

***NECESSARY CONDITIONS FOR BILEVEL
DYNAMIC OPTIMIZATION PROBLEMS***

JANE J. YE

DMS-626-IR

**March 1993
[Revised July 1993]**

NECESSARY CONDITIONS FOR BILEVEL DYNAMIC OPTIMIZATION PROBLEMS*

Jane J. Ye[†]

July 29, 1993

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 which 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.

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

AMS(MOS) subject classification: 90D65, 49K40

Abbreviated Title: Bilevel Dynamic Optimization Problems

*This work was supported by NSERC under grant WFA 0123160

[†]Department of Mathematics and Statistics, University of Victoria, Victoria, B.C., Canada V8W

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 to select his decision vector $z \in Z$ first and the follower to select his decision control function $u(\cdot) \in \mathcal{U}$ after that, 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 :

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

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

$$\begin{aligned} P_1 \quad \min \quad & J_1(z, u_z) = \int_{t_0}^{t_1} F(t, x_z(t), z, u_z(t))dt + f(x_z(t_1)) \\ \text{over} \quad & z \in Z \text{ and all optimal pairs } (x_z, u_z) \text{ of } P_2(z). \end{aligned}$$

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 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 and follower’s decisions are control functions were 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 literature such as Stackelberg game, principal-agent problem, bilevel programming problem and two level hierarchical optimization problem, and so on. 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 $V : Z \rightarrow \bar{R}$ defined by

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

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 the following single level optimal control problem:

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

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 derive 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) $Z \subset \mathbb{R}^n$ and C_1 are closed;
- (A2) $U(t) : [t_0, t_1] \rightarrow \mathbb{R}^m$ 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 $\text{Gr}U$, is $\mathcal{L} \times \mathcal{B}$ measurable, where $\mathcal{L} \times \mathcal{B}$ denotes the σ -algebra of subsets of $[t_0, t_1] \times \mathbb{R}^m$ 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 \mathbb{R}^m ;
- (A3) There exists an $\mathcal{L} \times \mathcal{B}$ measurable real-valued function k defined on $\text{Gr}U$ such that for each $(t, u) \in \text{Gr}U$, the functions $\phi(t, \cdot, \cdot, u)$, $F(t, \cdot, \cdot, u)$, $G(t, \cdot, \cdot, u)$ are locally Lipschitz of rank $k(t)$, where $k(t)$ is an integrable function. For each $(x, z) \in \mathbb{R}^d \times \mathbb{R}^n$, the functions $\phi(\cdot, x, z, \cdot) : [t_0, t_1] \times \mathbb{R}^m \rightarrow \mathbb{R}^d$, $F(\cdot, x, z, \cdot) : [t_0, t_1] \times \mathbb{R}^m \rightarrow \mathbb{R}$, $G(\cdot, x, z, \cdot) : [t_0, t_1] \times \mathbb{R}^m \rightarrow \mathbb{R}$ are $\mathcal{L} \times \mathcal{B}$ measurable;
- (A4) The functions $f, g : \mathbb{R}^d \rightarrow \mathbb{R}$ 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 $x(\cdot) : [t_0, t_1] \rightarrow \mathbb{R}^d$ is an arc which 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 *admissible trajectory* and *admissible control* respectively. A *solution* to problem $P_2(z)$ is an admissible pair which 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 section 2, we give background material on nonsmooth analysis which will be referred to in the following sections. In section 3, we study generalized differentiability of the value function $V(z)$. The necessary condition for optimality is given in section 4. In section 5, we consider an extension to the bilevel dynamic optimization problem defined in section 1 to allow opportunity costs and 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 which will be required.

Let C be a nonempty closed set in \mathbb{R}^n . A vector $\zeta \in \mathbb{R}^n$ 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 $x_k \in C, x_k \rightarrow \bar{x}$, and proximal normals ζ_k to C at x_k , such that $\zeta_k \rightarrow \zeta$. 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

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

Consider now a lower semicontinuous function $\phi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ 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

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

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

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

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

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

The limiting subgradient is a smaller object than the *Clarke generalized gradient*. In fact, if ϕ is Lipschitz continuous near x , we have $\partial\phi(x) = \text{co}\hat{\partial}\phi(x)$, where ∂ and co denote the Clarke generalized gradient and convex hull of 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] for more details.

