Dynamic Programming Approach to Discrete Time Dynamic Feedback Stackelberg Games with Independent and Dependent Followers [1]

Pu-yan Nie [2 3] Li-hua Chen [4] Masao Fukushima[5]

October 24, 2003
Revised March 19, 2004

Abstract Stackelberg games play extremely important roles in such fields as economics, management, politics and behavioral sciences. Stackelberg game can be modelled as a bilevel optimization problem. There exists extensive literature about static bilevel optimization problems. However, the studies on dynamic bilevel optimization problems are relatively scarce in spite of the importance in explaining and predicting some phenomena rationally. In this paper, we consider discrete time dynamic Stackelberg games with feedback information. Dynamic programming algorithms are presented for the solution of discrete time dynamic feedback Stackelberg games with multiple players both for independent followers and for dependent followers. When the followers act dependently, the game in this paper is a combination of Stackelberg game and Nash one.

AMS Subject classification (2000): 49L20, 91A10, 91A25, 91A65

Key words: Discrete time dynamic model, Stackelberg game, feedback information, dynamic programming

1 Introduction

In many decision processes there exists a hierarchy of decision makers. Decisions are made at different levels with different goals in this hierarchy. Moreover, those decision makers often cannot act independently of each other but have to take into account decisions made by players of different levels. Especially, in the bilevel case, the optimal strategies chosen by the lower level players

[1] This work was partially supported by the Informatics Research Center for Development of Knowledge Society Infrastructure, Graduate School of Informatics, Kyoto University, Japan
[2] Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto, 606-8501, Japan
[3] Department of Mathematics, College of Information Science and Technology, Jinan University, Guangzhou, 510632, China
[4] Guanghua School of Management, Peking University, Beijing, 100871, China, chenlh@gsm.pku.edu.cn
[5] Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto, 606-8501, Japan, fuku@i.kyoto-u.ac.jp.

(hereafter the "followers") depend on the strategy selected by the upper level player (hereafter the "leader"). Furthermore, the objective function of the leader may depend not only on his/her own decisions but also on the followers'. Then, the leader is able to make his/her decisions by estimating the followers' rational reactions, assuming that they behave in such a way that they optimize their objective functions given the leader's actions. This is the static bilevel model introduced by Von Stackelberg [24]. There exists extensive research on bilevel optimization [2, 3, 11, 20]. Under certain conditions, a bilevel optimization problem can be reformulated as a mathematical program with equilibrium constraints (MPEC) [17], which has recently drawn much attention in the optimization community [13, 14, 16, 18, 19]. However, the studies on dynamic bilevel optimization are relatively scarce. Dynamic bilevel optimization was first introduced by Chen and Cruz [10], and Simaan and Cruz [21, 22] in nonzero-sum dynamic games. Dynamic bilevel optimization problems have subsequently been studied by a number of authors [1, 4, 5, 8, 9, 12, 15]. In particular, Ye [25, 26] recently studies optimality conditions for continuous time dynamic bilevel optimization problems.

The discrete time dynamic optimization problems have many applications in economics and management sciences [1, 12]. See an excellent monograph [4] on dynamic games. Linear-quadratic (LQ) systems are considered in [5, 6, 8, 9, 10, 15, 21]. In [8], stochastic dynamic Stackelberg games with two players are introduced and explicit solutions are given when the information sets are nested. In [10, 21], necessary and sufficient conditions for the Stackelberg games with open-loop information have been obtained and explicit solutions are given. In [23], Stackelberg solution is extended to multi-players and necessary conditions for the existence of an open-loop Stackelberg games are shown. In [22], Stackelberg games with feedback information are considered and necessary conditions for the existence of Stackelberg strategies are obtained. In [5], explicit solutions are also given for deterministic games with multiple players under closed-loop information, in particular, feedback information structure. More recently, in [15], an incentive strategy for discrete time LQ state feedback Stackelberg games is developed. Moreover, in [1], recursive methods are presented for dynamic Stackelberg games with two players. In [9], a new feedback solution called anticipative feedback solution is introduced to cope with the infinite-horizon, linear-quadratic, dynamic, Stackelberg games. All methods to deal with LQ systems and dynamic Stackelberg games with two players are based on the special structure of the problems. In [6], continuous time dynamic Stackelberg games with closed-loop information are considered. In [12], pricing and advertising models in a market are modelled as continuous time dynamic Stackelberg games.

