdot))$ for which the initial state of $x(\cdot)$ coincides with that of $\hat{x}(\cdot)$. The estimate is a linear estimate in terms of the state constraint violation $h^+(\hat{x}(\cdot))$ of the state trajectory $\hat{x}(\cdot)$. The distance of two control functions $u(.)$ and $v(.)$ on the same interval $[S', T']$ is taken to be the Ekeland distance:

$$d_{[S', T']}(u(.), v(.)) = \text{meas } \{ t \in [S', T'] \mid u(t) \neq v(t) \}.$$  

The state constraint violation index $h^+(x(\cdot))$ of a state trajectory $x(\cdot)$ on $[S', T']$ is defined to be

$$h^+(\hat{x}(\cdot)) = \max_{t \in [S', T']} \{ h(t, \hat{x}(t)) \vee 0 \}.$$  

We shall invoke the following hypotheses in which $r_0$ is a given positive number: there exist $c > 0$, $\gamma > 0$, $\rho > 0$, $k_f(.) \in L^1([S, T]; \mathbb{R})$, $k_h > 0$ and $k'_h > 0$ such that

(D1): $f(., x, .)$ is $\mathcal{L} \times \mathcal{B}^m$ measurable for each $x$, where $\mathcal{L}$ denotes the Lebesgue subsets of $\mathbb{R}$ and $\mathcal{B}^m$ the Borel subsets of $\mathbb{R}^m$.  

December 11, 2009
(D2): \(|f(t, x, u)| \leq c(1 + |x|)\) for all \((t, x) \in [S, T] \times \mathbb{R}^n, u \in U(t)\).

(D3): \(|f(t, x, u) - f(t, x', u)| \leq k_f(t)|x - x'|\) for all \(t \in [S, T], x, x' \in e^{e|T-S|}(1 + r_0)B\) and \(u \in U(t)\).

(D4): There exist a subset \(\mathcal{S} \subset [S, T]\) of full measure such that

(i): \(|h(t, x) - h(t', x')| \leq k_h |(t, x) - (t', x')|\) for all \((t, x), (t', x') \in [S, T] \times e^{e|T-S|}(1 + r_0)B\);

(ii): \(|\nabla h(t, x) - \nabla h(t, x')| \leq k'_h |x - x'|\) for all \(t \in \mathcal{S}\) and \(x, x' \in e^{e|T-S|}(1 + r_0)B\) such that \(\nabla h(t, x)\) and \(\nabla h(t, x')\) exist;

(iii):

\[
\min_{u \in U(t)} \nabla h(t, x) \cdot (1, f(t, x, u)) \leq -\gamma
\]  

for all \((t, x) \in [S, T] \times e^{e|T-S|}(1 + r_0)B\) for which \(\|h(t, x)\| \leq \rho\) and \(\nabla h(t, x)\) exists.

**Theorem 2.1:** Take any \(r_0 > 0\). Assume that hypotheses (D1)-(D4) above are satisfied. Then there exists a constant \(K\) (whose magnitude depends only on \(r_0, c, \gamma, \rho, k_F(.), k_h, k'_h\) and \(|T - S|\)) with the following property: given any \(S' \in [S, T]\) and process \((x(.), u(.))\) on \([S', T]\) such that \(\dot{x}(S') \in A(S') \cap r_0B\), there exists a process \((\hat{x}(.), \hat{u}(.))\) on \([S', T]\) with \(x(S') = \hat{x}(S')\) such that

\[x(t) \in A(t) \quad \text{for all } t \in [S', T]\]

and

\[d_{[S', T]}(u(.), \hat{u}(.)) \leq K h^+(\hat{x}(.))\]  

(7)

The theorem is proved in Section 5.

In earlier related results, the distance of the process \((x(.), u(.))\) (on \([S, T]\)) satisfying the state constraint to the reference trajectory \((\hat{x}(.), \hat{u}(.))\) is the \(L^\infty\) distance between the state trajectories or, in more refined estimates, the \(W^{1,1}\) distance. Thm. 2.1 strengthens such estimates. Indeed, under the stated hypotheses and for any fixed \(r_0 > 0\), there exist constants \(K'\) such that, given two processes \((x'(.), u'(.)\)) and \((x''(.), u''(.)\)) satisfying \(x'(S) = x''(S)\) and \(|x'(S)| \leq r_0\), we have

\[||x'(.) - x''(.)||_{L^\infty(S, T)} \leq ||x'(.) - x''(.)||_{W^{1,1}(S, T)} \leq K' d_{[S', T]}(u'(.), u''(.))\]

Furthermore, these inequalities cannot be reversed. In Section IV we shall illustrate the benefits of the new, refined estimate, using it to derive improved conditions for non-degeneracy of the state constrained Maximum Principle.
III. RELATED ESTIMATES FOR DIFFERENTIAL INCLUSIONS: A COUNTER-EXAMPLE.