The following proposition summarizes the requisites 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., any vector $\zeta \in \hat{N}_C(\bar{x})$ 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) For all $\lambda \geq 0$, one has

$$\hat{\partial}(\lambda\phi)(x) = \lambda\hat{\partial}\phi(x).$$

(d) [Proposition 1.5, Clarke [5]] Let ϕ and $\psi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be lower semi-continuous 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 subsets 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 $F : \mathbb{R}^n \rightarrow \mathbb{R}$ is Lipschitz on some neighbourhood of \bar{x} , $f : \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous with $F(\bar{x})$ in $\text{dom} f := \{y : f(y) \neq +\infty\}$. Then if

$$0 \notin \hat{\partial} \lambda F(\bar{x}) \quad \text{for all nonzero vector } \lambda \in \hat{\partial}^\infty f(F(\bar{x})),$$

then we have

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

3 Differentiability of the value function

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

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

(A6)' The functions ϕ and G are continuous 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

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

is convex.

Let Hamiltonian for $P_2(z)$ be 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

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

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 trajectory x to problem $P_2(z)$. Let

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

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 properties 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. Suppose $QM^0(Y) = \{0\}$, 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 differentiable inclusion problem and apply the sensitivity result in Clarke and Loewen [Theorem 3.3, [6]]. Before proving Theorem 3.1 under assumptions (A1)–(A5) and (A6)'–(A7)', we first give the following result:

Lemma 3.1 *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], Theorem 3.1.7] by studying differential inclusion

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

where $y \in \mathbb{R}$ 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. By Lemma 4.1, V is lower-semicontinuous.

Step 1. Let $\alpha \in Z$ and α is near z . Let $\zeta \in \partial^\pi V(\alpha)$ and let (x, u) be a solution of $P_2(\alpha)$ which exists by virtue of Lemma 4.1. 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

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

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

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

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 with out modification [cf. Clarke [5]]. Applying Theorem 5.2.1. of Clarke [4] with the Clarke generalized gradient replaced by the limiting subgradient in the transversality conditions to the above problem leads to the existence of a scalar $\lambda \geq 0$ and arcs p_2, q such that

$$-\dot{p}_2(t) = \nabla_x \phi(t, x(t), \alpha, u(t))^\top p_2(t) - \lambda \nabla_x G(t, x(t), \alpha, u(t)) \quad \text{a.e.} \quad (1)$$

$$-\dot{q}(t) = \nabla_{\alpha} \phi(t, x(t), \alpha, u(t))^{\top} p_2(t) - \lambda \nabla_{\alpha} G(t, x(t), \alpha, u(t)) \quad \text{a.e.} \quad (2)$$

$$\begin{aligned} \max_{u \in U(t)} \{p_2(t) \cdot \phi(t, x(t), \alpha, u) - \lambda G(t, x(t), \alpha, u)\} \\ = p_2(t) \cdot \phi(t, x(t), \alpha, u(t)) - \lambda G(t, x(t), \alpha, u(t)) \quad \text{a.e.} \end{aligned} \quad (3)$$

$$-p_2(t_1) \in \lambda \hat{\partial} g(x(t_1)) + \hat{N}_{C_1}(x(t_1))$$

$$q(t_0) = -\lambda \zeta, \quad q(t_1) = 0$$

$$\|p_2\|_{\infty} + \|q\|_{\infty} + \lambda > 0,$$

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

By [Theorem 2.8.2, Clarke [4]], (1) (2) and (3) implies that

$$(-\dot{p}_2(t), -\dot{q}(t), \dot{x}(t)) \in \partial_{(x, \alpha, p_2)} H_2(t, x(t), \alpha, p_2(t); \lambda) \quad \text{a.e.}$$

Step 2. For any $\zeta \in \hat{\partial} V(z)$, by definition, $\zeta = \lim_{i \rightarrow \infty} \zeta_i$ where $\zeta_i \in \partial^{\pi} V(\alpha_i)$, $\alpha_i \rightarrow z$ and $V(\alpha_i) \rightarrow V(z)$. 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 which solves $P_2(\alpha_i)$ such that

$$(-\dot{p}_2^i(t), -\dot{q}_i(t), \dot{x}_i(t)) \in \partial_{(x, \alpha, p_2)} H_2(t, x_i(t), \alpha_i, p_2^i(t); \lambda_i) \quad \text{a.e.}$$