In this paper, we consider a general dynamic Stackelberg game under feedback information structure that may be modelled as a discrete time dynamic bilevel optimization problem, where the upper level state variables are influenced by the decisions of the leader, and the lower level state variables are related to the decisions of the leader and the followers. Assuming the feedback information structure, we will apply dynamic programming algorithms to the game with dependent followers as well as the one with independent followers.

Let us give the formal statement of the problem. The discrete time periods are denoted $t = 0, 1, \cdots, T$, and N is the number of followers in the game.

The variables involved in the problem are listed as follows:

Vectors $x_t \in \mathcal{X} \subset R^{m_0}$ denote the state of the leader at time $t = 0, 1, \cdots, T$.

Vectors $y_t^\nu \in \mathcal{Y}^\nu \subset R^{m_\nu}$ denote the state of the νth follower at time $t = 0, 1, \cdots, T$. The followers' state variables at time t are collectively denoted $y_t := (y_t^1, y_t^2, \cdots, y_t^N) \in \mathcal{Y} := \mathcal{Y}^1 \times \mathcal{Y}^2 \times \cdots \times \mathcal{Y}^N \subset R^m$ with $m = m_1 + m_2 + \cdots + m_N$.

Vectors $u_t \in \mathcal{U} \in R^{n_0}$ denote the decision variables for the leader at time $t = 0, 1, \cdots, T-1$.

Vectors $v_t^\nu \in \Pi_t^\nu(x_t, y_t^\nu, u_t) \subset \mathcal{V}^\nu \subset R^{n_\nu}$ denote the decision variables for the νth follower at time $t = 0, 1, \cdots, T-1$. (The definition of $\Pi_t^\nu(x_t, y_t^\nu, u_t)$ is given below). The followers' decision variables at time t are collectively denoted $v_t := (v_t^1, v_t^2, \cdots, v_t^N) \in \mathcal{V} := \mathcal{V}^1 \times \mathcal{V}^2 \times \cdots \times \mathcal{V}^N \subset R^n$ with $n = n_1 + n_2 + \cdots + n_N$.

Moreover, we denote

$$\begin{array}{ll} u := (u_0, u_1, \cdots, u_{T-1}) \in R^{n_0 T}, & x := (x_0, x_1, \cdots, x_T) \in R^{m_0(T+1)}, \\ v^\nu := (v_0^\nu, v_1^\nu, \cdots, v_{T-1}^\nu) \in R^{n_\nu T}, & v := (v^1, v^2, \cdots, v^N) \in R^{nT}, \\ y^\nu := (y_0^\nu, y_1^\nu, \cdots y_T^\nu) \in R^{m_\nu(T+1)}, & y := (y^1, y^2, \cdots, y^N) \in R^{m(T+1)}. \end{array}$$

The state variables $\{x_t\}_{t=0}^T$ and $\{y_t\}_{t=0}^T$ are governed by the systems of state transition equations

$$x_{t+1} = F_t(x_t, u_t), \quad t = 0, 1, \cdots, T-1, \tag{1.1}$$

$$y_{t+1}^\nu = f_t^\nu(x_t, y_t^\nu, u_t, v_t^\nu), \quad t = 0, 1, \cdots, T-1, \quad \nu = 1, 2, \cdots, N \tag{1.2}$$

with x_0 and y_0^ν being given as the initial states of the leader and the νth followers, $\nu = 1, 2, \cdots, N$, respectively.

Let the set of admissible decisions of the leader be given by

$$\Pi^0(x_0) := \{u | h_t^0(x_t, u_t) \leq 0, t = 0, 1, \cdots, T-1\},$$

and the sets of admissible decisions of the followers $\nu = 1, 2, \cdots, N$ be given by

$$\Pi^\nu(x_0, y_0^\nu, u) := \{v^\nu | h_t^\nu(x_t, y_t^\nu, u_t, v_t^\nu) \leq 0, t = 0, 1, \cdots, T-1\},$$

where $h_t^0 : R^{m_0} \times R^{n_0} \longrightarrow R^{\bar{n}_0}$ and $h_t^\nu : R^{m_0} \times R^{m_\nu} \times R^{n_0} \times R^{n_\nu} \longrightarrow R^{\bar{n}_\nu}, \nu = 1, 2, \cdots, N$, are some functions used to specify admissible decisions at each period and $\{x_\tau\}_{\tau=0}^T$ and $\{y_\tau\}_{\tau=0}^T$ are governed by (1.1) and (1.2), where $\bar{n}_0$ and $\bar{n}_\nu$ for $\nu = 1, 2, \cdots, N$ are all integers. Note that the set $\Pi^0(x_0)$, which specifies the set of admissible decisions of the leader at periods $t = 1, \ldots, T-1$, only depends on the initial state x_0, since the subsequent states x_t are determined by (1.1). On the other hand, the set $\Pi^\nu(x_0, y_0^\nu, u)$ of admissible decisions of the νth follower also depends only on the initial state y_0^ν of the follower along with the initial state x_0 and the decision u of the leader.