A widely employed, alternative description of a control system to that of Section 1 incorporates, in place of a controlled differential equation, a differential inclusion:

\[
\dot{x}(t) \in F(t, x(t)) \quad \text{a.e. } t \in [S, T] \\
\h(t, x(t)) \leq 0.
\] (8)

Here, the new ingredient is the multifunction \( F(., .) : [S, T] \times \mathbb{R}^n \rightrightarrows \mathbb{R}^n \) taking values closed subsets of \( \mathbb{R}^n \). Absolutely continuous arcs on \([S, T]\) that satisfy the differential inclusion are called \( F\)-trajectories.

Under the usual ‘linear growth, measurably, Lipschitz’ hypotheses on the multifunction and under an ‘inward pointing’ condition near the boundary of the state constraint set, the differential inclusion (8) is known to have the following property [1]: Fix \( r_0 > 0 \). Then there exists a number \( K \) such that given any \( F\)-trajectory \( \hat{x}(.) \) such that \( \h(S, \hat{x}(S)) \leq 0 \), an \( F\)-trajectory \( x(.) \) with the same initial value and satisfying the state constraint such that

\[
||x(.) - \hat{x}(.)||_{W^{1,1}} \leq K \h^+(\hat{x}(.)).
\]

Thm. 2.1 would suggest a refinement of this estimate for differential inclusions, namely

\[
d_{[S,T]}(\dot{x}(.), \dot{x}(.)) \leq K \h^+(\dot{x}(.)).
\] (9)

The following example illustrates that such a refinement is not valid in the differential inclusions context.

**Example.** Take \([S, T] = [0, 1]\) and choose the autonomous differential inclusion in \( \mathbb{R}^2 \) and state constraint functional to be defined by:

\[
F(x = (x_1, x_2)) = \{x_2\} \times [-1, +1] \\
\h(x = (x_1, x_2)) = x_2.
\] (10)

For this choice of \( F \) the linear growth, measurably Lipschitz hypotheses are satisfied. Furthermore, \( F \) satisfies

\[
\min_{v \in F(x)} \nabla h(x) \cdot v = -1
\]
for all $x$, that is the ‘inward pointing condition’. For each $\epsilon \in (0,1]$ consider the $F$-trajectory $\hat{x}_\epsilon = (\hat{x}_1, \hat{x}_2)$:

$$
\begin{bmatrix}
\hat{x}_1(t) \\
\hat{x}_2(t)
\end{bmatrix} =
\begin{bmatrix}
1/2t^2\chi_{[0,\epsilon]} + (1/2\epsilon^2 + \epsilon(t - \epsilon))\chi_{(\epsilon,1]} \\
t\chi_{[0,\epsilon]} + \epsilon\chi_{(\epsilon,1]}
\end{bmatrix}.
$$

Notice that, for each $\epsilon$,

$$
h^+(\hat{x}_\epsilon(.)) = \epsilon.
$$

Yet for any $F$-trajectory such that $x(0) = \hat{x}_\epsilon(0) = (0, 0)$ and satisfying the state constraint, we have

$$
d_{[S,T]}(\hat{x}(.), \hat{x}_{\epsilon}) = 1 = (1/\epsilon) h^+(\hat{x}_\epsilon(.)).
$$

In view of the inequality, no $K > 0$ can be found such that (9) is satisfied.

