SIAM J. CONTROL AND OPTIMIZATION
Vol. 33, No. 4, pp. 1208–1223, July 1995

013

NECESSARY CONDITIONS FOR BILEVEL DYNAMIC OPTIMIZATION PROBLEMS*

JANE J. YE†

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

Key words. necessary conditions, bilevel dynamic optimization problems, sensitivity analysis, nonsmooth analysis, value function

AMS subject classifications. 90D65, 49K40

1. Introduction.

Let us consider a two-level hierarchical system where the higher level (hereafter the "leader") and the lower level (hereafter the "follower") must find vectors $z \in Z$ and control functions $u(\cdot)$, respectively, to minimize their individual objective functions $J_1(z,u)$ and $J_2(z,u)$. The leader is assumed first to select his decision vector $z \in Z$ and the follower next to select his decision control function $u(\cdot) \in \mathcal{U}$, where Z is a nonempty subset of R^n and $\mathcal{U}$ is the set of admissible controls. Under these assumptions on the order of play, the game will proceed as follows. Given any decision vector $z \in Z$ chosen by the leader, the follower will select his decision control function $u_z(\cdot) \in \mathcal{U}$ (depending on the decision vector z chosen by the leader) to minimize his objective $J_2(z, u_z)$. Assume that the game is cooperative, i.e., if the follower's problem has several optimal controls for a given parameter z, then the follower allows the leader to choose which of them is actually used. Thus the leader chooses his optimal decision vector $z \in Z$ to minimize the leader's objective $J_1(z, u_z)$. In other words, given any decision vector $z \in Z$ chosen by the leader, the follower faces the ordinary (*single-level*) optimal control problem involving a parameter z:

$$\begin{aligned} P_2(z) \qquad \min J_2(z,u) &= \int_{t_0}^{t_1} G(t,x(t),z,u(t))dt + g(x(t_1)),\\ \text{s.t. } \dot{x}(t) &= \phi(t,x(t),z,u(t)) \qquad \text{a.e.},\\ x(t_0) &= x_0, \quad x(t_1) \in C_1,\\ u(t) &\in U(t) \qquad \text{a.e.}, \end{aligned}$$

while the leader faces the *bilevel dynamic optimization problem*:

$$P_1 \qquad \min J_1(z,u_z) = \int_{t_0}^{t_1} F(t,x_z(t),z,u_z(t))dt + f(x_z(t_1))$$

over $z \in Z$ and all optimal pairs (x_z, u_z) of $P_2(z)$.

* Received by the editors May 28, 1993; accepted for publication (in revised form) February 8, 1994. This research was supported by Natural Sciences and Engineering Research Council of Canada grant WFA 0123160.

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

The bilevel dynamic optimization problem has many applications in economics and management science. For instance, the leader may be the government that sets up the taxation policy z and the follower may be a company that seeks the optimal policy $u_z(t)$ in reaction to the government's taxation policy.

The bilevel static problem where both leader's and follower's decisions are vectors instead of control functions was first introduced by von Stackelberg [10] for an economic model. The bilevel dynamic problem where both leader's and follower's decisions are control functions was first considered by Chen and Cruz in [2]. The bilevel dynamic optimization problem studied in this paper is a special case of the bilevel dynamic problem as in Zhang [13]. Several names for bilevel (static or dynamic) optimization problems have been used in the literature, such as Stackelberg game, principal-agent problem, bilevel programming problem, and two-level hierarchical optimization problem. Most of the bilevel (static or dynamic) problems are attacked by reducing the lower-level problem through first-order necessary conditions (cf. Bard and Falk [1] and Zhang [13], [14] for the bilevel static problem and Zhang [13] for the bilevel dynamic problem). The reduction is equivalent if and only if the lower-level problem satisfies certain convexity assumptions since in this case the first-order necessary condition is also sufficient. Apart from the strong convexity assumption, the resulting optimality conditions of the above approach involve second-order (generalized in nonsmooth case [13]) 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.

The purpose of this paper is to provide first-order necessary conditions for problem P_1 under very general assumptions (in particular, without convexity assumptions on the lower-level problem).

Define the *value function of the lower-level optimal control problem* as an extended-valued function $V : Z \to \bar{R}$ defined by

$$V(z) := \inf \left\{ \begin{array}{ll} \int_{t_0}^{t_1} G(t, x(t), z, u(t))dt + g(x(t_1)) : & \dot{x}(t) = \phi(t, x(t), z, u(t)) \text{ a.e.} \\ & u(t) \in U(t) \qquad \text{a.e.} \\ & x(t_0) = x_0, \quad x(t_1) \in C_1 \end{array} \right\},$$

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 in the following single-level optimal control problem:

$$\begin{aligned} \widetilde{P}_1 \qquad \min J_1(z, u) = \int_{t_0}^{t_1} & F(t, x(t), z(t), u(t))dt + f(x(t_1)), \\ \text{s.t. } \dot{x}(t) &= \phi(t, x(t), z(t), u(t)) \qquad \text{a.e.}, \\ \dot{z}(t) &= 0, \\ x(t_0) &= x_0, \quad x(t_1) \in C_1, \\ u(t) &\in U(t) \qquad \text{a.e.}, \\ \int_{t_0}^{t_1} G(t, x(t), z(t), u(t))dt &+ g(x(t_1)) \le V(z(t_1)). \end{aligned}$$

The above problem is obviously equivalent to the original bilevel dynamic optimization problem P_1 and is a standard optimal control problem except that the endpoint constraints involve the value function V of the lower-level optimal control problem. In general V is not an explicit function of the problem data and is nonsmooth even in the case where all problem data are smooth functions. Recent developments in nonsmooth analysis allow us to study the generalized derivatives of the value function V and relate them to the multiplier sets for the lower-level optimal control problem, hence deriving a necessary condition for optimality. This approach was first used by Ye and Zhu [12] to derive first-order necessary conditions for the static bilevel optimization problem. The following basic assumptions are in force throughout this paper:

(A1) $Z \subset I\!\!R^n$ and C_1 are closed.

(A2) $U(t) : [t_0, t_1] \to I\!\!R^m$ is a nonempty compact-valued set-valued map. The graph of $U(t)$ (i.e., the set $\{(s, r) : s \in [t_0, t_1], r \in U(s)\}$), denoted by Gr$U$, is $\mathcal{L} \times \mathcal{B}$ measurable, where $\mathcal{L} \times \mathcal{B}$ denotes the σ-algebra of subsets of $[t_0, t_1] \times I\!\!R^m$ 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 $I\!\!R^m$.

(A3) There exists an integrable function k defined on $[t_0, t_1]$ such that for each $(t, u) \in$ GrU, the functions $\phi(t, \cdot, \cdot, u), F(t, \cdot, \cdot, u), G(t, \cdot, \cdot, u)$ are locally Lipschitz of rank $k(t)$. For each $(x, z) \in I\!\!R^d \times I\!\!R^n$, the functions $\phi(\cdot, x, z, \cdot) : [t_0, t_1] \times I\!\!R^m \to I\!\!R^d$, $F(\cdot, x, z, \cdot) : [t_0, t_1] \times I\!\!R^m \to I\!\!R$, $G(\cdot, x, z, \cdot) : [t_0, t_1] \times I\!\!R^m \to I\!\!R$ are $\mathcal{L} \times \mathcal{B}$ measurable.