We also denote

$$\begin{aligned} &\Pi_t^0(x_t) := \{u_t | h_t^0(x_t, u_t) \leq 0\} \subset R^{n_0}, \\ &\Pi_t^\nu(x_t, y_t^\nu, u_t) := \{v_t^\nu | h_t^\nu(x_t, y_t^\nu, u_t, v_t^\nu) \leq 0\} \subset R^{n_\nu}, \\ &\Pi(x_0, y_0, u) := \Pi^1(x_0, y_0^1, u) \times \Pi^2(x_0, y_0^2, u) \times \cdots \times \Pi^N(x_0, y_0^N, u) \subset R^{nT}, \\ &\Pi_t(x_t, y_t, u_t) := \Pi_t^1(x_t, y_t^1, u_t) \times \Pi_t^2(x_t, y_t^2, u_t) \times \cdots \times \Pi_t^N(x_t, y_t^N, u_t) \subset R^n. \end{aligned}$$

The problem is then formally stated as follows: Given the initial state $(x_0, y_0) \in \mathcal{X} \times \mathcal{Y}$,

$$\begin{array}{ll} \text{minimize} & J^0(x_0, y_0, u, v) \\ \text{subject to} & V^\nu(x_0, y_0, u) = J^\nu(x_0, y_0, u, v^\nu), \\ & u \in \Pi^0(x_0), \\ & v^\nu \in \Pi^\nu(x_0, y_0^\nu, u), \quad \nu = 1, 2, \cdots, N, \end{array} \tag{1.3}$$

where

$$J^0(x_0, y_0, u, v) := G_T(x_T, y_T) + \sum_{t=0}^{T-1} G_t(x_t, y_t, u_t, v_t),$$

$$J^\nu(x_0, y_0, u, v^\nu) := g_T^\nu(x_T, y_T^\nu) + \sum_{t=0}^{T-1} g_t^\nu(x_t, y_t^\nu, u_t, v_t^\nu),$$

$\{x_t\}_{t=0}^T$ and $\{y_t\}_{t=0}^T$ are determined by (1.1) and (1.2), respectively, and $V^\nu(x_0, y_0, u)$ are the optimal value functions for followers $\nu = 1, 2, \cdots, N$ defined by

$$V^\nu(x_0, y_0, u) := \min\{J^\nu(x_0, y_0, u, v^\nu) | (1.1), (1.2) \text{ and } v^\nu \in \Pi^\nu(x_0, y_0^\nu, u)\}.$$

In the following, we will sometimes write

$$f_t := (f_t^1, f_t^2, \cdots, f_t^N), \quad h_t := (h_t^1, h_t^2, \cdots, h_t^N),$$

for $t = 0, 1, \cdots, T-1$.

We refer to problem (1.3) as a dynamic bilevel optimization problem or DBOP for short. The paper is organized as follows: A dynamic programming algorithm for DBOP is proposed and its validity is shown in Section 2. The results are extended to DBOP with dependent followers in Section 3. Some remarks are given in the final section.

2 Dynamic Programming Algorithm for DBOP with Independent Followers under Feedback Information Structure

The aim of this section is to develop a dynamic programming algorithm for DBOP (1.3) under feedback information structure, which is an extension of the one presented in [10, 21, 22, 23] (see also [4, Section 7.3]) and is based on the principle of optimality stated in Theorem 2.1 below. In this section, we assume that the followers act independently of each other. Namely, when a follower makes a decision, he/she only takes into account the leader's action and will not consider other followers. Here and throughout, we assume that an optimal response of the followers is uniquely determined for any decisions of the leader.