On the other hand consider the following control system in $\mathbb{R}^2$

$$
\dot{x}(t) = (\dot{x}_1(t), \dot{x}_2(t)) = (x_2(t), u(t))
$$

$$
u(t) \in [-1, +1]
$$

with the same constraint (that is defined by $h(x) = x_2$). Clearly (11) is equivalent to (10) and the “violating” process $(\hat{x}_\epsilon(.), \hat{u}_\epsilon(.))$ with initial data $\hat{x}_\epsilon(0) = (0, 0)$ and where

$$
\begin{align*}
\hat{u}_\epsilon &= \begin{cases} 
1 & \text{on } [0, \epsilon] \\
0 & \text{on } (\epsilon, 1]
\end{cases}
\end{align*}
$$

corresponds to the $F$-trajectory $\hat{x}_\epsilon(.)$ above. Now take the process $(\bar{x}_\epsilon(.), \bar{u}_\epsilon(.))$

$$
\begin{align*}
\bar{u}_\epsilon &= \begin{cases} 
-1 & \text{on } [0, \epsilon] \\
0 & \text{on } (\epsilon, 1]
\end{cases}
\end{align*}
$$

with the same initial data $\bar{x}_\epsilon(0) = (0, 0)$. It is easy to see that $(\bar{x}_\epsilon(.), \bar{u}_\epsilon(.))$ is feasible and

$$
d_{[0,1]}(\bar{u}_\epsilon(.), \bar{u}_\epsilon(.)) = \epsilon = h^+(\hat{x}_\epsilon(.)).
$$

**IV. An Application: Conditions for Normality of the State Constrained Maximum Principle.**

Consider the optimal control problem
Minimize \( g(x(S), x(T)) \)
subject to
\[
\begin{align*}
\dot{x}(t) &= f(t, x(t), u(t)) \quad \text{a.e. } t \in [S, T], \\
u(t) &\in U(t) \quad \text{a.e. } t \in [S, T], \\
h(t, x(t)) &\leq 0 \quad \text{for all } t \in [S, T], \\
x(S) &\in C,
\end{align*}
\]

in which \( f(\cdot, \cdot, \cdot), U(\cdot) \) and \( h(\cdot, \cdot) \) are as in the Introduction, \( g(\cdot, \cdot) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \) is a given locally Lipschitz continuous function and \( C \subset \mathbb{R}^n \) is a given closed set. We say that a process which minimizes \( g(x(S), x(T)) \) over all processes \((x(\cdot), u(\cdot))\) satisfying the constraints of \((P)\) is an optimal process.

Take an optimal process \((\bar{x}(\cdot), \bar{u}(\cdot))\). The state constrained Maximum Principle (see e.g. [12]) asserts under appropriate hypotheses ((D1),(D2), (D3) and (D4)(i) and (ii) in which \( r_0 \) is some number such that \( r_0 > |\bar{x}(S)| \) for example) \((\bar{x}(\cdot), \bar{u}(\cdot))\) is an extremal, in the sense that there exist \( p(\cdot) \in W^{1,1}([S, T]; \mathbb{R}^n), \mu \in NBV^+(S, T) \) and \( \lambda \in \{0, 1\} \) such that

\[
(p(\cdot), \mu, \lambda) \neq 0,
\]

\[
-\dot{p}(t) \in \text{co } \partial_x H \left( t, \bar{x}(t), p(t) + \int_{[S, t]} \nabla_x h(t, \bar{x}(t)) \, d\mu(t), \bar{u}(t) \right) \quad \text{a.e. } t \in [S, T],
\]

\[
H \left( t, \bar{x}(t), p(t) + \int_{[S, t]} \nabla_x h(t, \bar{x}(t)) \, d\mu(t), \bar{u}(t) \right) = \max_{u \in U(t)} H \left( t, \bar{x}(t), p(t) + \int_{[S, t]} \nabla_x h(t, \bar{x}(t)) \, d\mu(t), u \right),
\]

\[
\text{supp } \{\mu\} \subset \{ t \mid h(t, \bar{x}(t)) = 0 \},
\]

\[
\left( p(S), - \left( p(T) + \int_{[S, T]} \nabla_x h(t, \bar{x}(t)) \, d\mu(t) \right) \right) \in \lambda \partial g(\bar{x}(S), \bar{x}(T)) + N_C(\bar{x}(S)) \times \{0\}.
\]

Here, \( H(\cdot, \cdot, \cdot, \cdot) : [S, T] \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \) is the Hamiltonian function

\[
H(t, x, p, u) = p \cdot f(t, x, u),
\]

\( NBV^+([S, T]; \mathbb{R}^n) \) denotes the space of non-decreasing \( \mathbb{R}^n \) valued functions \( \mu \) of bounded variation on \([S, T] \) such that \( \mu(S) = 0 \), which are right continuous on \((S, T)\). ‘supp \( \mu \)’ denotes the support of the measure on the Borel subsets of \([S, T]\) induced by \( \mu \).
An extremal for which the above conditions are satisfied with $\lambda = 1$ is called a normal extremal. It is important to know that optimal processes are normal extremals, since, otherwise, the extremality conditions merely express relations between the optimal process and the constraints and make no reference to the cost function.

