Necessary Optimality Conditions for Nonconvex Differential Inclusion with Endpoint Constraints

Q. Zhu
Groupe d'études et de recherche en analyse des décisions
5255 avenue Decelles
Montréal, Quebec, H3T 1V6

Abstract. In this paper, we derive necessary optimality conditions for optimization problems defined by non-convex differential inclusions with endpoint constraints. We do this in terms of parametrizations of the convexified form of the differential inclusion and, under additional assumptions, in terms of the inclusion itself.
1 Introduction

Consider the optimization problem $\mathcal{P}$:
\[
\text{minimize} \quad g_0(x(1)) \\
\text{subject to} \quad \dot{x}(t) \in F(t, x(t)) \quad \text{a.e. in } [0, 1], \quad \!
\begin{align*}
    x(0) &= x_0, \\
    g_1(x(1)) &= 0,
\end{align*}
\]

where $F$ is a closed-valued multifunction measurable in $t$ and Lipschitzian in $x$, and $g = (g_0, g_1) : \mathbb{R}^N \to \mathbb{R} \times \mathbb{R}^M$ is locally Lipschitzian. (The endpoint constraint of the form $x(1) \in C$, where $C$ is a closed set, is a special case of the above obtained by setting $g_1(x(1)) = \text{dist}(x(1), C) := \inf\{|x(1) - y| : y \in C\}$.) We are concerned with necessary conditions for a solution to $\mathcal{P}$. If $F(t, x)$ admits a parametrization, i.e., there exist a set $U$ and a function $f(t, x, u)$ such that $F(t, x) = \{f(t, x, u) : u \in U\}$ and if certain regularity conditions are satisfied then the maximum principle and its various generalizations (see [16, 20]) concerning the control system defined by $(f, U)$ provide necessary conditions for a solution to $\mathcal{P}$. However, for a nonconvex-valued multifunction $F$ it is very difficult to determine whether such a parametrization exists. On the other hand, a variant of Lojasiewicz's parametrization theorem [13] shows that under fairly general conditions such a parametrization exists when $F$ is convex-valued (Section 4). Therefore, it would be useful to establish necessary conditions for a solution to (the original, unrelaxed) problem $\mathcal{P}$ in terms of parametrizations for the convex closure $\overline{\text{co}}F$ of $F$. This is the main purpose of this paper. Crudely speaking, our main result is that a solution to problem $\mathcal{P}$ satisfies the same generalized maximum principle as do optimal solutions to the problem with the relation $\dot{x} \in F(t, x)$ replaced by the control system determined by a parametrization of $\overline{\text{co}}F$.

Under an additional assumption of the existence of a $C^1$ representation of $\overline{\text{co}}F$ with a compact parameter set, this necessary condition will lead to an intrinsic (independent of parametrizations) necessary condition in terms of the generalized derivative of the Hamiltonian associated with $F$, and an example shows that without the $C^1$ assumption on the parametrization and the compactness assumption on the parameter set such a condition is not valid in general.

There is abundant literature on necessary conditions for problem $\mathcal{P}$ in terms of various generalized derivatives of $F$ or its associated Hamiltonian under additional assumptions on $F$ (usually convexity) or endpoint constraints. We refer to [2, 3, 6, 7, 14] for more details and the references to much of the recent literature.

The proof of our main result, crudely speaking, is as follows: consider a parametrization $(f(t, x, u), U)$ of $\overline{\text{co}}F(t, x)$. For any given solution $x(t)$ to the original
differential inclusion \( \dot{x}(t) \in F(t, x(t)) \), Filippov’s Lemma [5, 15] implies that there exists a measurable mapping \( u : [0, 1] \to U \) such that \( \dot{x}(t) = f(t, x(t), u(t)) \) a.e. in \([0, 1]\). Denote the collection of all such measurable controls \( u \) by \( U_F \). Then problem \( \mathcal{P} \) is equivalent to the following optimal control problem \( \mathcal{P}' \):

\[
\begin{align*}
\text{minimize} & \quad g_0(x(1)) \\
\text{subject to} & \quad \dot{x}(t) = f(t, x(t), u(t)) \quad \text{a.e. in} \ [0, 1], \\
& \quad x(0) = x_0, \ g_1(x(1)) = 0, \ u \in U_F.
\end{align*}
\]

We then follow the road mapped by Warga [16, 20]. Denote by \( \mathcal{S}(U) \) the (compact convex) set of relaxed controls corresponding to \( U \). Consider, at the same time, the corresponding convexified optimization problem with \( U_F \) replaced by \( \mathcal{S}(U) \). Then a controllability-extremality alternative can be deduced by (a) proving that \( U_F \) is an “abundant” subset of \( \mathcal{S}(U) \) and (b) applying the nonsmooth open covering theorem in [20] which, in particular, yields a necessary condition for a solution to problem \( \mathcal{P}' \). The property of \( U_F \) being an abundant subset of \( \mathcal{S}(U) \) is established by invoking a recent result of A. Fryszkowski and J. Rzezuchowski [8].

Similar arguments extend this result to problems with unilateral constraints and also yield a strengthened Kaskosz type necessary condition [10, 24]. After the completion of this paper the author noticed that H. D. Tuan proved result similar to Theorem 6.1 in Section 6 independently.