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

$$q_i(t_0) = -\lambda_i \zeta_i, \quad q_i(t_1) = 0$$

$$\|p_2^i\| + \|q_i\| + \lambda_i > 0.$$

Since $M^0(Y) = \{0\}$, we must indeed have $\lambda_i = 1$ for i sufficiently large and $|p_2^i(0)|$ bounded (cf. p.253 of Clarke and Loewen [6]). Passing to a uniformly convergent subsequence of $\{(p_2^i, q_i, x_i)\}$ by Lemma 3.1 and [Theorem 3.1.7, Clarke [4]] leads to an optimal trajectory x for $P_2(z)$ and an arc (p_2, q) such that

$$(-\dot{p}_2(t), -\dot{q}(t), \dot{x}(t)) \in \partial_{(x, \alpha, p_2)} H_2(t, x(t), \alpha, p_2(t); \lambda) \quad \text{a.e.}$$

$$-p_2(t_1) \in \hat{\partial} g(x(t_1)) + \hat{N}_{C_1}(x(t_1))$$

$$q(t_0) = -\zeta, \quad q(t_1) = 0.$$

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

Similarly as in 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 (\tilde{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 $t \in [t_0, t_1]$, $x, p_1 \in \mathbb{R}^d$, $z \in Z$, $\lambda, r \in \mathbb{R}$.

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. Suppose that $\hat{\partial}g(x(t_1)) \cap \hat{N}_{C_1}(x(t_1)) = \{0\}$. 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:*

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

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

$$\eta(t_0) = 0$$

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

$$\eta(t_1) \in r \partial V(z) \quad (7)$$

$$\|p_1\|_\infty + \|\eta\|_\infty + \lambda + r > 0. \quad (8)$$

The following results will be useful in proving Theorem 4.1.

Lemma 4.1 Let $C = \{x : \psi(x) \leq 0\}$, where $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$ is Lipschitz continuous on some neighborhood of $\bar{x} \in C$. Suppose that $0 \notin \hat{\partial}\psi(\bar{x})$. Then

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

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

$$0 \notin \hat{\partial}r\psi(\bar{x}) \quad \forall \text{ nonzero vector } r \in \hat{\partial}^\infty \Psi_{\mathbb{R}_-}(\psi(\bar{x})) = \hat{N}_{\mathbb{R}_-}(\psi(\bar{x})) = \mathbb{R}_+.$$

Since $\Psi_C(\bar{x}) = \Psi_{\mathbb{R}_-}(\psi(\bar{x}))$, by the chain rule [(g) of Proposition 2.1] we have

$$\hat{\partial}\Psi_C(\bar{x}) \subset \cup \{\hat{\partial}r\psi(\bar{x}) : r \in \hat{\partial}\Psi_{\mathbb{R}_-}(\psi(\bar{x}))\}, \quad (10)$$

which is the relation (9) since

$$\hat{\partial}\Psi_{\mathbb{R}_-}(\psi(\bar{x})) = \hat{N}_{\mathbb{R}_-}(\psi(\bar{x})) = \mathbb{R}_+, \quad \hat{\partial}\Psi_C(\bar{x}) = \hat{N}_C(\bar{x}).$$

■

The proof of the following result is straightforward.

Lemma 4.2 Let $F(x, y, z) : \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ 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 \tilde{P}_1 equivalently as the following problem

$$\begin{aligned} \hat{P}_1 \quad & \min \quad \int_{t_0}^{t_1} F(t, x(t), z(t), u(t)) dt + f(x(t_1)) \\ & \text{s.t.} \quad \dot{x}(t) = \phi(t, x(t), z(t), u(t)) \quad \text{a.e.} \\ & \quad \dot{y}(t) = G(t, x(t), z(t), u(t)) \quad \text{a.e.} \\ & \quad \dot{z}(t) = 0 \\ & \quad u(t) \in U(t) \quad \text{a.e.} \\ & \quad (x, y, z)(t_0) \in \{x_0\} \times \{0\} \times \mathbb{R}, \\ & \quad (x, y, z)(t_1) \in S := \{(x, y, z) : g(x) + y - V(z) \leq 0, x \in C_1\} \end{aligned}$$

