Variational Analysis in Non-reflexive Spaces and Applications to Control Problems with $L^1$ Perturbations

Jonathan M. Borwein and Qiji Zhu
Department of Mathematics and Statistics
Simon Fraser University
Burnaby, BC V5A 1S6
Canada

Research of the first author was supported by NSERC and by the Shrum Endowment at Simon Fraser University.

Acknowledgment: We thank Philip D. Loewen for careful and critical reading of an early version of this paper. We also thank Jane J. J. Ye for her comments.
Please direct all correspondence to:

Jonathan M. Borwein
Department of Mathematics and Statistics
Simon Fraser University
Burnaby, BC V5A 1S6
Canada
Variational Analysis in Non-reflexive Spaces and Applications to Control Problems with $L^1$ Perturbations

Jonathan M. Borwein and Qiji Zhu

Abstract. We provide a refined sensitivity analysis for finite and infinite horizon control problems where in both cases the perturbation space is $L^1$. Our underlying technique relies on a recent sequential description of both the generalized gradient of Clarke and of the approximate $G$–subdifferential of functions defined on a smooth Banach space. We also show that the proximal limit formula for the generalized gradient and its $L^p$ analogue are direct consequences of these sequential formulas. Related characterizations of Lipschitzness of a function on a smooth space are given.

Keywords: Weak-Hadamard sub-derivatives, Hölder sub-derivatives, variational principles, smooth renorms, Clarke–subdifferentials, $G$–subdifferentials, sensitivity analysis, control system, infinite horizon problems.

1 Introduction

It has long been recognized that the value function plays an important role in optimization. It measures the sensitivity of the problem to perturbations of the objective function and the various constraints. Particularly interesting is the derivative of the value function, a measure of so called “differential stability”. When the value function is differentiable, it plays the role of a multiplier. In the context of dynamic optimization this observation establishes an heuristic relationship between the Maximum principle and the Dynamic programming approaches. Generally, however, the value function of a constrained optimization problem is far from being differentiable. To obtain a rigorous treatment of these heuristic relations one needs to apply the techniques of nonsmooth analysis.

Consider the constrained finite dimensional optimization problem:

$$(P) : \text{minimize } f(x)$$

subject to $g(x) = 0$.

Suppose that we perturb the constraint and define the value function of the perturbed problem as

$$V(z) := \inf\{f(x) : g(x) = z\}.$$ 

Then a precise form of the heuristic relation previously alluded to is (cf. [6])

$$-\partial_C V(0) \subset clco M(\Sigma)$$

where $\partial_C$ is the Clarke generalized gradient, $\Sigma$ is the solution set of the optimization problem $(P)$ and $M(\Sigma)$ denotes the set of multipliers corresponding to all solutions in $\Sigma$. This inclusion has been extended to optimal control problems with finite dimensional or $L^2$ perturbations [7, 16, 23].

However, in the setting of dynamic optimization problems, in particular those discussed in [7, 16, 23] that involve control systems, the natural space for perturbations is $L^1$ rather than $L^2$ or a finite dimensional space. The reason that in previous research the perturbation spaces were chosen to be Hilbertian or finite dimensional is principally that the underlying techniques depended crucially on the analysis of nearest points and relied heavily on the inner product structure of the perturbation space to establish a limit expression for the generalized (Clarke) gradient of the value function. Whether
such a sensitivity analysis can be carried out in an $L^1$ space or even in an $L^p$, $p \in (1, 2)$, space is an interesting problem in its own right.

It has become clear recently (cf. Borwein and Ioffe [2]) that the limit expression for the generalized (Clarke) gradient or even for smaller objects such as the G-subdifferential defined by Ioffe [13] can be established without recourse to the inner product structure of the perturbation space. In fact all one needs is a smooth renorming for an appropriate bornology. This new technique enables us to perform sensitivity analysis in a much more general setting. In particular it allows us to discuss $L^1$ perturbations of control problems (which is the largest reasonable function space for perturbations in such problems). In this paper we establish some abstract sensitivity analysis results for optimization problems with perturbations in any weakly compactly generated (WCG) Banach space which possesses a weak Hadamard renorming. Combining these results with a recent weak Hadamard renorming theorem for $L^1$ due to Borwein-Fitzpatrick [3] allows for a sensitivity analysis of control problems with an $L^1$ perturbation. We thus improve the existing sensitivity results in two respects: a) we allow $L^1$ perturbations; b) we established a sharper formula:

$$-\partial_G V(0) \subset M(\Sigma)$$

where $\partial_G$ is the G-subdifferential introduced by Ioffe[14]. Since $cl^{\star}co \partial_G V(0) = \partial_C V(0)$, Clarke’s formula is a direct corollary of our formula. This is of interest even in the case of finite dimensional perturbations. Our main result is in Section 4 (Theorem 4.4 and Theorem 4.8).

Our research shows that with Borwein and Ioffe’s new sequential limit expressions of the G-subdifferential and of the Clarke subgradient, one can avoid using the proximal derivative which is by nature restricted to Hilbert spaces. Interestingly, we can easily re-derive the classical proximal limit formula for the Clarke generalized gradient by combining the Borwein-Ioffe formula with the Borwein-Preiss smooth variational principle [5]. In fact, the same argument allows us to establish a H"older subderivative analogue of the the proximal limit formula for the G-subdifferential. This illustrates that the Borwein-Ioffe formula contains more information than its proximal counterpart. We include these and other related results in the appendix.

We arrange the paper as follows: in the next section we introduce notations and facts from nonsmooth analysis that will be used in the sequel. We state our abstract results in Section 3. In Section 4, we apply these abstract results
to the $L^1$ sensitivity analysis of both finite and infinite horizon optimal control problems. Longer proofs are deferred to Section 5.

2 Preliminaries

Let $X$ be a Banach space with closed unit ball $B$ and with continuous real dual $X^*$. We write $\rho(S,x)$ for the distance from $x$ to $S$ in the given norm $\|\cdot\|$ and $clS$, $cl^*S$ and $coS$ for the closure, weak-star closure and convex hull of $S$, respectively. A bornology $\beta$ of $X$ is a family of closed bounded and centrally symmetric (convex) subsets of $X$ such that for any finite set of $x \in X$ there are a $U \in \beta$ and $\lambda > 0$ such that $x \in \lambda U$ for all $x$ of the finite set. (It follows, in particular, that any finite dimensional subspace belongs to the subspace spanned by some element of the bornology.) The most important bornology in this paper is the weak Hadamard bornology (denoted by WH) formed by all (convex symmetric) weak compact sets (see [5, 11, 18]).

By a function we always mean an extended-real-valued function, usually lower semicontinuous and proper (that is to say, not everywhere equal to $+\infty$ and nowhere to $-\infty$). Given a function $f$ on $X$, we say $f$ is $\beta$-differentiable and has a $\beta$-derivative $\nabla^\beta f(x)$ at $x$ if

$$t^{-1}(f(x + tu) - f(x) - t(\nabla^\beta f(x), u)) \to 0$$

as $t \to 0$ uniformly in $u \in V$ for every $V \in \beta$. We say that a function $f$ is $\beta$-smooth if $\nabla^\beta f(x) : X \to X^*_\beta$ is continuous, where $X^*_\beta$ is the dual space of $X$ endowed with the topology of uniform convergence on $\beta$-sets. It is not hard to check that for a convex function, $\beta$-smooth and $\beta$-differentiable on a convex set are equivalent.

**Definition 2.1** The $\beta$-subdifferential of rank $k$ of the function $f$ at $x$ is the set $\partial_k^\beta f(x)$ of vectors $x^* \in X^*$ with the property that there is a function $m(\cdot)$ (depending on $x^*$) which is $\beta$-smooth on a neighborhood $V$ of $x$ and satisfies the following conditions:

- $(m_1)$ $m$ satisfies a Lipschitz condition with constant $k$ on $V$;
- $(m_2)$ $m(v) \leq f(v)$ for all $v \in V$;
- $(m_3)$ $m(x) = f(x)$;
(m₄) \( x^* = \nabla_\beta m(x) \).

We define the \( \beta \)-subdifferential of the function \( f \) at \( x \) by

\[
\partial_\beta f(x) := \bigcup_{k=1}^{\infty} \partial_{k}^{\beta} f(x).
\]

**Definition 2.2** A vector \( x^* \) is a Clarke normal to \( S \) at \( x \) if for any \( \epsilon > 0 \) and any finite dimensional subspace \( L \subset X \) there are \( \lambda > 0 \) and \( u^* \in X^* \) such that

\[
|\langle u^* - x^*, h \rangle| \leq \epsilon \|h\|, \quad \forall h \in L, \quad \text{and} \quad \lambda \rho^\circ(S, x; h) \geq \langle u^*, h \rangle, \quad \forall h,
\]

where \( \rho^\circ(S, x; h) \) is the Clarke directional derivative of \( \rho(S, \cdot) \) at \( x \):

\[
f^\circ(x; h) = \limsup_{u \to x, t \to 0^+} t^{-1} (f(u + th) - f(u)).
\]

The collection of all Clarke normals to \( S \) at \( x \) is a convex weak-star closed cone denoted \( \mathcal{N}_C(S, x) \). A vector \( x^* \) is a \( G \)-normal to \( S \) at \( x \) if for any \( \epsilon > 0 \) and any finite dimensional subspace \( L \subset X \) there are \( \lambda > 0 \), \( u \in X \) with \( \|u - x\| \leq \epsilon \) and \( u^* \in X^* \) such that

\[
|\langle u^* - x^*, h \rangle| \leq \epsilon \|h\|, \quad \forall h \in L, \quad \text{and} \quad \lambda d^- \rho(S, u; h) \geq \langle u^*, h \rangle, \quad \forall h \in L,
\]

where \( d^- \rho(S, x; h) \) is the lower Dini directional derivative of \( \rho(S, \cdot) \) at \( x \):

\[
d^- f(x; h) = \liminf_{e \to h, t \to 0^+} t^{-1} (f(x + te) - f(x)).
\]

The collection of all \( G \)-normals to \( S \) at \( x \) is a weak-star closed (generally non-convex) cone denoted \( \mathcal{N}_G(S, x) \).

**Definition 2.3** Let \( f \) be a lower semi-continuous function on \( X \) which is finite at \( x \). The \( G \)-subdifferential of \( f \) at \( x \) is

\[
\partial_G f(x) = \{ x^* \in X^* : (x^*, -1) \in \mathcal{N}_G(\text{epi} f, (x, f(x))) \}.
\]

Replacing the \( G \)-normal cone by Clarke’s normal cone, we obtain the definitions of the generalized gradient of Clarke \( \partial_C f(x) \).
Theorem 2.1 [2] Let $X$ be a Banach space whose norm is $\beta$-differentiable away from the origin. Let $f$ be a locally Lipschitz function on $X$. Then for any $x \in X$
\[
\partial_G f(x) = \bigcup_{k=1}^{\infty} \{ w^* - \lim_{n \to \infty} x_n^*: x_n^* \in \partial_{\beta}^k f(x_n), x_n \to_f x \},
\]
and
\[
\partial_C f(x) = \text{cl}^* \text{co} \partial_G f(x).
\]

We refer the reader to [2, 3, 4] for additional definitions and discussion. We will frequently need the following form of Gronwall’s inequality.