The remainder of this paper is arranged as follows: In Section 2 we briefly describe notations and background material that will be used later; in Section 3 we state and prove the main results; in Section 4 we discuss a parametrization theorem; Section 5 contains some comments and examples; and we discuss an extension of the main result in Section 6.

## 2 Preliminaries

Let \( \mathbb{R}^n \) be the usual \( n \)-dimensional Euclidean space with the inner product \( \langle \cdot , \cdot \rangle \) and norm \( | \cdot | \). An element \( x \in \mathbb{R}^n \) is represented by an \( n \)-dimensional column vector and we denote its transpose by \( x^\top \). We denote by \( 2\mathbb{R}^n \) the collection of all subsets of \( \mathbb{R}^n \). For \( x \in \mathbb{R}^n \) and \( A \subset \mathbb{R}^n \), we write \( \text{dist}(x, A) = \inf \{|x - y| : y \in A\} \) and \( |A| = \sup \{|a| : a \in A\} \). We define the Hausdorff distance between two subsets \( A \) and \( B \) of \( \mathbb{R}^n \) by

\[
\rho(A, B) = \max \{ \sup_{x \in A} d(x, B), \sup_{y \in B} d(x, A) \}
\]
2.1 Multifunctions and Parametrizations

Let $V \subset \mathbb{R}^n$ be an open set and $F : [0, 1] \times V \to 2^{\mathbb{R}^n}$ a multifunction. We refer to a pair $(f, U)$ as a parametrization of $F$ if $U$ is a set, $f : [0, 1] \times V \times U \to \mathbb{R}^n$ and $F(t, x) = f(t, x, U)$ for all $(t, x) \in [0, 1] \times V$. A parametrization will be called Lipschitzian or Lipschitz if

1. $U$ is a compact metric space;
2. for all $(x, u) \in V \times U$, $f(\cdot, x, u)$ is measurable;
3. for almost all $t \in [0, 1]$, $f(t, \cdot, \cdot)$ is continuous;
4. there exists an integrable function $k(t)$ such that, for all $(t, u) \in [0, 1] \times U$,

$$|f(t, x, u) - f(t, y, u)| \leq k(t)|x - y| \quad \forall x, y \in V.$$

A parametrization is called $C^1$ if it is Lipschitzian, $f(t, \cdot, u)$ is differentiable for all $(t, u) \in [0, 1] \times U$, and the function $(x, u) \mapsto f_x(t, x, u)$ is continuous for almost all $t$.

2.2 Differential Inclusions

Let $F : [0, 1] \times V \to 2^{\mathbb{R}^n}$ be a multifunction. An absolutely continuous function $x$ that satisfies the differential inclusion

$$\dot{x}(t) \in F(t, x(t)) \quad \text{a.e. in } [0, 1],$$

$$x(0) = x_0 \quad (1)$$

is called a solution to (1). We denote the set of all solutions to (1) by $S_F$.

2.3 Generalized Derivatives

Definition 2.3.1. [18, 19, 20] Let $V \subset \mathbb{R}^n$ be open and $h : V \to \mathbb{R}^m$ locally Lipschitzian. A bounded collection

$$\{\Lambda^\epsilon h(v) : \epsilon > 0, v \in V\}$$

of nonempty closed subsets of $L(\mathbb{R}^n, \mathbb{R}^m)$, also referred to as $\Lambda^\epsilon h$, is a derivate container for $h$ if

$$\Lambda^\epsilon h(v) \subset \Lambda^\epsilon' h(v) \quad \text{for} \quad \epsilon' > \epsilon$$

and for every compact $V^* \subset V$ there exists a neighborhood $\tilde{V}$ of $V^*$ in $V$ and a sequence of $C^1$ functions $h_i : \tilde{V} \to \mathbb{R}^m, i = 1, 2, ..., \epsilon$ such that

$$\lim h_i = h \quad \text{uniformly on } V^*$$
and for every \( \epsilon > 0 \) there exist

\[
i^* = \hat{i}(\epsilon, V^*) > 0 \text{ and } \delta^* = \delta(\epsilon, V^*) > 0
\]

such that

\[
h_i'(v) \in \Lambda^*h(w) \text{ for } i \geq i^*, w \in V^*, |v - w| \leq \delta^*.
\]

We write

\[
\Lambda h(v) := \bigcap_{\epsilon > 0} \Lambda^*h(v)
\]

and refer to \( \Lambda h(v) \) as well as to \( \Lambda^*h(v) \) as a derivate container of \( h \) at \( v \).

**Definition 2.3.2.**[2] Clarke’s Generalized Jacobian. If \( h \) is Lipschitzian, the sets

\[
\partial^*h(v) := \overline{co}\{h'(x) : |x - v| \leq \epsilon, \ h'(x) \text{exists}\}
\]

define a derivate container[19]. We write

\[
\partial h(v) := \bigcap_{\epsilon > 0} \partial^*h(v)
\]

and and observe that \( \partial h(v) \) is Clarke’s generalized Jacobian of \( h \) at \( v \). When \( h \) depends on additional arguments, say \( h = h(x, y) \), we write \( \partial_x h(x, y) \) to represent the generalized Jacobian of \( h \) with respect to \( x \).