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 $t \in [t_0, t_1]$, $x, p_1 \in \mathbb{R}^d$, $y, p_2, \eta, \lambda \in \mathbb{R}$, $z \in Z$. 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 $\lambda \geq 0$ and an arc (p_1, p_2, η) such that

$$-(\dot{p}_1(t), \dot{p}_2(t), \dot{\eta}(t)) \in \partial_{(x, y, z)} H(t, x(t), y(t), z(t), p_1(t), p_2(t), \eta(t), u(t), \lambda) \quad \text{a.e.} \quad (11)$$

$$\begin{aligned} & \max_{u \in U(t)} H(t, x(t), y(t), z(t), p_1(t), p_2(t), \eta(t), u, \lambda) \\ & = H(t, x(t), y(t), z(t), p_1(t), p_2(t), \eta(t), u(t), \lambda) \quad \text{a.e.} \end{aligned} \quad (12)$$

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

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

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

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

Let $\hat{F}(x, y, z) = g(x) + y - V(z)$. Then by Lemma 4.2, one has

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

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

Let $S_1 := \{(x, y, z) : g(x) + y - V(z) \leq 0\}$ and $S_2 := C_1 \times \mathbb{R} \times \mathbb{R}$. By Lemma 4.1 and inclusion (16), one has

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

Since $\Psi_{S_2}(x, y, z) = \Psi_{C_1}(x) + \Psi_{\mathbb{R}}(y) + \Psi_{\mathbb{R}}(z)$, by Lemma (4.2) and (e) of Proposition 2.1 one has

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

By assumption $\hat{\partial} g(x(t_1)) \cap \hat{N}_{C_1}(x(t_1)) = \{0\}$, hence $\hat{N}_{S_1}(x(t_1), y(t_1), z(t_1)) \cap (-\hat{N}_{S_2}(x(t_1), y(t_1), z(t_1))) = \{0\}$. Applying (f) of Proposition 2.1, we have

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

By Lemma 4.2, 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

$$-(p_1(t_1), p_2(t_1), \eta(t_1)) \in \lambda \hat{\partial} f(x(t_1)) \times \{0\} \times \{0\}$$

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

from which the transversality condition (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

$$-(\dot{p}_1(t), \dot{\eta}(t)) \in \partial_{(x,z)} H(t, x(t), y(t), z(t), p_1(t), p_2(t), \eta(t), u(t); \lambda) \quad \text{a.e..} \quad (17)$$

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 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 4.2 *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 $\hat{\partial} g(x(t_1)) \cap \hat{N}_{C_1}(x(t_1)) = \{0\}$ and $QM^0(Y) = \{0\}$. 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, η, p_{ij}, q_{ij} such that*

$$-(\dot{p}(t), \dot{\eta}(t)) \in \partial_{(x,z)} H_1(t, x(t), z, p(t), u(t); \lambda, r) \quad \text{a.e.} \quad (18)$$

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

$$\eta(t_0) = 0$$

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

$$\eta(t_1) = r \sum_{ij} \lambda_{ij} q_{ij}(t_0)$$

$$(-\dot{p}_{ij}(t), -\dot{q}_{ij}(t), \dot{x}_i(t)) \in \partial_{(x,z,p)} H_2(t, x_i(t), z, p_{ij}(t); 1) \quad \text{a.e.} \quad (19)$$

$$q_{ij}(t_1) = 0$$

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

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

Remark 4.1 A sufficient condition for both $QM^0(Y) = \{0\}$ and $\hat{\partial}g(x) \cap N_{C_1}(x) = \{0\}$, for all $x \in C_1$ to hold is $C_1 = \mathbb{R}^d$. Indeed, in the case $C_1 = \mathbb{R}^d$, the index 0 multiplier set are the set of arcs (p_2, q) such that

$$(-\dot{p}_2(t), -\dot{q}(t), \dot{x}(t)) \in \partial_{(x,z,p_2)} H_2(t, x(t), z, p_2(t); 0) \quad \text{a.e.} \quad (20)$$

$$p_2(t_1) = 0 \quad (21)$$

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

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 $QM^0(Y) = \{0\}$.

