PERTURBED DIFFERENTIAL INCLUSION PROBLEM WITH NONADDITIVE L^1 PERTURBATIONS AND APPLICATIONS

J.J. Ye and Q.J. Zhu

DMS-668-IR **June 1994**

PERTURBED DIFFERENTIAL INCLUSION PROBLEM WITH NONADDITIVE L^1 PERTURBATIONS AND APPLICATIONS *

J. J. Ye and Q. J. Zhu†

June 21, 1994

Abstract

In this paper we derive an estimate for the G-subderivative of the value function associated with a perturbed optimization problem with differential inclusion constraints. We then apply this result to obtain a necessary condition for a solution to a bilevel dynamic problem.

Key words: sensitivity analysis, nonsmooth analysis, value function, weak Hadamard differentiability, bilevel dynamic problems, necessary conditions.
AMS(MOS) subject classification: 49K40, 90D65.
Abbreviated Title: Perturbed inclusion problems

---

*This work was supported by NSERC under grant WFA 0123160

†Department of Mathematics and Statistics, University of Victoria, Victoria, B.C., Canada V8W 3P4.

1 Introduction

The sensitivity analysis of the value function associated with a perturbed optimization problem is important both in theory and practice. It has been studied by many researchers over the past decade, (e.g. [4, 7, 8, 9, 12, 14, 15, 16]). Among the various models of the perturbed problems, the differential inclusion model considered in Loewen [12] was potentially most general. However, in that paper the perturbation on the differential inclusion must be additive and belong to L^2 space. The additivity assumption on the perturbation is too restrictive in some applications (e.g. in applications to bilevel dynamic problems). Also, according to the setting of the problem the natural space for such a perturbation should be L^1 instead of L^2. Our main objective in this paper is to consider a perturbation model similar to that of [12] with nonadditive L^1 perturbations on the inclusion and derive an estimate of the G-subderivative of the value function. We then apply this estimate to obtain necessary conditions for a solution to a bilevel dynamic problem.

The sensitivity analysis of perturbed dynamic problems was initiated in Clarke's paper [7]. For control problems with smooth data general L^2 perturbations has been considered in Ye [16] and L^1 additive perturbations in Borwein and Zhu [4]. In extending these results to differential inclusion setting the main difficulty is the intrinsic nonsmoothness involved.

In dealing with the bilevel dynamic problem discussed in §3, we need a necessary optimality condition for a nonstandard optimization problem that involves differential inclusion constraints. This result can be proved by following closely Clarke's proof of the standard case [6]. We include the proof in the Appendix for completeness.

The paper is arranged as follows. In §2 we state the problem and our main results. In §3 we consider an application to a bilevel dynamic problem. §4 contains preliminary materials and some technical results that will be used in the proofs of the main results and we prove our main results in §5. Finally in the Appendix we prove a necessary optimality condition for a nonstandard optimization problem that involves differential inclusion constraints following Clarke's proof of the standard case.

2 Statement of the problem and the main result.

We consider the following perturbed optimization problem with differential inclusion constraints:

$$\begin{array}{lll} P(\beta) & \text{minimize} & g(x(1)) \\ & \text{subject to} & \dot{x}(s) \in F(s, x(s), \beta(s)) \quad \text{a.e. in } [0,1], \\ & & x(0) = 0, \end{array}$$

where $g : R^n \to R, F : [0,1] \times R^n \times R^m \longrightarrow 2^{R^n}$ and $\beta \in L^1([0,1], R^m)$. We will often use the following hypotheses.

(H1) The objective function g is Lipschitz of rank L_g.

(H2) The multifunction F has nonempty, compact, convex values, and is $\mathcal{L} \times \mathcal{B}$ measurable where $\mathcal{L} \times \mathcal{B}$ denote the σ-algebra of subsets of $[0,1] \times R^n \times R^m$ generated by product sets $M \times N$ where M is a Lebesgue measurable subset of $[0,1]$ and N a Borel subset of $R^n \times R^m$.

(H3) There is a nonnegative function $k \in L^\infty[0,1]$ such that, for almost all $s \in [0,1]$,

$$F(s, x_1, \beta_1) \subseteq F(s, x_2, \beta_2) + k(s)(\| x_1 - x_2 \| + \| \beta_1 - \beta_2 \|)B \qquad \forall x_1, x_2 \in R^n, \beta_1, \beta_2 \in R^m$$

where B is the closed unit ball.

(H4) For each $\beta \in L^1([0,1], R^m)$ there exists a nonnegative function $\phi_\beta \in L^1[0,1]$ such that

$$F(s, x, \beta(s)) \subseteq \phi_\beta(s) B \quad \forall x \in R^n.$$

(H5) For any $(s, x, p, \beta) \in [0,1] \times R^n \times R^n \times R^m$, define

$$H(s, x, p, \beta) := \sup\{\langle p, y \rangle : y \in F(s, x, \beta)\}.$$

We assume that

$$\partial_{x,p} H(s, x, p, \beta) \text{ and } \partial_\beta H(s, x, p, \beta)$$

are upper semicontinuous with respect to (x, p, β), where ∂ (∂_z) signifies the Clarke generalized gradient (with respect to z).

Remark 2.1. Assumptions (H1)-(H4) are standard. Assumption (H5) is required because we need to take limits from a sequence of "approximate" multipliers. This assumption is satisfied, for example, when the inclusion has a smooth parametrization, i.e. there exist a compact metric space V and a function $\phi : [0,1] \times R^n \times V \times R^m \to R^n$ such that, for all $(s, x, \beta) \in [0,1] \times R^n \times R^m$,

$$F(s, x, \beta) = \{\phi(s, x, v, \beta) : v \in V\}$$

and $\phi(s, x, v, \beta)$ is measurable in s, continuous in (x, v, β), Lipschitz in (x, β) and C^1 in x; or when the perturbation is "separated" from the inclusion, i.e.,

$$F(s, x, \beta) = G(s, x) + h(s, \beta),$$

where G is a multifunction and h a scalar function, including the additive perturbation ($h(s, \beta) = \beta$) as a special case.