Theorem 2.2 (Gronwall’s inequality) Let $u$ and $v$ be nonnegative continuous functions and $k$ a nonnegative integrable function defined on $[0, T]$ such that
\[
u(t) \leq \int_0^t k(s)u(s)ds + v(t), \forall t \in [0, T].
\]
Then, for any $t \in [0, T]$,
\[
u(t) \leq v(t) + \int_0^t k(s)v(s)e^{\int_0^r k(r)dr} ds.
\]
In particular, if $\overline{v}$ is an upper bound for $v(t)$ on $[0, T]$ then
\[
u(t) \leq \overline{v} e^{\int_0^r k(s)ds}.
\]

3 Abstract Multiplier Results and Maximum Principles

Let $X, X_0, Z$ be Banach spaces and $U$ a non-empty convex compact subset of a locally convex linear space $Y$. Assume that $X$ is continuously embedded in $X_0$, and that $Z$ is WCG (contains a weakly compact set whose span is dense) and possesses an equivalent weak Hadamard renorming. We will use
the same notation for elements in $X$ and $X_0$ when it does not cause confusion. Consider the following (global) optimization problem with parameter $z \in Z$:

$$P(z) \text{ minimize } F(x, u)$$
subject to $G(x, u) = z$

where $F : X_0 \times U \to R$ and $G : X \times U \to Z$. We will need the following assumptions.

(H1) $G$ maps $X \times U$ onto $Z$, is Lipschitz in $X$ uniformly in $U$, is affine for $u \in U$ and, for fixed $x$, $G(x, \cdot)$ is a continuous map from $U$ to $Z$ with the weak topology.

(H2) $G$ is Fréchet differentiable with respect to $x$, $(\nabla^F_x G(x, u))^{-1}$ exists (as a continuous linear operator) for each $(x, u) \in X \times U$ and, for each $u$, $(\nabla^F_x G(x, u))^{-1}$ is continuous in $x$.

(H3) $F$ is Lipschitz in $X_0$ uniformly in $U$, is Fréchet differentiable with respect to $x$, is affine in $u$ and, for fixed $x$, $F(x, \cdot)$ is continuous.

**Theorem 3.1** (Maximum Principle) Assume that $F$ and $G$ satisfy conditions (H1), (H2) and (H3). Let $(x_z, u_z)$ be a solution to $P(z)$. Then there exist a $z^* \in Z^*$ such that

i) $z^* = \nabla^F_x F(x_z, u_z) \cdot (\nabla^F_x G(x_z, u_z))^{-1}$

and

ii) $(z^*, G(x_z, u_z)) - F(x_z, u_z) = \max_{u \in U} \{ (z^*, G(x_z, u)) - F(x_z, u) \}$.

**Remark 3.1** If we consider the parameter $u$ as a control then Theorem 3.1 is an abstract form of the familiar Maximum Principle in optimal control theory.

**Definition 3.1** For a solution $(x_z, u_z)$ to $P(z)$, we denote all $z^*$ which satisfy i) and ii) in our Maximum Principle by $M(x_z, u_z)$. Let $\Sigma_z$ be the solution set of $P(z)$. Define

$$M(\Sigma_z) := \bigcup_{(x, u) \in \Sigma_z} M(x, u).$$
Let

\[ V(z) := \inf\{ F(x, u) : G(x, u) = z \} \]

denote the associated (global) value function.

**Theorem 3.2** Assume that \( G \) and \( F \) satisfy conditions \((H1),(H2)\) and \((H3)\).
If \( V \) is weak Hadamard subdifferentiable at \( z \) of order \( k \) and \( \Sigma_z \neq \emptyset \) then, for any \( z^* \in \partial_{W}^{k}V(z) \), there exists \( (x_z, u_z) \in \Sigma_z \) such that

\[ i) \quad \langle z^*, G(x_z, u_z) \rangle - F(x_z, u_z) = \max_{u \in U} \{ \langle z^*, G(x_z, u) \rangle - F(x_z, u) \} \]

and

\[ ii) \quad z^* = \nabla^F_x F(x_z, u_z) \cdot (\nabla^F_x G(x_z, u_z))^{-1}. \]

**Remark 3.2** Theorem 3.2 is a type of regularity result for the weak Hadamard subderivatives of the value function. It asserts that any such subderivative satisfies additional structural conditions (the Maximum Principle i) and ii)) which typically force that subderivative to lie in a better behaved space.

### 4 Applications to Control Problems

In this section we apply the abstract results of section 3 and the Borwein-Fitzpatrick \( L^1 \) renorming theorem [3] to the sensitivity analysis of \( L^1 \) perturbations of both finite and infinite horizon optimal control problems.

#### 4.1 A finite horizon optimal control problem

Let \( U \) be a compact subset of \( \mathbb{R}^m \). A version of the Dunford-Pettis theorem asserts that \( L^1([0, 1], C(U))^* \) is isomorphic to the set \( N \) of measurable functions \( \nu : [0, 1] \to (B(U), | \cdot |_w) \) such that \( \operatorname{ess sup}|\nu(\cdot)||U| < \infty \) where \( (B(U), | \cdot |_w) \) is the vector space of all Borel measures in \( U \) endowed the norm induced by the weak-star topology. Let \( \mathcal{U} \) be the subset of \( N \) of all measurable mappings from \([0, 1]\) to the Borel probability measures in \( U \), i.e., \( \mathcal{U} := \{ u : u \text{ is a measurable mapping from } [0, 1] \text{ to } prob(U) \} \). Then \( \mathcal{U} \)
is a convex compact subset of $\mathcal{N}$ and, therefore, is a complete metric space under the metric induced by $|\cdot|_w$ (cf. [22] for details).

Consider the optimal control problem

$$
P(0) \text{ minimize } \int_0^1 f(s, x(s), u(s)) ds
$$

subject to

$$
\dot{x}(s) = g(s, x(s), u(s))
$$

$$
x(0) = 0, \ u \in \mathcal{U}
$$

where $g : [0, 1] \times \mathbb{R}^n \times U \to \mathbb{R}^n$ and $f : [0, 1] \times \mathbb{R}^n \times U \to \mathbb{R}$ are measurable in $s$, continuous in $(x, u)$, continuously (Fréchet) differentiable in $x$ and

(C) there exists $k \in L^\infty([0, 1], \mathbb{R}^n)$ such that

$$
||h(s, x, u)|| \leq k(s) \ \forall x \in \mathbb{R}^n, \ u \in \mathcal{U}
$$

where $h$ represents each of $f, g, f_x$ and $g_x$. For $u \in \text{prob}(U)$, $h(s, x, u) := \int_{\mathcal{U}} h(s, x, r) u(\ dr)$. The control set $\mathcal{U}$ defined here is the relaxed control set (cf. [22]). It is known that

$$
\{(g(s, x, u), f(s, x, u)) : u \in \text{prob}(U)\} = \text{co}\{ (g(s, x, u), f(s, x, u)) : u \in U \}
$$

$$
= \{ \sum_{i=1}^{n+2} \lambda_i (g(s, x, u_i), f(s, x, u_i)) : u_i \in U, \ \sum_{i=1}^{n+2} \lambda_i = 1, \lambda_i \geq 0 \}.
$$

When $\{ (g(s, x, u), f(s, x, u)) : u \in U \}$ is convex as assumed in [7] and [23] the relaxed control set $\mathcal{U}$ coincides with the usual control set which consists of all measurable mappings from $[0, 1]$ to $U$. We also consider the perturbed problem:

$$
P(z) \text{ minimize } \int_0^1 f(s, x(s), u(s)) ds
$$

subject to

$$
\dot{x}(s) = g(s, x(s), u(s)) + z^1(s)
$$

$$
x(0) = z^0, \ u \in \mathcal{U}
$$

where $z = (z^0, z^1) \in \mathbb{R}^n \times L^1([0, 1], \mathbb{R}^n)$. We denote the (global) optimal value of problem $P(z)$ by $V(z)$. By the compactness of $\mathcal{U}$, we have

**Theorem 4.1** [22, Section V.1, page 296] *For every* $z \in \mathbb{R}^n \times L^1([0, 1], \mathbb{R}^n)$, *the solution set of problem* $P(z)$ *is not empty.*
We now formulate problem \( P(z) \) as an abstract functional optimization problem of the type described in Section 3.

Define \( X := W^{0,1} := \{ x : x \in L^1([0, 1], R^n), x \in C([0, 1], R^n) \} \) with norm
\[
\|x\|_W := \|x\|_\infty + \|x\|_{L^1}, \quad X_0 := L^1([0, 1], R^n), \quad Z := R^n \times L^1([0, 1], R^n) \]
with norm \( \|z\|_Z := \|(z^0, z^1)\|_Z := \|z^0\| + \|z^1\|_{L^1}, \quad \mathcal{U} := \mathcal{U} \subset Y := L^1([0, 1], C(U))^*, \)
\( F(x, u) := \int_0^1 f(s, x(s), u(s))ds \) and \( G(x, u) := (x(0), \dot{x}(\cdot) - g(\cdot, x(\cdot), u(\cdot))) \).

Then problem \( P(z) \) can be written as:
\[
P(z) \quad \text{minimize} \quad F(x, u) \\
\text{subject to} \quad G(x, u) = z.
\]

We now check \( F \) and \( G \) satisfy condition (H1)-(H3). We check condition (H1) first. The affine dependence on \( U \) follows directly from the definition of the relaxed control set \( U \) and since \( g \) is (Fréchet) differentiable use of the (finite dimensional) mean value theorem and condition (C) yields the Lipschitzness of \( G \). To see that \( G \) maps \( X \times U \) onto \( Z \), observe that, for any \( z \in Z \) and a fixed \( u \in U = \mathcal{U} \), the initial value problem
\[
\dot{x}(s) = g(s, x(s), u(s)) + z^1(s) \\
x(0) = z^0
\]
has a unique local solution. Under condition (C) this solution extends to \([0, 1]\) and, thus, belongs to \( X = W^{0,1} \). Since \( z \in Z \) is arbitrary, we have proved that \( G \) maps onto \( Z \). It follows from the definition of the metric on \( U \) that, for a fixed \( x \), \( G(x, \cdot) \) is a continuous map from \( U \) to \( Z \) with the weak topology.

