SIAM J. CONTROL OPTIM.
Vol. 35, No. 2, pp. 512–531, March 1997

OPTIMAL STRATEGIES FOR BILEVEL DYNAMIC PROBLEMS*

JANE J. YE†

Abstract. In this paper we study the bilevel dynamic problem, which is a hierarchy of two dynamic optimization problems, where the constraint region of the upper level problem is determined implicitly by the solutions to the lower level optimal control problem. To obtain optimality conditions, we reformulate the bilevel dynamic problem as a single level optimal control problem that involves the value function of the lower-level problem. Sensitivity analysis of the lower-level problem with respect to the perturbation in the upper-level decision variable is given and first-order necessary optimality conditions are derived by using nonsmooth analysis. A constraint qualification of calmness type and a sufficient condition for the calmness are also given.

Key words. necessary conditions, bilevel dynamic problems, sensitivity analysis, nonsmooth analysis, value function, constraint qualification, calmness condition

AMS subject classifications. 90D65, 49K4

PII. S0363012993256150

1. Introduction.

Let us consider a two-level hierarchical system where two decision makers try to find best decisions with respect to certain, but generally different, goals. Moreover, assume that these decision makers cannot act independently of each other but only according to a certain hierarchy whereby the optimal strategy chosen by the lower level (hereafter the "follower") depends on the strategy selected by the upper level (hereafter the "leader"). On the other hand, let the objective function of the leader depend not only on his own decision but also on the reaction of the follower. Then while having the first choice, the leader is able to evaluate the true value of his own selection only after knowing the follower's possible reactions. Assume that the game is cooperative; i.e., if the follower's problem has several optimal decisions for a given leader's decision, then the follower allows the leader to choose which of them is actually used. Thus the leader will choose his optimal decision among all decisions available and the follower's optimal decision to minimize his objective. In particular, we consider a hierarchical dynamical system, where the state $x(t) \in R^d$ is influenced by the decisions of both leader and follower $u(\cdot)$ and $v(\cdot)$. The state $x(t) \in R^d$ is described by

$$\dot{x}(t) = \phi(t, x(t), u(t), v(t)) \qquad \text{almost everywhere (a.e.) } t \in [t_0, t_1],$$
$$x(t_0) = x_0,$$

where $u(t) \in U$, a closed subset of R^n and $v(t) \in W(t) \subset R^m$ for almost all $t \in [t_0, t_1]$. In mathematical terms, given any control function $u(\cdot)$ selected by the leader, the follower faces the ordinary (single-level) optimal control problem involving a parameter u,

$$P_2(u) \qquad \min\ J_2(x, u, v) = \int_{t_0}^{t_1} G(t, x(t), u(t), v(t))dt + g(x(t_1))$$

*Received by the editors September 24, 1993; accepted for publication (in revised form) January 30, 1996. This research was supported in part by the National Sciences and Engineering Research Council of Canada.
http://www.siam.org/journals/sicon/35-2/25615.html

†Department of Mathematics and Statistics, University of Victoria, Victoria, BC, Canada V8W 3P4 (janeye@uvaix.uvic.ca).

$$\text{subject to (s.t.)} \quad \begin{aligned} &\dot{x}(t) = \phi(t, x(t), u(t), v(t)) \qquad \text{a.e.}, \\ &x(t_0) = x_0, \\ &v(t) \in W(t) \qquad \text{a.e.}, \end{aligned}$$

while the leader faces the *bilevel dynamic problem*,

$$P_1 \qquad \min \ J_1(x, u, v) = \int_{t_0}^{t_1} F(t, x(t), u(t), v(t)) dt + f(x(t_1))$$
$$\text{over } u \in L^2([t_0, t_1], U) \text{ and all solutions } (x, v) \text{ of } P_2(u).$$

The bilevel static problem, where both the leader's and the follower's decisions are vectors instead of control functions, was first introduced by von Stackelberg [14] for an economic model. The bilevel dynamic problems were first considered by Chen and Cruz in [2]. Most of the bilevel (static or dynamic) problems are attacked by reducing the bilevel problem to a single-level problem with the first-order necessary optimality conditions for the lower-level problem as additional constraints (cf. Bard and Falk [1] and Zhang [20], [21] for bilevel static problems, Chen and Cruz [2] and Zhang [20] for bilevel dynamic problems). The reduction is equivalent provided the lower-level optimal control problem is convex, since in this case the first-order necessary optimality condition is also sufficient. Apart from the strong convexity assumption, the resulting optimality conditions of the above approach involve second-order derivatives and a larger system, since the reduced problem minimizes over the set of original decision variables as well as the set of multipliers of the lower-level problem.

To our knowledge, there is no optimality condition for a general bilevel dynamic problem to date. The necessary condition obtained by Chen and Cruz in [2] holds in the case where Pontryagin's maximum principle for the lower-level optimal control problem is sufficient for optimality and no bounds are allowed for the control functions. The necessary condition was stated in a normal form (i.e., the multiplier for the objective function of the upper-level problem is 1) that holds only when the reduced single-level optimal control problem is calm (see [3] for definition). The necessary condition obtained by Zhang in [20] is only for a bilevel dynamic problem in which the dynamics are linear in the state and control variables and require convexity assumptions on the objective function of the lower-level problem. The purpose of this paper is to provide first-order necessary optimality conditions for problem P_1 under *very general* assumptions (in particular, without convexity assumptions and with bounds on the control functions).

Define the *value function of the lower-level optimal control problem* as an extended-valued functional $V(u) : L^2([t_0, t_1], U) \to \bar{R}$ defined by

$$V(u) := \inf \left\{ \begin{array}{ll} \int_{t_0}^{t_1} G(t, x(t), u(t), v(t)) dt + g(x(t_1)) : & \dot{x}(t) = \phi(t, x(t), u(t), v(t)) \ \text{a.e.} \\ & v(t) \in W(t) \qquad \text{a.e.} \\ & x(t_0) = x_0 \end{array} \right\},$$

where $\bar{R} := R \cup \{-\infty\} \cup \{+\infty\}$ is the extended real line and $\inf \emptyset = +\infty$ by convention. Our approach is to reformulate P_1 as the following single-level optimal control problem:

$$\widetilde{P}_1 \qquad \min \ J_1(u, v) = \int_{t_0}^{t_1} F(t, x(t), u(t), v(t)) dt + f(x(t_1))$$

$$\begin{aligned}\text{s.t.}\quad & \dot{x}(t) = \phi(t, x(t), u(t), v(t)) \qquad \text{a.e.},\\ & x(t_0) = x_0,\\ & u(\cdot) \in L^2([t_0, t_1], U), v(t) \in W(t) \qquad \text{a.e.},\end{aligned}$$

$$\int_{t_0}^{t_1} G(t, x(t), u(t), v(t))dt + g(x(t_1)) - V(u) = 0. \tag{1}$$

The above problem is obviously equivalent to the original bilevel dynamic problem P_1 and is a *nonstandard optimal control problem* since the constraint (1) involves a functional defined by the value function $V(u)$ of the lower-level optimal control problem. In general $V(u)$ is not an explicit function of the problem data and is nonsmooth even in the case where all problem data are smooth functions. To derive a necessary condition for optimality for problem P_1, one needs to study Lipschitz continuity and generalized gradients of the value function $V(u)$ and develop a necessary optimality condition for the nonstandard optimal control problem with functional constraints (1). Recent developments in nonsmooth analysis allow us to study Lipschitz continuity and generalized gradients of the value function $V(u)$ with respect to a *nonadditive* infinite-dimensional perturbation u. We then reformulate the nonstandard optimal control problem as an infinite-dimensional optimization problem and use a result due to Ioffe [8] to derive a necessary optimality condition for the nonstandard optimal control problem with functional constraints.

The approach of reducing a bilevel problem to a single-level problem using the value function was used in the literature (see [11], [12]) for numerical purposes and for deriving first-order necessary conditions for the static bilevel optimization problem [17], [18]. The essential issue in the static case is the constraint qualification since the generalized differentiability of the value function in the finite-dimensional case is well known and the resulting equivalent single-level problem is an ordinary mathematical programming problem. It was shown in [17] and [18] that bilevel problems always have abnormal multipliers, and the right constraint qualification for ensuring the existence of a normal multiplier is the calmness condition. In Ye [16], a bilevel dynamic optimization problem where the lower level is an optimal control problem while the upper-level decision variable is a vector is considered. Although the bilevel dynamic optimization problem considered in [16] is a special case of the problem we study in this paper, it deserves special attention since it reduces to a single-level optimal control problem with end point constraints involving a value function that is a function of the upper-level decision vector. Fritz John–type necessary optimality conditions were derived under more general assumptions.

The following basic assumptions are in force throughout this paper:

(A1) $W(t) : [t_0, t_1] \to R^m$ is a nonempty, compact-valued, set-valued map. The graph of $W(t)$ (i.e., the set $\{(s, r) : s \in [t_0, t_1], r \in W(s)\}$), denoted by $\mathrm{Gr}W$, is $\mathcal{L} \times \mathcal{B}$ measurable, where $\mathcal{L} \times \mathcal{B}$ denotes the σ-algebra of subsets of $[t_0, t_1] \times R^m$ generated by product sets $M \times N$ where M is a Lebesgue measurable subset of $[t_0, t_1]$ and N is a Borel subset of R^m.