For problem $P(\beta)$, we denote its solution set by Σ_β. Then the following result is well known.

Theorem 2.1 [6] *Assume that g and F satisfy assumptions* (H1)-(H4). *If $x \in \Sigma_\beta$ then there exists an absolutely continuous mapping p such that*

(i)

$$\begin{pmatrix} -\dot{p}(s) \\ \dot{x}(s) \end{pmatrix} \in \partial_{x,p} H(s, x(s), p(s), \beta(s)) \quad \text{a.e. in } [0,1];$$

(ii)

$$-p(1) \in \partial g(x(1)).$$

We call such p a Hamiltonian multiplier of x corresponding to β and denote by $M_\beta(x)$ the set of all Hamiltonian multipliers of x corresponding to perturbation β. Let $V(\beta) : L^1 \to R \cup \{+\infty\}$ be the value function corresponding to problem $P(\beta)$ defined by

$$V(\beta) := \inf\{g(x(1)) : \dot{x}(s) \in F(s, x(s), \beta(s)) \text{ a.e. in } [0,1], x(0) = 0\}$$

where by convention $\inf\{\emptyset\} = +\infty$. Then we can state our main result as:

Theorem 2.2 *Assume that g and F satisfy assumptions* (H1)-(H5). *Then V is a (locally) Lipschitz function and*

$$-\partial_C V(\beta) \subset \{\partial_\beta H(\cdot, x(\cdot), p(\cdot), \beta(\cdot)) : p \in M_\beta(x), x \in \Sigma_\beta\}.$$

where ∂_G is the G-subdifferential defined in Ioffe [11] *(see also* [1, 5] *and Theorem* 4.1 *for equivalent definitions) and*

$$\partial_\beta H(\cdot, x(\cdot), p(\cdot), \beta(\cdot)) := \{q \in L^1 : q(s) \in \partial_\beta H(s, x(s), p(s), \beta(s)) \text{ a.e. in } [0,1]\}.$$

Since V is Lipschitzian, $cl^* co \partial_G V = \partial V$ (cf. [1]). Thus, we have

Corollary 2.1 *Under the assumptions of Theorem 2.2*

$$-\partial V(\beta) \subset cl^* co\{\partial_\beta H(\cdot, x(\cdot), p(\cdot), \beta(\cdot)) : p \in M_\beta(x), x \in \Sigma_\beta\},$$

where cl^*co is the weak star convex closure.

Remark 2.2. When the perturbation is additive, i.e. the multifunction F is of the form

$$F(s, x, \beta) = G(s, x) + \beta,$$

we obtain that $\partial_\beta H(\cdot, x(\cdot), p(\cdot), \beta(\cdot)) = p$. Thus, Theorem 2.2 and Corollary 2.1 extend the corresponding results for additive perturbations in [4, 12] when there are no perturbations on the endpoints. (Although the multiplier in [4] is defined as a solution of a relaxed adjoint equation in the sense of Warga [13], under the smoothness assumption of the data in [4], it is actually the same as the Hamiltonian multiplier in this paper.) We should note that how to handle perturbed problems with both endpoint perturbations and L^1 perturbations on the differential inclusion remains open to us.

Remark 2.3. The perturbed control problem with nonadditive perturbations considered in [16, §2] can be reformulated as a perturbed differential inclusion problem with convexity assumptions on the velocity set. In [16, Theorem 2.1], by considering the control problem a result similar to Corollary 2.1 was proved under weaker convexity assumptions on the velocity set and a set of slightly different assumptions.

3 Application to bilevel dynamic problems

Consider the following bilevel dynamic problem:

$$\begin{array}{lll} (BL) & \text{minimize} & f(x(1)) \\ & \text{subject to} & \beta \in L^1([0,1], R^m) \quad x \in \Sigma_\beta \subset W^{1,1}, \end{array}$$

where $W^{1,1}$ is the Sobolev space of absolutely continuous functions with norm

$$||x||_{W^{1,1}} := ||\dot{x}||_{L^1} + ||x||_\infty$$

and Σ_β is the solution set of

$$\begin{array}{ll} \text{minimize} & g(x(1)) \\ \text{subject to} & \dot{x}(s) \in F(s, x(s), \beta(s)) \quad \text{a.e. in } [0,1], \\ & x(0) = 0, \end{array}$$

where f is a Lipschitz function of rank L_f, and g and F are as in §2. Simple in its form, (BL) incorporates many different models of bilevel dynamic problems (e.g. see [15, 16, 17]). In this section we will deduce a necessary condition for a solution to (BL) and illustrates the key difficulty in handling bilevel problems as well as the role of Theorem 2.2 in this process. For each $\beta \in L^1$ define the value function corresponding to the lower level problem by $V(\beta)$. Then following the discussions in [15, 16, 17] problem (BL) can be rewritten as the following equivalent single level problem:

$$\begin{array}{lll} \text{(SL)} & \text{minimize} & f(x(1)) \\ & \text{subject to} & \dot{x}(s) \in F(s, x(s), \beta(s)) \quad \text{a.e. in } [0,1], \\ & & x(0) = 0, \\ & & g(x(1)) - V(\beta) = 0, \quad \beta \in L^1. \end{array}$$

This is a *"nonstandard"* optimization problem with differential inclusion and endpoint constraints involving parameter β. However, following closely Clarke's proof of the Hamiltonian necessary conditions for a solution to the standard optimization problem with differential inclusion constraints [6] we can prove the following theorem (We give the full proof in the Appendix for completeness):

Theorem 3.1 *Assume that g and F satisfy assumptions* (H1)-(H4) *and f is a Lipschitz function of rank L_f. If (x, β) is an optimal pair for problem* (SL) *then there exist $\lambda \in [0,1]$, an absolutely continuous mapping p and $r \in (L^1)^*$ such that*

$$\begin{pmatrix} -\dot{p}(s) \\ \dot{x}(s) \\ -r(s) \end{pmatrix} \in \partial H(s, x(s), p(s), \beta(s)) \quad \text{a.e. in } [0,1],$$

$$
\begin{aligned}
-p(1) &\in \lambda\partial f(x(1)) + (1-\lambda)\partial g(x(1)), \\
r &\in (1-\lambda)\partial V(\beta),
\end{aligned}
$$

where ∂H signifies the generalized gradient of H with respect to (x, p, β).