(A4) The functions $f, g : I\!\!R^d \to I\!\!R$ are locally Lipschitz continuous.

(A5) For any $z \in Z$, $P_2(z)$ has an admissible pair (whose definition is given below).

A *control function* is a (Lebesgue) measurable selection $u(\cdot)$ for $U(\cdot)$, that is, a measurable function satisfying $u(t) \in U(t)$ a.e. $t \in [t_0, t_1]$. An *arc* is an absolutely continuous function. An *admissible pair* for $P_2(z)$ is a pair of functions $(x(\cdot), u(\cdot))$ on $[t_0, t_1]$ of which $u(\cdot)$ is a control function and $x(\cdot) : [t_0, t_1] \to I\!\!R^d$ is an arc that satisfies the differential equation $\dot{x}(t) = \phi(t, x(t), z, u(t))$ a.e., together with the initial condition $x(t_0) = x_0$ and the endpoint constraint $x(t_1) \in C_1$. The first and the second components of an admissible pair are called an *admissible trajectory* and an *admissible control*, respectively. A *solution* to problem $P_2(z)$ is an admissible pair that minimizes the value of the cost functional $J_2(z, u)$ over all admissible pairs. An *admissible strategy* for P_1 includes a vector $z \in Z$ and an optimal control u_z for $P_2(z)$. The strategy (z, u_z) is *optimal* for the bilevel dynamic optimization problem P_1 if (z, u_z) minimizes the value of the cost functional $J_1(z, u_z)$ among all admissible strategies for P_1.

A plan of the paper is as follows. In §2, we give background material on nonsmooth analysis that will be referred to in the following sections. In §3, we study generalized differentiability of the value function $V(z)$. The necessary condition for optimality is given in §4. In §5, we consider an extension to the bilevel dynamic optimization problem defined in §1 to allow opportunity costs; a fishery regulation problem is used to demonstrate applications of the necessary condition for optimality derived.

2. Nonsmooth analysis background. In this section we shall give a concise review of the material on nonsmooth analysis that will be required.

Let C be a nonempty closed set in $I\!\!R^n$. A vector $\zeta \in I\!\!R^n$ is a *proximal normal* to C at point $\bar{x} \in C$ if for $t > 0$ sufficiently small, the unique point of C nearest to $\bar{x} + t\zeta$ (in the Euclidean norm) is $\bar{x}$. It is a *limiting proximal normal* if there exist points $x_k \in C, x_k \to \bar{x}$, and proximal normals ζ_k to C at x_k, such that $\zeta_k \to \zeta$. Let the *limiting proximal normal cone to C at $\bar{x}$* be the set

$$\hat{N}_C(\bar{x}) := \{\zeta : \zeta \text{ is a limiting proximal normal to } C \text{ at } \bar{x}\}$$

and the *Clarke normal cone to C at $\bar{x}$* to be the set

$$N_C(\bar{x}) := \text{clco}\hat{N}_C(\bar{x}).$$

Now consider a lower semicontinuous function $\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and a point $\bar{x} \in \mathbb{R}^n$ where ϕ is finite. A vector $\zeta \in \mathbb{R}^n$ is called a *proximal subgradient* of $\phi(\cdot)$ at $\bar{x}$ provided that there exist $M > 0, \delta > 0$ such that

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

where $\langle a, b \rangle$ denotes the inner product of vectors a and b. The set of all proximal subgradients of $\phi(\cdot)$ at $\bar{x}$ is denoted $\partial^\pi \phi(\bar{x})$. The *limiting subgradient* of ϕ at $\bar{x}$ is the set

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

The *singular limiting subgradient* of ϕ at $\bar{x}$ is the set

$$\hat{\partial}^\infty \phi(\bar{x}) := \left\{ \lim_{k\to\infty} t_k \zeta_k : \zeta_k \in \partial^\pi \phi(x_k), x_k \to \bar{x}, \phi(x_k) \to \phi(\bar{x}), t_k \downarrow 0 \right\}.$$

The limiting subgradient is a smaller object than the *Clarke generalized gradient.* In fact, if ϕ is Lipschitz continuous near x, we have $\partial\phi(x) = \mathrm{co}\hat{\partial}\phi(x)$, where $\partial\phi$ and coA denote the Clarke generalized gradient of ϕ and the convex hull of the set A, respectively. For the definition and the precise relation between the limiting subgradient and the Clarke generalized gradient, the reader is referred to Clarke [5] and Rockafellar [9].

The following proposition summarizes the prerequisites regarding limiting subgradients and limiting proximal normal cones.

PROPOSITION 2.1. (a) *If C is a nonempty closed convex set, the limiting proximal normal cone to C coincides with the normal cone in the sense of convex analysis, i.e., one has $\zeta \in \hat{N}_C(\bar{x})$ if and only if*

$$\langle \zeta, x - \bar{x} \rangle \leq 0 \quad \forall x \in C.$$

(b) *The function $\phi(\cdot)$ is Lipschitz near x if and only if $\hat{\partial}^\infty \phi(x) = \{0\}$.*

(c) *If $\hat{\partial}\phi(x) \neq \emptyset$, then*

$$\hat{\partial}(s\phi)(x) = s\hat{\partial}\phi(x) \qquad \forall s \geq 0.$$

(d) (*Clarke* [5, *Prop.* 1.5]) *Let ϕ and $\psi : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ be lower semicontinuous functions finite at x, with $\hat{\partial}^\infty\phi(x) \cap (-\hat{\partial}^\infty\psi(x)) = \{0\}$. Then we have*

$$\hat{\partial}(\phi + \psi)(x) \subset \hat{\partial}\phi(x) + \hat{\partial}\psi(x).$$

(e) *Let $\Psi_C(x)$ be the indicator function of the set C. Then*

$$\hat{N}_C(x) = \hat{\partial}\Psi_C(x) = \hat{\partial}^\infty \Psi_C(x).$$

(f) *Let S_1 and S_2 be closed subets of $\mathbb{R}^n$ and let $\bar{x} \in S_1 \cap S_2$. If $\hat{N}_{S_1}(\bar{x}) \cap (-\hat{N}_{S_2}(\bar{x})) = \{0\}$, then we have*

$$\hat{N}_{S_1 \cap S_2}(\bar{x}) \subset \hat{N}_{S_1}(\bar{x}) + \hat{N}_{S_2}(\bar{x}).$$

(g) (*chain rule*) *Let $\phi(x) := f(F(x))$ where $F : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz on some neighbourhood of $\bar{x}$, while $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous with $F(\bar{x})$ in $\mathrm{dom} f := \{y : f(y) \neq +\infty\}$. Then if*

$$0 \notin \hat{\partial}(\zeta F)(\bar{x}) \qquad \forall \text{ nonzero vectors } \zeta \in \hat{\partial}^\infty f(F(\bar{x})),$$

we have

$$\hat{\partial}\phi(\bar{x}) \subset \cup\{\hat{\partial}(\zeta F)(\bar{x}) : \zeta \in \hat{\partial} f(F(\bar{x}))\}.$$

3. Differentiability of the value function. To discuss generalized differentiability of the value function $V(z)$, we will refer to the following assumptions:

(A6) For some $\alpha \geq 0, \beta \geq 0$, the function $(\phi(t,\cdot,z,u), G(t,\cdot,z,u))$ satisfies the following growth condition: for all $z \in Z$, $(t,u) \in \mathrm{Gr}U$, one has

$$|(\phi(t,x,z,u), G(t,x,z,u))| \leq \alpha|x| + \beta.$$

(A6)′ The functions ϕ and G are continuously differentiable in x and z and lower semicontinuous in u. There exists an integrable function $k(t)$ such that

$$|\phi| + |\nabla_x \phi| + |G| + |\nabla_x G| \leq k(t).$$

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

$$\{(\phi(t,x,z,u), G(t,x,z,u)) : u \in U(t)\}$$

is convex.