(A2) The function $F(t, x, u, v) : [t_0, t_1] \times R^d \times R^n \times R^m \to R$ is $\mathcal{L} \times \mathcal{B}$ measurable in (t, v) and continuously differentiable in x and u. The functions $\phi(t, x, u, v) : [t_0, t_1] \times R^d \times R^n \times R^m \to R^d$, $G(t, x, u, v) : [t_0, t_1] \times R^d \times R^n \times R^m \to R$ are measurable in t, continuously differentiable in x and u, and lower semicontinuous in v.

(A3) There exists an integrable function $\psi : [t_0, t_1] \to R$ such that

$$|\nabla_{(x,u)} F| + |\nabla_{(x,u)} G| + |\nabla_{(x,u)} \phi| \le \psi(t) \quad \forall (t, x, u, v) \in [t_0, t_1] \times R^d \times U \times W(t).$$

(A4) The function $f(x) : \mathbb{R}^d \to R$ is locally Lipschitz continuous, and the function $g(x) : R^d \to R$ is Lipschitz continuous of rank $L_g \geq 0$.

(A5) For any $u \in L^2([t_0, t_1], U)$, $P_2(u)$ has an admissible pair (whose definition is given below).

A *control function* for $P_2(u)$ is a (Lebesgue) measurable selection $v(\cdot)$ for $W(\cdot)$, that is, a measurable function satisfying $v(t) \in W(t)$ a.e. $t \in [t_0, t_1]$. An *arc* is an absolutely continuous function. An *admissible pair* for $P_2(u)$ is a pair of functions $(x(\cdot), v(\cdot))$ on $[t_0, t_1]$ of which $v(\cdot)$ is a control function for $P_2(u)$ and $x(\cdot) : [t_0, t_1] \to R^d$ is an arc that satisfies the differential equation $\dot{x}(t) = \phi(t, x(t), u(t), v(t))$ a.e., together with the initial condition $x(t_0) = x_0$. The first and the second components of an admissible pair are called an *admissible trajectory* and *admissible control*, respectively. A *solution* to problem $P_2(u)$ is an admissible pair for $P_2(u)$ that minimizes the value of the cost functional $J_2(x, u, v)$ over all admissible pairs for $P_2(u)$. An *admissible strategy* for P_1 includes $u \in L^2([t_0, t_1], U)$ and an optimal control v for $P_2(u)$. The strategy (u, v) and the corresponding trajectory x are *optimal* for the bilevel dynamic problem P_1 if (x, u, v) minimizes the value of the cost functional $J_1(x, u, v)$ among all admissible strategies and the corresponding trajectories for P_1.

The plan of the paper is as follows. In section 2, we study generalized differentiability of the value function $V(u)$. In section 3, under a calmness-type constraint qualification, we derive a Kuhn–Tucker–type necessary optimality condition for the bilevel dynamic problem. It is also shown that the existence of a uniformly weak sharp minimum is a sufficient condition for the calmness, and a sufficient condition for existence of a weak sharp minimum is given. Finally, three examples are given in section 3 to illustrate applications of the constraint qualification and the necessary optimality conditions.

2. Differentiability of the value function. Let X be a Hilbert space. Consider a lower semicontinuous functional $\phi : X \to R \cup \{+\infty\}$ and a point $\bar{x} \in X$, where ϕ is finite. A vector $\zeta \in X$ is called a *proximal subgradient* of $\phi(\cdot)$ at $\bar{x}$ provided that there exist $M > 0, \delta > 0$ such that

$$\phi(x') - \phi(\bar{x}) + M\|x' - \bar{x}\|^2 \geq \langle \zeta, x' - \bar{x} \rangle, \qquad x' \in \bar{x} + \delta B.$$

The set of all proximal subgradients of $\phi(\cdot)$ at $\bar{x}$ is denoted $\partial^\pi \phi(\bar{x})$. A *limiting subgradient* of ϕ at $\bar{x}$ is the set

$$\hat{\partial}\phi(\bar{x}) := \{\text{weak} \lim_{k \to \infty} \zeta_k : \zeta_k \in \partial^\pi \phi(x_k), x_k \to \bar{x}, \phi(x_k) \to \phi(\bar{x})\}.$$

The limiting subgradient is a smaller object than the *Clarke generalized gradient* (see Clarke [3] for definition). In fact, if ϕ is Lipschitz continuous near $\bar{x}$, we have $\partial\phi(\bar{x}) = \text{clco}\hat{\partial}\phi(\bar{x})$, where ∂ and clcoA denote the Clarke generalized gradient and closed convex hull of set A, respectively. For the definition and more details of the precise relation between the limiting subgradient and the Clarke generalized gradient, the reader is referred to Clarke [4] and Rockafellar [13].

The following result concerning the compactness of trajectories of a differential inclusion is slightly different from [3, Theorem 3.1.7] and will be used repeatedly. We omit the proof here since it can be proved similarly to [3, Theorem 3.1.7].

PROPOSITION 2.1. *Let $\Gamma : [t_0, t_1] \times R^d \times R^n \to R^n \times R^n$ be a set-valued map. We suppose that Γ is integrably bounded (i.e., there exists an integrable function $k(t)$ such that $|v| \leq k(t) \forall v \in \Gamma(t, x, u)$) and that Γ is nonempty, compact, and convex. We suppose that for every $(t, x, u) \in [t_0, t_1] \times R^d \times R^n$ the set-valued map $t' \to \Gamma(t', x, u)$*

is measurable and $\forall [t_0, t_1] \times R^d \times R^n$*, the set-valued map* $(x', u') \to \Gamma(t, x', u')$ *is upper semicontinuous. Let* Γ *be* $\mathcal{L} \times \mathcal{B}$ *measurable, where* $\mathcal{L} \times \mathcal{B}$ *denotes the* σ*-algebra of subsets of* $[t_0, t_1] \times R^d \times R^n$ *generated by product sets* $M \times N$*, where* M *is a Lebesque measurable subset of* $[t_0, t_1]$ *and* N *is a Borel subset of* $R^d \times R^n$*.*

Let $\{x_i\}$ *be a sequence of arcs on* $[t_0, t_1]$ *and* $\{\zeta_i\}$ *be a sequence of functions in* $L^2([t_0, t_1], R^n)$ *satisfying*

(i) $(\dot{x}_i(t), \zeta_i(t)) \in \Gamma(t, x_i(t), u_i(t))$ *a.e.* $t \in [t_0, t_1]$,

(ii) $\zeta_i \to \zeta$ *weakly in* L^2,

(iii) $u_i \to u$ *in* L^2,

(iv) $\{x_i(t_0)\}$ *is bounded.*

Then there exists a subsequence of $\{x_i\}$ *that converges uniformly to an arc* x *such that*

$$(\dot{x}(t), \zeta(t)) \in \Gamma(t, x(t), u(t)) \qquad a.e.\ t \in [t_0, t_1].$$

To discuss generalized differentiability of the value function $V(u)$, we will need the following assumptions:

(A6) There exists $k(t) \in L^2([t_0, t_1], R)$ such that

$$|\phi| + |\nabla_{(x,u)}\phi| + |G| + |\nabla_{(x,u)}G| \le k(t) \quad \forall (t, x, u, v) \in [t_0, t_1] \times R^d \times U \times W(t).$$

(A7) For any $(t, x, u) \in [t_0, t_1] \times R^d \times R^n$ the set

$$\{(\phi(t, x, u, v), G(t, x, u, v) + r) : v \in W(t), r \ge 0\}$$

is convex.

(A8) $|\nabla_u \phi| \le M \quad \forall (t, x, u, v) \in [t_0, t_1] \times R^d \times U \times W(t)$, where $M > 0$ is a constant.

Remark 2.2. Assumption (A7) is standard in control theory to ensure the existence of an optimal control for the lower-level problem. In the case where this assumption is not satisfied, the standard procedure is to go for the relaxed control (see, e.g., [19] and [22]).

Let the Hamiltonian for $P_2(u)$ be the function defined by

$$H_2(t, x, u, p_2) := \sup\{p_2 \cdot \phi(t, x, u, v) - G(t, x, u, v) : v \in W(t)\}$$

and Y_u be the set of all optimal trajectories x to problem $P_2(u)$.

The following result gives the Lipschitz continuity of the value function and characterizes the generalized gradient of the value function. It extends the result of Clarke [5] to allow general *nonadditive* perturbations in both the dynamics and the objective function.

THEOREM 2.3. *Suppose that assumptions* (A1)–(A8) *hold. Then* V *is Lipschitz continuous near* u *and*

$$\partial V(u) \subset clco \cup_{x \in Y_u} \{\zeta : \exists \text{ arc } p_2 \text{ s.t. } (-\dot{p_2}, -\zeta, \dot{x}) \in \partial H_2(t, x, u, p_2) \text{ a.e.} \\ -p_2(t_1) \in \hat{\partial} g(x(t_1))\},$$

where ∂H_2 *denotes the Clarke generalized gradient with respect to* (x, u, p_2).

Before proving Theorem 2.3, we first give the following result.