While given a necessary condition the term $\partial V(\beta)$ in the theorem is very hard to calculate. Now Theorem 2.2 helps in a crucial way. Combining Theorem 3.1 and Theorem 2.2 we are able to get the following useful necessary condition:

Theorem 3.2 *Assume that g and F satisfy assumptions* (H1)-(H5) *and f is a Lipschitz function of rank L_f. If (x, β) is an optimal pair for problem* (SL) *then there exists $\lambda \in [0,1]$, an absolutely continuous mapping p and $r \in (L^1)^*$ such that*

$$
\begin{aligned}
\begin{pmatrix} -\dot{p}(s) \\ \dot{x}(s) \\ -r(s) \end{pmatrix} &\in \partial H(s, x(s), p(s), \beta(s)) \quad \text{a.e. in } [0,1], \\
-p(1) &\in \lambda\partial f(x(1)) + (1-\lambda)\partial g(x(1)), \\
-r &\in (1-\lambda)cl^* co\{\partial_\beta H(\cdot, y(\cdot), q(\cdot), \beta(\cdot)) : q \in M_\beta(y), y \in \Sigma_\beta\}.
\end{aligned}
$$

Remark 3.1. In Theorem 3.1 and Theorem 3.2, $\lambda = 0$ corresponds to the abnormal case. Usually for the purposes of applications one is interested in the normal case $(\lambda > 0)$. For examples that illustrate how to apply these kind of necessary conditions we refer the reader to Ye [16].

4 Preliminaries and technical results.

4.1 Equivalent definition of G-subdifferential in weak Hadamard smooth space.

Let X be a Banach space with continuous real dual X^*. 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 and a functional $x^* \in X^*$, we say that x^* is the weak Hadamard *derivative* of f at x if

$$
t^{-1}(f(x+tu) - f(x) - t\langle x^*, u\rangle) \to 0
$$

as $t \to 0$ uniformly in $u \in U$ for every weakly compact subset $U \subset X$. We will denote the weak Hadamard derivative of f at x by $\nabla^{WH} f(x)$.

Definition 4.1 The *weak Hadamard subdifferential of rank* k of the function f at x is the set $\partial^k_{WH} f(x)$ of vectors $x^* \in X^*$ with the property that there is a function $m(\cdot)$ (depending on x^*) which is weak Hadamard smooth on a neighborhood U of x and satisfies the following conditions:

(m_1) m satisfies a Lipschitz condition with constant k *on* U*;*

(m_2) $m(u) \leq f(u)$ for all $u \in U$*;*

(m_3) $m(x) = f(x)$;

(m_4) $x^* = \nabla^{WH} m(x)$.

The following theorem is a special case of [1, Theorem 2] due to J. M. Borwein and A. Ioffe.

Theorem 4.1 *Let X be a weakly compactly generated Banach space whose norm is weak Hadamard differentiable away from the origin. Let f be a locally Lipschitz function on X. Then for any $x \in X$ the following formula gives an equivalent definition of the G-subderivative introduced in* [11]

$$\partial_G f(x) = \bigcup_{k=1}^{\infty} \{\mathrm{w}^* - \lim_{n \to \infty} x_n^* : x_n^* \in \partial^k_{WK} f(x_n),\ x_n \to x\},$$

and

$$\partial_C f(x) = \mathrm{cl}^* co \partial_G f(x).$$

This theorem in particular applies to the weakly compactly generated Banach space L^1 by the weak Hadamard smooth renorming theorem of Borwein and Fitzpartrick [2]. The general form of the sequential limit formula given in Theorem 3.1 can be found in [1]. We refer the reader to [2, 3, 4, 5] for additional discussions and applications.

4.2 Existence of optimal solutions

The following existence result on the optimal solution of problem $P(\beta)$ is standard and can be deduced, for example, directly from [6, Theorem 3.1.7].

Theorem 4.2 *For any* $\beta \in L^1([0,1], R^m)$, $\Sigma_\beta \neq \emptyset$.

4.3 A stability result about the Hamiltonian

Lemma 4.1 *Let* $f_i : [0,1] \times [0,1] \times R^m \to R$ *be a sequence of continuous mappings such that* f_i *uniformly converges to* f *and* r_i *a sequence of positive numbers converges to* 0. *Let* β *be a fixed function. Define, for* $(x,p) \in R^n \times R^n$,

$$H_i(s,x,p) := \sup\{H(s,x,p,y) + f_i(r,s,y) : y \in \beta(s) + rB, r \in [0, r_i]\}.$$

Then, for any sequence p_i *of continuous mappings uniformly converges to* p *and continuous mapping* x *we have*

$$\limsup_{i\to\infty} \partial_{x,p} H_i(s, x(s), p_i(s)) \subseteq \partial_{x,p} H(s, x(s), p(s), \beta(s)).$$

Proof. By the pointwise maxima formula [6, Theorem 2.8.2] for the generalized gradient,

$$\partial_{x,p} H_i(s, x(s), p_i(s)) \subseteq co\{\partial_{x,p} H(s, x(s), p_i(s), \beta') : \beta' \in \beta(s) + r_i B\}.$$