### 2.4 Extremality

Let \( V \subset \mathbb{R}^n \) be an open set, \( U \) a compact metric space and \( f : [0,1] \times V \times U \rightarrow \mathbb{R}^n \). Assume that \( f(\cdot, v, u) \) is measurable, \( f(t, \cdot, \cdot) \) continuous and \( f(t, \cdot, u) \) Lipschitzian with a Lipschitz constant \( k(t) \) independent of \( u \) and integrable on \([0,1] \). Denote by \( \mathcal{S}(U) \) the set of relaxed controls, i.e. Lebesgue measurable functions \( \sigma(t) \in rpm(U) \) on \([0,1] \), where \( rpm(U) \) is the set of Radon probability measures on \( U \) with the weak star topology of \( C(U)^* \). An original (ordinary) control function \( u : [0,1] \rightarrow U \) is embedded in \( \mathcal{S}(U) \) through the mapping \( u(\cdot) \rightarrow \delta_{u(\cdot)} \), where \( \delta_u \) is the Dirac measure concentrated at \( u \). We now consider the (relaxed) control system

\[
\dot{x}(t) = f(t, x(t), \sigma(t)) := \int_U f(t, x(t), r)\sigma(t)(dr), \quad x(0) = x_0, \ \sigma \in \mathcal{S}(U)
\]  \( (2) \)

Under our assumptions on \( f \), for each \( \sigma \in \mathcal{S}(U) \) there corresponds a unique solution of \( (2) \). We denote such a solution by \( x(\sigma, f)(t) \).

Let \( g : \mathbb{R}^n \rightarrow \mathbb{R}^m \) be a locally Lipschitzian function. Following Warga [21], we define the concept of \( g \)-extremal for \( \tilde{\sigma} \) (relative to \( f \)).
**Definition 2.4.1.** Let $(f, U)$ satisfy the conditions described above. A relaxed or unrelaxed control $\tilde{\sigma} \in \mathcal{S}(U)$ is called a $g$-extremal with respect to $f$ if there exist $l \in \mathbb{R}^m$, $h \in \Lambda g(\tilde{x}(1))$, $M(t) \in \partial f(t, \tilde{x}(t), \tilde{\sigma}(t))$ a.e. in $[0, 1]$ and a nontrivial absolutely continuous function $p(t)$ such that

1. $|l| > 0$,
2. $p(1) = h^\top l$
3. $\dot{p}(t)^\top = -p(t)^\top M(t)$ for almost all $t \in [0, 1]$,
4. For almost all $t \in [0, 1]$,
   $$p(t)^\top f(t, \tilde{x}(t), \tilde{\sigma}(t)) = \max_{u \in U} p(t)^\top f(t, \tilde{x}(t), u)$$

where $\tilde{x}(t) := x(\tilde{\sigma}, f)(t)$.

### 3 Main Results

Throughout this section we assume that $g = (g_0, g_1) : \mathbb{R}^n \rightarrow \mathbb{R} \times \mathbb{R}^m$ is locally Lipschitzian and $F : [0, 1] \times V \rightarrow 2\mathbb{R}^n$ satisfies the following conditions:

(H1) $F(t, x, u)$ is measurable in $t$ and continuous in $(x, u)$;

(H2) there exists an integrable function $k(\cdot)$ such that, for any $x, y \in V$,
   $$\rho(F(t, x), F(t, y)) \leq k(t)|x - y| \text{ a.e. in } [0, 1];$$

(H3) $F$ is integrably bounded, i.e. there exists an integrable function $m(\cdot)$ on $[0, 1]$ such that
   $$|F(t, x)| \leq m(t) \quad \forall x \in V \text{ a.e. in } [0, 1].$$

We associate with a multifunction $F$ a Hamiltonian $H$ defined by
   $$H(t, x, p) = \max\{ < p, v > : v \in F(t, x) \}$$

Consider the differential inclusion
   $$\dot{x}(t) \in F(t, x(t)), \quad x(0) = x_0$$
Definition 3.1. We say that \( x \in S_F \) is a \( g \)-extremal(\( L \)) (for Lojasiewicz extremal[12]) if for any Lipschitz parametrization \( (f(t, x, u), U) \) of \( \overline{\partial}_t F(t, x) \) and any measurable function \( u : [0, 1] \rightarrow U \) such that \( \dot{x}(t) = f(t, x(t), u(t)) \), \( u \) is a \( g \)-extremal with respect to \( f \).

Our main result is:

Theorem 3.1. Let \( x \in S_F \). Then either \( g(x(1)) \in \text{int} \{g(y(1)) : y \in S_F\} \) or \( x \) is a \( g \)-extremal(\( L \)).

Remark. As we verify in Section 4, there always exists (at least one) Lipschitzian parametrization for \( \overline{\partial}_t F \) when \( F \) satisfies assumptions \( (H_1) \), \( (H_2) \) and \( (H_3) \).

We will first discuss some corollaries of Theorem 3.1 and defer the proof of Theorem 3.1 to the end of this section. The following corollary is obvious.

Corollary 3.1.1. Let \( x \in S_F \) be a solution to the optimization problem \( \mathcal{P} \). Then \( x \) is a \( g \)-extremal(\( L \)).