For convenience, we will use the following notation for $t = 0, 1, \cdots, T-1$:

$$\begin{aligned}
&x_{t,T-1} := (x_t, \cdots, x_{T-1}), \quad y_{t,T-1} := (y_t, \cdots, y_{T-1}), \quad y_{t,T-1}^\nu := (y_t^\nu, \cdots, y_{T-1}^\nu),\\
&u_{t,T-1} := (u_t, \cdots, u_{T-1}), \quad v_{t,T-1} := (v_t, \cdots, v_{T-1}), \quad v_{t,T-1}^\nu := (v_t^\nu, \cdots, v_{T-1}^\nu),\\
&x_{0,t} := (x_0, \cdots, x_t), \quad y_{0,t}^\nu := (y_0^\nu, \cdots, y_t^\nu), \quad v_{0,t}^\nu := (v_0^\nu, \cdots, v_t^\nu),\\
&J_{T-t}^0(x_t, y_t, u_{t,T-1}, v_{t,T-1}) := G_T(x_T, y_T) + \sum_{\tau=t}^{T-1} G_\tau(x_\tau, y_\tau, u_\tau, v_\tau),\\
&J_{T-t}^\nu(x_t, y_t^\nu, u_{t,T-1}, v_{t,T-1}^\nu) := g_T(x_T, y_T^\nu) + \sum_{\tau=t}^{T-1} g_\tau(x_\tau, y_\tau^\nu, u_\tau, v_\tau^\nu),\\
&\Pi_{t,T-1}^0(x_t) := \{u_{t,T-1} | h_\tau^0(x_\tau, u_\tau) \le 0, \tau = t, t+1, \cdots, T-1\},\\
&\Pi_{t,T-1}^\nu(x_t, y_t^\nu, u_{t,T-1}) := \{v_{t,T-1}^\nu | h_\tau^\nu(x_\tau, y_\tau^\nu, u_\tau, v_\tau^\nu) \le 0, \tau = t, t+1, \cdots, T-1\},\\
&\Pi_{t,T-1}(x_t, y_t, u_{t,T-1}) := \Pi_{t,T-1}^1(x_t, y_t^1, u_{t,T-1}) \times \cdots \times \Pi_{t,T-1}^N(x_t, y_t^N, u_{t,T-1}),
\end{aligned}$$

$$\begin{aligned}
\tilde{V}_{T-t}^\nu(x_t, y_t^\nu, u_{t,T-1}) \;&:=\; \min\{J_{T-t}^\nu(x_t, y_t^\nu, u_{t,T-1}, v_{t,T-1}^\nu) :\\
&\qquad (v_t^\nu, v_{t+1}^\nu, \cdots, v_{T-1}^\nu) \in \Pi_{t,T-1}^\nu(x_t, y_t^\nu, u_{t,T-1})\}. \qquad (2.1)
\end{aligned}$$

For each $t = 0, 1, \cdots, T-1$, consider the subproblem

$$\begin{array}{ll} \text{minimize} & J^0_{T-t}(x_t, y_t, u_{t,T-1}, v_{t,T-1}) \\ \text{subject to} & \tilde{V}^\nu_{T-t}(x_t, y^\nu_t, u_{t,T-1}) = J^\nu_{T-t}(x_t, y^\nu_t, u_{t,T-1}, v^\nu_{t,T-1}) \\ & u_{t,T-1} \in \Pi^0_{t,T-1}(x_t), \\ & v^\nu_{t,T-1} \in \Pi^\nu_{t,T-1}(x_t, y^\nu_t, u_{t,T-1}) \\ & \nu = 1, 2, \cdots, N, \end{array} \tag{2.2}$$

where the initial state (x_t, y_t) is given, and $\{x_\tau\}^T_{\tau=t}$ and $\{y_\tau\}^T_{\tau=t}$ are determined by (1.1) and (1.2), respectively. This problem is referred to as $P_{T-t}(x_t, y_t)$.

Theorem 2.1 *Let $(u^\star_0, u^\star_1, \cdots, u^\star_{T-1})$ and $(v^\star_0, v^\star_1, \cdots, v^\star_{T-1})$ constitute an optimal policy for DBOP (1.3) under feedback information structure with corresponding optimal trajectories $(x_0, x^\star_1, \cdots, x^\star_T)$ and $(y_0, y^\star_1, \cdots, y^\star_T)$. Consider the subproblem $P_{T-t}(x^\star_t, y^\star_t)$ for every $t = 1, 2, \cdots, T-1$ with the initial state $(x^\star_t, y^\star_t)$. Then, the truncated policy*

$$\{(u^\star_t, u^\star_{t+1}, \cdots, u^\star_{T-1}), (v^\star_t, v^\star_{t+1}, \cdots, v^\star_{T-1})\}$$

is optimal for the subproblem $P_{T-t}(x^\star_t, y^\star_t)$.