Since $r_i \to 0, p_i(s) \to p(s)$, assumption (H5) implies that

$$\begin{aligned}\limsup_{i\to\infty} \partial_{x,p} H_i(s, x(s), p_i(s)) &\subseteq \limsup_{i\to\infty}\ co\{\partial_{x,p} H(s, x(s), p_i(s), \beta') : \beta' \in \beta(s) + r_i B\} \\ &\subseteq \partial_{x,p} H(s, x(s), p(s), \beta(s)).\end{aligned}$$

4.4 A relation between subderivative and Hamiltonian multiplier

Lemma 4.2 *Let* k *be an integer,* $q \in (L^1)^*$ *and* $\beta \in L^1$ *such that*

$$-q \in \partial^k_{WH} V(\beta).$$

Then there exist $x \in \Sigma_\beta$ *and* $p \in M_\beta(x)$ *such that*

$$q(s) \in \partial_\beta H(s, x(s), p(s), \beta(s)) \quad \text{a.e. in } [0,1].$$

Proof. Define

$$W := \{y \in L^1 : y(s) \in B \text{ a.e. in } [0,1]\}.$$

Then W is a weakly compact set. Since $-q \in \partial^k_{WH} V(\beta)$, by Definition 4.1, there exists a weak Hadamard smooth Lipschitz function m of rank k such that $V - m$ attains minimum 0 at β. Since W is a weakly compact set, for any integer i, there exists $r_i < i^{-1}$ such that

$$m(\beta + tw) - m(\beta) - \langle -q, tw\rangle \geq -i^{-1}t \quad \forall w \in W, t \in [0, r_i].$$

Observing that $V(\beta') \geq m(\beta')$ and $V(\beta) = m(\beta)$ we obtain

$$V(\beta') - V(\beta) + \langle q, \beta' - \beta\rangle \geq -i^{-1}t \quad \forall \beta' \in \beta + tW, t \in [0, r_i].$$

Let $x \in \Sigma_\beta$. Then $V(\beta) = g(x(1))$ and $V(\beta') \leq g(y(1))$, for any solution y of

$$\begin{aligned} &\dot{y}(s) \in F(s, y(s), \beta'(s)) \quad \text{a.e. in } [0,1], \\ &y(0) = 0, \;\; \beta' \in \beta + tW \;\; t \in [0, r_i]. \end{aligned} \tag{1}$$

Therefore, $(y, t, \beta') = (x, 0, \beta)$ is a solution to the following optimization problem:

$$\begin{aligned} \text{minimize} \quad & g(y(1)) + \int_0^1 [\langle q(s), \beta'(s)\rangle + i^{-1}t]ds \\ \text{subject to} \quad & \dot{y}(s) \in F(s, y(s), \beta'(s)) \quad \text{a.e. in } [0,1], \\ & y(0) = 0, \;\; \beta'(s) \in \beta(s) + tB \;\; t \in [0, r_i]. \end{aligned}$$

Define $\bar{y} = (y, \xi)$ and $\bar{g}(\bar{y}) := g(y) + \xi$. Then $\bar{x}(s) := (x(s), \int_0^s \langle q(\tau), \beta(\tau)\rangle d\tau)$ and $\beta' = \beta \in L^1$ is a solution pair of the following optimization problem:

$$\begin{aligned} \text{minimize} \quad & \bar{g}(\bar{y}(1)) \\ \text{subject to} \quad & \dot{\bar{y}}(s) \in F(s, y(s), \beta'(s)) \times \{i^{-1}t + \langle q(s), \beta'(s)\rangle\} \quad \text{a.e. in } [0,1], \\ & \bar{y}(0) = 0, \;\; \beta'(s) \in \beta(s) + tB, \;\; t \in [0, r_i]. \end{aligned}$$

Set $C_i^s := \{(t, \beta') : \beta' \in \beta(s) + rB, t \in [0, r_i]\}$ and

$$G_i(s, \bar{y}) := \bigcup_{(t,\beta') \in C_i^s} F(s, y, \beta') \times \{i^{-1}t + \langle q(s), \beta'\rangle\}.$$

Then $\bar{x}(s)$ is a solution of

$$\begin{aligned} &\text{minimize} \quad && \bar{g}(\bar{y}(1)) \\ &\text{subject to} \quad && \dot{\bar{y}}(s) \in G_i(s, \bar{y}(s)) \quad \text{a.e. in } [0,1], \\ & && \bar{y}(0) = 0. \end{aligned}$$

It is easy to check that G_i is Lipschitz in $\bar{y}$. By the relaxation theorem (cf. [18]), any trajectory of the convexified differential inclusion

$$\begin{aligned} &\dot{\bar{y}}(s) \in \overline{co} G_i(s, \bar{y}(s)) \quad \text{a.e. in } [0,1], \\ &\bar{y}(0) = 0. \end{aligned}$$

can be approximated uniformly by trajectories of the original inclusion

$$\begin{aligned} &\dot{\bar{y}}(s) \in G_i(s, \bar{y}(s)) \quad \text{a.e. in } [0,1], \\ &\bar{y}(0) = 0. \end{aligned}$$

Therefore $\bar{x}(s)$ is also a solution of the convexified problem:

$$\begin{aligned} &\text{minimize} \quad && \bar{g}(\bar{y}(1)) \\ &\text{subject to} \quad && \dot{\bar{y}}(s) \in \overline{co} G_i(s, \bar{y}(s)) \quad \text{a.e. in } [0,1], \\ & && \bar{y}(0) = 0. \end{aligned}$$