To check (H2) we make the following calculation. Define, for any \( y \in W^{0,1} \),
\[
\mathcal{L}y := (y(0), \dot{y}(\cdot) - g(\cdot, x(\cdot), u(\cdot))y(\cdot)).
\]
Let \( W \) be any bounded convex subset of \( W^{0,1} \). Then, for \( y \in W \), \( \|y\|_\infty \) is bounded, say by \( M \) and, for some \( \theta(t) \) in \((0, 1)\),
\[
\|t^{-1}(G(x + ty, u) - G(x, u)) - \mathcal{L}y\|_Z \\
= \int_0^1 |t^{-1}(g(s, x(s) + ty(s), u(s)) - g(s, x(s), u(s))) - g_x(s, x(s), u(s))y(s)|ds \\
= \int_0^1 |(g_x(s, x(s) + \theta(t)ty(s), u(s)) - g_x(s, x(s), u(s)))y(s)|ds \\
\leq M\int_0^1 |g_x(s, x(s) + \theta(t)ty(s), u(s)) - g_x(s, x(s), u(s))|ds.
\]
By condition (C) the integrand in the right-hand-side is bounded by $2k(s)$ and pointwise converges to 0 when $t \to 0$ uniformly for $y \in W$. Thus, by the dominated convergence theorem, $G$ is Fréchet differentiable with respect to $x$ and $\nabla^F_x G(x, u) = \mathcal{L}$ that is, for any $y \in W^{0,1}$,

$$\nabla^F_x G(x, u)y = (y(0), y(\cdot) - g_x(\cdot, x(\cdot), u(\cdot)y(\cdot))).$$

Observe that, for any $z = (z^0, z^1) \in R^n \times L^1([0, 1], R^n)$, $\nabla^F_x G(x, u)y = z$ is, in terms of a differential equation,

$$\dot{y}(s) - g_x(s, x(s), u(s))y(s) = z^1(s), \quad y(0) = z^0.$$  \hfill (1)

Condition (C) implies that, for any $x$ and $u$, this linear differential equation has an unique solution on $[0, 1]$. Since this is true for any $z \in Z$, $\nabla^F_x G(x, u)$ is 1-1 onto $Z$ for any $(x, u)$. Thus $(\nabla^F_x G(x, u))^{-1}$ is a bounded linear operator from $Z$ to $W^{0,1}$. It remains to show that $(\nabla^F_x G(x, u))^{-1}$ is continuous in $x$ (as an operator from $Z$ to $W^{0,1}$). Fix $u \in \mathbf{U}$, consider $z \in B_Z$, for a neighborhood $W$ of $x$ in $W^{0,1}$, let $y(s, x)$ be the solution of (1). By condition (C) $|g_x(s, x(s), u(s))| \leq k(s)$. Using Gronwall’s inequality we obtain, for all $x \in W$,

$$|y(s, x)| \leq ||z||_Z e^{\|k\|_\infty} \leq e^{\|k\|_\infty},$$ \hfill (2)

i.e., $||y(\cdot, x)||_\infty$ is uniformly bounded (independent of $z \in B_Z$). Now, for $x' \in W$,

$$|y(t, x') - y(t, x)|$$

\begin{align*}
&= \left| \int_0^t (g_x(s, x'(s), u(s))y(s, x') - g_x(s, x(s), u(s)))y(s, x)ds \right| \\
&\leq \int_0^t \left| (g_x(s, x'(s), u(s)) - g_x(s, x(s), u(s)))y(s, x) \right| ds \\
&\quad + \int_0^t |g_x(s, x'(s), u(s)) - g_x(s, x(s), u(s))| |y(s, x') - y(s, x)| ds \\
&\leq \phi(x') + \int_0^t k(s)|y(s, x') - y(s, x)| ds \\
\end{align*}

where

$$\phi(x') = \int_0^t \left| (g_x(s, x'(s), u(s)) - g_x(s, x(s), u(s)))y(s, x) \right| ds.$$
Applying Gronwall’s inequality yields that
\[ |y(t, x') - y(t, x)| \leq \phi(x') \exp(\int_0^t k(r) dr) \leq \phi(x') \exp(\int_0^1 k(r) dr). \]
Since the right-hand-side does not depend on \( t \), we have
\[ \|y(\cdot, x') - y(\cdot, x)\|_\infty \leq \phi(x') \exp(\int_0^1 k(r) dr). \]
When \( x' \to x \) in \( W^{0, 1} \), \( \|(g_x(s, x'(s), u(s)) - g_x(s, x(s), u(s)))y(s, x)\| \to 0 \) pointwise. Using condition (C) and the dominated convergence theorem we have
\[ \phi(x') := \int_0^1 |(g_x(s, x'(s), u(s)) - g_x(s, x(s), u(s)))y(s, x)| ds \to 0. \]
Therefore, \( \|y(\cdot, x') - y(\cdot, x)\|_\infty \to 0 \). For the derivative of \( y \) we have the following relation
\[ \|\dot{y}(\cdot, x') - \dot{y}(\cdot, x)\|_{L^1} = \int_0^1 |(g_x(s, x'(s), u(s))y(s, x') - g_x(s, x(s), u(s)))y(s, x)| ds \]
Observe that the integrand is bounded by \( 2k(s)e^{\exp(\|k\|_\infty)} \) (condition (C) and (2)) and converges to 0 pointwise when \( x' \to x \), again by the dominated convergence theorem
\[ \|\dot{y}(\cdot, x') - \dot{y}(\cdot, x)\|_{L^1} \to 0, \quad \text{as} \ x' \to x. \]
Thus, \( \|y(\cdot, x') - y(\cdot, x)\|_{W^{0, 1}} \to 0 \). Since all the estimates are independent of \( z \in B_z, (\nabla^E G(x, u))^{-1} \) is continuous in \( x \). We can check condition (H3) similarly.

Thus, Theorem 3.1 yields:

**Theorem 4.2** (Maximum Principle) Let \((x_z, u_z)\) be a local solution to the problem \( P(z) \). Then there exists an absolutely continuous function \( p \) such that
\[
\text{i)} \quad -\dot{p}(s) = (g_x(s, x_z(s), u_z(s)))^\top p(s) - f_x(s, x_z(s), u_z(s)), \quad p(1) = 0
\]
and
\[
\text{ii)} \quad \langle p(s), g(s, x_z(s), u_z(s)) \rangle + f(s, x_z(s), u_z(s)) = \max_{u \in \mathcal{U}} \{\langle p(s), g(s, x_z(s), u) \rangle + f(s, x_z(s), u)\}.
\]
**Proof.** See Section 5. ☹️

Also, we have:

**Theorem 4.3** \( V \) is (globally) Lipschitz.

The proof of this theorem is similar to and easier than the proof of Theorem 4.7 in the sequel and is, therefore, omitted. ☹️

We call an absolutely continuous function satisfying i) and ii) in the maximum principle a **multiplier** of the problem \( P(z) \) and we will use the notations \( M(x_z, u_z) \) and \( M(\Sigma_z) \) in the same way as in Section 3. Then we have the following main result:

**Theorem 4.4**

\[-\partial_G V(0) \subset \{(p(0), p(\cdot)) : p \in M(\Sigma_0)\}.

**Proof.** See Section 5. ☹️

Notice that \( cl^*co \partial_G V(0) = \partial_G V(0) \) because \( V \) is locally Lipschitz. As a direct consequence of Theorem 4.4 we have:

**Corollary 4.1**

\[-\partial_G V(0) \subset cl^*co\{(p(0), p(\cdot)) : p \in M(\Sigma_0)\}.

**Remark 4.1**

a) If we consider perturbations in \( L^p \) instead of \( L^1 \) then in condition (C) \( k \in L^\infty([0, 1], R^a) \) can be replaced by \( k \in L^q([0, 1], R^a) \) where \( 1/q + 1/p = 1 \). Our results then extend the results in [7] given for \( p = 2 \).

b) In contrast to [7], we have perturbed both the control equation and the initial condition. When there is only a perturbation on the initial value, \( \tilde{V}(z^0) := V(z^0, 0) \) is a function defined on \( R^n \). Since \( \partial_G V(0) = \partial_{G,z} V(0, 0) \) is contained in the projection of \( \partial_G V(0, 0) \) to \( R^n \), we obtain

\[-\partial_G \tilde{V}(0) \subset \{p(0) : p \in M(\Sigma_0)\}.

We thus get information on the initial condition perturbation at no extra cost.
c) One reason for establishing the formula given in Theorem 4.4 and Corollary 4.1 is to use the inclusions to derive properties of the value function because the generalized subdifferential of $V$ is very hard to calculate while properties of the multipliers can be derived through relations in the Maximum Principle. In finite-dimensional space it has long been known that the boundness of $\partial_c V$ implies the Lipschitzness of $V$. This was extended to Hilbert space recently in [10] by using the Borwein-Preiss smooth variational principle. A similar argument works in more general settings. We include such an extension to smooth spaces in the Appendix.

d) We should note that in [7, 16] the terminology multiplier is refer to a Halmilton multiplier which is different from ours. Under the assumption that $f$ and $g$ are smooth with respect to $x$ as assumed here and in [7], it is not hard to show (see the proof of [26, Corollary 3.1.2]) that our multiplier set is no bigger than the Halmilton multiplier set defined in [7, 16]. Thus our inclusion in Corollary 4.1 is potentially sharper than the corresponding results in [7, 16].

As shown by the example in the end of [7] it is impossible to replace the relaxed multiplier set in Theorem 4.4 and Corollary 4.1 by the (generally) strictly smaller original multiplier set that defined by conditions i) and ii) in Theorem 4.2 with the relaxed control set replaced by the original one.

Similar remark applies to Theorem 4.8 and Corollary 4.2 in the next subsection.

### 4.2 An infinite horizon optimal control problem

Let $\mathcal{U}$ be the set of all measurable mapping from $[0, \infty)$ to the Borel probability measure in $U \subset \mathbb{R}^n$, where $U$ is a compact set. Then $\mathcal{U}$ is a convex set. For $0 \leq a < b$, define $\mathcal{U}_{[a,b]} := \{u|_{[a,b]} : u \in \mathcal{U}\}$. Then $\mathcal{U}_{[a,b]}$ is a weak-star compact subset of $L^1([a, b]; C(U))^*$ (cf. [22] for details).

Consider the infinite horizon optimal control problem

\[
\begin{align*}
P(0) \quad & \text{minimize} & & J(x, u) := \int_0^\infty e^{-Ls} f(s, x(s), u(s)) ds \\
\text{subject to} & & \dot{x}(s) & = g(s, x(s), u(s)) \\
& & x(0) & = 0, \ u \in \mathcal{U}
\end{align*}
\]

where $g : [0, \infty) \times \mathbb{R}^n \times U \to \mathbb{R}^n$ and $f : [0, \infty) \times \mathbb{R}^n \times U \to \mathbb{R}$ are measurable in $s$, continuous in $(x, u)$, continuously (Fréchet) differentiable with respect
to \( x \) and there exists a constant \( N \) such that \( N \leq L \) and

\[ (D) \quad \begin{align*}
(i) \quad & |h(s,0,u)| \leq N, \quad \forall u \in U, \ s \in [0,\infty) \\
(ii) \quad & |h(s,x,u) - h(s,y,u)| \leq N|x-y| \quad \forall x, y \in \mathbb{R}^n, \ u \in U, \ s \in [0,\infty)
\end{align*} \]

where \( h \) represents either \( f \) or \( g \). Notice that the Lipschitz condition (ii) in (D) implies that

\[ |h_x(s,x,u)| \leq N \quad \forall x, y \in \mathbb{R}^n, \ u \in U, \ s \in [0,\infty) \]

for \( h = f \) or \( g \). For \( u \in \text{prob}(U) \), \( h(s,x,u) := \int_U h(s,x,r)u(dr) \).