We may establish, with the help of the estimates of Section 2, conditions for minimizers to be normal extremals, which are weaker, in some significant respects, than earlier conditions. These are summarized as the following proposition:

**Proposition 4.1:** Let $(\bar{x}(\cdot), \bar{u}(\cdot))$ be an optimal process for (P). Assume that, for some $r_0 > |\bar{x}(S)|$, (D1)-(D4) are satisfied. Then $(\bar{x}(\cdot), \bar{u}(\cdot))$ is a normal extremal.

**Comments** It is known (see [9], [8] and [3]) that estimates on the distance of a given trajectory to the class of trajectories with a given initial condition and satisfying a state constraint lead to conditions for non-degeneracy of the state constrained Maximum Principle. Previous work is based on the use of $L^\infty$ or $W^{1,1}$ distance functions in these estimates and, for this reason, requires (among other hypotheses) that

$U(\cdot)$ takes values closed sets and $f(t, x, \cdot)$ is continuous.

The fact that, in our analysis, we use sharper estimates, based on the Ekeland distance function, permits us to obtain conditions for non-degeneracy even in situations when $U(\cdot)$ fails to take closed values and $f(t, x, \cdot)$ is no longer assumed continuous.

**Proof.** Let $(\bar{x}(\cdot), \bar{u}(\cdot))$ be an optimal process for (P), can be expressed as

\[
\begin{align*}
&\text{Minimize } g(x(S), x(T)) \\
&\quad \text{over processes } (x(\cdot), u(\cdot)) \text{ on } [S, T] \text{ such that} \\
&\quad h(t, x(t)) \leq 0 \quad \text{for all } t \in [S, T] \\
&\quad x(S) \in C.
\end{align*}
\]

We claim that, as a consequence of Thm. 2.1, $(\bar{x}(\cdot), \bar{u}(\cdot))$ is also an optimal process for the following optimal control problem, in which the state constraint is replaced by a penalty term.
But then, since this confirms the claim.

But, from the Gronwall’s Inequality (cf. for instance [12, Lemma 2.4.4]) we also derive
\[
R = e^{\epsilon(T-S)}(1 + r_0).
\]
Take any process \((x(.), u(.))\) satisfying the constraints of \((Q)\). It follows from Thm. 2.1 that there exists a process \((x',(.), u'(.))\) such that \(x'(S) = x(S)\) and that satisfies the constraints of \((P)\) and also the condition:
\[
d_{[S,T]}(u(.), u'(.)) \leq K\epsilon^+(x(.)).
\]

But, from the Gronwall’s Inequality (cf. for instance [12, Lemma 2.4.4]) we also derive
\[
||x'(.)-x(.)||_{L^\infty(S',T)} \leq 2c(1 + R) e^{\int_{S}^{T} k_f(s) ds} K\epsilon^+(x(.)).
\]

It follows that
\[
g(x(S), x(T)) \geq g(x'(S), x'(T)) - K'\epsilon^+(x(.))
\]
where
\[
K' = 2c(1 + R)k_gK e^{\int_{S}^{T} k_f(s) ds}.
\]

But then, since \((\bar{x}(.), \bar{u}(.))\) is an optimal process (for \((P)\)) and \((\bar{x}(.), \bar{u}(.))\) satisfies the constraints of \((P)\), we have, by optimality,
\[
\begin{align*}
g(x(S), x(T)) + K'\epsilon^+(x(.)) &\geq g(x'(S), x'(T)) \\
&\geq g(\bar{x}(S), \bar{x}(T)) = g(\bar{x}(S), \bar{x}(T)) + K'\epsilon^+(\bar{x}(.)).
\end{align*}
\]

This confirms the claim.