(A7)′ For any $(t,x,z) \in [t_0,t_1] \times \mathbb{R}^d \times \mathbb{R}^n$, the set

$$\{(\phi(t,x,z,u), G(t,x,z,u) + \delta) : u \in U(t), \delta \geq 0\}$$

is convex.

The Hamiltonian for $P_2(z)$ is the function defined by

$$H_2(t,x,z,p_2;\lambda) := \sup\{p_2 \cdot \phi(t,x,z,u) - \lambda G(t,x,z,u) : u \in U(t)\}.$$

An index λ multiplier corresponding to an admissible trajectory x for $P_2(z)$ is an arc (p_2, q) such that

$$\begin{aligned} (-\dot{p}_2(t), -\dot{q}(t), \dot{x}(t)) &\in \partial_{(x,z,p_2)} H_2(t,x(t),z,p_2(t);\lambda) \qquad \text{a.e.} \\ -p_2(t_1) &\in \lambda\hat{\partial} g(x(t_1)) + \hat{N}_{C_1}(x(t_1)), \\ q(t_1) &= 0. \end{aligned}$$

The collection of all such arcs is the set $M^\lambda(x)$, the index λ multiplier set corresponding to x. Let Y be the set of all optimal trajectories x to problem $P_2(z)$. Let

$$M^\lambda(Y) := \bigcup_{x \in Y} M^\lambda(x).$$

For any index λ multiplier $(p_2,q) \in M^\lambda(x)$, we define $Q(p_2,q) = -q(t_0)$. The notation $QM^\lambda(x)$ designates the set of all possible values of $-q(t_0)$ obtained in this way, and $Q(M^\lambda(Y))$ denotes $\cup_{x\in Y} Q(M^\lambda(x))$. The following result relates the differential properities of V to the arcs q in the multiplier sets introduced above.

THEOREM 3.1. *In addition to assumptions* (A1)–(A5), *suppose either* (A6)–(A7) *or* (A6)′–(A7)′ *hold. If* $QM^0(Y) = \{0\}$, *then* V *is Lipschitz continuous near* z *and one has*

$$\hat{\partial} V(z) \subset QM^1(Y).$$

Theorem 3.1 under assumptions (A6)–(A7) can be obtained by reducing the original optimal control problem to an differential inclusion problem and applying the sensitivity result in Clarke and Loewen [6, Thm. 3.3]. Before proving Theorem 3.1 under assumptions (A1)–(A5) and (A6)′–(A7)′, we first give the following result.

LEMMA 3.2. *Let α_i be a sequence converging to α, and let (x_i, u_i) be an admissible pair for $P_2(\alpha_i)$. Then there exists a subsequence of $\{x_i\}$ converging uniformly to an arc x and a control u with (x, u) being an admissible pair for $P_2(\alpha)$ such that*

$$J_2(x, u) \leq \liminf J_2(x_i, u_i).$$

The proof can be reduced to an application of [4, Thm. 3.1.7] by studying the differential inclusion

$$(\dot{x}(t), \dot{y}(t), \dot{\alpha}(t)) \in \Gamma(t, x(t), y(t), \alpha(t)) \qquad \text{a.e.},$$

where $y \in I\!R$ and the convex multifunction Γ is defined via

$$\Gamma(t, x, y, \alpha) := \{[\phi(t, x, \alpha, u), r, 0] : G(t, x, \alpha, u) \leq r \leq k(t) + 1, u \in U(t)\}.$$

The essential fact in the reduction is Filippov's lemma: (x, y, α) satisfies the above differential inclusion iff there is a control u for x such that (x, u) is an admissible pair for $P_2(\alpha)$ and y satisfies

$$G(t, x, \alpha, u) \leq \dot{y} \leq k(t) + 1.$$

We now turn to the proof of the theorem. By (A5), $P_2(z)$ has an admissible pair. So $V(z)$ is finite. It follows from Lemma 3.2 that V is lower-semicontinuous.

Step 1. Let $\alpha \in Z$ be a point near z. Let $\zeta \in \partial^\pi V(\alpha)$, and let (x, u) be a solution of $P_2(\alpha)$ that exists by virtue of Lemma 3.2. Then by definition, for some $M > 0$ and for all α' near α, we have

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

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

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

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

$$\begin{aligned} \min \int_{t_0}^{t_1} & G(t, x'(t), \alpha'(t), u'(t))dt + g(x'(t_1)) - \langle \zeta, \alpha'(t_0) \rangle, \\ \text{s.t. } & \dot{x}'(t) = \phi(t, x'(t), \alpha'(t), u'(t)) \qquad \text{a.e.}, \\ & \dot{\alpha}'(t) = 0, \\ & x'(t_0) = x_0, \quad x'(t_1) \in C_1, \\ & u'(t) \in U(t) \qquad \text{a.e.} \end{aligned}$$

In the proof of Theorem 5.2.1 of Clarke [4], if we replace the the Clarke generalized gradient ∂ by the limiting subgradient $\hat{\partial}$ in the transversality conditions, the argument goes through without modification (cf. Clarke [5]). It follows that there exist a scalar $\lambda \geq 0$ and arcs p_2, q such that

$$
\begin{aligned}
(1)\quad & -\dot{p_2}(t) = \nabla_x \phi(t, x(t), \alpha, u(t))^\top p_2(t) - \lambda \nabla_x G(t, x(t), \alpha, u(t)) \quad \text{a.e.;}\\
(2)\quad & -\dot{q}(t) = \nabla_\alpha \phi(t, x(t), \alpha, u(t))^\top p_2(t) - \lambda \nabla_\alpha G(t, x(t), \alpha, u(t)) \quad \text{a.e.;}\\
& \max_{u \in U(t)} \{p_2(t) \cdot \phi(t, x(t), \alpha, u) - \lambda G(t, x(t), \alpha, u)\}\\
(3)\quad & \qquad = p_2(t) \cdot \phi(t, x(t), \alpha, u(t)) - \lambda G(t, x(t), \alpha, u(t)) \quad \text{a.e.,}\\
& -p_2(t_1) \in \lambda \hat{\partial} g(x(t_1)) + \hat{N}_{C_1}(x(t_1)),\\
& q(t_0) = -\lambda \zeta, \qquad q(t_1) = 0,\\
& \|p_2\|_\infty + \|q\|_\infty + \lambda > 0,
\end{aligned}
$$

where ∂ denotes the Clarke generalized gradient, $\|\cdot\|_\infty$ denotes the supremum norm, and $^\top$ denotes the transpose.