A 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 If in addition to assumptions (A1)–(A5) and (A6)'–(A7)', we assume the set

$$\{(\nabla_{x,z}\phi(t, x, z, u)^\top(p_2, q), \nabla_{x,z}G(t, x, z, u)) : u \in U(t)\}$$

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

$$-\dot{p}_{ij}(t) = \nabla_x \phi(t, x_i(t), z, u_i(t))^\top p_{ij}(t) - \nabla_x G(t, x_i(t), z, u_i(t)) \quad \text{a.e.}$$

$$-\dot{q}_{ij}(t) = \nabla_z \phi(t, x_i(t), z, u_i(t))^\top p_{ij}(t) - \nabla_z G(t, x_i(t), z, u_i(t)) \quad \text{a.e.}$$

$$\begin{aligned} & \max_{u \in U(t)} \{p_{ij}(t) \cdot \phi(t, x_i(t), z, u) - G(t, x_i(t), z, u)\} \\ & p_{ij}(t) \cdot \phi(t, x_i(t), z, u_i(t)) - G(t, x_i(t), z, u_i(t)) \quad \text{a.e.,} \end{aligned}$$

$$\dot{x}_i(t) = \phi(t, x_i(t), z, u_i(t)) \quad \text{a.e.}$$

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 or equal the opportunity

cost $L \geq 0$ he may receive from somewhere else. In this case, the leader faces the following bilevel optimization problem:

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

The technique described in the previous section can be applied to this more general problem in exactly 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 \bar{P}_1 and $x(t)$ the corresponding trajectory. Suppose that for all $\hat{\partial}g(x(t_1)) \cap \hat{N}_{C_1}(x(t_1)) = \{0\}$ and $QM^0(Y) = \{0\}$. 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, η, p_{ij}, q_{ij} such that*

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

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 presense 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 fish population at time t follows the dynamic system

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

where $x(t)$ is the fish population at time t ; $F(x)$ is the natural growth rate; $E(t)$ is the fishing effort at time t ; and q is a positive constant. We assume that $F(x) > 0$ 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 arbitrary upper bound on $E(t)$. Suppose that the coastal state imposes a condition such that the at the terminal time T_1 , the fish population can not be less than $\tilde{x} \geq 0$.

For given unit tax on catch and effort n and m , the distant water fleet as a follower faces the following optimal control problem:

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

where p_0 and c_0 are the unit price on catch and unit cost on effort, respectively.

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

$$-\dot{p}_2(t) = p_2[F'(x) - qE] + e^{-\delta t}(p_0 - n)qE \quad (23)$$

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

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

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

$$\begin{aligned} p_2(t) &= e^{-\delta t}[(p_0 - n) - \frac{c_0 + m}{qx}] \\ \dot{p}_2(t) &= e^{-\delta t}[-\delta[(p_0 - n) - \frac{c_0 + m}{qx}] + \frac{c_0 + m}{qx^2} \frac{dx}{dt}]. \end{aligned}$$

From the adjoint equation (23), one has

$$\begin{aligned} \dot{p}_2(t) &= -p_2[F'(x) - qE] - e^{-\delta t}[(p_0 - n)qE] \\ &= -e^{-\delta t}\{[(p_0 - n) - \frac{c_0 + m}{qx}][F'(x) - qE] + (p_0 - n)qE\}, \end{aligned}$$

where (23) is used for $p_2(t)$. When these two expressions for $\dot{p}_2(t)$ are equated, the control variable E cancels out and the following equation emerges:

$$F'(x) + \frac{F(x)(c_0 + m)/qx^2}{p_0 - n - (c_0 + m)/qx} = \delta. \quad (25)$$

For fixed (n, m) , this equation gives a unique solution x_* which is the optimal biomass and the optimal trajectory is the one which 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 distinct 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.