When \( \overline{\partial}_t F \) has a \( C^1 \) parametrization we can derive an intrinsic necessary condition.

Corollary 3.1.2. Let \( x \in S_F \) and let \( \Lambda g(x(1)) \) be a derivate container of \( g \) at \( x(1) \). Assume that \( \overline{\partial}_t F \) has a \( C^1 \) parametrization. Then either \( g(x(1)) \in \text{int} \{g(y(1)) : y \in S_F\} \) or there exist \( l \in \mathbb{R}^m \), \( h \in \Lambda g(x(1)) \), and a nontrivial absolutely continuous function \( p(t) \) such that

1. \( |l| > 0 \),
2. \( p(1) = h^T l \),
3. \( -\dot{p}(t)^T \in \partial_x H(t, x(t), p(t)) \) for almost all \( t \in [0, 1] \),
4. For almost all \( t \in [0, 1] \),
   \[ p(t)^T \dot{x}(t) = H(t, x(t), p(t)) \]

Proof. Let \( (f, U) \) be a \( C^1 \) representation of \( \overline{\partial}_t F \) and \( \sigma \in S(U) \) satisfy \( \dot{x}(t) = f(t, x(t), \sigma(t)) \) a.e. in \( [0, 1] \). By Theorem 3.1, \( \sigma \) is a \( g \)-extremal (relative to \( f \)). Thus there exist \( l \in \mathbb{R}^m \), \( h \in \Lambda g(x(1)) \), and a nontrivial absolutely continuous function \( p(t) \) such that

1. \( |l| > 0 \),
2. \( p(1) = h^\top l \)

3. \( -\dot{p}(t)^\top = p(t)^\top f_x(t, x(t), \sigma(t)) \) for almost all \( t \in [0, 1] \),

4. For almost all \( t \in [0, 1] \),
   \[
   p(t)^\top \dot{x}(t) = p(t)^\top f(t, \bar{x}(t), \sigma(t))
   = \max_{u \in U} p(t)^\top f(t, x(t), u) = H(t, x(t), p(t))
   \]

We need only to show that \( -\dot{p}(t)^\top \in \partial_x H(t, x(t), p(t)) \). Observe that \(^2\)

\[
\partial_x H(t, x, p) = \{ \xi :< \xi, v > \leq H^x(t, x, p; v), \forall v \in \mathbb{R}^n \},
\]

where

\[
H^x(t, x, p; v) = \limsup_{h \to \sigma^\tau, y \to x(t)} \frac{H(t, y + hv, p) - H(t, y, p)}{h}
\]

We have, for any \( v \in \mathbb{R}^n \),

\[
H^x(t, x(t), p(t); v) = \limsup_{h \to \sigma^\tau, y \to x(t)} \frac{H(t, y + hv, p(t)) - H(t, y, p(t))}{h}
\]

\[
\geq \limsup_{h \to \sigma^\tau} \frac{H(t, x(t) + hv, p(t)) - H(t, x(t), p(t))}{h}
\]

\[
\geq \limsup_{h \to \sigma^\tau} \frac{p(t)^\top f(t, x(t) + hv, \sigma(t)) - p(t)^\top f(t, x(t), \sigma(t))}{h}
\]

\[
= p(t)^\top f_x(t, x(t), \sigma(t))v
\]

\[
= -\dot{p}(t)^\top v
\]

Therefore \( -\dot{p}(t)^\top \in \partial_x H(t, x(t), p(t)) \). \( \square \)

We now turn to the proof of Theorem 3.1. Set

\[
T_N' = \left\{ (\lambda_0, ..., \lambda_N) : \sum_{i=0}^N \lambda_i = 1, \lambda_i \geq 0 \right\}
\]

**Definition 3.2.** Let \( X \) be a linear space, \( \phi : X \to \mathbb{R}^\nu \) and \( K \subset X \) convex. We say that a subset \( U \) of \( K \) is \( \phi \)-abundant in \( K \) if, for any given \( x_1, ..., x_N \in K \), there exists a sequence of functions \( \lambda \to u^n(\lambda) \) from \( T_N' \) to \( U \) such that

i. \( \lim_n \phi(u^n(\lambda)) = \phi(\sum_{i=0}^N \lambda_i x_i) \) uniformly for \( \lambda \) in \( T_N' \).

ii. \( \lambda \to \phi(u^n(\lambda)) : T_N' \to \mathbb{R}^\nu \) is continuous for each \( n \).
The following proposition follows directly from the definition.

**Proposition 3.1.** Assume that $\mathcal{U}$ is $\phi$-abundant in $K$. Then, for any continuous function $\psi : \mathbb{R}^r \to \mathbb{R}^s$, $\mathcal{U}$ is $\psi \circ \phi$-abundant in $K$.

We need the following modification of Warga’s extremality-controllability alternative [20, Theorem 3.1].