By Clarke [4, Thm. 2.8.2], since ϕ and G are continuously differentiable in (x, z), $\partial_{(x,\alpha,p_2)} H_2(t, x, \alpha, p_2; \lambda)$ is the convex hull of all points of the form

$$[\nabla_x \phi(t, x, \alpha, u)^\top p_2 - \lambda \nabla_x G(t, x, \alpha, u), \nabla_\alpha \phi(t, x, \alpha, u)^\top p_2 - \lambda \nabla_\alpha G(t, x, \alpha, u), \phi(t, x, \alpha, u)],$$

where u in $U(t)$ is any point at which the maximum defining $H_2(t, x, \alpha, p_2; \lambda)$ is achieved. Hence (1), (2), and (3) imply that

$$(-\dot{p_2}(t), -\dot{q}(t), \dot{x}(t)) \in \partial_{(x,\alpha,p_2)} H_2(t, x(t), \alpha, p_2(t); \lambda) \qquad \text{a.e.}$$

Step 2. For any $\zeta \in \hat{\partial} V(z)$, by definition, $\zeta = \lim_{i\to\infty} \zeta_i$ where $\zeta_i \in \partial^\pi V(\alpha_i)$, $\alpha_i \to z$, and $V(\alpha_i) \to V(z)$. By Step 1, for each ζ_i, there exists an arc (p_2^i, q_i), a scalar λ_i, and an arc x_i that solves $P_2(\alpha_i)$ such that

$$
\begin{gathered}
(-\dot{p_2}^i(t), -\dot{q}_i(t), \dot{x}_i(t)) \in \partial_{(x,\alpha,p_2)} H_2(t, x_i(t), \alpha_i, p_2^i(t); \lambda_i) \qquad \text{a.e.,}\\
-\, p_2^i(t_1) \in \lambda_i \hat{\partial} g(x_i(t_1)) + \hat{N}_{C_1}(x_i(t_1)),\\
q_i(t_0) = -\lambda_i \zeta_i, \qquad q_i(t_1) = 0,\\
\|p_2^i\| + \|q_i\| + \lambda_i > 0.
\end{gathered}
$$

Since $M^0(Y) = \{0\}$, we must indeed have $\lambda_i = 1$ for i sufficiently large and $|p_2^i(0)|$ bounded (cf., Clarke and Loewen [6, p. 253]). Passing to a uniformly convergent subsequence of $\{(p_2^i, q_i, x_i)\}$ by Lemma 3.2 and Clarke [4, Thm. 3.1.7] leads to an optimal trajectory x for $P_2(z)$ and an arc (p_2, q) such that

$$
\begin{gathered}
(-\dot{p_2}(t), -\dot{q}(t), \dot{x}(t)) \in \partial_{(x,\alpha,p_2)} H_2(t, x(t), \alpha, p_2(t); \lambda) \qquad \text{a.e.,}\\
-\, p_2(t_1) \in \hat{\partial} g(x(t_1)) + \hat{N}_{C_1}(x(t_1)),\\
q(t_0) = -\zeta, \qquad q(t_1) = 0.
\end{gathered}
$$

That is, $(p_2, q) \in QM^1(Y)$.

Similarly to Ye [11], one can show $\hat{\partial}^\infty V(z) \subset QM^0(Y)$ using results from Step 2. The Lipschitz continuity of V near z then follows by virtue of assumption $M^0(Y) = \{0\}$ and (b) of Proposition 2.1. The proof of Theorem 3.1 is now complete.

4. Necessary conditions for optimality. Define the *pseudo-Hamiltonian* for problem $(\widetilde{P}_1)$ as

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

for $t \in [t_0, t_1]$, $x, p_1 \in I\!R^d$, $z \in Z$, $\lambda, r \in I\!R$.

THEOREM 4.1. *Assume assumptions* (A1)–(A4) *hold. Let* $(z, u(t))$ *be an optimal strategy of the bilevel dynamic optimization problem* P_1 *and* $x(t)$ *the corresponding trajectory. Assume that the value function for the lower-level problem* V *is locally Lipschitz continuous. Then there exist* $\lambda \geq 0, r \geq 0$ *and arcs* p_1, η *such that*:

$$-(\dot{p}_1(t), \dot{\eta}(t)) \in \partial_{(x,z)} H_1(t, x(t), z, p_1(t), u(t); \lambda, r) \qquad a.e., \tag{4}$$

$$\max_{u \in U(t)} H_1(t, x(t), z, p_1(t), u; \lambda, r) = H_1(t, x(t), z, p_1(t), u(t); \lambda, r) \qquad a.e., \tag{5}$$

$$\eta(t_0) = 0,$$

$$-p_1(t_1) \in \lambda \hat{\partial} f(x(t_1)) + r\hat{\partial} g(x(t_1)) + \hat{N}_{C_1}(x(t_1)), \tag{6}$$

$$\eta(t_1) \in r\partial V(z), \tag{7}$$

$$\|p_1\|_\infty + \|\eta\|_\infty + \lambda + r > 0. \tag{8}$$

The following result, which is a limiting subgradient version of Corollary 1 of Theorem 2.4.7 in Clarke [4], will be useful in proving Theorem 4.1. We should prove it by using a chain rule.

LEMMA 4.2. *Let* $C = \{x : \psi(x) \leq 0\}$, *where* $\psi : I\!R^n \to I\!R$ *is Lipschitz continuous on some neighborhood of* $\bar{x} \in C$. *Suppose that* $0 \notin \hat{\partial}\psi(\bar{x})$. *Then*

$$\hat{N}_C(\bar{x}) \subset \bigcup_{r \geq 0} r\hat{\partial}\psi(\bar{x}). \tag{9}$$

Proof. If $\bar{x}$ is in the interior of C, then $\hat{N}_C(\bar{x}) = \{0\}$ and the above relation is trivially satisfied. Suppose $\bar{x}$ is in the boundary of C. By virtue of (a), (c), and (e) of Proposition 2.1, $0 \notin \hat{\partial}\psi(\bar{x})$ implies

$$0 \notin \hat{\partial} r\psi(\bar{x}) \qquad \forall \text{ nonzero scalars } r \in I\!R_+ = \hat{\partial}^\infty \Psi_{I\!R_-}(\psi(\bar{x})) = \hat{N}_{I\!R_-}(\psi(\bar{x})).$$

Since $\Psi_C(\bar{x}) = \Psi_{I\!R_-}(\psi(\bar{x}))$. by the chain rule ((g) of Proposition 2.1) we have

$$\hat{\partial}\Psi_C(\bar{x}) \subset \cup\{\hat{\partial}(r\psi)(\bar{x}) : r \in \hat{\partial}\Psi_{I\!R_-}(\psi(\bar{x}))\}, \tag{10}$$

which is the relation (9) thanks to Proposition 2.1(e). ◻

The proof of the following result is straightforward.