LEMMA 2.4. *Let* u_i *be a sequence converging (in* L^2*) to* u *and let* (x_i, v_i) *be an admissible pair for* $P_2(u_i)$*. Then there exist a subsequence of* $\{x_i\}$ *converging* *uniformly to an arc x and a control v with (x, v) being an admissible pair for* $P_2(u)$ *such that*

$$J_2(x,u,v) \le \liminf J_2(x_i,u_i,v_i).$$

Proof. Let

$$\dot{y}_i(t) := G(t, x_i(t), u_i(t), v_i(t)).$$

Then

$$(\dot{x}_i(t), \dot{y}_i(t)) \in \Gamma(t, x_i(t), y_i(t), u_i(t)), \tag{2}$$

where

$$\Gamma(t,x,y,u) := \{(\phi(t,x,u,v), r) : G(t,x,u,v) \le r \le k(t)+1, v \in V(t)\}.$$

The proof can be reduced to an application of Proposition 2.1 by studying the differential inclusion (2). The essential fact in the reduction is Fillipov's lemma: an (extended) arc (x, y) satisfies the differential inclusion iff there is a control function v for x such that (x, v) is feasible for $P_2(u)$ and y satisfies $G(t,x,u,v) \le \dot{y} \le k(t)+1$. ∎

We now turn to the proof of the theorem. By (A5), $P_2(u)$ has an admissible pair. So $V(u)$ is finite. By Lemma 2.4, V is (strongly) lower semicontinuous.

Step 1. Let $u \in L^2([t_0,t_1], U)$ and $\zeta \in \partial^\pi V(u)$. Let (x, v) be a solution of $P_2(u)$ that exists by virtue of Lemma 2.4. Then by definition, for some $M > 0$ and $\forall\ u'$ near u (in the L^2 norm), we have

$$\begin{aligned} V(u') - \langle \zeta, u' \rangle + M\|u'-u\|_2^2 &\ge V(u) - \langle \zeta, u \rangle \\ &= \int_{t_0}^{t_1} G(t,x(t),u(t),v(t))dt + g(x(t_1)) - \int_{t_0}^{t_1} \langle \zeta(t), u(t) \rangle dt. \end{aligned}$$

Let (x', v') be an admissible pair for $P_2(u')$. Then

$$\begin{aligned} &\int_{t_0}^{t_1} G(t,x'(t),u'(t),v'(t))dt + g(x'(t_1)) - \int_{t_0}^{t_1} \langle \zeta(t), u'(t) \rangle dt + M\|u'-u\|_2^2 \\ &\qquad \ge \int_{t_0}^{t_1} G(t,x(t),u(t),v(t))dt + g(x(t_1)) - \int_{t_0}^{t_1} \langle \zeta(t), u(t) \rangle dt. \end{aligned}$$

Hence (x, u, v) is a solution of the following optimal control problem:

$$\begin{aligned} \min &\int_{t_0}^{t_1} [G(t,x'(t),u'(t),v'(t)) - \langle \zeta(t), u'(t) \rangle]dt + g(x'(t_1)) + M\|u'-u\|_2^2 \\ \text{s.t.}\ \ & \dot{x}'(t) = \phi(t,x'(t),u'(t),v'(t)) \qquad \text{a.e.,} \\ & x'(t_0) = x_0, \\ & v'(t) \in W(t) \qquad \text{a.e.,} \\ & u'(t) \in U(t) := \{u' \in R^n : |u' - u(t)| \le \epsilon\}. \end{aligned}$$

Applying Theorem 5.2.1 of Clarke [3] with the Clarke generalized gradient replaced by the limiting subgradient in the transversality conditions (cf. [4, 10, 9]) to the above optimal control problem with free end points leads to the existence of an arc p_2 such that

$$(3) \qquad -\dot{p_2}(t) = \nabla_x \phi(t, x(t), u(t), v(t))^\top p_2(t) - \nabla_x G(t, x(t), u(t), v(t)) \qquad \text{a.e.,}$$

$$\max_{u \in U(t), v \in W(t)} \{p_2(t) \cdot \phi(t, x(t), u, v) - G(t, x(t), u, v) + \langle \zeta(t), u \rangle\}$$

$$(4) \qquad = p_2(t) \cdot \phi(t, x(t), u(t), v(t)) - G(t, x(t), u(t), v(t)) + \langle \zeta(t), u(t) \rangle \text{ a.e.,}$$

$$(5) \qquad -p_2(t_1) \in \hat{\partial} g(x(t_1)),$$

where $^\top$ denotes the transpose. Equation (4) implies that

$$\max_{v \in W(t)} \{p_2(t) \cdot \phi(t, x(t), u(t), v) - G(t, x(t), u(t), v)\}$$
$$= p_2(t) \cdot \phi(t, x(t), u(t), v(t)) - G(t, x(t), u(t), v(t)) \text{ a.e.}$$

and

$$(6) \qquad -\zeta(t) = \nabla_u \phi(t, x(t), u(t), v(t))^\top p_2(t) - \nabla_u G(t, x(t), u(t), v(t)).$$

Step 2. For any $\zeta \in \hat{\partial} V(u)$ by definition $\zeta = \text{weak}\lim_{i \to \infty} \zeta_i$, where $\zeta_i \in \partial^\pi V(u_i)$, $u_i \to u$ in L^2 and $V(u_i) \to V(u)$. By Step 1, for each u_i there exists an arc p_2^i and an arc x_i that solves $P_2(u_i)$ (along with v_i) such that

$$(7) \qquad -\dot{p_2}^i(t) = \nabla_x \phi(t, x_i(t), u_i(t), v_i(t))^\top p_2^i(t) - \nabla_x G(t, x_i(t), u_i(t), v_i(t)) \quad \text{a.e.,}$$

$$\max_{v \in W(t)} \{p_2^i(t) \cdot \phi(t, x_i(t), u_i(t), v) - G(t, x_i(t), u_i(t), v)\}$$

$$(8) \qquad = p_2^i(t) \cdot \phi(t, x_i(t), u_i(t), v_i(t)) - G(t, x_i(t), u_i(t), v_i(t)) \text{ a.e.,}$$

$$(9) \qquad -\zeta_i(t) = \nabla_u \phi(t, x_i(t), u_i(t), v_i(t))^\top p_2^i(t) - \nabla_u G(t, x_i(t), u_i(t), v_i(t)),$$

$$(10) \qquad -p_2^i(t_1) \in \hat{\partial} g(x_i(t_1)).$$

By [3, Theorem 2.8.2], (7), (8), and (9) imply that

$$(11) \qquad (-\dot{p}_2^i(t), -\zeta_i(t), \dot{x}_i(t)) \in \partial H_2(t, x_i(t), u_i(t), p_2^i(t)) \qquad \text{a.e.}$$

From (7)

$$p_2^i(t) = p_2^i(t_1) - \int_{t_1}^t [\nabla_x \phi(s, x_i(s), u_i(s), v_i(s))^\top p_2^i(s) - \nabla_x G(s, x_i(s), u_i(s), v_i(s))] ds.$$

By assumption (A4) and inclusion (10), the norm of $p_2^i(t_1)$ is bounded by L_g. Assumption (A3) implies that the norms of $\nabla_x \phi$ and $\nabla_x G$ are bounded by the integrable function ψ. Thus

$$|p_2^i(t)| \le (L_g + \int_{t_0}^{t_1} \psi(s) ds) + \int_t^{t_1} \psi(s) |p_2^i(s)| ds$$
$$= K + \int_t^{t_1} \psi(s) |p_2^i(s)| ds,$$

where $K := L_g + \int_{t_0}^{t_1} \psi(s)ds$. Invoking Gronwall's inequality, we conclude that

$$|p_2^i(t)| \le K e^{\int_t^{t_1} \psi(s)ds},$$

which implies that $\|p_2^i\|_\infty$ is bounded. It follows that the set-valued map ∂H_2 is integrably bounded. Applying Proposition 2.1 to differential inclusion (11) with boundary condition (10), we conclude that there exists a convergent subsequence of $\{x_i, p_2^i\}$ that converges to the arcs x, p_2 such that

$$(-\dot{p_2}(t), -\zeta(t), \dot{x}(t)) \in \partial H_2(t, x(t), u(t), p_2(t)) \qquad \text{a.e.}$$

Note that by Lemma 2.4 we may suppose $x \in Y_u$ since x_i is an optimal trajectory of $P_2(u_i)$. From the upper semicontinuity of the limiting subgradients

$$-p_2(t_1) \in \hat{\partial} g(x(t_1)).$$

Therefore we conclude that

$$\hat{\partial} V(u) \subset \cup_{x \in Y_u} \{\zeta : \exists \text{ arc } p_2 \text{ s.t. } (-\dot{p_2}, -\zeta, \dot{x}) \in \partial H_2(t, x, u, p_2) \text{ a.e.}, -p_2(t_1) \in \hat{\partial} g(x(t_1))\}.$$

Step 3. To complete the proof of the theorem, we only have to show that V is Lipschitz near u. By [6, Theorem 3.6], V is Lipschitz near u of rank C iff

$$\sup\{\|\zeta\|_2 : \zeta \in \partial^\pi V(u')\} \le C \qquad \forall u' \text{ in a neighborhood of } u.$$