Applying Clarke's Hamiltonian necessary condition [6], we obtain that there exists an arc $\bar{p}_i = (p_i, \eta_i)$ such that

$$\begin{pmatrix} -\dot{\bar{p}}_i(s) \\ \dot{\bar{x}}(s) \end{pmatrix} \in \partial_{\bar{x},\bar{p}} \bar{H}_i(s, \bar{x}(s), \bar{p}_i(s)) \quad \text{a.e. in } [0,1], \tag{2}$$

$$-\bar{p}_i(1) \in \partial \bar{g}(\bar{x}(1)) \tag{3}$$

where

$$\begin{aligned} \bar{H}_i(s, \bar{x}, \bar{p}_i) := \sup\{ & \langle p_i, y \rangle + \eta_i (i^{-1} t + \langle q(s), \beta' \rangle) \\ & : y \in F(s, x, \beta'), \quad (t, \beta') \in C_i^s \} \end{aligned}$$

Since $\bar{H}_i$ does not depend on ξ, inclusion (2) implies that $\dot{\eta}_i(s) = 0$ and

$$\begin{pmatrix} -\dot{p}_i(s) \\ \dot{\bar{x}}(s) \end{pmatrix} \in \partial_{\bar{x},p} \bar{H}_i(s, \bar{x}(s), \bar{p}_i(s)) \quad \text{a.e. in } [0,1]. \tag{4}$$

In considering relation (3) $\eta_i(s) = -1$ and (4) becomes

$$\begin{pmatrix} -\dot{\bar{p}}_i(s) \\ \dot{\bar{x}}(s) \end{pmatrix} \in \partial_{\bar{x},p}\bar{H}_i(s,\bar{x}(s),p_i(s),-1) \quad \text{a.e. in } [0,1]. \tag{5}$$

Thus, using an argument as in page 173 of [6], one has

$$\begin{pmatrix} -\dot{p}_i(s) \\ \dot{x}(s) \end{pmatrix} \in \partial_{x,p}H_i(s,x(s),p_i(s)), \quad \text{a.e. in } [0,1], \tag{6}$$

$$-p_i(1) \in \partial g(x(1)), \tag{7}$$

where

$$\begin{aligned} H_i(s,x,p_i) &:= \sup\{\langle p_i, y\rangle - (i^{-1}t + \langle q(s),\beta'\rangle) : y \in F(s,x,\beta'), \quad (t,\beta') \in C_i^s\} \\ &= \sup\{H(s,x,p_i,\beta') - (i^{-1}t + \langle q(s),\beta'\rangle) : (t,\beta') \in C_i^s\}. \end{aligned}$$

By relation (3) $p_i(1)$ are uniformly bounded. Using (2) and Gronwall's inequality we may conclude that $\dot{p}_i$ is uniformly bounded. Thus, we may assume that there exists an absolutely continuous mapping p such that p_i uniformly converges to p and $\dot{p}_i$ weakly converges to $\dot{p}$ in L^1. Taking limits in (6) and (7) by Lemma 4.1 we have

$$\begin{pmatrix} -\dot{p}(s) \\ \dot{x}(s) \end{pmatrix} \in \partial_{x,p}H(s,x(s),p(s),\beta(s)) \quad \text{a.e. in } [0,1],$$

$$-p(1) \in \partial g(x(1)).$$

Also (2) implies that (see [6, Proposition 3.2.4])

$$\langle \bar{p}_i(s), \dot{\bar{x}}(s)\rangle = \bar{H}_i(s,\bar{x}(s),\bar{p}_i(s)) \quad \text{a.e. in } [0,1]. \tag{8}$$

Since $\dot{x}(s) \in F(s,x(s),\beta(s))$ a.e. in $[0,1]$, we clearly have, for almost all $s \in [0,1]$,

$$\begin{aligned} \langle \bar{p}_i(s), \dot{\bar{x}}(s)\rangle &= \langle p_i(s), \dot{x}(s)\rangle - \langle q(s),\beta(s)\rangle \\ &\leq H(s,x(s),p_i(s),\beta(s)) - \langle q(s),\beta(s)\rangle. \end{aligned} \tag{9}$$

Combining (8) and (9) we obtain that

$$\phi^s(t,\beta') := H(s,x(s),p_i(s),\beta') - (i^{-1}t + \langle q(s),\beta'\rangle)$$

restricted to C_i^s attains a maximum at $(0,\beta(s))$ for almost all $s \in [0,1]$. Thus,

$$0 \in -\partial\phi^s(0,\beta(s)) + N_{C_i^s}(0,\beta(s)), \quad \text{a.e. in } [0,1], \tag{10}$$

where $N_{C_i^s}(0,\beta(s))$ is the normal cone (in the sense of convex analysis) of C_i^s at $(0,\beta(s))$ and, by direct calculation,

$$N_{C_i^s}(0,\beta(s)) = \{(r,y) : y \in -rB, r \leq 0\}.$$

Therefore relation (10) implies that, for almost all $s \in [0,1]$, there exists an $h(s) \in \partial_\beta H(s,x(s),p_i(s),\beta(s))$ such that

$$(-i^{-1}, h(s) - q(s)) \in \{(r,y) : y \in -rB, r \leq 0\}.$$

That is to say

$$-q(s) \in -\partial_\beta H(s,x(s),p_i(s),\beta(s)) + i^{-1}B \quad \text{a.e. in } [0,1].$$

Upon taking limits we obtain

$$q(s) \in \partial_\beta H(s,x(s),p(s),\beta(s)) \quad \text{a.e. in } [0,1].$$

5 Proof of the main result.

Proof of Theorem 2.2.

Step 1. Lipschitz continuity of the value function V. The proof of V is Lipschitz with respect to β is in the same spirit of [18, Theorem 4.7]. We include details here for completeness. We will need the following special case of Filippov's result on the relation of approximate solutions and solutions of a differential inclusion (cf. [10, 18]).