LEMMA 4.3. *Let* $F(x, y, z) : I\!R^d \times I\!R^m \times I\!R^n \to I\!R \cup \{+\infty\}$ *be a lower semicontinuous function and* $(\bar{x}, \bar{y}, \bar{z}) \in \text{dom}F$. *Suppose* $F(x, y, z) = F_1(x) + F_2(y) + F_3(z)$. *Then*

$$\hat{\partial}F(\bar{x}, \bar{y}, \bar{z}) \subset \hat{\partial}F_1(\bar{x}) \times \hat{\partial}F_2(\bar{y}) \times \hat{\partial}F_3(\bar{z}).$$

Proof of Theorem 4.1. We pose the optimal control problem $\widetilde{P}_1$ equivalently as the problem

$$\widehat{P}_1 \qquad \min \int_{t_0}^{t_1} F(t, x(t), z(t), u(t))dt + f(x(t_1))$$

$$\begin{aligned} \text{s.t. } \dot{x}(t) &= \phi(t, x(t), z(t), u(t)) \quad \text{a.e.},\\ \dot{y}(t) &= G(t, x(t), z(t), u(t)) \quad \text{a.e.},\\ \dot{z}(t) &= 0,\\ u(t) &\in U(t) \quad \text{a.e.},\\ (x, y, z)(t_0) &\in \{x_0\} \times \{0\} \times I\!R,\\ (x, y, z)(t_1) &\in S := \{(x, y, z) : g(x) + y - V(z) \le 0, x \in C_1\}. \end{aligned}$$

The problem above is exactly in the form described in §5.2.1 of Clarke [4]. The pseudo-Hamiltonian is the function

$$H(t, x, y, z, p_1, p_2, \eta, u, \lambda) = p_1 \cdot \phi(t, x, z, u) + p_2 G(t, x, z, u) - \lambda F(t, x, z, u),$$

for $t \in [t_0, t_1], x, p_1 \in I\!R^d, y, p_2, \eta, \lambda \in I\!R, z \in Z$. Applying Theorem 5.2.1 of Clarke [4] with the generalized gradient replaced by the limiting subgradient in the transversality conditions leads to the existence of a scalar $\lambda \ge 0$ and an arc (p_1, p_2, η) such that

(11) $-(\dot{p_1}(t), \dot{p_2}(t), \dot{\eta}(t)) \in \partial_{(x,y,z)} H(t, x(t), y(t), z(t), p_1(t), p_2(t), \eta(t), u(t), \lambda)$ a.e.,

$$\max_{u \in U(t)} H(t, x(t), y(t), z(t), p_1(t), p_2(t), \eta(t), u, \lambda)$$

(12) $\quad = H(t, x(t), y(t), z(t), p_1(t), p_2(t), \eta(t), u(t), \lambda)$ a.e.,

(13) $(p_1(t_0), p_2(t_0), \eta(t_0)) \in \hat{N}_{\{x_0\} \times \{0\} \times I\!R}(x(t_0), y(t_0), z(t_0))$,

(14) $-(p_1(t_1), p_2(t_1), \eta(t_1)) \in \lambda \hat{\partial} \hat{f}(x(t_1), y(t_1), z(t_1)) + \hat{N}_S(x(t_1), y(t_1), z(t_1))$,

(15) $\|p_1\|_\infty + \|p_2\|_\infty + \|\eta\|_\infty + \lambda > 0$,

where $\hat{f}(x, y, z) := f(x)$.

Let $\hat{F}(x, y, z) = g(x) + y - V(z)$. Then by Lemma 4.3, one has

$$\hat{\partial} \hat{F}(x, y, z) \subset \hat{\partial} g(x) \times \{1\} \times \hat{\partial}(-V(z)). \tag{16}$$

Therefore $0 \notin \hat{\partial} \hat{F}(x, y, z)$.

Let $S_1 := \{(x, y, z) : g(x) + y - V(z) \le 0\}$ and $S_2 := C_1 \times I\!R \times I\!R$. By Lemma 4.2 and inclusion (16), one has

$$\begin{aligned} \hat{N}_{S_1}(x, y, z) &\subset \bigcup_{r \ge 0} r \hat{\partial} \hat{F}(x, y, z)\\ &\subset \bigcup_{r \ge 0} r[\hat{\partial} g(x) \times \{1\} \times \hat{\partial}(-V)(z)]. \end{aligned}$$

Since $\Psi_{S_2}(x, y, z) = \Psi_{C_1}(x) + \Psi_{I\!R}(y) + \Psi_{I\!R}(z)$, by Lemma 4.3 and (e) of Proposition 2.1 one has

$$\hat{N}_{S_2}(x, y, z) \subset \hat{N}_{C_1}(x) \times \{0\} \times \{0\} \qquad \forall (x, y, z) \in C_1 \times I\!R \times I\!R.$$

It follows that the second component of any triple in the set $-\hat{N}_{S_2}(x, y, z)$ is 0. The only vectors in $\hat{N}_{S_1}(x, y, z)$ that share this property are among those for which $r = 0$ in the estimate above. Thus, $\hat{N}_{S_1}(x, y, z) \cap (-\hat{N}_{S_2}(x, y, z)) = \{0\}$ and Proposition 2.1 (f) gives

$$\begin{aligned} \hat{N}_S(x(t_1), y(t_1), z(t_1)) &\subset \hat{N}_{S_1}(x(t_1), y(t_1), z(t_1)) + \hat{N}_{S_2}(x(t_1), y(t_1), z(t_1))\\ &\subset \bigcup_{r \ge 0} r[\hat{\partial} g(x(t_1)) \times \{1\} \times \hat{\partial}(-V)(z)]\\ &\quad + \hat{N}_{C_1}(x(t_1)) \times \{0\} \times \{0\}. \end{aligned}$$

By Lemma 4.3, one has

$$\hat{\partial}\hat{f}(x(t_1), y(t_1), z(t_1)) \subset \hat{\partial} f(x(t_1)) \times \{0\} \times \{0\}.$$

Hence from (14), one has

$$\begin{aligned}
-(p_1(t_1), p_2(t_1), \eta(t_1)) \in\ & \lambda\hat{\partial} f(x(t_1)) \times \{0\} \times \{0\} \\
& + \bigcup_{r\geq 0} r[\hat{\partial} g(x(t_1)) \times \{1\} \times \hat{\partial}(-V)(z)] \\
& + \hat{N}_{C_1}(x(t_1)) \times \{0\} \times \{0\} \\
\subset\ & \lambda\hat{\partial} f(x(t_1)) \times \{0\} \times \{0\} \\
& + \bigcup_{r\geq 0} r[\hat{\partial} g(x(t_1)) \times \{1\} \times (-\partial V(z))] \\
& + \hat{N}_{C_1}(x(t_1)) \times \{0\} \times \{0\},
\end{aligned}$$

from which the transversality conditions (6) and (7) follow and one has $p_2(t_1) = -r$, where $r \geq 0$. Since H is independent of y, (11) implies that $\dot{p_2}(t) = 0$ and

$$(17) \quad -(\dot{p_1}(t), \dot{\eta}(t)) \in \partial_{(x,z)} H(t, x(t), y(t), z(t), p_1(t), p_2(t), \eta(t), u(t); \lambda) \qquad \text{a.e.}$$

Hence $p_2 \equiv -r$, where $r \geq 0$; and (4), (5), and (8) follow from (17), (12), and (15), respectively. From (13), one has $\eta(t_0) = 0$. The proof of the theorem is thus complete. □

Combining Theorem 4.1 and Theorem 3.1, one has the following necessary conditions for optimality for the general bilevel dynamic optimization problem.