Indeed, by Step 1, for any u and any $\zeta \in \partial^\pi V(u)$ there exists an arc p_2 along with a solution (x, v) of $P_2(u)$ such that (3), (5), and (6) hold. Therefore

$$|\zeta(t)| \le M(|p_2(t)| + |\nabla_u G|). \tag{12}$$

Since $\forall$ such p_2, $\|p_2\|_\infty \le K e^{\int_{t_0}^{t_1} \psi(s)ds}$, it then follows from (12) that all $\zeta \in \partial^\pi V(u), \forall u \in L^2([t_0, t_1], U)$ are bounded in L^2. Hence V is Lipschitz continuous, and the proof of Theorem 2.3 is now complete. □

3. Necessary conditions for optimality. As in the static case (cf. [17, 18]), it is easy to show that the equivalent single-level optimal control problem $\widetilde{P}_1$ always has a nontrivial abnormal multiplier; i.e., there always exists (λ, r, p_1) not all equal to zero with $\lambda = 0$ satisfying (13), (14), (15), and (16). Hence the traditional technique of concluding the existence of a normal multiplier from the nonexistence of a nontrivial abnormal multiplier will not work for the bilevel dynamic problem, and the calmness is the right constraint qualification (see more discussion in [17, 18]). The purpose of this section is to derive a Kuhn–Tucker–type necessary optimality condition for the bilevel dynamic problem under a *calmness*-type constraint qualification. Our approach is to reformulate the original problem as an infinite-dimensional optimization problem and derive the desired result from the necessary optimality condition for this infinite-dimensional optimization problem. Formulation as an infinite-dimensional optimization problem takes care of the functional constraints. However, the usual Lagrange multiplier rule for infinite-dimensional optimization problems cannot be used here since the problem data are not Lipschitz in the control variable in the lower-level optimal control problem. Ioffe [8] derived a very general maximum principle for the standard optimal control problem by reduction to an infinite-dimensional optimization problem. We will use the result and approach of Ioffe to derive the necessary optimality condition of the maximum principle type for the bilevel dynamic problem.

DEFINITION 3.1. *Let* (u^*, v^*) *be an optimal strategy of* P_1 *(equivalently* $\widetilde{P}_1$*) and* x^* *the corresponding trajectory.* $\widetilde{P}_1$ *is said to be* partially calm *at* (x^*, u^*, v^*) *with modulus* $\mu \geq 0$ *if* $\forall (x, u, v)$ *satisfying*

$$\begin{aligned} &\dot{x}(t) = \phi(t, x(t), u(t), v(t)) \quad \text{a.e.,} \\ &x(t_0) = x_0, \\ &u(\cdot) \in L^2([t_0, t_1], U), v(\cdot) \in \mathcal{V}, \end{aligned}$$

we have

$$J_1(x, u, v) - J_1(x^*, u^*, v^*) + \mu(J_2(x, u, v) - V(u)) \geq 0,$$

where $\mathcal{V}$ *denotes the collection of all admissible control functions for* $P_2(u)$.

Define the *pseudo Hamiltonian* for problem $(\widetilde{P}_1)$ as

$$H_1(t, x, u, v, p_1; \lambda, r) := p_1 \cdot \phi(t, x, u, v) - rG(t, x, u, v) - \lambda F(t, x, u, v),$$

for $t \in [t_0, t_1]$, $x, p_1 \in \mathbb{R}^d$, $u \in R^n$, $v \in R^m$, $\lambda, r \in \mathbb{R}$.

THEOREM 3.2. *Assume that* (A1)–(A5) *hold. Let* (x^*, u^*, v^*) *be an optimal solution of* P_1. *Suppose that* $\widetilde{P}_1$ *is* partially calm *at* (x^*, u^*, v^*) *with modulus* $\mu \geq 0$. *Assume that the value function for the lower-level problem* V *is locally Lipschitz continuous near* u^*. *Then there exist* $\lambda > 0, r = \lambda\mu$, *and an arc* p_1 *such that*

$$-\dot{p_1}(t) = \nabla_x H_1(t, x^*(t), u^*(t), v^*(t), p_1(t); \lambda, r) \quad \text{a.e.,} \tag{13}$$

$$\max_{v \in W(t)} H_1(t, x^*(t), u^*(t), v, p_1(t); \lambda, r)$$

$$= H_1(t, x^*(t), u^*(t), v^*(t), p_1(t); \lambda, r) \quad \text{a.e.,} \tag{14}$$

$$-p_1(t_1) \in \lambda \partial f(x^*(t_1)) + r \partial g(x^*(t_1)), \tag{15}$$

$$\nabla_u H_1(\cdot, x^*(\cdot), u^*(\cdot), v^*(\cdot), p_1(\cdot); \lambda, r) \in -r\partial V(u^*) + N_{L^2([t_0, t_1], U)}(u^*). \tag{16}$$

Proof. Since $\widetilde{P}_1$ is partially calm at (x^*, u^*, v^*) with modulus μ, it is easy to see that (x^*, u^*, v^*) is also optimal for the following penalized problem:

$$\begin{aligned} P(\mu) \qquad &\min \; J_1(x, u, v) + \mu(J_2(x, u, v) - V(u)) \\ &\text{s.t.} \; \dot{x}(t) = \phi(t, x(t), u(t), v(t)) \quad \text{a.e.,} \\ &\quad x(t_0) = x_0, \\ &\quad u(\cdot) \in L^2([t_0, t_1], U), \quad v(t) \in W(t) \quad \text{a.e.,} \end{aligned}$$

which can be equivalently posed as the following problem:

$$\begin{aligned} \widehat{P}_1 \qquad &\min \; f(x(t_1)) + z(t_1) + \mu(g(x(t_1)) + y(t_1) - V(u)) \\ &\text{s.t.} \; \dot{x}(t) = \phi(t, x(t), u(t), v(t)) \quad \text{a.e.,} \\ &\quad \dot{y}(t) = G(t, x(t), u(t), v(t)) \quad \text{a.e.,} \\ &\quad \dot{z}(t) = F(t, x(t), u(t), v(t)) \quad \text{a.e.,} \\ &\quad v(t) \in W(t) \quad \text{a.e.,} \\ &\quad (x, y, z)(t_0) \in \{x_0\} \times \{0\} \times \{0\}. \end{aligned}$$

We now reformulate the above problem as an infinite-dimensional optimization problem. Let $C([t_0, t_1], R^n)$ be the space of continuous mappings from $[t_0, t_1]$ into R^n with the usual supremum norm. Set

$$\widetilde{x} := (x, y, z), \qquad \widetilde{\phi} := (\phi, G, F).$$

For $v(\cdot) \in \mathcal{V}$, the mapping $(\widetilde{x}(\cdot), u(\cdot)) \to F_0(\widetilde{x}(\cdot), u(\cdot), v(\cdot))$ from $X := C([t_0, t_1], R^{d+2}) \times L^2([t_0, t_1], U)$ into $Y := C([t_0, t_1], R^{d+2})$:

$$F_0(\widetilde{x}(\cdot), u(\cdot), v(\cdot))(t) := \widetilde{x}(t) - \widetilde{x}(t_0) + \int_{t_0}^{t} \widetilde{\phi}(s, \widetilde{x}(s), u(s), v(s))ds$$

is well defined, continuously differentiable in $\widetilde{x}(\cdot)$, and Lipschitz continuous in $u(\cdot)$. Finally, let

$$f_0(\widetilde{x}(\cdot)) := f(x(t_1)) + z(t_1), \tag{17}$$

$$G_0(\widetilde{x}(\cdot), u(\cdot)) := y(t_1) + g(x(t_1)) - V(u), \tag{18}$$

$$S := \{\widetilde{x} \subset Y : x(t_0) = x_0, y(t_0) = 0, z(t_0) = 0\}.$$

Then problem $\widehat{P}_1$ is equivalent to the following infinite-dimensional optimization problem:

$$\begin{array}{lrl} {P_1}' & \min & f_0(\widetilde{x}) + \mu G_0(\widetilde{x}, u) \\ & \text{s.t.} & F_0(\widetilde{x}, u, v) = 0, \\ & & (\widetilde{x}, u) \in S \times L^2([t_0, t_1], U), \\ & & v \in \mathcal{V}. \end{array}$$

The above problem is in the form of a very general problem in section 4 of Ioffe [8]. Let the Lagrangian of the above problem be

$$L(\lambda, \alpha, \widetilde{x}, u, v) := \lambda(f_0(\widetilde{x}) + \mu G_0(\widetilde{x}, u)) + \langle \alpha, F_0(\widetilde{x}, u, v)\rangle.$$