Lemma 5.1 *Assume that F satisfies assumptions* (H2)-(H4). *For any absolutely continuous mapping x on $[0,1]$ and $\beta' \in L^1$, there exists a solution y to the differential inclusion*

$$\dot{y}(s) \in F(s,y(s),\beta'(s)) \quad \text{a.e. in } [0,1],$$

$$y(0) = x(0).$$

such that

$$||y - x||_{W^{1,1}} \leq exp(\int_0^1 k(s)ds) \cdot \int_0^1 d(\dot{x}(s), F(s,x(s),\beta'(s)))dt$$

where $d(b,A) := \inf\{||b-a|| : a \in A\}$.

Consider $\beta, \beta' \in L^1$. Let $x \in \Sigma_\beta$. Then $V(\beta) = f(x(1))$. By Lemma 5.1 there exists a solution y to the differential inclusion

$$\dot{y}(s) \in F(s, y(s), \beta'(s)) \quad \text{a.e. in } [0,1],$$

$$y(0) = x(0) = 0$$

such that

$$||y(1) - x(1)|| \leq exp(\int_0^1 k(s)ds) \cdot \int_0^1 d(\dot{x}(s), F(s, x(s), \beta'(s)))dt$$

Observing that $\dot{x}(s) \in F(s, x(s), \beta(s))$ almost everywhere and using assumption (H3), we obtain

$$\int_0^1 d(\dot{x}(s), F(s, x(s), \beta'(s)))dt \leq \int_0^1 k(s)\|\beta'(s) - \beta(s)\|ds \leq ||k||_\infty \cdot ||\beta' - \beta||_{L^1}.$$

Thus,

$$\begin{aligned} V(\beta') - V(\beta) &\leq g(y(1)) - g(x(1)) \\ &\leq L_g \cdot ||y(1) - x(1)|| \\ &\leq L_g \cdot exp(\int_0^1 k(s)ds) \cdot ||k||_\infty \cdot ||\beta' - \beta||_{L^1}. \end{aligned}$$

Since we can interchange the positions of β and β' in the above arguments V is Lipschitz with respect to β.

Step 2. Let $-q \in \partial_G V(\beta)$. Then, by the sequential limit formula stated in Theorem 4.1, there exist integer $k, q_n \in (L^1)^*$ and $\beta_n \in L^1$ such that

$$-q_n \in \partial^k_{WH} V(\beta_n),$$

where $\beta_n \to \beta$ in L^1 and $q_n \xrightarrow{w^*} q$ in $(L^1)^*$.

Step 3. By Lemma 4.2, for each n, there exist $x_n \in \Sigma_{\beta_n}$ and an absolutely continuous function p_n satisfies, for almost all $s \in [0,1]$,

$$\begin{pmatrix} -\dot{p}_n(s) \\ \dot{x}_n(s) \end{pmatrix} \in \partial_{x,p} H(s, x_n(s), p_n(s), \beta_n(s)) \tag{11}$$

$$-p_n(1) \in \partial g(x_n(1)) \tag{12}$$

$$q_n(s) \in \partial_\beta H(s, x_n(s), p_n(s), \beta_n(s)). \tag{13}$$

Since g is Lipschitz with rank L_g, $p_n(1)$ are uniformly bounded by relation (12). Using (11) and the Gronwall's inequality we conclude that $\dot{p}_n$ is uniformly bounded. Thus, we may assume without loss of generality that p_n uniformly converges to an absolutely continuous mapping p. Similarly we may assume that x_n converges uniformly to an absolutely continuous mapping x. Taking limits in (11), (12) and (13) yields, for almost all $s \in [0, 1]$,

$$\begin{pmatrix} -\dot{p}(s) \\ \dot{x}(s) \end{pmatrix} \in \partial_{x,p} H(s, x(s), p(s), \beta(s))$$

$$-p(1) \in \partial g(x(1))$$

$$q(s) \in \partial_\beta H(s, x(s), p(s), \beta(s)).$$

It is easy to see that x is a solution of the differential inclusion

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

$$x(0) = 0.$$

Since V is Lipschitz, $x \in \Sigma_\beta$. Thus, $p \in M_\beta(x)$ and the proof is completed.

6 Appendix

In this appendix we give a proof of Theorem 3.1 following Clarke's proof of Theorem 2.1 (cf. [6, page 123-130]). We will keep the notation and exposition close to that of [6].

Proof of Theorem 3.1 To keep notation simple we will omit the variable s when this does not cause confusion. For any $a > 0$ and $p \in R^n$ define

$$F_a(y, \gamma) := F(y, \gamma) + aB,$$

$$\rho(y, v, \gamma) := d(v, F(y, \gamma)),$$

$$\rho_a(y, v, \gamma) := d(v, F_a(y, \gamma)),$$

and

$$H_a(y, p, \gamma) := \sup\{\langle p, v\rangle : v \in F_a(y, \gamma)\}.$$

We denote by $S_a(\gamma)$ the solution set of

$$\dot{y} \in F_a(y, \gamma), \quad y(0) = 0,$$

and define

$$A_a := \{(y,\gamma) : y \in S_a(\gamma), ||\gamma - \beta|| \le a\}.$$

Then the metric

$$\Delta((y,\gamma),(z,\alpha)) := ||y - z||_{L^1} + ||\gamma - \alpha||_{L^1}$$

turns A_a into a complete metric space.

Let (x,β) be a solution pair to (SL).

Step 1. Define

$$\phi_\varepsilon(y,\gamma) := \max\{g(y(1)) - V(\gamma), f(y(1)) - f(x(1)) + \varepsilon^2\}.$$

Lemma 6.1 *Given $\varepsilon > 0$, then for all $a > 0$ sufficiently small and for all $(y,\alpha) \in A_a$ one has $\phi_\varepsilon(y,\alpha) > 0$.*

If this were false, there would exist a sequence $a_i \to 0$ and $(y_i,\alpha_i) \in A_{a_i}$ with $\phi_\varepsilon(y_i,\alpha_i) \le 0$. Without loss of generality we may assume that y_i uniformly converges to $y \in S_0(\beta)$. It follows that

$$\phi_\varepsilon(y,\beta) \le 0$$

so that $g(y(1)) - V(\beta) \le 0$ and also $f(y(1)) \le f(x(1)) - \varepsilon^2$. contradicts the fact that (x,β) is an optimal pair.