THEOREM 4.4. *In addition to assumptions* (A1)–(A5)*, suppose either assumptions* (A6)–(A7) *or* (A6)′–(A7)′ *hold. Let* (z, u) *be an optimal strategy of the bilevel dynamic optimization problem* P_1 *and* $x(t)$ *the corresponding trajectory. Suppose that* $QM^0(Y) = \{0\}$*. Then there exist scalars* $\lambda \geq 0, r \geq 0$*, integers* I, J*,* $\lambda_{ij} \geq 0$*,* $\sum_{i=1}^{I}\sum_{j=1}^{J}\lambda_{ij} = 1$*, optimal trajectories* $x_i(t)$ *of the lower-level problem* $P_2(z)$*, and arcs* $p_1, \eta, p_2^{ij}, q^{ij}$ *such that*

$$\begin{aligned}
(18) \qquad & -(\dot{p}_1(t), \dot{\eta}(t)) \in \partial_{(x,z)} H_1(t, x(t), z, p_1(t), u(t); \lambda, r) \qquad \text{a.e.}, \\
& \max_{u\in U(t)} H_1(t, x(t), z, p_1(t), u; \lambda, r) = H_1(t, x(t), z, p_1(t), u(t); \lambda, r) \qquad \text{a.e.}, \\
& \eta(t_0) = 0, \\
& -p_1(t_1) \in \lambda\hat{\partial} f(x(t_1)) + r\hat{\partial} g(x(t_1)) + \hat{N}_{C_1}(x(t_1)), \\
& \eta(t_1) = r\sum_{ij}\lambda_{ij} q^{ij}(t_0); \\
(19) \qquad & (-\dot{p}_2^{ij}(t), -\dot{q}^{ij}(t), \dot{x}_i(t)) \in \partial_{(x,z,p_2)} H_2(t, x_i(t), z, p_2^{ij}(t); 1) \qquad \text{a.e.}, \\
& q^{ij}(t_1) = 0, \\
& -p_2^{ij}(t_1) \in \hat{\partial} g(x_i(t_1)) + \hat{N}_{C_1}(x_i(t_1)), \\
& \|p_1\|_\infty + \|\eta\|_\infty + \lambda + r > 0.
\end{aligned}$$

Remark 4.1. A sufficient condition for $QM^0(Y) = \{0\}$ to hold is $C_1 = \mathbb{R}^d$. Indeed, in this case, the index 0 multiplier set consists of all arcs (p_2, q) such that

$$(20) \qquad (-\dot{p_2}(t), -\dot{q}(t), \dot{x}(t)) \in \partial_{(x,z,p_2)} H_2(t, x(t), z, p_2(t); 0) \qquad \text{a.e.}$$

(21) $$p_2(t_1) = 0,$$

(22) $$q(t_1) = 0.$$

Due to the Lipschitz continuity of ϕ in (x, z), by virtue of Theorem 2.8.2 of Clarke [4], (20) implies that

$$\|\dot{p_2}(t)\| \leq k(t)\|p_2(t)\|.$$

By Gronwall's Lemma, the above inequality implies that p_2 is either identically 0 or nonvanishing on $[t_0, t_1]$. Therefore (21) implies that $p_2 \equiv 0$. Hence $\dot{q}(t) = 0$ by virtue of (20). But q satisfies (22), therefore $q \equiv 0$. That is $QM^0(Y) = \{0\}$.

Another sufficient condition for $M^0(Y) = \{0\}$ to hold is that $\phi(t, x, z, u)$ be independent of z since in this case $q(t) \equiv 0$.

Remark 4.2. By Clarke [4, Thm. 2.8.2], $\partial_{(x,\alpha,p_2)} H_2(t, x, \alpha, p_2; 1)$ is the convex hull of all points of the form

$$[\nabla_x \phi(t, x, \alpha, u)^\top p_2 - \nabla_x G(t, x, \alpha, u), \nabla_\alpha \phi(t, x, \alpha, u)^\top p_2 - \nabla_\alpha G(t, x, \alpha, u), \phi(t, x, \alpha, u)],$$

where u in $U(t)$ is any point at which the maximum defining $H_2(t, x, \alpha, p_2; 1)$ is achieved. Therefore if in addition to assumptions (A1)–(A5) and (A6)$'$–(A7)$'$, we assume the set

$$\{(\nabla_x \phi(t, x, \alpha, u)^\top p_2 - \nabla_x G(t, x, \alpha, u), \nabla_\alpha \phi(t, x, \alpha, u)^\top p_2 - \nabla_\alpha G(t, x, \alpha, u) : u \in U(t)\}$$

is convex for any t, x, z, p_2, then the inclusion (19) becomes the following equations:

$$\begin{aligned}
-\dot{p}_2^{ij}(t) &= \nabla_x \phi(t, x_i(t), z, u_i(t))^\top p_2^{ij}(t) - \nabla_x G(t, x_i(t), z, u_i(t)) \qquad \text{a.e.},\\
-\dot{q}^{ij}(t) &= \nabla_z \phi(t, x_i(t), z, u_i(t))^\top p_2^{ij}(t) - \nabla_z G(t, x_i(t), z, u_i(t)) \qquad \text{a.e.},\\
&\max_{u \in U(t)} \{p_2^{ij}(t) \cdot \phi(t, x_i(t), z, u) - G(t, x_i(t), z, u)\}\\
&\qquad = p_2^{ij}(t) \cdot \phi(t, x_i(t), z, u_i(t)) - G(t, x_i(t), z, u_i(t)) \qquad \text{a.e.},\\
&\qquad \dot{x}_i(t) = \phi(t, x_i(t), z, u_i(t)) \qquad \text{a.e.},
\end{aligned}$$

where $u_i(t)$ is an optimal control function associated with trajectory $x_i(t)$.

5. Extensions and an example. There are many situations where an opportunity cost exists for the follower. That is, the follower will participate only if his optimal cost is less than or equal to the opportunity cost $L \geq 0$ that he may receive from somewhere else. In this case, the leader faces the following bilevel optimization problem:

$$\begin{aligned}
\overline{P}_1 \qquad \min J_1(z, u) &= \int_{t_0}^{t_1} F(t, x(t), z(t), u(t))dt + f(x(t_1)),\\
\text{s.t. } \dot{x}(t) &= \phi(t, x(t), z(t), u(t)) \qquad \text{a.e.},\\
\dot{z}(t) &= 0,\\
x(t_0) &= x_0, \quad x(t_1) \in C_1,\\
u(t) &\in U(t) \qquad \text{a.e.},\\
\int_{t_0}^{t_1} G(t, x(t), z(t), u(t))dt &+ g(x(t_1)) \leq V(z),\\
\int_{t_0}^{t_1} G(t, x(t), z(t), u(t))dt &+ g(x(t_1)) \leq L.
\end{aligned}$$

The technique described in the previous section can be applied to this more general problem in exactly the same way, and one obtains the following necessary conditions for optimality.