The associated perturbed problem is:

\[
P(z) \quad \text{minimize} \quad \int_0^\infty e^{-Ls}f(s,x(s),u(s))ds
\]

subject to

\[
\begin{align*}
\dot{x}(s) &= g(s,x(s),u(s)) + z^1(s) \\
x(0) &= z^0, \ u \in \mathcal{U}
\end{align*}
\]

where \( z = (z^0,z^1) \in \mathbb{R}^n \times L^1([0,\infty), \mathbb{R}^n) \). The problem is well defined. In fact, for any \( z \), consider the solution of

\[
\dot{x}(s) = g(s,x(s),u(s)) + z^1(s), \ x(0) = z^0.
\]

Using Gronwall’s inequality and the Lipschitz condition on \( g \) we have

\[
|x(t)| \leq (Nt + |z^0| + \int_0^\infty |z^1(s)|ds)e^{Nt}.
\]

Thus, condition (D) shows that \( \int_0^\infty e^{-Ls}f(s,x(s),u(s))ds \) is always defined.

J. Ye discussed perturbed infinite horizon control problems (with finite dimensional perturbations) in [23] and proved a maximum principle for such systems in [24]. We will need two results from [23, 24]. While the assumptions in [23, 24] are slightly different from ours (we have posed more restrictions on \( g \) while in [23, 24] a stronger condition was required for \( f \)), the proofs can be easily adapted to suit our case. A complete proof of Lemma 4.1 below is included in Section 5 and the adaptation for deriving a proof for Theorem 4.5 is similar.

**Lemma 4.1** [23] Let \( z_n \) be a sequence converging to \( z \) in \( \mathbb{R}^n \times L^1([0,\infty), \mathbb{R}^n) \) and let \( (x_n,u_n) \) be feasible for \( P(z_n) \). Then there exists a subsequence \((x_i,u_i)\)
of \((x_n, u_n)\) such that \(x_i\) converges to \(x\) uniformly on any finite interval \([a, b]\) and \(u_i\) converges to \(u\), in any \(U_{[a, b]}\). Moreover, \((x, u)\) is feasible for \(P(z)\) and

\[
J(x, u) \leq \liminf_{n \to \infty} J(x_n, u_n).
\]

**Proof.** See Section 5. 

The following maximum principle for the problem \(P(z)\) is an adaptation of [24, Theorem 2.1].

**Theorem 4.5** Let \((x_z, u_z)\) be a local solution to the problem \(P(z)\). Then there exists an absolutely continuous function \(p\) such that

i)

\[
\dot{p}(s) = g_z(s, x_z(s), u_z(s)) + p(s) L^s f_z(s, x_z(s), u_z(s))
\]

and

ii)

\[
\langle p(s), g(s, x_z(s), u_z(s)) \rangle + e^{-L^s f(s, x_z(s), u_z(s))}
\]

\[= \max_{u \in U} \{ \langle p(s), g(s, x_z(s), u) \rangle + e^{-L^s f(s, x_z(s), u)} \}. \]

Lemma 4.1 immediately leads to:

**Theorem 4.6** For any \(z \in R^c \times L^1([0, \infty), R^c)\), the solution set of \(P(z)\) is not empty.

We again denote the optimal value of problem \(P(z)\) by \(V(z)\) and the solution set of \(P(z)\) by \(\Sigma_z\). Then we have:

**Theorem 4.7** \(V\) is (globally) Lipschitz.

**Proof.** Consider \(y, z \in Z\). By Theorem 4.6, let \((x_y, u_y)\) be an optimal solution pair corresponding to \(y\), i.e.,

\[
\dot{x}_y(s) = g(s, x_y(s), u_y(s)) + y^1(s), \quad x_y(0) = y^0
\]
and
\[ V(y) = \int_0^\infty e^{-L_s} f(s, x_y(s), u_y(s)) ds. \]

Consider the solution \( x \) of
\[ \dot{x}(s) = g(s, x(s), u_y(s)) + z^1(s), \quad x(0) = z^0. \]

Then
\[ |x(t) - x_y(t)| \leq \int_0^t |g(s, x(s), u_y(s)) - g(s, x_y(s), u_y(s))| ds \]
\[ + \int_0^t |z^1(s) - y^1(s)| ds + |z^0 - y^0| \]
\[ \leq \int_0^t N|x(s) - x_y(s)| ds + \|z - y\|_\mathcal{L}. \]

By Gronwall’s inequality
\[ |x(t) - x_y(t)| \leq \|z - y\|_\mathcal{L} e^{-M t}, \quad t \in [0, \infty). \]

Thus,
\[ V(z) - V(y) \leq \int_0^\infty e^{-L_s} (f(s, x(s), u_y(s)) - f(s, x_y(s), u_y(s))) ds \]
\[ \leq \int_0^\infty e^{-L_s} N|x(s) - x_y(s)| ds \leq N\|z - y\|_\mathcal{L} \int_0^\infty e^{-(L-N)s} ds. \]

Since the roles of \( y \) and \( z \) may be reversed in the preceding arguments \( V \) is Lipschitz (of rank \( N \int_0^\infty e^{-(L-N)s} ds \)).

We again define an absolutely continuous function \( p \) satisfying i) and ii) in Theorem 4.5 to be a multiplier corresponding to \((x_z, u_z)\). We will use the notations \( M(x_z, u_z) \) and \( M(\Sigma_z) \) as defined in Section 3. Then we have:

**Theorem 4.8**

\[ -\partial_c V(0) \subset \{(p(0), p(\cdot)) : p \in M(\Sigma_0)\}. \]

**Proof.** See Section 5.

Finally Theorem 2.1 yields:

**Corollary 4.2**

\[ -\partial_c V(0) \subset cl^* co\{(p(0), p(\cdot)) : p \in M(\Sigma_0)\}. \]
5 Proofs of Technical Assertions

5.1 Proofs for results in Section 3.

In this section, with a little abuse of notations we will often extend affine mappings on \( U \) to the span of \( U \) (perhaps discontinuously) without further comment. This allows us to discuss derivatives of functions with respect to \( U \).

**Proposition 5.1** Assume that \( G \) satisfies conditions (H1) and (H2). For any \( u \in U \) and \( z \in Z \) there exists an unique \( x(u, z) \) such that \( G(x(u, z), u) = z \), \( x : U \times Z \to X \) is Gateaux differentiable and, for any \( u' \in U \) and \( z' \in Z \),

\[
\nabla^G G(u, z)(u' - u, z' - z) = (\nabla^F G(u, z), u)^{-1}(G(x(u, z), u) - G(x(u, z), u') + z' - z). \tag{3}
\]

Moreover, for any fixed \( u \), \( x(u, \cdot) \) is Fréchet differentiable with respect to \( z \) and, thus,

\[
\nabla^F_x u, z) = (\nabla^F G(u, z), u)^{-1}. \tag{4}
\]

In particular, \( x(u, \cdot) \) is locally Lipschitz in \( z \).

**Proof.** Consider \( H(x, u, z) := G(x, u) - z \). By condition (H1) and (H2), when \( u \) is restricted to a finite dimensional space spanned by a finite number of elements in \( U \), \( \nabla^F_x H(x, u, z) = \nabla^F_x G(x, u) \) is bijective and continuous in a neighbourhood of \( (x, u, z) \). The classical Hildebrandt-Graves implicit function theorem [25] implies that there exist \( x(u, z) \) such that

\[
G(x(u, z), u) = z
\]

and \( x(u, z) \) is Fréchet differentiable (when \( u \) is restricted to this finite dimensional space). In particular, for a fixed \( u \), \( x(u, z) \) is Fréchet differentiable with respect to \( z \).

Now fix \( z', z, u' \) and \( u \) and consider \( t \in [0, 1] \). From \( G(x(u + t(u' - u), z + t(z' - z)), u + t(u' - u)) = z + t(z' - z) \) and \( G(x(u, z), u) = z \) we obtain

\[
\begin{align*}
(G(x(u + t(u' - u), z + t(z' - z)), u) & - G(x(u, z), u)) \\
& = t[G(x(u + t(u' - u), z + t(z' - z)), u) \\
& - G(x(u + t(u' - u), z + t(z' - z)), u') + (z' - z)]. \tag{5}
\end{align*}
\]
Since $x$ is continuous in $z$ and is continuous in $u$ restricted to any polytope in $U$, $x(u+t(u'-u),z+t(z'-z)) \to x(u,z)$ when $t \to 0$. Since $G$ is Fréchet differentiable in $x$ we have

$$\phi(t) := (G(x(u+t(u'-u),z+t(z'-z)),u) - G(x(u,z),u))$$

$$- \nabla_x^F G(x(u,z),u)(x(u+t(u'-u),z+t(z'-z)) - x(u,z))$$

$$= o(x(u+t(u'-u),z+t(z'-z)) - x(u,z)).$$

Invoking (5) we get

$$\left(\nabla_x^F G(x(u,z),u)\right)^{-1}\phi(t)/t$$

$$= \left(\nabla_x^F G(x(u,z),u)\right)^{-1}[G(x(u+t(u'-u),z+t(z'-z)),u)$$

$$- G(x(u+t(u'-u),z+t(z'-z)),u') + (z'-z)]$$

$$- (x(u+t(u'-u),z+t(z'-z)) - x(u,z))/t.$$  (6)

Since $(\nabla_x^F G(x(u,z),u))^{-1}\phi(t)/t = o((x(u+t(u'-u),z+t(z'-z)) - x(u,z))/t)$, for $t$ sufficiently small,

$$\left\|\left(\nabla_x^F G(x(u,z),u)\right)^{-1}[G(x(u+t(u'-u),z+t(z'-z)),u)$$

$$- G(x(u+t(u'-u),z+t(z'-z)),u') + (z'-z)]\right\|$$

$$= \left\|\left(x(u+t(u'-u),z+t(z'-z)) - x(u,z)\right)/t\right\|$$

$$+ \left(\nabla_x^F G(x(u,z),u))^{-1}\phi(t)/t\right\|$$

$$\geq \frac{1}{2}\left\|\left(x(u+t(u'-u),z+t(z'-z)) - x(u,z)\right)/t\right\|.$$  

Observing that

$$\left(\nabla_x^F G(x(u,z),u)\right)^{-1}[G(x(u+t(u'-u),z+t(z'-z)),u)$$

$$- G(x(u+t(u'-u),z+t(z'-z)),u') + (z'-z)]$$

is bounded, $(x(u+t(u'-u),z+t(z'-z)) - x(u,z))/t$ is bounded when $t \to 0$. Therefore, $\phi(t)/t \to 0$ when $t \to 0$. Taking limits in (6) shows that the directional derivative

$$d^+ x(u,z;u'-u,z'-z)$$

$$= (\nabla_x^F G(x(u,z),u))^{-1}(G(x(u,z),u) - G(x(u,z),u') + z'-z).$$

The expression in the right-hand-side does not depend on the direction $(u'-u,z'-z)$. Therefore, $x$ is Gateaux differentiable and we obtain (3). The remaining part of the proposition is obvious.  