Step 2. Choose a as in the statement of Lemma 6.1 with $a < \varepsilon$, and we note that $\phi_\varepsilon(x,\beta) = \varepsilon^2$. By Lemma 6.1 it follows that

$$\phi_\varepsilon(x,\beta) < \inf_{A_a} \phi_\varepsilon + \varepsilon^2.$$

By Ekeland's variational principle, there exist $(z,\gamma) \in A_a$ minimizes

$$\phi_\varepsilon(y,\alpha) + \varepsilon\Delta((y,\alpha),(z,\gamma))$$

over $(y,\alpha) \in A_a$ and such that

$$\Delta((x,\beta),(z,\gamma)) \le \varepsilon, \phi_\varepsilon(z,\gamma) \le \varepsilon^2.$$

Lemma 6.2 *For some $\eta > 0$, the pair (z,γ) minimizes*

$$\phi_\varepsilon(y,\alpha) + \varepsilon\Delta((y,\alpha),(z,\gamma)) + K\int_0^1 \rho_a(y,\dot{y},\alpha)ds$$

over all pairs (y,α) satisfying

$$\Delta((y,\alpha),(z,\gamma)) < \eta,$$

where $K := [\max(L_g, L_f) + \varepsilon]exp(\int_0^1 k(s)ds)$.

Suppose the assertion to be false. Then there is a sequence (y_i, α_i) of pairs such that y_i uniformly converges to z and $\alpha_i \to \gamma$ in L^1 for which the expression in the lemma is less than its value at (z, γ). It follows that

$$\int_0^1 \rho_a(y_i, \dot{y}_i, \alpha_i) ds \to 0.$$

We are now able to apply Filippov's result (Lemma 5.1) to deduce the existence of a trajectory $z_i \in S_a(\alpha_i)$ such that $z_i(0) = y_i(0)$ and

$$||z_i - y_i|| \leq exp(\int_0^1 k(s)ds) \int_0^1 \rho_a(y_i, \dot{y}_i, \alpha_i)ds.$$

Therefore we have

$$\begin{aligned} & \phi_\varepsilon(z_i, \alpha_i) + \varepsilon\Delta((z_i, \alpha_i), (z, \gamma)) \\ \leq \ & \phi_\varepsilon(y_i, \alpha_i) + \varepsilon\Delta((y_i, \alpha_i), (z, \gamma)) + (\max(L_g, L_f) + \varepsilon)||z_i - y_i|| \\ \leq \ & \phi_\varepsilon(y_i, \alpha_i) + \varepsilon\Delta((y_i, \alpha_i), (z, \gamma)) + K \int_0^1 \rho_a(y_i, \dot{y}_i, \alpha_i)ds \\ < \ & \phi_\varepsilon(z, \gamma). \end{aligned}$$

The upshot is that the element $(z_i, \alpha_i) \in A_a$ assigns a lower value to

$$\phi_\varepsilon(\cdot, \cdot) + \varepsilon\Delta((\cdot, \cdot), (z, \gamma))$$

than does (z, γ). This contradiction of the optimality of (z, γ) completes the proof of the Lemma 6.2.

Step 3. Define, for $(y, \alpha) \in W^{1,1} \times L^1$,

$$h(y, \alpha) := K \int_0^1 \rho_a(y, \dot{y}, \alpha)ds.$$

By Lemma 6.2.

$$0 \in \partial\phi_\varepsilon(z, \gamma) + \partial\varepsilon\Delta((z, \gamma), (z, \gamma)) + \partial h(z, \gamma).$$

Now we turn to the calculation of these generalized gradients.

a. By the integral formula of the generalized gradients, for any $\xi \in \partial h(z, \gamma)$ there exists

$$(q, p, r)(s) \in \partial K \rho_a(z, \dot{z}, \gamma)(s)$$

such that for any $(y, u) \in W^{1,1} \times L^1$,

$$\xi(y, u) = \int_0^1 \{\langle q, y\rangle + \langle p, \dot{y}\rangle + \langle r, u\rangle\}ds.$$

b. Similarly, any element $\xi \in \partial\varepsilon\Delta((z,\gamma),(z,\gamma))$ corresponding to a pair of functions (θ_1,θ_2) with $(\theta_1,\theta_2)(s) \in \varepsilon(B_{R^n} \times B_{R^n})$ such that for any $(y,u) \in W^{1,1} \times L^1$,

$$\xi(y,u) = \int_0^1 \{\langle\theta_1, y\rangle + \langle\theta_2, u\rangle\} ds.$$

c. Now we turn to $\partial\phi_\varepsilon(z,\gamma)$. If ξ lies in the generalized gradient of $(y,\alpha) \to f(y(1))$ at (z,γ) then

$$\xi(y,u) = \langle\xi_1, y(1)\rangle$$

for some element $\xi_1 \in \partial f(z(1))$. If ξ lies in the generalized gradient of $(y,\alpha) \to g(y(1)) - V(\alpha)$ at (z,γ) then there exist $\xi_2 \in \partial g(z(1))$ and $\xi_3 \in \partial V(\gamma)$ such that for any $(y,u) \in W^{1,1} \times L^1$,

$$\xi(y,u) = \langle\xi_2, y(1)\rangle - \int_0^1 \langle\xi_3, u\rangle ds.$$

Then by the pointwise maxima formula (cf. [6, Theorem 2.8.2]) generally there exists a $\lambda \in [0,1]$ such that

$$\xi(y,u) = \lambda\langle\xi_1, y(1)\rangle + (1-\lambda)(\langle\xi_2, y(1)\rangle - \int_0^1 \langle\xi_3, u\rangle ds)$$

for some $\xi_1 \in \partial f(z(1)), \xi_2 \in \partial g(z(1))$ and $\xi_3 \in \partial V(\gamma)$. In summary we have