As in section 5 of Ioffe [8], the assumptions for [8, Theorem 2] can be verified. By [8, Theorem 2], if (x^*, u^*, v^*) is a local solution to ${P_1}'$, then there exist Lagrange multipliers $\lambda \geq 0$, $\alpha \in Y^*$ not all equal to zero such that

$$0 \in \partial_{(\widetilde{x},u)} L(\lambda, \alpha, \widetilde{x}^*, u^*, v^*) + N_S(\widetilde{x}^*) \times N_{L^2([t_0,t_1],U)}(u^*), \tag{19}$$

$$L(\lambda, \alpha, \widetilde{x}^*, u^*, v^*) = \min_{v \in \mathcal{V}} L(\lambda, \alpha, \widetilde{x}^*, u^*, v), \tag{20}$$

where Y^* denotes the space of continuous linear functions on Y. Since f_0, G_0 are separable functions of $(\widetilde{x}, u)$ (f_0 is independent of u and G_0 is the sum of a function independent of $\widetilde{x}$ and a function independent of u), by [15, Proposition 1.8], (19) implies that

$$\begin{aligned} 0 \in\ & \lambda \partial f_0(\widetilde{x}^*) \times \{0\} + (\lambda\mu \partial_{\widetilde{x}} G_0(\widetilde{x}^*, u^*)) \times (-\lambda\mu\partial V(u^*)) \\ & + \partial_{(\widetilde{x},u)} \langle \alpha, F_0(\widetilde{x}^*, u^*, v^*)\rangle + N_S(\widetilde{x}^*) \times N_{L^2([t_0,t_1],U)}(u^*). \end{aligned} \tag{21}$$

Notice that $\langle \alpha, F_0(\widetilde{x}, u, v)\rangle$ can be represented as an integral functional on $X \times L^2$ by

$$\begin{aligned}&\langle \alpha, F_0(\widetilde{x}, u, v)\rangle \\ &= \int_{t_0}^{t_1} \langle \widetilde{x}(s) - \widetilde{x}(t_0), \xi(s)\rangle d\mu - \int_{t_0}^{t_1} \left\langle \int_t^{t_1} \xi(\tau) d\mu, \widetilde{\phi}(t, \widetilde{x}(t), u(t), v(t)) \right\rangle dt,\end{aligned}$$

where the pair $(\mu, \xi(\cdot))$ represents the functional $\alpha \in Y^*$ (μ being a nonnegative Radon measure on $[t_0, t_1]$ and $\xi(\cdot) : [t_0, t_1] \to R^{d+2}$, μ-integrable); i.e.,

$$\int_{t_0}^{t_1} \langle \xi(t), y(t)\rangle d\mu = \langle \alpha, y(\cdot)\rangle \qquad \forall y(\cdot) \in Y.$$

Hence by Theorems 2.7.4 and 2.7.5 of [3] it is regular. Therefore, by [3, Proposition 2.3.15], (21) implies that

$$0 \in \lambda \partial f_0(\widetilde{x}^*) + \lambda\mu \partial_{\widetilde{x}} G_0(\widetilde{x}^*, u^*) + D_{\widetilde{x}} \langle \alpha, F_0(\widetilde{x}^*, u^*, v^*)\rangle + N_S(\widetilde{x}^*), \tag{22}$$

$$0 \in -\lambda\mu \partial V(u^*) + \partial_u \langle \alpha, F_0(\widetilde{x}^*, u^*, v^*)\rangle + N_{L^2([t_0,t_1],U)}(u^*), \tag{23}$$

where $D_{\widetilde{x}} \langle \alpha, F_0(\widetilde{x}, u, v)\rangle$ denotes the Gâteaux derivative of the functional $\langle \alpha, F_0(\widetilde{x}, u, v)\rangle$ with respect to $\widetilde{x}$.

Now let us analyze (22). We have that $\partial f_0(\widetilde{x}(\cdot))$ contains those $\beta \in Y^*$ that can be represented in the form

$$\langle \beta, h(\cdot)\rangle = \langle a, h(t_1)\rangle$$

for some $a \in \partial f(x(t_1)) \times \{0\} \times \{1\}$.

Similarly, $\partial_{\widetilde{x}} G_0(\widetilde{x}, u)$ contains those $\beta \in Y^*$ that can be represented in the form

$$\langle \beta, h(\cdot)\rangle = \langle b, h(t_1)\rangle$$

for some $b \in \partial g(x(t_1)) \times \{1\} \times \{0\}$.

Let $p(t) := \int_t^{t_1} \xi(\tau) d\mu$. Then p is an arc. For any $h \in X$,

$$\begin{aligned}\langle D_{\widetilde{x}} \langle \alpha, F_0(\widetilde{x}, u, v)\rangle, h(\cdot)\rangle &= \int_{t_0}^{t_1} \langle h(t) - h(t_0), \xi(t)\rangle d\mu \\ &\quad - \int_{t_0}^{t_1} \langle \nabla_{\widetilde{x}} \widetilde{\phi}(t, \widetilde{x}(t), u(t), v(t))^\top p(t), h(t)\rangle dt.\end{aligned}$$

$N_S(\widetilde{x})$ contains those $\beta \in Y^*$ that can be represented in the form

$$\langle \beta, h(\cdot)\rangle = \langle c, h(t_0)\rangle$$

for some $c \in N_{\{x_0\}\times\{0\}\times\{0\}}(\widetilde{x}(t_0))$.

Inclusion (22) yields the existence of

$$a \in \partial f(x^*(t_1)) \times \{0\} \times \{1\}, b \in \partial g(x^*(t_1)) \times \{1\} \times \{0\}, c \in N_{\{x_0\}\times\{0\}\times\{0\}}(\widetilde{x}^*(t_0))$$

such that

$$\begin{aligned}0 &= \lambda \langle a, h(t_1)\rangle + \lambda\mu \langle b, h(t_1)\rangle + \int_{t_0}^{t_1} \langle h(t) - h(t_0), \xi(t)\rangle d\mu \\ &\quad - \int_{t_0}^{t_1} \langle \nabla_{\widetilde{x}} \widetilde{\phi}(t, \widetilde{x}^*(t), u^*(t), v^*(t))^\top p(t), h(t)\rangle dt + \langle c, h(t_0)\rangle \qquad \forall h \in X.\end{aligned}$$

Let us denote $h = (h_1, h_2, h_3)$, $\xi = (\xi_1, \xi_2, \xi_3)$, $p = (p_1, p_2, p_3)$, where subscript 1 corresponds to vectors in R^d and subscripts 2, 3 to vectors in R. In particular, if we choose $h(\cdot)$ that are absolutely continuous with $h(t_0) = 0, h_i(\cdot) = 0$ for $i = 1, 3$, we have

$$0 = \lambda\mu h_2(t_1) + \int_{t_0}^{t_1} h_2(t)\xi_2(t)d\mu,$$

which is equal to

$$0 = \int_{t_0}^{t_1} \left(\int_t^{t_1} \xi_2(s)d\mu + \lambda\mu \right) dh_2(t),$$

which implies that $p_2(t) = -\lambda\mu$.

Similarly, if we choose $h(\cdot)$ that are absolutely continuous with $h(t_0) = 0, h_i(\cdot) = 0$ for $i = 1, 2$, we have

$$0 = \lambda h_3(t_1) + \int_{t_0}^{t_1} h_3(t)\xi_3(t)d\mu,$$

which implies that $p_3(t) = -\lambda$.

If we choose $h(\cdot)$ that are absolutely continuous with $h(t_0) = 0, h_i(\cdot) = 0$ for $i = 2, 3$, we have

$$0 = \lambda\langle a_1, h_1(t_1)\rangle + \lambda\mu\langle b_1, h_1(t_1)\rangle + \int_{t_0}^{t_1} \langle h_1(t), \xi_1(t)\rangle d\mu$$
$$- \int_{t_0}^{t_1} \langle \nabla_x \widetilde{\phi}(t, \widetilde{x}^*(t), u^*(t), v^*(t))^\top p(t), h_1(t)\rangle dt.$$

Setting $-q = \lambda a_1 + \lambda\mu b_1$ and changing the order of integration, we obtain

$$0 = \int_{t_0}^{t_1} \Bigg\langle \int_t^{t_1} \xi_1(t)d\mu + \nabla_x\phi(t, x^*(t), u^*(t), v^*(t))^\top p_1(t)$$
$$- \lambda\mu\nabla_x G(t, x^*(t), u^*(t), v^*(t)) - \lambda\nabla_x F(t, x^*(t), u^*(t), v^*(t)) - q, k(t) \Bigg\rangle dt,$$

where $k(t) = \dot{h}(t)$ is an arbitrary integrable mapping. In view of the definition of $p_1(t)$, this implies

$$p_1(t) - q = -\int_t^{t_1} (\nabla_x\phi(s, x^*(s), u^*(s), v^*(s))^\top p_1(s)$$
$$+\lambda\mu\nabla_x G(s, x^*(s), u^*(s), v^*(s)) + \lambda\nabla_x F(s, x^*(s), u^*(s), v^*(s)))ds,$$

from which we derive (13).

Let us now analyze (23). Since $\langle \alpha, F_0(\widetilde{x}, u, v)\rangle$ is an integral functional of u on L^2, it is not Gâteaux differentiable. However, under our assumptions, [3, Theorem 2.7.5] applies. Therefore, for $\beta \in \partial_u\langle \alpha, F_0(\widetilde{x}^*, u^*, v^*)\rangle$,

$$\langle \beta, h(\cdot)\rangle = -\int_{t_0}^{t_1} \langle \nabla_u \widetilde{\phi}(t, \widetilde{x}^*(t), u^*(t), v^*(t))^\top p(t), h(t)\rangle dt$$

for any $h \in L^2([t_0, t_1], R^n)$. Hence (23) implies (16).

We also have

$$p(t_1) = q \in -\lambda\partial f(x^*(t_1)) - \lambda\mu\partial g(x^*(t_1)).$$

That is (15).