We conclude that \(((\bar{z}(.), \bar{e}(.), \bar{u}(.))\) is an optimal process for the problem
\[
\begin{align*}
(Q') \quad &\text{Minimize} \ g(x(S), x(T)) + K'(z(T) \lor 0) \\
&\text{subject to} \\
&d/dt(z(t), x(t)) = (0, f(t, x(t), u(t))), \ u(t) \in U(t) \\
h(t, x(t)) - z(t) \leq 0 \\
x(S) \in C, z(S) \geq 0 \text{ and } |x(S) - \bar{x}(S)| < \epsilon'.
\end{align*}
\]
Now apply the non-smooth state constrained Maximum Principle. Let \( \lambda \) and \( \mu \) be the cost and ‘measure’ multipliers respectively, and let \( p(.) \) and \( q(.) \equiv -c \) be the costate arcs associate with the \( x \) and \( z \) variables. We deduce the usual Maximum Principle conditions for \( (P) \) in relation to \( \bar{x}(.) \) and \( p(.) \). But the transversality conditions (16) (in relation to \( \bar{z}(.) \) and \( q(.) \)) yield the additional information that \( c \geq 0 \) and \( c + \int_{[S,T]} d\mu(t) \leq K'\lambda \).

If \( \lambda = 0 \), we would have, by the preceding condition, \( \mu = 0 \) and \( q(.) \equiv 0 \). But also, in consequence of the adjoint inclusion and the transversality condition for \( p(.) \), we would also have \( p(.) \equiv 0 \). From this contradiction of the non-triviality of the multipliers, we conclude that \( \lambda = 1 \).

V. PROOF OF THEOREM 2.1

We begin by proving the following local version of the theorem:

**Fix** \( r_0 > 0 \) **and assume** (D1)-(D4). **Then there exists** \( \delta(r_0) > 0 \) **and** \( M(r_0) > 0 \) **with the following property:** take any \( S' \in [S,T] \) **and any process** \( (\hat{x}(.), \hat{u}(.)) \) **on** \( [S', T] \) **with** \( \hat{x}(S') \in r_0B \cap A(S') \). **Then, there exists a process** \( (x(.), u(.)) \) **on** \( [S', T] \) **such that** \( \hat{x}(S') = x(S') \),

\[
x(t) \in A(t) \quad \text{for all} \quad t \in [S', (S' + \delta(r_0)) \wedge T] 
\]

and

\[
d_{[S',T]}(u(.), \hat{u}(.)) \leq M(r_0) h^+(\hat{x}(.)). 
\]

Fix \( r_0 > 0 \). Let \( R := e^{c|T-S|}(1 + r_0) \). (Note that \( R \) is an upper bound on the magnitude of all state trajectories \( x(.) \) on \( [S', T] \) such that \( x(S') \in r_0B \). Therefore from (D2) \( c(1 + R) \) is an upper bound for its velocities \( \hat{x}(.) \). Take \( \gamma, \rho, k_f(.) \) as in (D1)-(D4) (for the parameter \( r_0 \)). Let \( k_h \) and \( k'_h \) be Lipschitz constants for \( h(\cdot, \cdot) \) and \( \nabla h(\cdot, \cdot) \) respectively, on \( [S, T] \times RB \). Let \( \omega_f(.) \) be a modulus of continuity for \( t \to \int_{S}^{t} k_f(s')ds' \).

An analysis based on hypotheses (D1)-(D4) and application of Aumann’s measurable selection theorem establishes the existence of \( \epsilon' > 0 \) and \( \bar{\delta} > 0 \) with the following property: for any
$S' \in [S, T]$ and any process $(z(.), v(.))$ on $[S', (S' + \delta) \land T]$ such that $z(S') \in A(S') \cap r_0 B$, either

(A): $h(S', z(S')) \leq -\epsilon'$; in this case $z(t) \in A(t)$ for all $t \in [S', (S' + \delta) \land T]$

or

(B): $-\epsilon' < h(S', z(S')) \leq 0$; in this case there exists $\bar{u}(.) : [S', (S' + \delta) \land T] \to \mathbb{R}^n$ such that

$$\nabla h(t, z(t)) \cdot (1, f(t, z(t), \bar{u}(t))) \leq -\gamma \quad \text{a.e.} \quad [S', (S' + \delta) \land T].$$

Define the monotone function $\phi(.)$

$$\phi(\delta) = 2c(1 + R) \left( k'^h [1 + c(1 + R)] \delta + k_h \omega(\delta) \right) e^{\int_0^T k_f(t) \, dt}.$$ 

Since $\phi(.)$ is monotone and $\phi(0) = 0$, we can choose $\delta(r_0) \in (0, \delta)$ such that $\phi(\delta(r)) < \gamma$. Now choose $M(r_0) > 0$ such that

$$M(r_0) > (\gamma - \phi(\delta(r)))^{-1}. \quad (17)$$

Write for simplicity $M = M(r_0)$. Call

$$k(r_0) = 2c(1 + R) M e^{\int_0^T k_f(t) \, dt}.$$ 

and $\bar{\sigma} = (S' + \delta(r_0)) \land T$.