**Theorem 3.2.** (Warga [20]) Consider the control problem defined in Section 2.4. Let $\mathcal{U}$ be a $g(x(1, \cdot, f))$-abundant subset of $\mathcal{S}(U)$. Then, for any $u \in \mathcal{U}$, either

$$g(x(u, f)(1)) \in \text{int} \{g(x(v, f)(1)) : v \in \mathcal{U}\}$$

or there exist $l \in \mathbb{R}^m$, $h \in \Lambda g(x(u, f)(1))$, $M(t) \in \partial_x f(t, x(u, f)(t), u(t))$ a.e. in $[0, 1]$ and a nontrivial absolutely continuous function $p(t)$ such that

1. $|l| > 0$,
2. $p(1) = h^T l$
3. $\dot{p}(t)^T = -p(t)^T M(t)$ for almost all $t \in [0, 1]$,
4. $$\int_0^1 p(t)^T f(t, x(u, f)(t), u(t)) dt = \max_{\sigma \in \mathcal{S}(U)} \int_0^1 p(t)^T f(t, \tilde{x}(t), \sigma(t)) dt$$

**Proof.** The theorem is the same as [20, Theorem 3.1] except that the assumption that $\mathcal{U}$ is an abundant subset of $\mathcal{S}(U)$ in [20] is replaced by the weaker assumption that $\mathcal{U}$ is a $g(x(\cdot, f)(1))$-abundant subset of $\mathcal{S}(U)$ and the extremal condition in pointwise form in [20] is replaced by the extremal condition in integral form. The proof of Theorem 3.1 in [20] is valid here up to the point where the extremal condition in integral form is derived since only the weaker assumption that $\mathcal{U}$ is a $g(x(\cdot, f)(1))$-abundant subset of $\mathcal{S}(U)$ is needed in the proof.

**Proof of Theorem 3.1.** Let $\bar{x} \in S_F$. Assume that $g(\bar{x}(1)) \not\in \text{int} \{g(x(1)) : x \in S_F\}$. We proceed to show that $\bar{x}$ is a $g$-extremal($L$).

Let $(f, U)$ be a Lipschitzian representation of $\mathcal{F}(t, x)$ and $\bar{u} : [0, 1] \to U$ a measurable function such that

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

We need to show that $\bar{u}$ is a $g$-extremal (relative to $f$). To this end, first we define $\mathcal{U}_F$ to be the collection of all measurable functions from $[0, 1]$ to $U$ such that there exists
an $x \in S_F$ with $\dot{x}(t) = f(t, x(t), u(t))$ and show that $\mathcal{U}_F$ is $g(x(\cdot, f)(1))$-abundant in $\mathcal{S}(U)$. Let $\sigma_0, \ldots, \sigma_N \in \mathcal{S}(U)$. Then, for any $\lambda \in \mathcal{T}_N^\lambda$, $\sum_{i=0}^{N} \lambda_i \sigma_i \in \mathcal{S}(U)$. It is easy to see that $\lambda \rightarrow x(\cdot, \sum_{i=0}^{N} \lambda_i \sigma_i, f) : \mathcal{T}_N^\lambda \rightarrow C([0, 1], V)$ is continuous. By a result of A. Fryszkowski and T. Rzezuchowski [8, Theorem 2], for each $i$ there exists $x_i(\lambda)(\cdot) \in S_F$ such that $x_i(\lambda)(\cdot)$ is continuous in $\lambda$ and
\[
|x_i(\lambda)(1) - x(\sum_{i=0}^{N} \lambda_i \sigma_i, f)(1)| \leq 1/i
\]
Let $u_i(\lambda)(\cdot) \in \mathcal{U}_F$ be such that
\[
\dot{x}_i(\lambda)(t) = f(t, x_i(\lambda)(t), u_i(\lambda)(t))
\]
Then the uniqueness of the solution to (2) implies that
\[
x_i(\lambda)(t) = x_i(u_i(\lambda), f)(t)
\]
Thus $\mathcal{U}_F$ is $x(\cdot, f)(1)$-abundant in $\mathcal{S}(U)$ and, therefore, $g(x(\cdot, f)(1))$-abundant in $\mathcal{S}(U)$.

Next we observe that
\[
\{g(x(1)) : x \in S_F\} = \{g(x(u, f)(1)) : u \in \mathcal{U}_F\}.
\]
Thus,
\[
g(x(\bar{u}, f)(1)) = g(\bar{x}(1)) \not\subset \text{int} \{g(x(1)) : x \in S_F\} = \text{int} \{g(x(u, f)(1)) : u \in \mathcal{U}_F\}
\]
It follows, invoking Theorem 3.2, that there exist $l \in \mathbb{R}^m$, $h \in \Lambda g(x(\bar{u}, f)(1))$,
\[
M(t) \in \partial_x f(t, x(\bar{u}, f)(t), \bar{u}(t)) \text{ a.e. in } [0, 1]
\]
and a nontrivial absolutely continuous function $p(t)$ such that
\begin{enumerate}
\item $|l| > 0$,
\item $p(1) = h^T l$
\item $\dot{p}(t)^T = -p(t)^T M(t)$ for almost all $t \in [0, 1]$,
\item \[
\int_0^1 p(t)^T f(t, x(\bar{u}, f)(t), \bar{u}(t)) dt = \max_{\sigma \in \mathcal{S}(U)} \int_0^1 p(t)^T f(t, x(\bar{u}, f)(t), \sigma(t)) dt
\]
\end{enumerate}
Since all measurable functions $u : [0, 1] \rightarrow U$ belong to $\mathcal{S}(U)$, relation 4. and a standard argument (see e.g. [16]) lead to the pointwise form of maximum principle:
\[
p(t)^T f(t, x(\bar{u}, f)(t), \bar{u}(t)) = \max_{u \in \mathcal{U}} p(t)^T f(t, x(\bar{u}, f)(t), u)
\]
\[\square\]
4 Existence of Lipschitz Parametrization

We prove in this section a variant of Lojasiewicz’s parametrization theorem [13] stating that a convex-valued multifunction satisfying assumptions (H1), (H2) and (H3) (in Section 3) admits a Lipschitz parametrization.