Equation (20) implies that

$$-\int_{t_0}^{t_1} \langle p(t), \widetilde{\phi}(t, \widetilde{x}^*(t), u^*(t), v^*(t))\rangle dt \le -\int_{t_0}^{t_1} \langle p(t), \widetilde{\phi}(t, \widetilde{x}^*(t), u^*(t), v(t))\rangle dt.$$

Since $-\lambda = p_2(t), \lambda\mu = -p_3(t)$, the above inequality implies that

$$\int_{t_0}^{t_1} H_1(t, x^*(t), u^*(t), v^*(t), p_1(t); \lambda, \lambda\mu)dt \ge \int_{t_0}^{t_1} H_1(t, x^*(t), u^*(t), v(t), p_1(t); \lambda, \lambda\mu)dt$$

for any $v(\cdot) \in \mathcal{V}$. Since for any measurable set $E \subset [t_0, t_1]$,

$$v(\cdot) = \chi_E(\cdot)v(\cdot) + (1 - \chi_E(\cdot))v^*(\cdot),$$

where χ_E denotes the characteristic function of E, and belongs to $\mathcal{V}$ whenever $v(\cdot) \in \mathcal{V}$, it follows that

$$H_1(t, x^*(t), u^*(t), v^*(t), p_1(t); \lambda, \lambda\mu) \ge H_1(t, x^*(t), u^*(t), v(t), p_1(t); \lambda, \lambda\mu) \text{ a.e.}$$

for any $v(\cdot) \in \mathcal{V}$. From measurable selection theory, (14) follows.

Now we need to show that $\lambda \neq 0$. From the fact that λ and α are not all equal to zero, it follows easily that

$$\|p_1\|_\infty + \lambda > 0. \tag{24}$$

This condition prevents λ becoming zero. Indeed if $\lambda = 0$, then the transversality condition (15) would imply that $p_1(t_1) = 0$. This in turn implies that $p_1 \equiv 0$, which contradicts (24). The proof of the theorem is now complete. □

Combining Theorems 3.2 and 2.3, the following Kuhn–Tucker–type necessary optimality condition for the general bilevel dynamic problem is obtained.

THEOREM 3.3. *Assume* (A1)–(A8) *hold. Let* $(u^*(t), v^*(t))$ *be an optimal strategy of the bilevel dynamic problem* P_1 *and* $x^*(t)$ *the corresponding optimal trajectory. Suppose that* $\widetilde{P}_1$ *is partially calm at* (x^*, u^*, v^*) *with modulus* $\mu \ge 0$. *Then there exists arc* p_1 *such that*

$$-\dot{p_1}(t) = \nabla_x H_1(t, x^*(t), u^*(t), v^*(t), p_1(t); 1, \mu), \tag{25}$$

$$\max_{v \in W(t)} H_1(t, x^*(t), u^*(t), v, p_1(t); 1, \mu) = H_1(t, x^*(t), u^*(t), v^*(t), p_1(t); 1, \mu) \text{ a.e.}, \tag{26}$$

$$-p_1(t_1) \in \partial f(x^*(t_1)) + \mu\partial g(x^*(t_1)), \tag{27}$$

$$\begin{aligned}\nabla_u H_1(\cdot, x^*(\cdot), u^*(\cdot), v^*(\cdot), p_1(\cdot); 1, \mu) &\in \mu\, clco \cup_{x \in Y_{u^*}} \{\zeta : \exists \text{ arc } p_2 \text{ s.t. } (-\dot{p_2}, \zeta, \dot{x}) \in \partial H_2(t, x, u^*, p_2) \quad a.e., \\ &\qquad -p_2(t_1) \in \hat{\partial} g(x(t_1))\} \\ &\quad + N_{L^2([t_0,t_1],U)}(u^*).\end{aligned} \tag{28}$$

It is clear that in the minimax case (i.e., when $J_1 = -J_2$) and the trivial case (i.e., when $J_1 = J_2$), the calmness condition always holds with $\mu = 1$ and $\mu = 0$, respectively. We now give an example that satisfies the partial calmness condition.

Example 1. Consider the following bilevel dynamic problem:

$$\begin{aligned} &\min\ (x_1(1))^2 + (x_2(1))^2 \\ &\text{s.t. } u(t) \geq 0, (x, v) \in S(u), \end{aligned}$$

where $S(u)$ is the solution set of

$$\begin{aligned} &\min\ (x_1(1) + x_2(1))^3 \\ &\text{s.t. } \dot{x}_1(t) = u(t), \\ &\qquad \dot{x}_2(t) = v(t),\ \ v(t) \geq 0, \\ &\qquad x_1(0) = x_2(0) = 0. \end{aligned}$$

The solution of the above problem is $x^* = 0, u^* = 0, v^* = 0$. Since $J_1(x, u, v) \geq 0$ for all (x, u, v) that are admissible for $P(\mu)$ and $J(x^*, u^*, v^*) = 0$, it is easy to see that the above problem is partially calm.

As seen in Example 1, the calmness condition depends on knowledge of the optimal value of the dynamic bilevel problem. It is therefore important to find sufficient conditions for the calmness condition. For the static case, [17] identifies the existence of a uniformly weak sharp minimum as a sufficient condition for the calmness. It is shown in that paper that the bilevel programming problem in which the lower-level problem is linear is always calm, and sufficient conditions for the calmness of the bilevel problem where the lower-level problem is a linear quadratic problem are given.

To extend the definition of a uniform weak sharp minimum to our dynamic setting, we introduce the following notation. Given u, a control function for the upper level, let $\Omega(u)$ denote

$$\Omega(u) = \{(x, v) \in C([t_0, t_1], R^d) \times \mathcal{V} : \dot{x} = \phi(t, x, u, v), x(t_0) = x_0\}.$$

Let $S(u)$ denote the set of all solutions to problem $P_2(u)$. We say that the family of optimal control problems $\{P_2(u) : u \in L^2([t_0, t_1], U)\}$ has a uniformly weak sharp minimum with modulus $\alpha > 0$ if

$$d_{S(u)}(x, v) \leq \alpha(J_2(x, u, v) - V(u)) \qquad \forall (x, v) \in \Omega(u), u \in L^2([t_0, t_1], U),$$

where $d_{S(u)}(x, v)$ denotes the distance from (x, v) to the set $S(u)$. As in [17], we can show that a uniformly weak sharp minimum is a sufficient condition for partial calmness.

PROPOSITION 3.4. *In addition to* (A1) and (A7)*, assume that for any $u(\cdot)$ there exists $k(\cdot) \in L^1([t_0, t_1])$ such that*

$$|F(t, x', u(t), v') - F(t, x'', u(t), v'')| \leq k(t)\|(x', v') - (x'', v'')\| \qquad \forall t \in [t_0, t_1], x', x'' \in R^d, v', v'' \in R^m$$

and that f is Lipschitz continuous with constant $L_f > 0$. That $\{P_2(u) : u \in L^2([t_0, t_1], U)\}$ has a uniformly weak sharp minimum with modulus α implies that $\widetilde{P}_1$ is partially calm with modulus $\mu \geq \alpha(\|k\|_1 + L_f)$ at any solution of the problem.

Proof. By the definition of a uniformly weak sharp minimum, there exists $\alpha > 0$ such that $\forall\ (x, v) \in \Omega(u), u \in L^2([t_0, t_1], U)$,

$$\begin{aligned} J_2(x, u, v) - V(u) &\geq (1/\alpha) d_{S(u)}(x, v) \\ &= (1/\alpha)|(x, v) - (x(u), v(u))|, \end{aligned}$$

where $(x(u), v(u))$ is the metric projection of (x, v) onto the set $S(u)$. Let (x^*, u^*, v^*) be any solution of the problem P_1. The assumptions imply that $J_1(x, u, v)$ is Lipschitz continuous in (x, v) uniformly in u with constant $L_1 = \|k\|_1 + L_f$. It follows that

$$\begin{aligned} J_2(x, u, v) - V(u) &\geq \frac{1}{\alpha} d_{S(u)}(x, v) \\ &= \frac{1}{\alpha}|(x, v) - (x(u), v(u))| \\ &\geq \frac{1}{\alpha L_1}(J_1(x, u, v) - J_1(x(u), u, v(u))) \\ &\geq \frac{1}{\alpha L_1}(J_1(x, u, v) - J_1(x^*, u^*, v^*)) \\ &\geq \frac{1}{\mu}(J_1(x, u, v) - J_1(x^*, u^*, v^*)). \end{aligned}$$

Therefore, we see that $\widetilde{P}_1$ is partially calm at any solution of the problem with modulus $\mu \geq \alpha L_1$. ☐

The following result is a sufficient condition for a uniformly weak sharp minimum. The proof technique follows from a result about regular points due to Ioffe (Theorem 1 and Corollary 1.1 of [7]).