21
Proof of Theorem 3.1. For any \( t \in (0, 1) \) and \( u \in U \) we have
\[
0 \leq t^{-1}(F(x(u_z + t(u - u_z), z), u_z + t(u - u_z)) - F(x, u_z)) \\
= t^{-1}(F(x(u_z + t(u - u_z), z), u_z) - F(x, u_z)) \\
+ F(x(u_z + t(u - u_z), z), u) - F(x(u_z + t(u - u_z), z), u_z).
\]
Let \( t \to 0 \). We obtain
\[
\nabla_x^F F(x, u_z) \nabla_x^G x(u_z, z)(u - u_z, 0) + F(x, u) - F(x, u_z) \geq 0.
\]
By (3) of Proposition 5.1, this is
\[
\nabla_x^F F(x, u_z) (\nabla_x^F (G(x, u_z))^{-1}(G(x, u_z)) - F(x, u_z)) \\
\geq \nabla_x^F F(x, u_z) (\nabla_x^F (G(x, u_z))^{-1}(G(x, u)) - F(x, u)) \quad \forall u \in U.
\]
It remains to take \( z^* = \nabla_x^F F(x, u_z) (\nabla_x^F (G(x, u_z))^{-1}. \quad \Box \)

Proof of Theorem 3.2. Proof of (i): Let \( z^* \) be an arbitrary element of \( \partial_{W, H}^k V(z) \). We define
\[
W := W(x, u_z) = \{G(x, u) - G(x, u_z) : u \in U\}.
\]
Then \( W \) is a weakly compact subset of \( Z \) as the image of a continuous function from the compact metric space \( U \) to \( Z \) with the weak topology (note that \( W \) is not necessarily norm compact). Thus, for any \( \varepsilon > 0 \), there exists a \( \delta > 0 \) such that, for any \( w \in W \) and \( t \in (0, \delta) \),
\[
(V(z + tw) - V(z))/t - <z^*, w> > -\varepsilon.
\]
Setting \( w = G(x, u) - G(x, u_z) \) yields
\[
(V(z + tw) - V(z))/t - <z^*, G(x, u) - G(x, u_z)> > -\varepsilon.
\]
Observing that
\[
w = G(x, u) - G(x, u_z) = (G(x, u_z + t(u - u_z)) - G(x, u_z))/t
\]
or
\[
G(x, u_z + t(u - u_z)) = z + tw
\]
we obtain
\[
F(x, u_z + t(u - u_z)) \geq V(z + tw).
\]
Therefore

\[
(F(x_z, u_z + t(u - u_z)) - F(x_z, u_z))/t - \langle z^*, G(x_z, u) - G(x_z, u_z) \rangle
\]

\[
= F(x_z, u - u_z) - \langle z^*, G(x_z, u) - G(x_z, u_z) \rangle > -\epsilon.
\]

As \(\epsilon\) is arbitrary, we get

\[
F(x_z, u) - F(x_z, u_z) - \langle z^*, G(x_z, u) - G(x_z, u_z) \rangle \geq 0.
\]

Proof of ii): Let \(K\) be a weak compact convex symmetric set with a dense span in \(Z\). Let \(K_z = z + K\). Then by definition there exists a WH-differentiable function \(m\) of order \(k\) such that \(V(z) = m(z)\), \(v(z') \geq m(z')\) in a neighborhood \(O\) of \(z\) and \(z^* = \nabla^W H m(z)\). We may assume that \(K_z \subset O\). Then, for any \(\epsilon > 0\), there exists a \(\delta > 0\), such that \(t \in (0, \delta), z' \in K_z\) implies

\[
(V(z + t(z' - z)) - V(z))/t - \langle z^*, z' - z \rangle > -\epsilon.
\]

Let \((x_z, u_z)\) be a solution to \(P(z)\). Then \(V(z) = F(x_z, u_z)\) and \(V(z + t(z' - z)) \leq F(x(u_z, z + t(z' - z)), u_z)\); hence

\[
(F(x(u_z, z + t(z' - z)), u_z) - F(x_z, u_z))/t - \langle z^*, z' - z \rangle > -\epsilon.
\]

On denoting \(\phi(z') := F(x(u_z, z'), u_z)\) we have

\[
(\phi(z + t(z' - z)) - \phi(z))/t - \langle z^*, z' - z \rangle > -\epsilon. \tag{7}
\]

Since \(F\) is (Fréchet) differentiable in \(x\) and \(x\) is (Fréchet) differentiable in \(z\), \(\phi(z)\) is differentiable and, by Proposition 5.1,

\[
\nabla^F \phi(z) = \nabla_x^F F(x(u_z, z), u_z) \cdot \nabla^F x(u_z, z)
\]

\[
= \nabla_x^F F(x(u_z, z), u_z) (\nabla^F G(x(u_z, z), u_z))^{-1}.
\]

Taking limits in (7) when \(t \to 0\) we obtain

\[
\nabla^F \phi(z)(z' - z) - \langle z^*, z' - z \rangle \geq 0 \quad \forall z' \in K_z.
\]

As \(\text{span}(K)\) is dense in \(Z\), this amounts to saying that \(z^* = \nabla^F \phi(z)\). Therefore,

\[
z^* = \nabla_x^F F(x(u_z, z), u_z) \cdot (\nabla^F G(x, u_z))^{-1}
\]

as was to be shown. \(\Box\)
5.2 Proofs of results in Section 4

**Proof of Theorem 4.1.** Theorem 3.1 asserts that there exists a $q = (q^0, q^1) \in R^n \times L^\infty([0,1], R^n)$ such that

i) $q = \omega \cdot (\nabla_x^F G(x_z, u_z))^{-1}$

and

ii) $\langle q, G(x_z, u_z) \rangle - F(x_z, u_z) = \max_{u \in U} \{\langle q, G(x_z, u) \rangle - F(x_z, u)\}$

where $\omega = \nabla_x^F F(x_z, u_z)$. We will show that $p = -q^1$ has the required property. As with the calculation of $\nabla_x^F G(x_z, u_z)$ we can compute that, for any $y \in L^1([0,1], R^n)$,

$$\nabla_x^F F(x_z, u_z) y = \int_0^1 f_x(s, x_z(s), u_z(s)) y(s) ds.$$

Therefore, for any $y \in L^1([0,1], R^n)$,

$$\omega \cdot y = \int_0^1 f_x(s, x_z(s), u_z(s)) y(s) ds.$$

For any $r = (r^0, r^1) \in R^n \times L^1([0,1], R^n)$, let $y$ be the solution of

$$\dot{y}(s) = g_x(s, x_z(s), u_z(s)) y(s) + r^1(s), \quad y(0) = r^0.$$

Then $y = (\nabla_x^F G(x_z, u_z))^{-1} r$ and

$$\langle q, r \rangle = \int_0^1 \langle q^1(s), r^1(s) \rangle ds + \langle q^0, r^0 \rangle = \omega \cdot y = \int_0^1 f_x(s, x_z(s), u_z(s)) y(s) ds.$$

Thus

$$\int_0^1 \langle q^1(s), \dot{y}(s) - g_x(s, x_z(s), u_z(s)) y(s) \rangle ds + \langle q^0, r^0 \rangle$$

$$= \int_0^1 f_x(s, x_z(s), u_z(s)) y(s) ds. \quad (8)$$

In particular, if we let $r^0 = 0$ then

$$\int_0^1 \langle q^1(s), \dot{y}(s) - g_x(s, x_z(s), u_z(s)) y(s) \rangle ds$$

$$= \int_0^1 f_x(s, x_z(s), u_z(s)) y(s) ds. \quad (9)$$
Let $p$ be the solution of

$$-\dot{p}(s) = (g_x(s, x_z(s), u_z(s)), f_x(s, x_z(s), u_z(s))),$$

$$p(1) = 0.$$ 

Now we can check by direct calculation that $-p$ satisfies (9). As $r^1(s) = \dot{y}(s) - g_x(s, x(s), u(s))y(s)$ can be taken arbitrarily in $L^1$, $q = -p$. Obviously $p$ is absolutely continuous. Replacing $q$ with $-p$ in (ii) yields

$$\int_0^1 [(p(s), g(s, x_z(s), u_z(s))) + f(s, x_z(s), u_z(s))]ds$$

$$= \max_{v \in \mathcal{U}} \int_0^1 [(p(s), g(s, x_z(s), v(s))) + f(s, x_z(s), v(s))]ds$$

which is the integrated form of the maximum relation. A standard argument for problems without state constraint (cf. [22]) then yields the pointwise form of the maximum relation (ii) in the theorem.

**Remark 5.1.** Actually, using (8) we can show that $-p(0) = q^0$ which is to say $q^1(0) = q^0$. This fact will be useful in the proof of Theorem 4.4.

**Proof of Theorem 4.4.** Select $q = (q^0, q^1) \in \partial G V(0)$. By Theorem 4.3, $V$ is locally Lipschitz. Since $V$ is defined on the WCG Banach space $R^n \times L^1([0, 1], R^n)$ which has a smooth weak Hadamard renorming [3], Theorem 2.1 asserts that

$$\partial G V(0) = \bigcup_{k=1}^{\infty} \{w^* - \lim_{n \to \infty} z_n^* : z_n^* \in \partial W_H V(z_n), z_n \to 0\}.$$ 

Therefore, there exist sequences $q_n = (q_n^0, q_n^1) \in R^n \times L^\infty([0, 1], R^n)$ and $z_n = (z_n^0, z_n^1) \in R^n \times L^1([0, 1], R^n)$ such that $q_n \in \partial W_H V(z_n)$, $q_n \rightharpoonup q$ in $R^n \times L^\infty([0, 1], R^n)$ and $z_n \to 0$ in $R^n \times L^1([0, 1], R^n)$. By Theorem 4.1, for each $z_n$, $\Sigma z_n \neq \emptyset$. Using Theorem 3.2 and the concrete form of the Maximum Principle stated in Theorem 4.2, $p_n^1 = -q_n^1$ is the solution of

$$-\dot{p}_n(s) = (g_x(s, x_n(s), u_n(s)), f_x(s, x_n(s), u_n(s))), \quad p_n^1(1) = 0$$

(10)

where $(x_n, u_n)$ is a solution pair for problem $P(z_n)$ and satisfies

$$\int_0^1 [(p_n^1(s), g(s, x_n(s), u_n(s))) + f(s, x_n(s), u_n(s))]ds$$

$$= \max_{v \in \mathcal{U}} \int_0^1 [(p_n^1(s), g(s, x_n(s), v(s))) + f(s, x_n(s), v(s))]ds.$$ 