Proof First, we show that the policy $(v^\star_t, v^\star_{t+1}, \cdots, v^\star_{T-1})$ is the optimal response to the leader's decision $(u^\star_t, u^\star_{t+1}, \cdots, u^\star_{T-1})$. If it were not an optimal response of the followers, then there would exist some $\nu \in \{1, 2, \cdots, N\}$ and $(\hat{v}^\nu_t, \hat{v}^\nu_{t+1}, \cdots, \hat{v}^\nu_{T-1}) \in \Pi^\nu_{t,T-1}(x^\star_t, \hat{y}^\nu_t, u^\star_{t,T-1})$ with the corresponding sequences $\{\hat{y}^\nu_\tau\}^T_{\tau=t}$ such that $y^{\nu\star}_t = \hat{y}^\nu_t$ and

$$g^\nu_T(x^\star_T, \hat{y}^\nu_T) + \sum_{\tau=t}^{T-1} g^\nu_\tau(x^\star_\tau, \hat{y}^\nu_\tau, u^\star_\tau, \hat{v}^\nu_\tau) \ < \ g^\nu_T(x^\star_T, y^{\nu\star}_T) + \sum_{\tau=t}^{T-1} g^\nu_\tau(x^\star_\tau, y^{\nu\star}_\tau, u^\star_\tau, v^{\nu\star}_\tau),$$

where the sequences $\{x^\star_\tau\}^T_{\tau=t}$ and $\{y^{\nu\star}_\tau\}^T_{\tau=t}$ are generated by $\{u^\star_\tau\}^T_{\tau=t}$ and $\{v^{\nu\star}_\tau\}^T_{\tau=t}$, respectively. Thus, we have

$$(v^{\nu\star}_{0,t-1}, \hat{v}^\nu_{t,T-1}) \in \Pi^\nu_{0,T-1}(x^\star_0, y^{\nu\star}_0, u^\star_{0,T-1})$$

and

$$\begin{aligned} & g^\nu_T(x^\star_T, \hat{y}^\nu_T) + \sum_{\tau=t}^{T-1} g^\nu_\tau(x^\star_\tau, \hat{y}^\nu_\tau, u^\star_\tau, \hat{v}^\nu_\tau) + \sum_{\tau=0}^{t-1} g^\nu_\tau(x^\star_\tau, y^{\nu\star}_\tau, u^\star_\tau, v^{\nu\star}_\tau) \\ < \ & g^\nu_T(x^\star_T, y^{\nu\star}_T) + \sum_{\tau=0}^{T-1} g^\nu_\tau(x^\star_\tau, y^{\nu\star}_\tau, u^\star_\tau, v^{\nu\star}_\tau). \end{aligned}$$

This indicates that $(v^\star_0, v^\star_1, \cdots, v^\star_{T-1})$ is not an optimal response to $(u^\star_0, u^\star_1, \cdots, u^\star_{T-1})$, which contradicts the assumption of this theorem. Consequently, $(v^\star_t, v^\star_{t+1}, \cdots, v^\star_{T-1})$ is the optimal response to the leader's decision $(u^\star_t, u^\star_{t+1}, \cdots, u^\star_{T-1})$.

To prove the theorem by contradiction, suppose that $\{(u^\star_t, u^\star_{t+1}, \cdots, u^\star_{T-1}), (v^\star_t, v^\star_{t+1}, \cdots, v^\star_{T-1})\}$ is not an optimal solution to the subproblem $P_{T-t}(x^\star_t, y^\star_t)$. Then, there must exist another policy $\{(\bar{u}_t, \bar{u}_{t+1}, \cdots, \bar{u}_{T-1}), (\bar{v}_t, \bar{v}_{t+1}, \cdots, \bar{v}_{T-1})\}$ such that $(\bar{v}_t, \bar{v}_{t+1}, \cdots, \bar{v}_{T-1})$ is the optimal response of the followers to the leader's decision $(\bar{u}_t, \bar{u}_{t+1}, \cdots, \bar{u}_{T-1})$ and

$$G_T(\bar{x}_T, \bar{y}_T) + \sum_{\tau=t}^{T-1} G_\tau(\bar{x}_\tau, \bar{y}_\tau, \bar{u}_\tau, \bar{v}_\tau) \ < \ G_T(x^\star_T, y^\star_T) + \sum_{\tau=t}^{T-1} G_\tau(x^\star_\tau, y^\star_\tau, u^\star_\tau, v^\star_\tau), \tag{2.3}$$

where $\{(x_t^\star, x_{t+1}^\star, \cdots, x_T^\star), (y_t^\star, y_{t+1}^\star, \cdots, y_T^\star)\}$ and $\{(\bar{x}_t, \bar{x}_{t+1}, \cdots, \bar{x}_T), (\bar{y}_t, \bar{y}_{t+1}, \cdots, \bar{y}_T)\}$ are the sequences of the leader's and the followers' states generated by the above-mentioned policies, with the initial conditions $x_t = x_t^\star = \bar{x}_t$ and $y_t = y_t^\star = \bar{y}_t$.