**Theorem 4.1.** Let $F : [0, 1] \times V \to 2^{\mathbb{R}^n}$ be a convex-valued multifunction satisfying assumptions (H1), (H2) and (H3). Then $F$ has a Lipschitz parametrization.

**Proof.** By Lojasiewicz’s result [13, Section 3] there exists an integrable function $\theta(t)$ on $[0, 1]$ such that, for any $(t, x) \in [0, 1] \times V$ and $v \in F(t, x)$, there exists a selection $\phi$ of $F$ having the following properties:

i. $\phi$ is measurable in $t$;

ii. for all $x, y \in V$,

$$|\phi(t, x) - \phi(t, y)| \leq \theta(t)|x - y|$$

Let $U$ be the collection of all selections $\phi$ of $F$ that satisfy i. and ii. Consider any $\phi \in U$ as a mapping $t \mapsto \phi(t, \cdot) : [0, 1] \to C(V, \mathbb{R}^n)$. Since $\phi$ is a selection of $F$ and $F$ is integrably bounded, $\int_0^1 \|\phi(t, \cdot)\|_{sup} dt \leq M$, where $M = \int_0^1 m(t) dt$ and $m(\cdot)$ is as defined in (H3). Thus $U$ is embedded in $L_1([0, 1], C(V, \mathbb{R}^n))$ as a bounded subset. It is easy to check that $U$ is a closed convex set. Now let $U$ be provided with the weak topology of $L_1$. Since $m(\cdot)$ is integrable and $|\phi(t, x)| \leq m(t)$ for all $x \in V$ and all $\phi \in U$, it follows from [4, Theorem 9, p.292] that $U$ is sequentially compact and from [4, Theorem 3, p.434] that $U$ is metrizable. Thus, $U$ with the weak topology of $L_1$ is a compact metric space. We define $f : [0, 1] \times V \times U \to \mathbb{R}^n$ by

$$f(t, x, \phi) = \phi(t, x), \ \forall \phi \in U$$

Then $f$ is a Lipschitz parametrization of $F$. \qed

5 Examples and Comments

Theorem 3.1 may sometimes yield stronger results than the maximum principle in control formulation. The following is an example.

**Example 5.1.** Consider problem $\mathcal{P}$ in $\mathbb{R}^2$ defined by $F(x) = \{ f(x, u) : u \in \{-1, 1\} \times [-1, 1] \}$ with

$$f(x, u) = \begin{pmatrix} u_2 \\ x_1 u_1 + x_1 u_2 \end{pmatrix}.$$
\(g_0(x) = x_1\) and \(g_1(x) = x_2\). It is easy to see that \(x(t) = 0\) is not a solution to \(P\) because the control \((u_1(t), u_2(t)) = (-1, -1)\) yields the trajectory \((x_1(t), x_2(t)) = (-t, 0)\) and a negative value of \(g_0(x(1))\). However, applying the nonsmooth maximum principle (see e.g. [20]), we cannot rule out \(x(t) = 0\) as a candidate for extremal. Indeed, let \(\bar{u}\) be a control corresponding to \(x(t) = 0\). Then \(\bar{u}(t) = (u_1(t), 0)^T\). Thus the corresponding adjoint trajectory \(p^T = (0, 1)\) satisfies

\[
-p^T = p^T \partial_x f(0, \bar{u}(t)) = p^T \begin{pmatrix} 0 & 0 \\ -u_1(t) & u_1(t) \end{pmatrix}
\]

and

\[
p^T \cdot 0 = \max_{u \in [-1,1] \times [-1,1]} \{ p^T f(0, u) \}
\]

as well as the transversality conditions (when we take \(l = (0, 1)^T\)).

Now, observing that \(\overline{\mathcal{M}} F(x) = \{ \tilde{f}(x, w) \mid w \in [-1,1] \times [-1,1] \}\) where \(\tilde{f}(x, w) = (w_1, 2x_1w_2)\) we can use Theorem 3.1 to rule out \(x(t) = 0\) as a candidate for a solution to \(P\). Indeed \(\bar{u} = (0, 1)\) is a control corresponding to the solution \(x(t) = 0\). If \(x(t) = 0\) were a solution to problem \(P\) then there would exist a nontrivial adjoint trajectory \(p\) such that \(\dot{p}_1 = -2p_2, \dot{p}_2 = 0\) and

\[
0 = \max_{w \in [-1,1] \times [-1,1]} p^T \tilde{f}(0, w) \tag{3}
\]

But (3) implies that \(p_1 = 0\) which in turn implies \(p_2 = 0\), showing that \(p\) must be trivial.