(11)
Also by the Remark 5.1 $p_n^1(0) = -q_n^0$. Gronwall’s inequality and $z_n \to 0$ implies that $p_n^1$ and $x_n$ are bounded sequences in $W^{0,1}$ and, hence, norm precompact in $C([0,1], R^n)$. Thus, without loss of generality, we may assume that $x_n \to x, p_n^1 \to p$ in $C([0,1], R^n)$ and $u_n \to u$ in $U$. Since $-p_n^1 = q_n^1$ weak-star converges to $q^1$, $q^1 = -p$. Also $q^0 = -p(0) = q^1(0)$ follows by taking limits in $p_n^1(0) = -q_n^0$. Similarly
\[
x_n(t) = \int_0^t g(s, x_n(s), u_n(s)) ds + \int_0^t z_n(s) ds
\]
yields
\[
x(t) = \int_0^t g(s, x(s), u(s)) ds.
\]
Also
\[
V(0) = \lim_{n \to \infty} V(z_n) = \lim_{n \to \infty} \int_0^1 f(s, x_n(s), u_n(s)) ds = \int_0^1 f(s, x(s), u(s)) ds.
\]
Therefore, $(x, u)$ is a solution to problem $P(0)$. Again taking limits in the integrated form of (10) and (11) yields
\[
-p(s) = (g_x(s, x(s), u(s)))^\top p(s) - f_x(s, x(s), u(s)), \quad p(1) = 0
\]
and
\[
\int_0^1 [\langle p(s), g(s, x(s), u(s)) \rangle + f(s, x(s), u(s))] ds
\]
\[
= \max_{v \in U} \int_0^1 [\langle p(s), g(s, x(s), v(s)) \rangle + f(s, x(s), v(s))] ds.
\]
As before this yields a pointwise maximum relation and so $-q^1 \in M(\Sigma_0)$ as was to be shown. \(\Box\)

**Proof of Lemma 4.1.** For each integer $m$ consider the restriction of $(x_n, u_n)$ to $[m - 1, m]$. It is easy to see that $x_n$ is equi-continuous and uniformly bounded on $[m - 1, m]$ and, therefore, precompact in $C([m - 1, m]; R^n)$. Also, $u_n|_{[m - 1, m]}$ lies in the compact set $U|_{[m - 1, m]}$. Thus, for any $[m - 1, m]$ we can extract a convergent subsequence from $(x_n, u_n)|_{[m - 1, m]}$. Using the diagonal method we can choose a subsequence $(x_i, u_i)$ of $(x_n, u_n)$ such that $x_i$ converges to $x$ pointwise and, for any $m$, $x_i|_{[0,m]}$ uniformly converges to $x|_{[0,m]}$ and $u_i|_{[0,m]}$ converges to $u|_{[0,m]}$ in $U|_{[0,m]}$. For any $t$, taking limits in
\[
x_i(t) = \int_0^t [g(s, x_i(s), u_i(s)) + z_i^1(s)] ds + z_i^0
\]

26
leads to

$$x(t) = \int_0^t [g(s, x(s), u(s)) + z^1(s)]ds + z^0.$$ 

Thus, \((x, u)\) is feasible for \(P(z)\). It remains to show that

$$J(x, u) \leq \liminf_{n \to \infty} J(x_n, u_n).$$

Relabeling if needed we may arrange that

$$\liminf_{n \to \infty} J(x_n, u_n) = \lim_{i \to \infty} J(x_i, u_i).$$

Observe that by Gronwall’s inequality \(|x_n(s)e^{-Ns}|\) is uniformly bounded. Combined with the growth condition on \(f\), this shows that there exists a constant \(K > 0\) such that

$$f(s, x_i(s), u_i(s)) + Ke^{Ns} \geq 0.$$ 

Thus, for any \(t\),

$$\int_0^t e^{-Ls} f(s, x_i(s), u_i(s))ds$$

$$= \int_0^t e^{-Ls}[f(s, x_i(s), u_i(s)) + Ke^{Ns}]ds - \int_0^t Ke^{-(L-N)s} ds$$

$$\leq \int_0^\infty e^{-Ls}[f(s, x_i(s), u_i(s)) + Ke^{Ns}]ds - \int_0^t Ke^{-(L-N)s} ds$$

$$= \int_0^\infty e^{-Ls} f(s, x_i(s), u_i(s))ds + \int_t^\infty Ke^{-(L-N)s} ds.$$ 

Taking limits when \(i \to \infty\) we obtain

$$\int_0^t e^{-Ls} f(s, x(s), u(s))ds \leq \lim_{i \to \infty} J(x_i, u_i) + \int_t^\infty Ke^{-(L-N)s} ds.$$ 

Letting \(t \to \infty\), we are done. ☺️

**Proof of Theorem 4.8.** Let \(p \in \partial_G V(0)\). By Theorem 4.5, \(V\) is locally Lipschitz. Since \(V\) is defined on the WCG Banach space \(Z := R^n \times L^1\) which has a weak Hadamard smooth renorming \([3]\), Theorem 2.1 asserts that

$$\partial_G V(0) = \bigcup_{k=1}^\infty \{w^* - \lim_{n \to \infty} z^*_n : z^*_n \in \partial_{W^*} V(z_n), z_n \to 0\}.$$ 

27
Therefore, there exist sequences \( p_n \in (L^1)^* \) and \( z_n \in L^1 \) such that \( p_n \in \partial_{W,H}^k V(z_n) \), \( p_n \rightharpoonup p \) in \((L^1)^* \) and \( z_n \to 0 \) in \( L^1 \). Define \( W_\eta := \{ z = (z^0, z^1) \in R^n \times L^1 : z^0 \in z^0_n + \eta B, z^1(s) \in z^1_n(s) + \eta B \} \). Then, for each \( \eta > 0 \), \( W_\eta \) is a weak compact set that generates \( Z \). Since \( p_n \in \partial_{W,H}^k V(z_n) \), for any integer \( i \), there exists \( \eta(i) \) such that

\[
V(z) - V(z_n) + \eta^{-1} \| z - z_n \| \geq \langle p_n, z - z_n \rangle \quad \forall z \in W_{\eta(i)}.
\]

Let \((x_n, u_n) \in \Sigma_{z_n} \). Then

\[
V(z_n) = \int_0^\infty e^{-Ls} f_x(s, x_n(s), u_n(s)) \, ds
\]

and

\[
V(z) \leq \int_0^\infty e^{-Ls} f_x(s, x(s), u_n(s)) \, ds
\]

for any solution \( x \) of

\[
\dot{x}(s) = g(s, x(s), u_n(s)) + z^1(s)
\]

\[
x(0) = z^0, \quad u \in U.
\]

Therefore, \((x_n, z_n)\) is a solution to the infinite horizon optimal control problem

\[
\text{minimize} \quad \int_0^\infty \left[ e^{-Ls} f(s, x(s), u_n(s)) \, ds + \eta^{-1} \| z^1(s) - z^1_n(s) \| \right. \\
\left. - \langle p^1_n(s), z^1(s) \rangle \right] \, ds - \langle p^0_n, z^0 \rangle + \eta^{-1} \| z^0 - z^0_n \| \\
\text{subject to} \quad \dot{x}(s) = g(s, x(s), u_n(s)) + z^1(s) \\
x(0) = z^0, \quad u \in U, \quad (z^0, z^1) \in W_{\eta(i)}.
\]

or equivalently

\[
\text{minimize} \quad \int_0^\infty \left[ e^{-Ls} f(s, x(s), u_n(s)) \, ds + \eta^{-1} \| z^1(s) - z^1_n(s) \| \right. \\
\left. - \langle p^1_n(s), z^1(s) \rangle \right] \, ds - \langle p^0_n, x(0) \rangle + \eta^{-1} \| x(0) - z^0_n \| \\
\text{subject to} \quad \dot{x}(s) = g(s, x(s), u_n(s)) + z^1(s) \\
u \in U, \quad (x(0), z^1) \in W_{\eta(i)}.
\]

Observe that this is a free initial point problem. By [23, Proposition 3.2] there exists an absolutely continuous function \( q^i_n \) such that

\[
-q^i_n(s) = g_x(s, x_n(s), u_n(s))^\top q^i_n(s) - e^{-Ls} f_x(s, x_n(s), u_n(s)), \quad q^i_n(0) \in -p^0_n + \eta^{-1} B 
\]

(12)
and

$$\langle q_n^i(s), g(s, x_n(s), u_n(s)) + z^1_n(s) \rangle + \langle p_n^i(s), z^1_n(s) \rangle = \max_{z^1 \in {\mathbb R}^n, \eta > 0} \left\{ \langle g_n^i(s), g(s, x_n(s), u_n(s)) + z^1 \rangle + \langle p_n^i(s), z^1 \rangle - \frac{1}{i} |z^1 - z_n(s)| \right\}.$$  

The last relation implies that $|q_n^i(s) + p_n^i(s)| \leq (1/i)$. So if we denote $q_n := -p_n^0$ then $q_n^i$ converges to $q_n$ uniformly. Taking limits in (12) yields

$$-q_n(s) = g_x(s, x_n(s), u_n(s))^T q_n(s) - e^{-L^s} f_x(s, x_n(s), u_n(s)),$$

$$q_n(0) = -p_n^0.$$

We next prove that $q_n = -p_n^0$ satisfies

$$\langle q_n(s), g_x(s, x_n(s), u_n(s)) \rangle + e^{-L^s} f_x(s, x_n(s), u_n(s)) = \max_{u \in {\mathcal U}} \left\{ \langle q_n(s), g_x(s, x_n(s), u) \rangle + e^{-L^s} f_x(s, x_n(s), u) \right\}.$$  

To simplify notations we denote, for any absolutely continuous function $x$ such that $|x| e^{-L^s} \in {L}^1([0, \infty), {R}^n)$ and $u \in {\mathcal U}$,

$$G(x, u) := (x(0), \dot{x}(-) - g(\cdot, x(\cdot), u(\cdot)))$$  

and

$$F(x, u) := \int_0^\infty e^{-L^s} f(s, x(s), u(s))ds$$

We define

$$W := \{ G(x_n, u) - G(x_n, u_n) : u|_{(0, T]} \in {\mathcal U}|_{(0, T]} \text{ and } u|_{(T, \infty)} = u_n|_{(T, \infty)} \}.$$  

Then $W$ is a weakly compact subset of $Z = {R}^n \times {L}^1([0, \infty), {R}^n)$. Since $p_n \in \partial_W V(z_n)$, for any $\varepsilon > 0$, there exists a $\delta > 0$ such that, for any $w \in W$ and $t \in (0, \delta)$

$$(V(z_n + tw) - V(z_n))/t - \langle p_n, w \rangle > -\varepsilon.$$

Setting $w = G(x_n, u) - G(x_n, u_n)$ yields

$$(V(z_n + tw) - V(z_n))/t - \langle p_n, G(x_n, u) - G(x_n, u_n) \rangle > -\varepsilon.$$  