Note that the policies $\{(u_t^\star, u_{t+1}^\star, \cdots, u_{T-1}^\star), (v_t^\star, v_{t+1}^\star, \cdots, v_{T-1}^\star)\}$ and $\{(\bar{u}_t, \bar{u}_{t+1}, \cdots, \bar{u}_{T-1}), (\bar{v}_t, \bar{v}_{t+1}, \cdots, \bar{v}_{T-1})\}$ all satisfy the constraints

$$(v_t^\star, \cdots, v_{T-1}^\star) \in \Pi_{t,T-1}(x_t^\star, y_t^\star, u_{t,T-1}^\star),$$

$$(\bar{v}_t, \cdots, \bar{v}_{T-1}) \in \Pi_{t,T-1}(\bar{x}_t, \bar{y}_t, \bar{u}_{t,T-1})$$

with $x_t^\star = \bar{x}_t$ and $y_t^\star = \bar{y}_t$. Accordingly, we have

$$(v_{0,t-1}^\star, v_{t,T-1}^\star) \in \Pi_{0,T-1}(x_0^\star, y_0^\star, u_{0,t-1}^\star, u_{t,T-1}^\star),$$

$$(v_{0,t-1}^\star, \bar{v}_{t,T-1}) \in \Pi_{0,T-1}(x_0^\star, y_0^\star, u_{0,t-1}^\star, \bar{u}_{t,T-1})$$

with $x_t^\star = \bar{x}_t$ and $y_t^\star = \bar{y}_t$.

Furthermore, we show that $(v_{0,t-1}^\star, \bar{v}_{t,T-1})$ is the optimal response to $(u_{0,t-1}^\star, \bar{u}_{t,T-1})$ by contradiction. If this were false, from the above proof, there would exist an optimal response of the form $(\bar{v}_{0,t-1}, \bar{v}_{t,T-1})$, corresponding to $(u_{0,t-1}^\star, \bar{u}_{t,T-1})$, and some $\nu \in \{1, 2, \cdots, N\}$ satisfying

$$\begin{aligned} & g_T^\nu(\bar{x}_T, \bar{y}_T^\nu) + \sum_{\tau=t}^{T-1} g_\tau^\nu(\bar{x}_\tau, \bar{y}_\tau^\nu, \bar{u}_\tau, \bar{v}_\tau^\nu) + \sum_{\tau=0}^{t-1} g_\tau^\nu(x_\tau^\star, \bar{y}_\tau^\nu, u_\tau^\star, \bar{v}_\tau^\nu) \\ < \; & g_T^\nu(\bar{x}_T, \bar{y}_T^\nu) + \sum_{\tau=t}^{T-1} g_\tau^\nu(\bar{x}_\tau, \bar{y}_\tau^\nu, \bar{u}_\tau, \bar{v}_\tau^\nu) + \sum_{\tau=0}^{t-1} g_\tau^\nu(x_\tau^\star, y_\tau^{\nu\star}, u_\tau^\star, v_\tau^{\nu\star}), \end{aligned} \tag{2.4}$$

where $(\bar{x}_t, \bar{y}_t) = (x_t^\star, y_t^\star)$ and $(x_{0,t-1}^\star, \bar{x}_{t,T}), (y_{0,t-1}^\star, \bar{y}_{t,T})$ and $\bar{y}_{0,T}$ are generated by $(u_{0,t-1}^\star, \bar{u}_{t,T})$, $(v_{0,t-1}^\star, \bar{v}_{t,T})$ and $\bar{v}_{0,T}$, respectively.