It is worth noting that even though, for this example, the conditions of Corollary 3.1.2 are satisfied, this Corollary cannot eliminate \(x(t) = 0\) as a candidate. In fact, we can calculate directly that \(H(x, p) = |p_1| + 2|x_1p_2|\) and \(\partial_x H(0, p(t)) = [-2|p_2|, 2|p_2|] \times \{0\}\). Thus, \(p^T = (0, 1)\) also satisfies

\[-\dot{p}^T \in p^T \partial_x H(0, p)\]

Therefore, the “intrinsic” condition in Corollary 3.1.2 is not better in this case than the parametrized condition in Theorem 3.1. Furthermore, this example shows the importance of choosing a proper form of parametrization when applying Theorem 3.1. For discussion about relations between extremals under different parametrizations we refer to [12, 25]. We ought to mention that a similar conclusion can be reached by applying the Kaskosz maximum principle [10].

The next example, adapted from [12, Example 3], will show that the assumption of compactness of \(U\) and continuity of \(f_x(t, \cdot, \cdot)\) in Corollary 3.1.2 cannot be
dispensed with. To show this, we shall apply a special case of a result of Ioffe [9] about the existence of a “semi-C^1” parametrization of a multifunction. This special case is obtained from Ioffe’s Theorem 1 in [9] by defining L in [9, Theorem 1] to be the set of all C^1 mappings from R^n to R^n.

**Theorem 5.1** (Ioffe [9, Theorem 1]) Let F : [0, 1] x V → 2R^n be a multifunction satisfying assumptions (H1), (H2) and (H3). If F is convex-valued and, for every (t_0, x_0) and v_0 ∈ F(t_0, x_0), there exists a selection φ of F such that φ is measurable in t, C^1 in x and φ(t_0, x_0) = v_0 then F has a “semi-C^1” representation (f(t, x, u), U) such that f is measurable in t, C^1 in x and continuous in (x, u).

**Example 5.2.** Consider problem P in R^2 defined by g_0(x) = x_1^2/2 - x_2, g_1(x) = x_2, x_0 = 0 and F(x) = { f(x, u) : u ∈ [-1, 1] x [-1, 0] } with

\[
f_1(x, u) = u_1, \quad f_2(x, u) = \begin{cases} x_1u_1 + u_2 & \text{if } x_1 \geq 0 \\ x_1 + u_2 & \text{if } x_1 \leq 0 \end{cases}
\]

We now prove that F(x) has a semi-C^1 parametrization (with noncompact U). By Ioffe’s Theorem, we need only show that, for every x and v ∈ F(x), there exists a C^1 selection \( \phi \) of F such that \( \phi(x) = v \). To this end, let \( \psi : \mathbb{R} \to \mathbb{R} \) be a C^1 function such that \( \psi(t) = 0 \) for \( t \leq 0 \), \( \psi(t) = 1 \) for \( t \geq 1 \) and \( 0 \leq \psi(t) \leq 1 \) for \( t \in [0, 1] \) (e.g. \( \psi(t) := [\sin(\pi(t - 1/2)) + 1]/2 \) for \( t \in [0, 1] \)). Let \(( \bar{x}, \bar{v} )\) be an arbitrary pair satisfying \( \bar{v} \in F(\bar{x}) \). We define \( \phi \) by different formulas depending on the sign of \( \bar{x}_1 \).

1. \( \bar{x}_1 > 0 \). Then \( \bar{v} = (u_1, \bar{x}_1 u_1 + u_2) \) for some \( u_1 \in [-1, 1] \) and \( u_2 \in [-1, 0] \). We define

\[
\phi(x) = \begin{cases} 1 + \psi(\frac{x_1}{\bar{x}_1})(u_1 - 1) & \text{if } x_1 \geq 0 \\ x_1 [1 + \psi(\frac{x_1}{\bar{x}_1})(u_1 - 1)] + u_2 & \text{if } x_1 \leq 0 \end{cases}
\]

2. \( \bar{x}_1 < 0 \). Then \( \bar{v} = (u_1, \bar{x}_1 + u_2) \) for some \( u_1 \in [-1, 1] \) and \( u_2 \in [-1, 0] \). We define

\[
\phi(x) = \begin{cases} \psi(\frac{x_1}{\bar{x}_1})u_1 & \text{if } x_1 \geq 0 \\ x_1\psi(\frac{x_1}{\bar{x}_1}) + u_2 & \text{if } x_1 \leq 0 \end{cases}
\]

3. \( \bar{x}_1 = 0 \). Then \( \bar{v} = (u_1, u_2) \) for some \( u_1 \in [-1, 1] \) and \( u_2 \in [-1, 0] \). We define

\[
\phi(x) = \begin{cases} \psi(x_1)u_1 & \text{if } x_1 \geq 0 \\ x_1\psi(x_1) + u_2 & \text{if } x_1 \leq 0 \end{cases}
\]
It is easy to check that, in each of these three cases, \( \phi \) is a \( C^1 \) selection of \( F \) that satisfies \( \phi(\tilde{x}) = \tilde{v} \). Thus, \( F \) has a parametrization \((\tilde{f}(x,u),U)\) such that \( \tilde{f} \) is continuous and \( C^1 \) in \( x \).