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

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

$$-\dot{p}_1 = p_1[F'(x) - qE] + e^{-\delta t}[r(p_0 - n) + \lambda n]qE \quad (26)$$

$$\dot{\eta}_1 = (r - 1)e^{-\delta t}qx E$$

$$\dot{\eta}_2 = (r - 1)e^{-\delta t}E$$

$$\begin{aligned} & \max_{E \in [0, E_{\max}]} \{p_1(t)[F(x(t)) - qEx(t)] + e^{-\delta t}[r[(p_0 - n)qx(t) - (c_0 + m)] + \lambda(nqx(t) + m)]E\} \\ & = p_1(t)[F(x(t)) - qE(t)x(t)] + e^{-\delta t}[r[(p_0 - n)qx(t) - (c_0 + m)] + \lambda(nqx(t) + m)]E(t) \end{aligned} \quad (27)$$

$$(\eta_1, \eta_2)(0) = (0, 0)$$

$$p_1(T_1) \geq 0 \quad (28)$$

$$(\eta_1, \eta_2)(T_1) = \hat{r}q(0)$$

$$-\dot{p}_2 = p_2[F'(x) - qE] + e^{-\delta t}(p_0 - n)qE \quad (29)$$

$$\dot{q}_1 = e^{-\delta t}qx E$$

$$\dot{q}_2 = e^{-\delta t}E$$

$$\begin{aligned} & \max_{E \in [0, E_{\max}]} \{p_2(t)[F(x(t)) - qEx(t)] + e^{-\delta t}[(p_0 - n)qx(t) - (c_0 + m)]E\} \\ & = p_2(t)[F(x(t)) - qE(t)x(t)] + e^{-\delta t}[(p_0 - n)qx(t) - (c_0 + m)]E(t) \end{aligned} \quad (30)$$

$$(q_1, q_2)(0) = (0, 0)$$

$$p_2(T_1) \geq 0 \quad (31)$$

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

Without loss of generality, take $\lambda = 1$. From (26) and (27), we can show that the steady state (n, m, x_*) for problem P_1 is a solution of the following equation:

$$F'(x_*) + \frac{F(x_*)(r(c_0 + m) - m)/qx_*^2}{r(p_0 - n) + n - (r(c_0 + m) - m)/qx_*} = \delta. \quad (32)$$

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 (25). Combining two equations (25) and (32), one has

$$n = \rho p_0$$

$$m = -\rho c_0,$$

where ρ is some constant. As shown in Clarke and Munro [7], the optimal tax (n, m) must be such that $V(n, m) = L$ from which it follows that $\rho = (V_0 - L)/V_0$ where V_0 is the net global returns from the fishery

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

Clarke and Munro [7] have also shown that the optimal effort policy $E(t)$ will maximize the global net returns from the fishery. Hence the above necessary condition for optimality is indeed satisfied by $\lambda = 1$, $r = 1$, $\hat{r} = 0$ since equations (26), (27) and (28) is the necessary condition for $E(t)$ to maximize the net global returns from the fishery, (29), (30) and (31) is the necessary condition for the lower level problem and the rest of equations are easily seen to hold.

6 Acknowledgements

The author would like to thank Philip Loewen for many helpful suggestions which lead to improvements 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), 77-100.
- [2] C.I. Chen and J.B. Cruz Jr., Stackelberg solution for two-person games with biased information patterns, *IEEE trans. Auto. Control*, **6**(1972), 791-798.
- [3] C.W. Clark, *Mathematical Bioeconomics: The Optimal Management of Renewable Resources*, Second Edition, John Wiley and Sons, New York, 1990.
- [4] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley-Interscience, New York, 1983.
- [5] F.H. Clarke, *Methods of Dynamic and Nonsmooth Optimization*, NSF-CBMS Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics, Philadelphia, vol 57, 1989.
- [6] F.H. Clarke and P.D. Loewen, The value function in optimal control: sensitivity, controllability, and time-optimality, *SIAM. J. Control and Optimization*, **24**(1986), 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), 87-107.
- [8] F.H. Clarke and G.R. Munro, Coastal states and distant water fishing nations: conflicting views of the future, To appear in *Natural Resource Modeling*.
- [9] R.T. Rockafellar, Extensions of subgradient calculus with applications to optimization, *Nonlinear Analysis, Theory, Methods and Applications*, **9**(1985), 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, To appear in *Journal of Mathematical Analysis and Its Applications*.
- [12] J.J. Ye and D.L. Zhu, Optimality conditions for bilevel programming problems, Submitted for publication.
- [13] Roxin Zhang, Problems of Hierarchical Optimization: Nonsmoothness and Analysis of Solutions, University of Washington, Ph.D. Thesis, 1990.
- [14] Roxin Zhang, Problems of hierarchical optimization in finite dimensions, To appear in *SIAM J. Optimization*.