However, (2.4) implies

$$\sum_{\tau=0}^{t-1} g_\tau^\nu(x_\tau^\star, \bar{y}_\tau^\nu, u_\tau^\star, \bar{v}_\tau^\nu) < \sum_{\tau=0}^{t-1} g_\tau^\nu(x_\tau^\star, y_\tau^{\nu\star}, u_\tau^\star, v_\tau^{\nu\star}),$$

$v_{0,T-1}^{\nu\star} \in \Pi_{0,T-1}^\nu(x_0, y_0^\nu, u_{0,T-1}^\star)$ and $(\bar{v}_{0,t-1}^\nu, v_{t,T-1}^{\nu\star}) \in \Pi_{0,T-1}^\nu(x_0, y_0^\nu, u_{0,T-1}^\star)$ with $x_0 = x_0^\star$ and $y_0^\nu = y_0^{\nu\star} = \bar{y}_0^\nu$. We therefore have

$$\begin{aligned} & g_T^\nu(x_T^\star, y_T^\nu) + \sum_{\tau=0}^{t-1} g_\tau^\nu(x_\tau^\star, \bar{y}_\tau^\nu, u_\tau^\star, \bar{v}_\tau^\nu) + \sum_{\tau=t}^{T-1} g_\tau^\nu(x_\tau^\star, y_\tau^{\nu\star}, u_\tau^\star, v_\tau^{\nu\star}) \\ < \; & g_T^\nu(x_T^\star, y_T^{\nu\star}) + \sum_{\tau=0}^{T-1} g_\tau^\nu(x_\tau^\star, y_\tau^{\nu\star}, u_\tau^\star, v_\tau^{\nu\star}), \end{aligned}$$

which contradicts the hypothesis that $(v_0^\star, v_1^\star, \cdots, v_{T-1}^\star)$ is the optimal response to $(u_0^\star, u_1^\star, \cdots, u_{T-1}^\star)$. Consequently, $(v_{0,t-1}^\star, \bar{v}_{t,T-1})$ is the optimal response to $(u_{0,t-1}^\star, \bar{u}_{t,T-1})$.

Moreover, from (2.3) we have

$$G_T(\bar{x}_T, \bar{y}_T) + \sum_{\tau=t}^{T-1} G_\tau(\bar{x}_\tau, \bar{y}_\tau, \bar{u}_\tau, \bar{v}_\tau) + \sum_{\tau=0}^{t-1} G_\tau(x_\tau^\star, y_\tau^\star, u_\tau^\star, v_\tau^\star)$$

$$< \; G_T(x_T^\star, y_T^\star) + \sum_{\tau=t}^{T-1} G_\tau(x_\tau^\star, y_\tau^\star, u_\tau^\star, v_\tau^\star) + \sum_{\tau=0}^{t-1} G_\tau(x_\tau^\star, y_\tau^\star, u_\tau^\star, v_\tau^\star). \tag{2.5}$$

This contradicts the optimality of the policy $\{(u_0^\star, u_1^\star, \cdots, u_{T-1}^\star), (v_0^\star, v_1^\star, \cdots, v_{T-1}^\star)\}$ in DBOP (1.3). The proof is complete . □

The principle of optimality shown in Theorem 2.1 suggests that an optimal policy of DBOP can be constructed in a piecemeal manner. First, optimal policies are found for subproblems involving only the last stage. Then, utilizing these results, we obtain optimal policies for the last two stages. We repeat this procedure step by step backward until an optimal policy for the entire problem is constructed. The dynamic programming algorithm is presented as follows:

Algorithm 2.2 *(DP Algorithm for DBOP with Independent Followers)*
Step 1: Set $t := T$ and let for each $(x_T, y_T) \in \mathcal{X} \times \mathcal{Y}$

$$V_0^0(x_T, y_T) := G_T(x_T, y_T), \quad V_0^\nu(x_T, y_T^\nu) := g_T^\nu(x_T, y_T^\nu) \quad \nu = 1, 2, \cdots, N.$$

Step 2: Set $t := t-1$ and solve the following problem for each $(x_t, y_t) \in \mathcal{X} \times \mathcal{Y}$

$$\begin{array}{l} \min_{u_t \in \Pi_t^0(x_t)} \{G_t(x_t, y_t, u_t, v_t(x_t, y_t, u_t)) \\ +V_{T-t-1}^0(F_t(x_t, u_t), f_t(x_t, y_t, u_t, v_t(x_t, y_t, u_t)))\}, \end{array} \tag{2.6}$$

where $v_t(x_t, y_t, u_t) := (v_t^1(x_t, y_t^1, u_t), v_t^2(x_t, y_t^2, u_t), \cdots, v_t^N(x_t, y_t^N, u_t))$ comprises the solutions of the lower level problems

$$\min_{v_t^\nu \in \Pi_t^\nu(x_t, y_t^\nu, u_t)} \quad g_t^\nu(x_t, y_t^\nu, u_t, v_t^\nu) + V_{T-t-1}^\nu(F_t(x_t, u_t), f_t(x_t, y_t, u_t, v_t)) \tag{2.7}$$