THEOREM 5.1. *Assume that in addition to* (A1)–(A5), *either assumptions* (A6)–(A7) *or* (A6)′–(A7)′ *hold. Let* (z,u) *be an optimal strategy of the bilevel dynamic optimization problem* $\overline{P}_1$ *and* $x(t)$ *the corresponding trajectory. Suppose that* $QM^0(Y) = \{0\}$. *Then there exist scalars* $\lambda \geq 0, r \geq 0, 0 \leq \hat{r} \leq r$, *integers* I, J, $\lambda_{ij} \geq 0$, $\sum_{i=1}^{I}\sum_{j=1}^{J}\lambda_{ij} = 1$, *optimal trajectories* $x_i(t)$ *of the lower-level problem* $P_2(z)$, *and arcs* $p_1, \eta, p_2^{ij}, q^{ij}$ *such that*

$$-(\dot{p}_1(t), \dot{\eta}(t)) \in \partial_{(x,z)} H_1(t, x(t), z, p_1(t), u(t); \lambda, r) \qquad a.e.,$$

$$\max_{u \in U(t)} H_1(t, x(t), z, p_1(t), u; \lambda, r) = H_1(t, x(t), z, p_1(t), u(t); \lambda, r) \qquad a.e.,$$

$$\eta(t_0) = 0,$$

$$-p_1(t_1) \in \lambda \hat{\partial} f(x(t_1)) + r\hat{\partial} g(x(t_1)) + \hat{N}_{C_1}(x(t_1)),$$

$$\eta(t_1) = \hat{r} \sum_{ij} \lambda_{ij} q^{ij}(t_0),$$

$$(-\dot{p}_2^{ij}(t), -\dot{q}^{ij}(t), \dot{x}_i(t)) \in \partial_{(x,z,p_2)} H_2(t, x_i(t), z, p_2^{ij}(t); 1) \qquad a.e.,$$

$$q^{ij}(t_1) = 0,$$

$$-p_2^{ij}(t_1) \in \hat{\partial} g(x_i(t_1)) + \hat{N}_{C_1}(x_i(t_1)),$$

$$\|p_1\|_\infty + \|\eta\|_\infty + \lambda + r > 0.$$

The following example is a simplified and finite horizon version of a fishery regulation problem first formulated and solved by Clarke and Munro using principal and agent analysis (see Clarke and Munro [7] and [8] for details).

Example. It has now been generally agreed that the fishery resources within the 200-mile zones are the property of the adjacent coastal states. For those coastal states opting to permit a distant water presence in their 200-mile zones, one of the problems they face is devising optimum terms and conditions of access to the Coastal State Exclusive Economic Zones to be imposed upon the distant water fleets.

Assume that the fish population follows the dynamic system

$$\dot{x}(t) = F(x(t)) - qE(t)x(t),$$

where $x(t)$ is the fish population at time t; $F(x)$ is the rate of natural growth; and $qE(t)x(t)$ is the rate of catch at time t, where $E(t)$ is the fishing effort at time t and q is a positive constant. We assume that $F(x)$ is a twice continuously differentiable function satisfying $F(x) > 0$ for $0 < x < \bar{x}, F(0) = F(\bar{x}) = 0$ and $F''(x) < 0$ for all $x > 0$, where $\bar{x}$ denotes the carrying capacity of the resource. It is also assumed that

$$0 \leq E(t) \leq E_{\max},$$

where $E_{\max}$ is an arbitary upper bound on $E(t)$. Suppose that the coastal state imposes the condition that at the terminal time T_1, the fish population cannot be less than $\widetilde{x} \geq 0$.

Suppose that the coastal states as a leader impose a unit tax n on catch $qE(t)x(t)$ and a unit tax m on effort $E(t)$. Then the distant water fleet would receive the profit in time period $[0, T_1]$

$$\int_0^{T_1} e^{-\delta t}[(p_0 - n)qx(t) - (c_0 + m)]E(t)dt$$

if he decided to use the fishing effort $E(\cdot)$ where p_0 and c_0 are the unit price on catch and unit cost on effort, respectively, $\delta > 0$ is the discount rate, and $x(\cdot)$ is the fish population corresponding to the fishing effort $E(\cdot)$. Hence for the given unit tax on catch and effort n and m, the distant water fleet as a follower faces the following optimal control problem:

$$\begin{aligned} P_2(n,m) \qquad & \max \int_0^{T_1} e^{-\delta t}[(p_0 - n)qx(t) - (c_0 + m)]E(t)dt, \\ & \text{s.t. } \dot{x}(t) = F(x(t)) - qE(t)x(t), \\ & x(0) = x_0, \quad x(T_1) \geq \widetilde{x}, \\ & E(t) \in [0, E_{\max}]. \end{aligned}$$

The optimal control problem $P_2(n, m)$ is linear. The necessary condition for (x, E) to solve $P_2(n, m)$ is the existence of an arc p_2 such that

$$-\dot{p_2}(t) = p_2(t)[F'(x(t)) - qE(t)] + e^{-\delta t}(p_0 - n)qE(t), \tag{23}$$

$$\begin{aligned} & \max_{E\in[0,E_{\max}]} \{p_2(t)[F(x(t)) - qEx(t)] + e^{-\delta t}[(p_0 - n)qx(t) - (c_0 + m)]E\} \\ & \quad = p_2(t)[F(x(t)) - qE(t)x(t)] + e^{-\delta t}[(p_0 - n)qx(t) - (c_0 + m)]E(t), \end{aligned} \tag{24}$$

$$p_2(T_1) \geq 0.$$

Since $E(t)$ has to maximize the Hamiltonian (see (24)), $E(t)$ must be either the singular control or else $E(t) = 0$ or $E_{\max}$. The singular control arises when the coefficient of E in the Hamiltonian is zero, implying that

$$p_2(\dot{t}) = e^{-\delta t}\left[(p_0 - n) - \frac{c_0 + m}{qx}\right] \tag{25}$$

$$\dot{p_2}(t) = e^{-\delta t}\left[-\delta\left[(p_0 - n) - \frac{c_0 + m}{qx}\right] + \frac{c_0 + m}{qx^2}\frac{dx}{dt}\right]. \tag{26}$$

From the adjoint equation (23), one has

$$\begin{aligned} \dot{p_2} &= -p_2[F'(x) - qE] - e^{-\delta t}[(p_0 - n)qE \\ &= -e^{-\delta t}\left\{\left[(p_0 - n) - \frac{c_0 + m}{qx}\right][F'(x) - qE] + (p_0 - n)qE\right\}, \end{aligned} \tag{27}$$

where (25) is used for p_2. When the two expressions for $\dot{p_2}(t)$, (26) and (27), are equated, the control variable E cancels out and the following equation emerges:

$$F'(x) + \frac{F(x)(c_0 + m)/qx^2}{p_0 - n - (c_0 + m)/qx} = \delta. \tag{28}$$

For fixed (n, m), this equation gives a unique solution x_* that is the optimal biomass and the optimal trajectory is the one that takes the most rapid path to the optimal biomass x_* (cf. Clark [3]).

Let $V(n,m)$ be the optimal value of the above problem. The distant water fleet will participate only when $V(n,m) \geq L$, the alternative remuneration from some other coastal state.