Observe that, for any \( x \in S_F \), we have
\[
\dot{x}_2 \leq x_1 \dot{x}_1
\]
Thus
\[
g_0(x(1)) = \int_0^1 (x_1(t) \dot{x}_1(t) - \dot{x}_2(t)) dt \geq 0
\]
Therefore, \( x(t) = 0 \) is a solution to problem \( \mathcal{P} \). We show that \( x(t) = 0 \) does not satisfy the necessary condition given in Corollary 3.1.2. We calculate first that
\[
H(x,p) = |p_1| + \frac{|p_2| - p_2^2}{2} + \left\{ \begin{array}{ll}
x_1|p_2| & \text{if } x_1 \geq 0 \\
x_1 p_2 & \text{if } x_1 \leq 0
\end{array} \right.
\]
(4)
It follows that if \( 0 = p(t)^\top \dot{x}(t) = H(0,p(t)) \) then \( p(t) = (0,p_2(t))^\top \) with \( p_2(t) \geq 0 \). Notice that, for such a \( p(t) \),
\[
\partial_x H(0,p(t)) = (p_2(t),0)^\top,
\]
and therefore, the adjoint inclusion in Corollary 3.1.2 becomes \( \dot{p}_1 = -p_2 \) and \( \dot{p}_2 = 0 \). Since \( p_1 = 0 \) we also have \( p_2 = -\dot{p}_1 = 0 \) which shows that \( p \) must be trivial.

We conclude this section by remarking that (as pointed out in section 3 of [24]), using nonuniquely determined (but not necessarily convex-valued) \( \Lambda g \) instead of uniquely determined convex-valued \( \partial g \) in Theorem 3.1 may, in some cases, yield more accurate results.

6 Extension

Theorem 3.1 can be extended to problem involving unilateral constraints and strengthened to yield a Kaskosz-type necessary condition by using the extension of Kaskosz’ maximum principle in Warga [24, Theorem 2.2]. The modification of Theorem 2.2 in [24] to suit our purpose is similar to our argument in the proof of Theorem 3.1. Therefore, we only state the result.

Consider the optimization problem \( \mathcal{P}_1 \):
\[
\begin{align*}
\text{minimize} & \quad g_0(x(1)) \\
\text{subject to} & \quad \dot{x}(t) \in F(t,x(t)) \quad \text{a.e. in } [0,1], \\
& \quad x(0) = x_0, \ g_1(x(1)) = 0, \ g_2(t,x(t)) \in A,
\end{align*}
\]
where \( g_2 : [0, 1] \times V \to \mathbb{R}^p \) is continuous and bounded, \( g_2(t, \cdot) \) Lipschitz and \( A \) a closed convex set in \( \mathbb{R}^p \) with a nonempty interior. Let \( g = (g_0, g_1) \) and assume that \( F \) satisfies the conditions described in Section 3.

Let \( r : [0, 1] \times V \to \mathbb{R}^n \) be a selection of \( \overline{co}F' \) and \( x : [0, 1] \to V \) absolutely continuous and such that
\[
\dot{x}(t) = r(t, x(t)), \quad x(0) = x_0
\]
\[
g_2(t, x(t)) \in A \quad \forall t \in [0, 1]
\]

Then we have

**Theorem 6.1.** Assume that \( x \) is a solution to problem \( P_1 \). Then there exist
\[ l \in \mathbb{R}^n, \quad h \in \Lambda g(x(1)), \quad h_1 : [0, 1] \to \mathbb{R}^n, \quad G : [0, 1] \to L(\mathbb{R}^n, \mathbb{R}^n) \]
and a nonnegative Radon measure \( \mu \) on \([0, 1]\) such that \( G \) and \( h_1 \) are bounded and Borel measurable,

\[ i. \quad |l| + \mu([0, 1]) > 0; \]
\[ ii. \quad \mu\{ t \in [0, 1] : g_2(t, x(t)) \in \text{int}A \} = 0; \]
\[ iii. \quad h_1(t) \in \overline{co}\{ L_1 L_2 : L_1 g_2(t, x(t)) = \max_{a \in A} L_1 a, |L_1| \geq 1, \}
\]
\[ L_2 \in \bigcap_{\epsilon > 0} \overline{co} \bigcup_{|v-x(t)| \leq \epsilon} \partial^r g_2(s, v), |v-x(t)| \leq \epsilon \} \mu \text{ a.e.}; \]
\[ iv. \quad G(t) \in \partial r(t, x(t)) \text{ a.e. in } [0, 1]; \]
\[ v. \quad p(t)^T \dot{x}(t) = H(t, x(t), p(t)), \text{ a.e. in } [0, 1], \quad \text{where} \]
\[
p(t)^T = [l^T h + \int_t^1 h_1(s)\Phi(s)^{-1}\mu(ds)]\Phi(t),
\]
\[
I \text{ is the unit } n \times n \text{ matrix, and } \Phi(\cdot) \text{ is the solution of}
\]
\[
\Phi(t) = I + \int_t^1 \Phi(s)G(s)ds \quad \forall t \in [0, 1].
\]

**Remark.** As shown in [24, 10, 11] by examples, the Kaskosz-type maximum principles sometimes yield better results than the usual maximum principle.

**Acknowledgment:** I am indebted to J. Warga for his help on the proof of Theorem 4.1 and for his comments on an early version of the paper that helped to improve the exposition.
References