Lemma 6.3 *There exists a $\lambda \in [0,1]$, elements ξ_1, ξ_2 and ξ_3 of $\partial f(z(1)), \partial g(z(1))$ and $\partial V(\gamma)$ respectively, a selection (q,p,r) of $\partial K\rho_a(z,\dot{z},\gamma)$ and a selection (θ_1,θ_2) of $\varepsilon(B_{R^n} \times B_{R^n})$ such that, for any pair $(y,u) \in W^{1,1} \times L^1$, one has*

$$\begin{aligned} 0 &= \lambda\langle\xi_1, y(1)\rangle + (1-\lambda)(\langle\xi_2, y(1)\rangle - \int_0^1 \langle\xi_3, u\rangle ds) \\ &+ \int_0^1 \{\langle\theta_1, y\rangle + \langle\theta_2, u\rangle\} ds + \int_0^1 \{\langle q, y\rangle + \langle p, \dot{y}\rangle + \langle r, u\rangle\} ds. \qquad (14) \end{aligned}$$

Separating y and u and using standard variational arguments the identity of the lemma yields

$$\begin{aligned} r + \theta_2 &= (1-\lambda)\xi_3 \\ p(s) &= \int_0^s (q(\tau) + \theta_1(\tau)) d\tau \\ p(1) &= -\lambda\xi_1 - (1-\lambda)\xi_2. \end{aligned}$$

Thus,

$$r + \theta_2 \in (1-\lambda)\partial V(\gamma)$$
$$(\dot p - \theta_1, p, r)(s) \in K\partial\rho_a(z, \dot z, \gamma)(s)$$
$$p(1) \in -\lambda\partial f(z(1)) - (1-\lambda)\partial g(z(1)).$$

Step 4. We now wish to bring in the Hamiltonian H.

Lemma 6.4 *The relations* $(q, p, r) \in \partial K\rho_a(y, v, \gamma)$ *and* $\rho_a(y, v, \gamma) = 0$ *imply*

$$(-q, v, -r) \in \partial H_a(y, p, \gamma).$$

This lemma is the same as Lemma 4 in Clarke's proof except for the extra variable pair r and γ. Observing that r and γ behave in the same way as q and y do, respectively, the proof of Lemma 4 in [6] applies here with obvious notational change.

Using this last lemma we can rewrite the (approximate) necessary condition as

$$r \in (1-\lambda)\partial V(\gamma) + \varepsilon B$$
$$(-\dot p, \dot z, -r)(s) \in \partial H(s, z(s), p(s), \gamma(s)) + \varepsilon(1 + |p(s)|)B$$
$$-p(1) \in \lambda\partial f(z(1)) + (1-\lambda)\partial g(z(1)).$$

Now we can take limits (as $(z, \gamma) \to (x, \beta)$) as in Clarke's proof to complete the proof.

References

[1] J. M. Borwein and A. Ioffe, Proximal analysis in smooth spaces, CECM Research Report 93-04 (1993).

[2] J. M. Borwein and S. Fitzpatrick, A weak Hadamard smooth renorming of $L_1(\Omega, \mu)$, *Canadian Math. Bull.* **36** (1993), 407-413.

[3] J. M. Borwein and S. Fitzpatrick, Weak-star sequential compactness and bornological limit derivatives, CECM Research Report 93-02 (1993).

[4] J. M. Borwein and Q. J. Zhu, Variational analysis in non-reflexive spaces and applications to control problems with L^1 perturbations, CECM Research Report 94-10 (1994).

[5] J. M. Borwein and Q. J. Zhu, Viscosity solutions and viscosity subderivatives in smooth Banach spaces with applications to metric regularity, CECM Research Report 94-12 (1994).

[6] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley 1983.

[7] F. H. Clarke, Perturbed optimal control problems, *IEEE Trans. AC* Vol. AC-31, (1986), 535-542.

[8] F. H. Clarke and F. D. Loewen, The value function in optimal control: sensitivity, controllability, and time-optimality, *SIAM J. Contr. and Optim.* **24** (1986), 243-263.

[9] F. H. Clarke and F. D. Loewen, State constraints in optimal control: A case study in proximal normal analysis, *SIAM J. Contr. and Optim.* **25** (1987), 1440-1456.

[10] A. F. Filippov, Classical solutions of differential equations with multivalued right-hand side, *SIAM J. Control* **5** (1967), 609-621.

[11] A. Ioffe, Approximate subdifferentials and applications 3: the metric theory, *Mathematika* **36** Part 1 (1989), 1-38.

[12] P. D. Loewen, Perturbed differential inclusion problems, in *Nonsmooth Optimization and related Topics*, eds. F. H. Clarke, V. F. Dem'yanov and F. Giannessi, Plenum Press, New York, 1989, 255-263.

[13] J. Warga, *Optimal controls of differential and functional equations*, Academic Press, New York, 1972.

[14] J. J. Ye, Perturbed infinite horizon optimal control problems, *J. Math. Anal. Appl.* **182** (1994), 90-112.

[15] J. J. Ye, Necessary conditions for bilevel dynamic optimization problems, To appear in *SIAM J. Contr. and Optim.*

[16] J. J. Ye, Optimal Strategies for bilevel dynamical programming problems, Research Report DMS-645-IR, University of Victoria, 1993.

[17] J. J. Ye and D. L. Zhu, Optimality conditions for bilevel programming problems, to appear in *Optimization*.

[18] Q. J. Zhu, On the solution set of differential inclusions in Banach space, *J. Diff. Equs.* **93** (1991), 213-237.