Relations

$$w = G(x_n, u) - G(x_n, u_n) = (G(x_n, u_n + t(u - u_n)) - G(x_n, u_n))/t$$
and \( G(x_n, u_n) = z_n \) yields

\[
G(x_n, u_n + t(u - u_n)) = z_n + tw.
\]

Thus we obtain

\[
F(x_n, u_n + t(u - u_n)) \geq V(z_n + tw).
\]

Therefore

\[
(F(x_n, u_n + t(u - u_n)) - F(x_n, u_n))/t - \left\langle p_n, G(x_n, u) - G(x_n, u_n) \right\rangle
= F(x_n, u) - F(x_n, u_n) - \left\langle p_n, G(x_n, u) - G(x_n, u_n) \right\rangle > -\epsilon.
\]

As \( \epsilon \) is arbitrary, we get

\[
F(x_n, u) - F(x_n, u_n) - \left\langle p_n, G(x_n, u) - G(x_n, u_n) \right\rangle \geq 0 \quad \forall u \in W.
\]

Observing that, for any \( u \in W \), \( u(s) = u_n(s) \) when \( s \geq T \) and noticing that \( q_n^a = -p_n^1 \) we obtain

\[
\int_0^T \left\{ (q_n(s), g_x(s, x_n(s), u_n(s))) + e^{-L_s} f_x(s, x_n(s), u_n(s)) \right\} ds
= \max_{w \in \Pi_{[0,T]}} \left\{ \int_0^T \left\{ (q_n(s), g_x(s, x_n(s), u(s))) + e^{-L_s} f_x(s, x_n(s), u(s)) \right\} ds \right\}.
\]

Using standard arguments (cf. [22]) we obtain the maximum relation in pointwise form for \( s \in [0,T] \). Since \( T \) is arbitrary we have proved (14).

Applying Lemma 4.1, without loss of generality, we may assume that \( x_n \) convergence to an \( x \) uniformly in any finite interval and \( u_n \) converges to \( u \) in any \( \mathcal{U}_{[a,b]} \). Furthermore, \( (x, u) \) is feasible for \( P(0) \) and

\[
J(x, u) \leq \liminf_{n \to \infty} J(x_n, u_n) = \liminf_{n \to \infty} V(z_n) = V(0).
\]

Therefore, \( (x, u) \in \Sigma_0 \). We claim that \(-p_1 \in M(x, u) \) and \( p_1(0) = p^0 \). Indeed, equation (12) implies that in any finite interval \([0,T] \), \( q_n = -p_n^1 \) has a subsequence convergent in \( C[0,T] \). As \( p_n \) weak-star converges to \( p \) the limit of this subsequences must be \(-p_1 |_{[0,T]} \). For each finite interval taking limit in the integrated form of (13) and (14) leads to the result. 😊
6 Appendix: Proximity and Lipschitzness

In the first part of this appendix we will deduce a proximal limit description of the G-subdifferential. We also establish a corresponding Hölder subdifferential generalization. Both will follow from the smooth variational principle in tandem with the Borwein-Ioffe formula stated in Theorem 2.1 and its normal cone variant. In the second part of the appendix, we show that a lower semicontinuous function on a \( \beta \)-smooth space is Lipschitz precisely when its \( \beta \)-subderivatives are bounded.

6.1 Sequential limit formulas for generalized derivatives

**Definition 6.1** [5] Let \( X \) be a Banach space, let \( f : X \to [-\infty, \infty] \) be lower semicontinuous, and suppose \( f(x) \) is finite. Then \( f \) is \( s \)-Hölder subdifferentiable at \( x \) \((s \in (0, 1])\) with subderivative \( x^* \in X^* \) if there exists a positive constant \( C_x \) such that

\[
 f(y) - f(x) - \langle x^*, y - x \rangle \geq -C_x \| y - x \|^{1+s}
\]

in a neighbourhood of \( x \).

We denote the set of \( s \)-Hölder subderivatives of \( f \) at \( x \) by \( \partial^{H(s)} f(x) \). When \( s = 1 \) such subderivatives are called Lipschitz smooth and in Hilbert space they coincide with Rockafellar’s proximal subderivatives, written \( \partial^\circ f(x) \) [20].

**Definition 6.2** [15] Let \( X \) be a Banach space with dimension at least 2. We say that \( X \) has a power modulus of smoothness \( t^p \) if

\[
 \sup\{ \frac{\|x + y\| + \|x - y\|}{2} - 1 : x, y \in X, \|x\| = 1, \|y\| = t \} \leq K t^p
\]

for some constant \( K \).

**Lemma 6.1** Let \( X \) be a Banach space with a power modulus of smoothness \( t^{1+s} \), \( s > 0 \) and \( f \) a lower semicontinuous proper function on \( X \). Let \( x^* \) be a Fréchet subderivative of \( f \) at \( x \). Then, for any \( \varepsilon > 0 \), there exist \( y \) and \( y^* \) such that \( \|y - x\| < \varepsilon \), \( \|y^* - x^*\| < \varepsilon \), \( |f(y) - f(x)| < \varepsilon \) and \( y^* \in \partial^{H(s)} f(y) \).
Proof. Without loss of generality we may assume that $x = 0$ and $x^* = 0$. Then, for $\varepsilon > 0$, there exists a $\eta \in (0, \varepsilon)$ such that, for all $||x|| \leq \eta$,

$$f(x) \geq -\eta \varepsilon / 2.$$  

Define

$$g(x) := f(x) + \delta_A(x)$$

where $\delta_A$ is the characteristic function of $A$. Then $g(x)$ is lower semicontinuous and

$$\inf_{x \in X} g(x) \geq g(0) - \eta \varepsilon / 2.$$  

Applying the Borwein-Preiss smooth variational principle [5, Theorem 2.6, (d),(e)] to $g$ with $p = 2$ and $\lambda = \eta$ there exist $y$ such that, $||y|| < \eta < \varepsilon$,

$$g(y) < \inf_{X} g + \eta \varepsilon / 2 < \inf_{X} g + \varepsilon$$

and

$$0 \in \partial^{H(s)} g(y) + \varepsilon B^*$$

i.e., there exists a $y^* \in \partial^{H(s)} g(y)$ such that $||y^*|| \leq \varepsilon$. Since $||y|| < \eta$, $\partial^{H(s)} g(y) = \partial^{H(s)} f(y)$. Also observe that $g(y) = f(y)$ and $g(x) = f(x)$ we obtain $|f(y) - f(x)| \leq \varepsilon$. ☺

Remark 6.1. $L^p$, $p > 1$ has a power modulus of smoothness $t^{\min(p,2)}$.

Combining Lemma 6.1 and Corollary 6.1 with [2, Theorem 1.2 and 3] we have

**Theorem 6.1** Let $X$ be a Banach space with a power modulus of smoothness $t^{1+s}$, $s > 0$. Let $f$ be a lower semicontinuous proper function on $X$. Then for any $x \in X$

$$\partial_G f(x) = \text{cl}^* \{ w^* - \lim_{n \to \infty} x_n^ : x_n^* \in \partial^{H(s)} f(x_n), x_n \to x \},$$

and

$$\partial_C f(x) = \text{cl}^* \text{co} \{ w^* - \lim_{n \to \infty} x_n^ : x_n^* \in \partial^{H(s)} f(x_n), x_n \to x \} + \partial^{\infty}_H f(x).$$

**Theorem 6.2** Let $X$ be a Banach space with a power modulus of smoothness $t^{1+s}$, $s > 0$. Let $S$ be a closed subset of $X$. Then for any $x \in S$
\[ N_G(S, x) = \operatorname{cl}^* \bigcup_{k=1}^{\infty} \{ w_n - \lim_{n \to \infty} x_n^+ : x_n^+ \in k \partial^H \rho(S, x_n), x_n \to S x \}, \]

and

\[ N_C(S, x) = \operatorname{cl}^* \operatorname{co} \bigcup_{k=1}^{\infty} \{ w_n - \lim_{n \to \infty} x_n^+ : x_n^+ \in k \partial^H \rho(S, x_n), x_n \to S x \}. \]

In particular, if \( X \) is a Hilbert space then it has a power modulus of smoothness of \( t^2 \) and we obtain the following corollaries:

**Corollary 6.1** Let \( X \) be a Hilbert space. Let \( f \) be a lower semicontinuous proper function on \( X \). Then for any \( x \in X \)

\[ \partial_G f(x) = \operatorname{cl}^* \{ w_n - \lim_{n \to \infty} x_n^+ : x_n^+ \in \partial^S f(x_n), x_n \to f x \}, \]

and

\[ \partial_C f(x) = \operatorname{cl}^* \operatorname{co} \{ w_n - \lim_{n \to \infty} x_n^+ : x_n^+ \in \partial^S f(x_n), x_n \to f x \} + \partial_C^\infty f(x). \]

**Corollary 6.2** Let \( X \) be a Hilbert space. Let \( S \) be a closed subset of \( X \). Then for any \( x \in S \)

\[ N_G(S, x) = \operatorname{cl}^* \bigcup_{k=1}^{\infty} \{ w_n - \lim_{n \to \infty} x_n^+ : x_n^+ \in k \partial^S \rho(S, x_n), x_n \to S x \}, \]

and

\[ N_C(S, x) = \operatorname{cl}^* \operatorname{co} \bigcup_{k=1}^{\infty} \{ w_n - \lim_{n \to \infty} x_n^+ : x_n^+ \in k \partial^S \rho(S, x_n), x_n \to S x \}. \]

### 6.2 A criterion for a function to be Lipschitz

**Theorem 6.3** Let \( X \) be a Banach space with a \( \beta \)-smooth renorming, \( U \) a convex open subset of \( X \) and \( f \) a lower semicontinuous proper function on \( W \). Let \( L \geq 0 \). Then \( f \) is Lipschitz of rank \( L \) if and only if

\[ \sup \{ \| x^* \| : x^* \in \partial_\beta f(x) \} \leq L, \forall x \in U \]

(where \( \sup \emptyset = -\infty \)). In particular, when \( L = 0 \) we obtain that \( f \) is a constant if and only if \( \partial_\beta f(x) = \{ 0 \} \) for all \( x \in W \).
Proof. The “only if” part is obvious. We prove the “if” part. If $f$ is nowhere finite on $U$ then there is nothing to prove. Let $x_0 \in U$ and $f(x) < \infty$. Since $f$ is lower semicontinuous we can choose a $\eta > 0$ such that $x_0 + 3\eta B \subset U$ and $f$ is bounded below on $x_0 + 3\eta B$. Let $M > L$, $y \in x_0 + \eta B$ and consider function

$$g(x) := \begin{cases} f(x) + M\|x - y\| & \text{if } \|x - y\| \leq \eta \\ f(x) + M\|x - y\| + \frac{(||x-\eta||^2-\eta^2)^2}{2\eta^2\|y-x\|} & \text{if } \eta \leq \|x - y\| \leq 2\eta \\ \infty & \text{otherwise.} \end{cases}$$