PROPOSITION 3.5. *Suppose that $J_2(x, u, v)$ is Lipschitz continuous in (x, v) uniformly in u with constant $L > 0$. If there exists a constant $c > 0$ such that $\|\xi + \eta\| \geq c$ whenever $\xi \in \partial_{(x,v)} J_2(x, u, v)$, $\eta \in (L + 1)\partial d_{\Omega(u)}(x, v)$ (or $\eta \in N_{\Omega(u)}(x, v)$)$\forall (x, v) \in \Omega(u)$ such that $(x, v) \notin S(u)$ $\forall$ admissible controls u, then*

$$d_{S(u)}(x, v) \leq (1/c)(J_2(x, u, v) - V(u)) \forall (x, v) \in \Omega(u).$$

Proof. Assume that the statement is false. Then there is $u \in L^2([t_0.t_1], U)$ and $(x, v) \in \Omega(u)$ such that

$$d_{S(u)}(x, v) > \frac{1}{c}(J_2(x, u, v) - V(u)).$$

We can obviously choose $\delta > 1$ to make the following inequality valid:

$$d_{S(u)}(x, v) > \frac{\delta}{c}(J_2(x, u, v) - V(u)) := \gamma. \tag{29}$$

It is also obvious that

$$J_2(x, u, v) - V(u) \leq \inf_{(x,v) \in \Omega(u)} (J_2(x, u, v) - V(u)) + \frac{c\gamma}{\delta}.$$

Let δ_S denote the indicator function of set S. Applying the Ekeland variational principle [3, Theorem 7.5.1] with $F(x', v') := J_2(x', u, v') - V(u) + \delta_{\Omega(u)}(x', v')$, $\epsilon = \gamma c/\delta$, and $\lambda = \gamma$, we find $(\widetilde{x}, \widetilde{v}) \in \Omega(u)$ such that

$$\|(\widetilde{x}, \widetilde{v}) - (x, v)\| \leq \gamma \tag{30}$$

and

$$\phi(x', v') := J_2(x', u, v') - V(u) + (c/\delta)\|(x', v') - (\widetilde{x}, \widetilde{v})\|$$

attains its minimum on $\Omega(u)$ at $(\widetilde{x}, \widetilde{v})$. It follows that

$$\begin{aligned} 0 &\in \partial\phi(\widetilde{x}, \widetilde{v}) + (L+1)\partial d_{S(u)}(\widetilde{x}, \widetilde{v}) \\ &\subset \partial_{(x,v)} J_2(\widetilde{x}, u, \widetilde{v}) + (c/\delta)B + (L+1)\partial d_{S(u)}(\widetilde{x}, \widetilde{v}). \end{aligned}$$

Thus there exist

$$\xi \in \partial_{(x,v)} J_2(\widetilde{x}, u, \widetilde{v}), \quad \eta \in (L+1)\partial d_{S(u)}(\widetilde{x}, \widetilde{v})$$

such that

$$\|\xi + \eta\| \leq c/\delta < c. \tag{31}$$

According to (29), (30), and $(\widetilde{x}, \widetilde{v}) \in \Omega(u)$, we have that

$$(\widetilde{x}, \widetilde{v}) \notin S(u).$$

Therefore (31) contradicts the assumption. The proof of the proposition is then complete. ◻

We now use an example to illustrate the application of the above result. It is different from Example 1 only in the lower-level objective function.

Example 2. Consider the following bilevel dynamic problem:

$$\begin{aligned} &\min\ (x_1(1))^2 + (x_2(1))^2 \\ &\text{ s.t. } u(t) \geq 0, (x, v) \in S(u), \end{aligned}$$

where $S(u)$ is the solution set of

$$\begin{aligned} \min\ & x_1(1) + x_2(1) + (x_1(1) + x_2(1))^3 \\ \text{s.t. } & \dot{x}_1(t) = u(t), \\ & \dot{x}_2(t) = v(t),\ v(t) \geq 0, \\ & x_1(0) = x_2(0) = 0. \end{aligned}$$

It is easy to see that $\Omega(u) = \{x_1 : x_1(t) = \int_0^t u(s)ds\} \times \{x_2 : x_2(t) \geq 0\} \times \{v : v(t) \geq 0\}$ and $S(u) = \{(x, v) : x_1(t) = \int_0^t u(s)ds, x_2 \equiv 0, v \equiv 0\}\ \forall\ u(t) \geq 0$. Since

$$\begin{aligned} \partial_{(x,v)} J_2(x, u, v) = \{(\xi_1, \xi_2, 0) :&\langle \xi_1, h(\cdot)\rangle = ((1 + 3(x_1(1) + x_2(1))^2)h(1), \\ &\langle \xi_2, h(\cdot)\rangle = (1 + 3(x_1(1) + x_2(1))^2)h(1)\ \forall\ h \in C[0,1]\}, \end{aligned}$$

and $N_{\Omega(u)}(x, v) = N_{\{x_1 : x_1(t) = \int_0^t u(s)ds\}}(x_1) \times \{0\} \times N_{\{v : v(t) \geq 0\}}(v)$ for any $(x_1, x_2, v) \notin S(u)$, it is easy to see that the assumptions in Proposition 3.5 are satisfied.

We now calculate that

$$H_1(t,x,u,v,p_1;1,\mu) = p_1^1 u + p_1^2 v, \qquad H_2(t,x,u,v,p_2) = \sup\{p_2^1 u + p_2^2 v : v \geq 0\}.$$

Since H_2 is independent of x, (28) implies that there exists an arc p_2 such that

$$\dot{p}_2(t) = 0, \tag{32}$$

$$-p_2(1) = (1 + 3(x_1^*(1) + x_2^*(1))^2, 1 + 3(x_1^*(1) + x_2^*(1))^2), \tag{33}$$

$$p_1^1 - \mu p_2^1 \in N_{\{u \in C[0,1]: u \geq 0\}}(u^*). \tag{34}$$

Observing that $x_1(0) = x_2(0) = 0$ and both $x_1(t)$ and $x_2(t)$ are nondecreasing, we derive from (32) and (33) that $p_2^1 = p_2^2 \leq -1$ are constants. Hence $H_2 = p_2^1 u$, which implies from (28) that $\dot{x}_2^*(t) = 0$. That is $x_2^* \equiv 0$. If $u^* \not\equiv 0$ then (34) implies that

$$p_1^1 = \mu p_2^1.$$

But by (25) and (27), p_1^1 and p_1^2 are nonpositive constants and

$$-2x_1^*(1) - \mu(1 + 3[x_1^*(1) + x_2^*(1)]^2) = -\mu(1 + 3[x_1^*(1) + x_2^*(1)]^2),$$

which implies that $x_1^*(1) = 0$. But this is a contradiction. Therefore $u^* \equiv 0, v^* \equiv 0$ is a candidate for an optimal solution since $x_1(0) = x_2(0) = 0$ and both $x_1(t)$ and $x_2(t)$ are nondecreasing. It is not hard to check that it is indeed a solution. Notice that in Example 2 the lower-level problem is not convex and hence is out of the scope of any currently available control theory.

Finally, we use another example to illustrate applications of Theorem 3.2 in solving bilevel dynamic problems in the absence of the calmness condition. Example 3 shows that even without the calmness condition, the necessary condition that we derived may be used to find condition for the existence of a normal multiplier.

Example 3. Consider the following bilevel dynamic problem with linear-quadratic cost functions on the interval $[0,1]$, where

$$F(t,x,u,v) = \frac{1}{2}[x^\top Q_1 x + u^\top R_{11} u + v^\top R_{12} v],$$

$$f(x) = \frac{1}{2} x^\top K_1 x,$$

$$G(t,x,u,v) = \frac{1}{2}[x^\top Q_2 x + u^\top R_{21} u + v^\top R_{22} v],$$

$$g(x) = \frac{1}{2} x^\top K_2 x,$$

$$\phi(t,x,u,v) = A(t)x + B(t)u + C(t)v,$$

where $x \in R^d$, $u \in R^n$, $v \in R^m$, Q_1, Q_2, K_1, K_2 are positive semidefinite matrices and $R_{22}, R_{11}, rR_{22} + R_{12}$, where $r \geq 0$ is any constant, are positive definite matrices with appropriate order; R_{21} is a $n \times n$ matrix; $A(t), B(t)$, and $C(t)$ are matrices with continuous components.