The coastal state as a leader now faces the following bilevel dynamic optimization problem:

$$
\begin{aligned}
P_1 \qquad & \max \int_0^{T_1} e^{-\delta t}(nqx(t)+m)E(t)dt,\\
& \text{s.t. } \dot{x}(t) = F(x(t)) - qE(t)x(t),\\
& x(0) = x_0, \quad x(T_1) \geq \widetilde{x},\\
& E(t) \in [0, E_{\max}] \qquad \text{a.e.},\\
& V(n,m) \leq \int_0^{T_1} e^{-\delta t}[(p_0-n)qx(t) - (c_0+m)]E(t)dt,\\
& V(n,m) \geq L.
\end{aligned}
$$

It is easy to show that all the conditions of Theorem 5.1 are satisfied. Notice that the lower-level problem $P(n,m)$ has a unique solution. By Theorem 5.1 and Remark 3.2, if (n,m,x,E) is an optimal solution to P_1, then there exist arcs $p_1, p_2, \eta_1, \eta_2, q_1, q_2$ and scalars $\lambda \geq 0$, $r \geq 0$, $0 \leq \hat{r} \leq r$ such that

$$
\begin{aligned}
(29) \qquad & -\dot{p_1} = p_1[F'(x) - qE) + e^{-\delta t}[r(p_0-n) + \lambda n]qE,\\
& \dot{\eta_1} = (r-1)e^{-\delta t}qxE,\\
& \dot{\eta_2} = (r-1)e^{-\delta t}E,\\
& \max_{E\in[0,E_{\max}]} \{p_1(t)[F(x(t)) - qEx(t)] + e^{-\delta t}[r[(p_0-n)qx(t) - (c_0+m)]\\
& \qquad\qquad +\lambda(nqx(t)+m)]E\}\\
& = p_1(t)[F(x(t)) - qE(t)x(t)] + e^{-\delta t}[r[(p_0-n)qx(t) - (c_0+m)]\\
(30) \qquad & \qquad\qquad +\lambda(nqx(t)+m)]E(t),\\
& (\eta_1,\eta_2)(0) = (0,0),\\
(31) \qquad & p_1(T_1) \geq 0,\\
& (\eta_1,\eta_2)(T_1) = \hat{r}q(0),\\
(32) \qquad & -\dot{p_2} = p_2[F'(x) - qE] + e^{-\delta t}(p_0-n)qE,\\
& \dot{q}_1 = e^{-\delta t}qxE,\\
& \dot{q}_2 = e^{-\delta t}E,\\
(33) \qquad & \max_{E\in[0,E_{\max}]} \{p_2(t)[F(x(t)) - qEx(t)] + e^{-\delta t}[(p_0-n)qx(t) - (c_0+m)]E\}\\
& = p_2(t)[F(x(t) - qE(t)x(t)] + e^{-\delta t}[(p_0-n)qx(t) - (c_0+m)]E(t),\\
& (q_1,q_2)(0) = (0,0),\\
(34) \qquad & p_2(T_1) \geq 0,\\
& \|p_1\|_\infty + \|\eta\|_\infty + \lambda + r > 0.
\end{aligned}
$$

Take $\lambda = 1$. As in the proof of (28), from (29) and (30) we can show that the steady state (n,m,x_*) for problem P_1 is a solution of the following equation:

$$
(35) \qquad F'(x_*) + \frac{F(x_*)(r(c_0+m)-m)/q{x_*}^2}{r(p_0-n)+n-(r(c_0+m)-m)/qx_*} = \delta,
$$

and the optimal trajectory for P_1 is the one that takes the most rapid path to the optimal biomass (n, m, x_*). Since (n, m, x, E) is an optimal solution of P_1, (x, E) must be the optimal solution of the lower-level problem $P_2(n, m)$. Therefore x_* must be the optimal biomass associated with (n, m) defined by (28). Combining equations (28) and (35), one has

$$\begin{aligned} n &= \rho p_0, \\ m &= -\rho c_0, \end{aligned}$$

where ρ is some constant to be determined. It is obvious that the optimal tax (n, m) must be such that $V(n, m) = L$. Let V_0 be the net global returns from the fishery, i.e.,

$$V_0 = \max \left\{ \int_0^{T_1} e^{-\delta t}(p_0 q x(t) - c_0) E(t) dt \right\}.$$

Then

$$\begin{aligned} (1-\rho)V_0 &= \max \left\{ \int_0^{T_1} e^{-\delta t}[(1-\rho)p_0 q x(t) - (1-\rho)c_0] E(t) dt \right\} \\ &= \max \left\{ \int_0^{T_1} e^{-\delta t}[(p_0 - n) q x(t) - (c_0 + m)] E(t) dt \right\} \\ &= V(n, m) = L, \end{aligned} \tag{36}$$

from which it follows that $\rho = (V_0 - L)/V_0$. (36) also indicates that $E(t)$ will maximize the global net returns from the fishery. Hence the above necessary condition for optimality is indeed satisfied by $\lambda = 1$, $r = 1$, $\hat{r} = 0$, $n = \rho p_0$, $m = -\rho c_0$, and the corresponding fishing effort $E(t)$ since equations (29), (30), and (31) are necessary for $E(t)$ to maximize the net global returns from the fishery; (32), (33), and (34) are the necessary optimality conditions for the lower-level problem; and the rest of equations are easily seen to hold. The results agree with the work of Clarke and Munro [7].

Acknowledgments. The author would like to thank Philip Loewen for many helpful suggestions that led to improvements in the results of this paper and two anonymous referees for comments that led to the improvements in the style of the presentation in this version.

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 baised information patterns*, IEEE Trans. Automat. Control, 6 (1972), pp. 791–798.

[3] C. W. CLARK, *Mathematical Bioeconomics: The Optimal Management of Renewable Resources*, 2nd ed., John Wiley and Sons, New York, 1990.

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

[5] ———, *Methods of Dynamic and Nonsmooth Optimization*, NSF-CBMS Regional Conf. Ser. in Appl. Math. 57, Society for Industrial and Applied Mathematics, Philadelphia, 1989.

[6] F. H. CLARKE AND P. D. LOEWEN, *The value function in optimal control: Sensitivity, controllability, and time-optimality*, SIAM. J. Control Optim., 24 (1986), pp. 243–263.

[7] F. H. CLARKE AND G. R. MUNRO, *Coastal states, distant water fishing nations and extended jurisdiction: A principal-agent analysis*, Natural Resource Modeling, 2 (1987), pp. 87–107.

[8] F. H. CLARKE AND G. R. MUNRO, *Coastal states and distant water fishing nations: Conflicting views of the future*, Natural Resource Modeling, to appear.

[9] R. T. ROCKAFELLAR, *Extensions of subgradient calculus with applications to optimization*, Nonlinear Analysis, Theory, Methods Appl., 9 (1985), pp. 665–698.

[10] H. VON STACKELBERG, *The Theory of the Market Economy*, Oxford University Press, Oxford, 1952.

[11] J. J. YE, *Perturbed infinite horizon optimal control problems*, J. Math. Anal. Its Appl., 182 (1994), pp.90–112.

[12] J. J. YE AND D. L. ZHU, *Optimality conditions for bilevel programming problems*, Optimization, to appear.

[13] R. ZHANG, *Problems of Hierarchical Optimization: Nonsmoothness and Analysis of Solutions*, Ph.D. Thesis, University of Washington, 1990.

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