Now choose any $S' \in [S, T]$ and any process $(\hat{x}(.), \hat{u}(.))$ on $[S', T]$. Our aim is to construct a process $(x(.), u(.))$ on $[S', T]$ such that $x(S') = \hat{x}(S'),$

$$x(t) \in A(t) \quad \forall \ t \in [S', \bar{\sigma}] \quad (18)$$

and

$$d_{[S', T]}(u(.), \hat{u}(.)) \leq M h^+(\hat{x}(.)). \quad (19)$$

We may assume (B) and also that $h(t, \hat{x}(t)) > 0$ for some $t \in [S', \bar{\sigma}]$ since, otherwise, $(x(.), u(.)) = (\hat{x}(.), \hat{u}(.))$ has the asserted properties. It follows that we can define $\sigma \in [S', \bar{\sigma}]$ according to

$$\sigma = \inf \{ s \in [S', \bar{\sigma}] \mid h(s, \hat{x}(s)) > 0 \}.$$ 

Clearly $h(\sigma, \hat{x}(\sigma)) = 0.$
A straightforward analysis yields, for each $t \in [\sigma, \tilde{\sigma}]$

$$A = \left\{ t \in [\sigma, \tilde{\sigma}] \mid \frac{d}{ds} h(s, \hat{x}(s)) > 0 \right\}. $$

Define $\tau \in [\sigma, \tilde{\sigma}]$ as follows:

$$
\tau := \begin{cases} 
\bar{\sigma} & \text{if } \int_{[\sigma, \sigma] \cap A} ds \leq M h^+ (\hat{x}(\cdot)) \\
\inf \{ t \in [\sigma, \tilde{\sigma}] : \int_{[\sigma, t] \cap A} ds > M h^+ (\hat{x}(\cdot)) \} & \text{otherwise}
\end{cases}
$$

Let $(x(\cdot), u(\cdot))$ be the process on $[S', \tilde{\sigma}]$ such that $x(S') = \hat{x}(S')$ and

$$u(t) = \begin{cases} 
\bar{u}(t) & \text{for } t \in A \cap [\sigma, \tau] \\
\hat{u}(t) & \text{for } t \in [S', \tilde{\sigma}] \setminus (A \cap [\sigma, \tau])
\end{cases}.$$

We have

$$d_{[S', \tilde{\sigma}]}(u(\cdot), \hat{u}(\cdot)) \leq M h^+ (\hat{x}(\cdot)).$$

Making use of the Gronwall’s Inequality (cf. for instance [12, Lemma 2.4.4]), we can show that

$$||x - \hat{x}||_{L^\infty(S', t)} \leq 2c(1 + R) e^{\int_{S'} k_f(s) ds} \text{meas } \{ [\sigma, t \wedge \tau] \cap A \}.$$  \hspace{1cm} (20)

Write

$$E = \int_{[\sigma, t]} \nabla h(s, \hat{x}(s)) \cdot (1, f(s, \hat{x}(s), u(s))) \ ds.$$  \hspace{1cm} (21)

A straightforward analysis yields, for each $t \in [\sigma, \tilde{\sigma}]$, the relation

$$h(t, x(t)) \leq \int_{[\sigma, t]} \nabla h(s, \hat{x}(s)) \cdot (1, f(s, x(s), u(s))) ds$$

$$+ k_h^2 (1 + c(1 + R)) |t - \sigma| ||x(\cdot) - \hat{x}(\cdot)||_{L^\infty(S', t)}$$

$$\leq E + k_h^2 (1 + c(1 + R)) |t - \sigma| + k_h \omega_f (t - \sigma) ||x(\cdot) - \hat{x}(\cdot)||_{L^\infty(S', t)}.$$  \hspace{1cm} (22)

We must show that $h(t, x(t)) \leq 0$ for all $t \in [S', \tilde{\sigma}]$. But $h(t, x(t)) \leq 0$ for $t \leq \sigma$, by definition of $\sigma$. We may therefore assume that $t \geq \sigma$. There are two possible cases to consider.