We can calculate

$$\begin{aligned} H_1(t,x,u,v,p_1;1,\mu) &= p_1^\top \phi - \mu G - F \\ &= p_1^\top (A(t)x + B(t)u + C(t)v) \\ &\quad -\frac{1}{2}\mu[x^\top Q_2 x + u^\top R_{21} u + v^\top R_{22} v] \\ &\quad\; -\frac{1}{2}[x^\top Q_1 x + u^\top R_{11} u + v^\top R_{12} v], \\ H_2(t,x,u,p_2) &= \sup_v \{p_2^\top \phi - G\} \\ &= \sup_v \{p_2^\top (A(t)x + B(t)u + C(t)v) - \frac{1}{2}[x^\top Q_2 x + u^\top R_{21} u + v^\top R_{22} v]\} \\ &= p_2^\top (A(t)x + B(t)u + C(t)R_{22}^{-1} C(t)^\top p_2) \\ &\quad\; -\frac{1}{2}[x^\top Q_2 x + u^\top R_{21} u + p_2^\top C(t) R_{22}^{-1} C(t)^\top p_2]. \end{aligned}$$

Suppose that (u^*, v^*) is an optimal control pair and x^* is the corresponding trajectory. If the conclusion of Theorem 3.3 holds, then there exist adjoint arcs p_1, p_2 and constant $\mu \geq 0$ such that

$$\begin{aligned} -\dot{p}_1 &= A(t)^\top p_1 - [\mu Q_2 + Q_1] x^*, \\ -p_1(1) &= [\mu K_2 + K_1] x^*(1), \\ -\dot{p}_2 &= A(t)^\top p_2 - Q_2^\top x^*, \\ -p_2(1) &= K_2 x^*(1), \end{aligned}$$

$$\dot{x}^* = A(t)x^* + B(t)u^* + C(t)v^*, \tag{35}$$

$$B(t)^\top p_1 - R_{11} u^* = \mu B(t)^\top p_2, \tag{36}$$

$$v^*(t) = R_{22}^{-1} C(t)^\top p_2 = [\mu R_{22} + R_{12}]^{-1} C(t)^\top p_1. \tag{37}$$

Equation (36) implies that

$$u^*(t) = R_{11}^{-1} B(t)^\top (p_1 - \mu p_2). \tag{38}$$

Substituting (37) and (38) into (35) yields

$$\dot{x}^* = A(t)x^* + B(t)R_{11}^{-1} B(t)^\top (p_1 - \mu p_2) + C(t) R_{22}^{-1} C(t)^\top p_2. \tag{39}$$

Thus, for any $\mu \geq 0$ that satisfies (37), i.e.,

$$R_{22}^{-1} C(t)^\top p_2 = [\mu R_{22} + R_{12}]^{-1} C(t)^\top p_1, \tag{40}$$

we obtain a set of $3d$ equations with equal numbers of unknowns.

$$\begin{aligned} \dot{x}^* &= A(t)x^* + B(t)R_{11}^{-1} B(t)^\top (p_1 - \mu p_2) + C(t) R_{22}^{-1} C(t)^\top p_2, \\ -\dot{p}_1 &= A(t)^\top p_1 - [\mu Q_2^\top + Q_1^\top] x^*, \\ -\dot{p}_2 &= A(t)^\top p_2 - Q_2^\top x^*. \end{aligned}$$

Assume that

$$p_1(t) = \psi_1(t)x^*(t),$$
$$p_2(t) = \psi_2(t)x^*(t),$$

where ψ_i are matrices that satisfy the end point conditions

$$\psi_1(1) = -[K_1 + \mu K_2], \quad \psi_2(1) = -K_2$$

to be determined. Differentiating them with respect to t gives

$$\dot{p}_1 = \dot{\psi}_1 x^*(t) + \psi_1 \dot{x}^*(t),$$
$$\dot{p}_2 = \dot{\psi}_2 x^*(t) + \psi_2 \dot{x}^*(t).$$

Substituting for $\dot{x}^*, \dot{p}_1$, and $\dot{p}_2$ gives

$$\dot{\psi}_1 = -A(t)^\top \psi_1 - \psi_1 A(t) + \mu Q_2 + Q_1 - \psi_1 B(t) R_{11}^{-1} B(t)^\top (\psi_1 - \mu\psi_2)$$
$$-\psi_1 C(t) R_{22}^{-1} C(t)^\top \psi_2,$$
$$\dot{\psi}_2 = -A(t)^\top \psi_2 - \psi_2 A(t) + Q_2 - \psi_2 B(t) R_{11}^{-1} B(t)^\top (\psi_1 - \mu\psi_2) - \psi_2 C(t) R_{22}^{-1} C(t)^\top \psi_2.$$

Moreover,

$$u^*(t) = R_{11}^{-1} B(t)^\top (\psi_1(t) - \mu\psi_2(t)) x(t),$$
$$v^*(t) = R_{22}^{-1} C(t)^\top \psi_2(t) x(t).$$

Let $\psi_3 = \psi_1 - \mu\psi_2$. Then provided that there exists $\mu \geq 0$ that satisfies

$$R_{22}^{-1} C(t)^\top \psi_2 = [\mu R_{22} + R_{12}]^{-1} C(t)^\top (\psi_3 + \mu\psi_2) \tag{41}$$

we obtain

$$u^*(t) = R_{11}^{-1} B(t)^\top \psi_3(t) x(t), \tag{42}$$
$$v^*(t) = R_{22}^{-1} C(t)^\top \psi_2(t) x(t), \tag{43}$$

where ψ_3 and ψ_2 are solutions to

$$\dot{\psi}_3 = -A(t)^\top \psi_3 - \psi_3 A(t) + Q_1 - \psi_3 B(t) R_{11}^{-1} B(t)^\top \psi_3 - \psi_3 C(t) R_{22}^{-1} C(t)^\top \psi_2,$$
$$\dot{\psi}_2 = -A(t)^\top \psi_2 - \psi_2 A(t) + Q_2 - \psi_2 B(t) R_{11}^{-1} B(t)^\top \psi_3 - \psi_2 C(t) R_{22}^{-1} C(t)^\top \psi_2,$$

with end point conditions

$$\psi_3(1) = -K_1, \quad \psi_2(1) = -K_2.$$

It is clear that the existence of $\mu \geq 0$ that satisfies the equality (41) is a constraint qualification for ensuring the existence of normal multipliers for the class of linear-quadratic bilevel problems. Such $\mu \geq 0$ exists, for example, when

$$K_1 = 0, \quad Q_1 = 0, \quad R_{12} = 0$$

or

$$K_1 = K_2 = 0, \qquad Q_1 = Q_2 = 0.$$

Acknowledgments. The author would like to thank Qiji Zhu for suggestions which led to improvements in Theorem 2.3 of this paper. The author would also like to thank anonymous referees for comments on an early version of this paper that helped to improve the exposition.

REFERENCES

[1] J. F. Bard and J. E. Falk, *An explicit solution to the multi-level programming problem*, Oper. Res., 9 (1982), pp. 77–100.

[2] C. I. Chen and J. B. Cruz Jr., *Stackelberg solution for two-person games with biased information patterns*, IEEE Trans. Automat. Control, 6 (1972), pp. 791–798.

[3] F. H. Clarke, *Optimization and Nonsmooth Analysis*. Wiley-Interscience, New York, 1983.

[4] F. H. Clarke, *Methods of Dynamic and Nonsmooth Optimization*, NSF-CBMS Regional Conference Series in Applied Mathematics 57, SIAM, Philadelphia, PA, 1989.

[5] F. H. Clarke, *Perturbed optimal control problems*, IEEE Trans. Automat. Control, 6 (1986), pp 535–542.

[6] F. H. Clarke, R. J. Stern, and P. R. Wolenski, *Subgradient criteria for monotonicity, the Lipschitz condition and convexity*, Canad. J. Math., 45 (1993), pp. 1167–1183.

[7] A. D. Ioffe, *Regular points of Lipschitz functions*, Trans. Amer. Math. Soc., 251 (1979), pp. 61–69.

[8] A. D. Ioffe, *Necessary conditions in nonsmooth optimization*, Math. Oper. Res., 9 (1984), pp. 159–189.

[9] A. Y. Kruger and B. S. Mordukhovich, *Minimization of nonsmooth functionals in optimal control problems*, Engrg. Cybernetics, 16 (1978), pp. 126–133.

[10] B. S. Mordukhovich, *Maximum principle in problems of time optimal control with nonsmooth constraints*, J. Appl. Math. Mech., 40 (1976), pp. 960–969.

[11] J. Outrata, *A note on the usage of nondifferentiable exact penalties in some special optimization problems*, Kybernetika, 24 (1988), pp. 251–258.

[12] J. Outrata, *On the numerical solution of a class of Stackelberg problems*, Z. Oper. Res., 34 (1990), pp. 255–277.

[13] R. T. Rockafellar, *Extensions of subgradient calculus with applications to optimization*, Nonlinear Anal., 9 (1985), pp. 665–698.

[14] H. von Stackelberg, *The Theory of the Market Economy*, Oxford University Press, Oxford, UK, 1952.

[15] J. J. Ye, *Optimal Control of Piecewise Deterministic Markov Processes*, Ph.D. thesis, Department of Mathematics and Statistics, Dalhousie University, Halifax, Canada, 1990.

[16] J. J. Ye, *Necessary conditions for bilevel dynamic optimization problems*, SIAM J. Control Optim., 33 (1995), pp. 1208–1223.

[17] J. J. Ye and D. L. Zhu, *Optimality conditions for bilevel programming problems*, Optimization, 33 (1995), pp. 9–27.

[18] J. J. Ye, D. L. Zhu, and Q. J. Zhu, *Exact penalization and necessary optimality conditions for generalized bilevel programming problems*, SIAM J. Optim., 7 (1997), to appear.

[19] L. C. Young, *Lectures on the Calculus of Variations and Optimal Control Theory*, W. B. Saunders, Philadelphia, PA, 1969.

[20] R. Zhang, *Problems of Hierarchical Optimization: Nonsmoothness and Analysis of Solutions*, Ph.D. thesis, Department of Applied Mathematics, University of Washington, Seattle, WA, 1990.

[21] R. Zhang, *Problems of hierarchical optimization in finite dimensions*, SIAM J. Optim., 4 (1994), pp. 521–536.

[22] J. Warga, *Optimal Control of Differential and Functional Equations*, Academic Press, New York, 1972.