Then $g$ is lower semicontinuous and bounded below on $y + 2\eta B$. For any $\varepsilon > 0$, by the Borwein-Preiss smooth variational principle [5, Theorem 2.6], there exists a $v$ such that $g(v) \leq \inf_X g + \varepsilon$ and

$$g(x) + \varepsilon \Delta(x) \geq g(v) + \varepsilon \Delta(v)$$

and

$$0 \in \partial_\beta g(v) + 2\varepsilon B$$

where $\Delta(x) = \sum_{n=1}^{\infty} \mu_n\|x - v_n\|^2$, $\sum_{n=1}^{\infty} \mu_n = 1$, $\mu_n \geq 0$ and $\{v_n\}$ converges in norm to some element in $X$. Obviously $\|v - y\| < 2\eta$. If $v \neq y$ then $\|x - y\|$ and $\frac{(||x-\eta||^2-\eta^2)^2}{2\eta^2\|y-x\|}$ are $\beta$-differentiable at $v$ and, therefore, $f$ is $\beta$-subdifferentiable at $x$. When $\|v - y\| \leq \eta$,

$$-M\nabla_\beta\|v - y\| + d_\varepsilon \in \partial_\beta f(v);$$

otherwise

$$-(M + 2\eta(\|v - y\| - \eta)/(2\eta - \|v - y\|^2))\nabla_\beta\|v - y\| + d_\varepsilon \in \partial_\beta f(v)$$

where $d_\varepsilon \in \varepsilon B$. Observe that if $v \neq y$ the norm of $\nabla_\beta\|v - y\|$ is 1. Then in both inclusions the norms of the left-hand-sides will exceed $L$ when $\varepsilon$ is sufficiently small which is absurd. Therefore, $v$ must equal to $y$. Then, for any $x \in y + \eta B$,

$$g(y) + \varepsilon \Delta(y) \leq g(x) + \varepsilon \Delta(x).$$

Letting $\varepsilon \to 0$ yields $g(y) \leq g(x)$ or

$$f(y) \leq f(x) + M\|x - y\|, \forall x \in y + \eta B.$$
We now extend this local Lipschitz condition to $U$. Let $x$ and $y$ be points in $W$ and we prove $|f(x) - f(y)| \leq L|x - y|$. Assume, say, $f(x) < \infty$ (otherwise there is nothing to prove). Define $\alpha := \max\{t \in [0, 1] : \|f(x + t(y - x)) - f(x)\| \leq Lt\|y - x\|\}$. Since $f(x) < \infty$, $f$ is locally Lipschitz around $x$ and, therefore, $\alpha \in (0, 1]$. If $\alpha < 1$ then since $f(x + \alpha(y - x)) < \infty$ there is a neighbourhood of $x + \alpha(y - x)$ on which $f$ is Lipschitz of rank $L$ by the local result. This contradicts the maximality of $\alpha$. Thus, $\alpha = 1$ and the proof is completed. \(\blacksquare\)

**Remark 6.2** For previous research in this direction we refer to [9, 10, 17, 21, 19]. In analogy to the classical smooth results one might guess that two functions with equal subdifferentials will differ only by a constant. However, this is not the case as shown by the following examples.

**Example 6.1** Let $f(x) := -\chi_{[0,1]}(x)$ and $g(x) := 2f(x)$. Then

$$
\partial_F f(x) = \partial_F g(x) = \begin{cases} (-\infty, 0] & x = 0 \\
[0, \infty) & x = 1 \\
0 & \text{elsewhere}
\end{cases}
$$

where $\partial_F$ is the Fréchet subdifferential, but $f(x) - g(x) = \chi_{[0,1]}(x)$ is not a constant.

In this example the problem occurs at the point where the functions are discontinuous. So it is natural to ask whether the statement is true if in addition we demand that the functions $f$ and $g$ are continuous. The following example, which was originally designed to show that the Newton-Leibniz formula (Fundamental Theorem of Calculus) fails without an absolute continuity assumption, shows that the answer is still negative.

**Example 6.2** Let $C$ be the Cantor ternary set on $[0,1]$ consisting of every ternary decimal involving only 0 and 2 in its expression (cf. [12, pp. 95-98]). As $C$ is closed, $[0,1]\setminus C$ is the union of numerable disjoint open intervals. We write

$$
[0,1]\setminus C := \bigcup_{k=1}^\infty (a_k, b_k).
$$

Consider the classical Cantor ternary function $h : C \to [0,1]$ defined as follows:

$$
h(x) := \sum_{n=0}^\infty \frac{x_n}{2^{n+1}}
$$
where \( x_n \) is the \( n \)th digit of the ternary decimal expression \( x = 0.x_1x_2... \) of \( x \). As, for each \( k \), \( a_k \) and \( b_k \) must have “dual” ternary expressions

\[
a_k = 0.c_1c_2...c_n0222... \quad \text{and} \quad b_k = 0.c_1c_2...c_n2000...,\]

we can check that \( h(a_k) = h(b_k) \). Thus, we can extend \( h \) to \([0, 1]\) by defining

\[
h(x) := h(a_k) = h(b_k), \quad \forall x \in (a_k, b_k).
\]

We further extend \( h \) to \( R \) by setting \( h(x) := 0, x < 0 \) and \( h(x) := 1, x > 1 \). It is well known that \( h \) is continuous.

**Claim:**

\[
\partial_F h(x) = \begin{cases} 
0 & x \in C \setminus (\{b_k\} \cup \{0\}) \\
[0, \infty) & x \in (\{b_k\} \cup \{0\}) \\
0 & x \in R \setminus C.
\end{cases}
\]

Now \( f = 2h \) and \( g = h \) have the identical subdifferentials everywhere yet their difference \( f - g = h \) is not a constant.

**Proof of the Claim.** For any \( x \in C \setminus (\{b_k\} \cup \{0\}) \), its ternary expression \( x = 0.x_1x_2... \) contains infinitely many \( 2 \)'s. If \( \xi \in \partial_F h(x) \neq \emptyset \) then

\[
h(y) - h(x) + o(|y - x|) \geq \xi(y - x)
\]

for \( y \) sufficiently close to \( x \). Consider \( y^m = 0.y_1y_2... \) such that \( y_i = x_i \) if \( i \neq m \) and \( y_m = 0 \). Then \( y^m \leq x \) and \( y^m \) converges to \( x \). Substituting \( y^m \) into the aforementioned inequality leads to

\[
-\frac{x_m}{2^{m+1}} + o(\frac{x_m}{3^m}) \geq -\xi \frac{x_m}{3^m}
\]

This is absurd since there are infinitely many \( x_m = 2 \). Therefore \( \partial_F h(x) = \emptyset \).

If \( x \in (\{b_k\} \cup \{0\}) \), say \( x = b_k \), then \( x \) has a finite ternary decimal expression \( x = b_k = 0.x_1x_2...x_m \). For any integer \( p > m \) and \( y = 0.y_1y_2... \in (x, x+2^{-p}) \), we must have \( y_i = x_i, i = 1,2,..., m \) and \( y_{m+1} = ... = y_p = 0 \). Thus,

\[
h(y) - h(x) \geq \sum_{n=p+1}^{\infty} \frac{y_n}{2^n+1}.
\]

Therefore,

\[
\frac{h(y) - h(x)}{y - x} \geq \frac{3p}{2^{p+1}}.
\]
Set \( \eta := \min \{b_k - a_k, 2^{-n}\} \). Then, for any \( y \in (b_k - \eta, b_k + \eta) \) and \( \xi \in [0, \frac{3p}{2^{n+1}}] \), observing that \( h \) is a constant on \((a_k, b_k)\), we have

\[
h(y) - h(x) \geq \xi(y - x)
\]

and hence \( \partial_F h(x) \supset [0, \frac{3p}{2^{n+1}}] \). As \( p \) can be taken arbitrarily large, we obtain \( \partial_F h(x) = [0, \infty) \).

It is obvious that \( \partial_F h(x) = 0 \) when \( x \in R \setminus C \). \( \square \)

**Remark 6.3** Replacing \( \partial_F \) with \( \partial^\sigma \) and \( o(\cdot, \cdot) \) with \( \sigma \cdot \cdot \mid \mid^2 \) for some proper constant \( \sigma \), both Example 6.1 and Example 6.2 as well as their arguments remain valid. In particular, the proximal version of Example 6.2 gives a negative answer to the following question: Given two continuous functions \( f \) and \( g \), satisfying \( f(0) = g(0) = 0 \) and \( \partial^\sigma f(x) = \partial^\sigma g(x) \) for all \( x \in \mathbb{R}^n \), must \( f = g \)?

Both examples essentially use the fact that functions involved have subdifferentials with either 0 or half line as their values. Consider functions with everywhere bounded subdifferentials which by Theorem 6.3 are Lipschitz. When \( X \) is finite dimensional (in which case all \( \beta \)-subdifferential coincide), observing that subderivatives equals derivatives when the later exists, by the Radamacher theorem we deduce that if two function have the same subdifferential everywhere then their derivatives exist and are equal almost everywhere. Thus, the Newton-Leibniz formula (Fundamental Theorem of Calculus) and the Fubini theorem lead to the following positive result:

**Theorem 6.4** Let \( f \) and \( g \) be Lipschitz functions on a finite dimensional Banach space \( X \). Then \( f - g \) is a constant if and only if the (Fréchet) subdifferentials of \( f \) and \( g \) coincide everywhere.

For infinite dimensional space, the problem becomes much more complicated and remains open in general. A fairly general result in this respect is [1, Proposition 4.4] which asserts that if \( f \) and \( g \) are locally Lipschitz functions with identical subdifferentials then \( f - g \) is a constant if and only if the subdifferential of \( f - g \) is a minimal \( w^* \) compact-valued convex upper-semicontinuous multifunction \( (w^*-cuso) \). For detailed discussion of various conditions that guarantee’s the subdifferential of \( f - g \) to be a \( w^*-cuso \) we refer the reader to [1, Section 3].
References


38
[13] A. Ioffe, Necessary and sufficient conditions for a local minimum, 1: a
reduction theorem and first order conditions, *SIAM Control and Opti-
mization* **17** (1979), 245-265.

[14] A. Ioffe, Approximate subdifferentials and applications 3: the metric


[16] P. D. Loewen, Perturbed differential inclusion problems, in *Nonsmooth
Optimization and Related Topics*, eds. F. H. Clarke, V. F. Dem’yanov

Notes Series, Amer. Math. Soc., Summer School on Control, CRM,

[18] R.R. Phelps, *Convex functions, Monotone operators and differentiabil-
ity*, Lecture Notes in Mathematics, No. 1364, Springer Verlag, N.Y.,

[19] R. A. Poliquin, Integration of subdifferentials of nonconvex functions,

[20] R. T. Rockafellar, Proximal subgradients, marginal values and aug-


[24] J. Ye, A nonsmooth Maximum Principle for infinite horizon problems,
preprint.