(1): $t \in [\sigma, \tau]$. In this case

$$E \leq \int_{[\sigma, t] \setminus A} \nabla h(s, \hat{x}(s)) \cdot (1, f(s, \hat{x}(s), \bar{u}(s))) ds$$

$$+ \int_{[\sigma, t] \cap A} \nabla x h(s, \hat{x}(s)) \cdot (1, f(s, \hat{x}(s), \bar{u}(s))) ds$$

$$\leq 0 - \gamma \text{meas } \{ [\sigma, t] \cap A \}. $$
We conclude from (20), (21) and (22) that
\[ h(t, x(t)) \leq - (\gamma - \phi(\delta(r_0))) \text{ meas } \{[\sigma, t] \cap A\} \ ds \leq 0. \]

(2): \( t > \tau \). In this case, since \( h(\sigma, \hat{x}(\sigma)) = 0 \) we have that
\[ E \leq \nabla h(s, \hat{x}(s)) \cdot (1, f(s, \hat{x}(s), \hat{u}(s))) \ ds \\
+ \int_{[\sigma, t]} \nabla_x h(s, \hat{x}(s)) \cdot \{ f(s, \hat{x}(s), u(s)) - f(s, \hat{x}(s), \hat{u}(s)) \} \ ds \\
\leq h(t, \hat{x}(t)) - \gamma \text{ meas } \{[\sigma, t] \cap A\}. \]

It follows from (20), (21) and (22) that
\[ h(t, x(t)) \leq h^+(\hat{x}(\cdot)) - (\gamma - \phi(\delta(r_0))) M h^+(\hat{x}(\cdot)) \leq 0. \]

This completes the proof of the local version of the theorem, in which the constructed process
\((x(\cdot), u(\cdot))\), with initial state in \(r_0 B \cap A(S')\) satisfies the state constraint on an interval \((S', (S' + \delta(r_0) \wedge T)]\), where \(\delta(r_0) > 0\) does not depend on the reference process \((\hat{x}(\cdot), \hat{u}(\cdot))\).

To obtain the global version, we apply the local version recursively, to construct a finite sequence of processes \((x_i(\cdot), u_i(\cdot))\) on \([S', T], j = 1, \ldots, N\), the \(j\)th process within which is an extension of the \((j - 1)\)th process and satisfies the state constraint on \([S', (S' + j\delta(r_0)) \wedge T]\). We choose \((x(\cdot), u(\cdot)) = (x_N(\cdot), u_N(\cdot))\), where \(N\) is the first index value \(j\) for which \(S' + (j\delta(r_0)) \geq T\). The process thus obtained satisfies \(x(S') = \hat{x}(S')\), \(x(t) \in A(t)\) for all \(t \in [S', T]\). Writing \((x_0(\cdot), u_0(\cdot)) = (\hat{x}(\cdot), \hat{u}(\cdot))\) we have, for \(j = 1, \ldots, N\)
\[ d_{[S', T]}(u_j, u_{j-1}) \leq M h^+(x_{j-1}(\cdot)) \]
\[ h^+(x_j(\cdot)) \leq h^+(x_{j-1}(\cdot)) + k h 2c(1 + R) e^{\int_{S'}^{T} k_f(s) \ ds} \ d_{[S', T]}(u_j, u_{j-1}). \]

Notice that \(N\) satisfies \(N \leq |T - S|/\delta(r_0)\), and therefore does not depend on the choice of reference process \((\hat{x}(\cdot), \hat{u}(\cdot))\).

Routine calculations yield the estimate:
\[ d_{[S', T]}(u_j(\cdot), u_{j-1}(\cdot)) \leq M (1 + \tilde{k} M)^{j-1} h^+(\hat{x}(\cdot)) \]
for \(j = 1, \ldots, N\), where \(\tilde{k} := k h 2c(1 + R) e^{\int_{S'}^{T} k_f(s) \ ds}\). It follows that
\[ d_{[S', T]}(u(\cdot), \hat{u}(\cdot)) \leq \sum_{j=1}^{N} M (1 + \tilde{k} M)^{j-1} h^+(\hat{x}(\cdot)) \]
\[ = ((1 + \tilde{k} M)^N - 1) \tilde{k}^{-1} h^+(\hat{x}(\cdot)). \]
We have confirmed the assertions of the theorem, with

\[ K = \left( (1 + \tilde{k}M)^{T-S/\delta(r_0)} - 1 \right) \tilde{k}^{-1}. \]

**REFERENCES**