for $\nu = 1, 2, \cdots, N$. Let $V_{T-t}^0(x_t, y_t)$ and $\hat{V}_{T-t}^\nu(x_t, y_t^\nu, u_t)$ denote the optimal values of problems (2.6) and (2.7), respectively. Let $u_t^\star(x_t, y_t)$ be an optimal solution of (2.6), and $v_t^\star(x_t, y_t) := v_t(x_t, y_t, u_t^\star(x_t, y_t))$. Define $V_{T-t}^\nu(x_t, y_t), \nu = 1, 2, \cdots, N$, by

$$V_{T-t}^\nu(x_t, y_t) := \hat{V}_{T-t}^\nu(x_t, y_t^\nu, u_t^\star(x_t, y_t)).$$

Step 3: If $t = 0$, a solution of DBOP (1.3) is obtained and stop. Otherwise, go to Step 2.

The next theorem shows that an optimal solution to DBOP (1.3) under feedback information structure is obtained with Algorithm 2.2.

Theorem 2.3 *For each $t = 0, 1, \cdots, T$, and $(x_t, y_t) \in \mathcal{X} \times \mathcal{Y}$, let $\{u_t^\star(x_t, y_t), v_t^\star(x_t, y_t)\}$ be an optimal solution to (2.6)–(2.7). Then, for any t and (x_t, y_t), an optimal solution to the subproblem $P_{T-t}(x_t, y_t)$ is given by*

$$\{(u_t^\star(x_t, y_t), u_{t+1}^\star(x_{t+1}, y_{t+1}), \cdots, u_{T-1}^\star(x_{T-1}, y_{T-1})),$$
$$(v_t^\star(x_t, y_t), v_{t+1}^\star(x_{t+1}, y_{t+1}), \cdots, v_{T-1}^\star(x_{T-1}, y_{T-1}))\}$$

with

$$x_{\tau+1} = F_\tau(x_\tau, u_\tau^\star), \quad y_{\tau+1}^\nu = f_\tau^\nu(x_\tau, y_\tau^\nu, u_\tau^\star, v_\tau^{\nu\star}) \tag{2.8}$$

for $\tau = t, t+1, \cdots, T-1$ and $\nu = 1, 2, \cdots, N$. In particular,

$$\{(u_0^\star(x_0, y_0), u_1^\star(x_1, y_1), \cdots, u_{T-1}^\star(x_{T-1}, y_{T-1})),$$
$$(v_0^\star(x_0, y_0), v_1^\star(x_1, y_1), \cdots, v_{T-1}^\star(x_{T-1}, y_{T-1}))\}$$

is an optimal solution to DBOP (1.3) under feedback information structure.

Proof We prove by induction in t. Apparently, the results hold for $t = T-1$. Assuming the conclusions are correct for all $t = T-1, \cdots, \bar{t}+1$, we show that the statement also holds for $t = \bar{t}$.

Since the results hold when $t = \bar{t}+1$, we see that, for any $(x_{\bar{t}+1}, y_{\bar{t}+1}) \in \mathcal{X} \times \mathcal{Y}$,

$$\{(u^\star_{\bar{t}+1}(x_{\bar{t}+1}, y_{\bar{t}+1}), u^\star_{\bar{t}+2}(x_{\bar{t}+2}, y_{\bar{t}+2}), \cdots, u^\star_{T-1}(x_{T-1}, y_{T-1})),$$

$$(v^\star_{\bar{t}+1}(x_{\bar{t}+1}, y_{\bar{t}+1}), v^\star_{\bar{t}+2}(x_{\bar{t}+2}, y_{\bar{t}+2}), \cdots, v^\star_{T-1}(x_{T-1}, y_{T-1}))\}$$

is an optimal solution to the subproblem $P_{T-\bar{t}-1}(x_{\bar{t}+1}, y_{\bar{t}+1})$. On one hand,

$$(v^\star_{\bar{t}+1}(x_{\bar{t}+1}, y_{\bar{t}+1}), v^\star_{\bar{t}+2}(x_{\bar{t}+2}, y_{\bar{t}+2}), \cdots, v^\star_{T-1}(x_{T-1}, y_{T-1}))$$

is the optimal response to the leader's decision

$$(u^\star_{\bar{t}+1}(x_{\bar{t}+1}, y_{\bar{t}+1}), u^\star_{\bar{t}+2}(x_{\bar{t}+2}, y_{\bar{t}+2}), \cdots, u^\star_{T-1}(x_{T-1}, y_{T-1})),$$

where $\{x_\tau\}_{\tau=\bar{t}+2}^T$ and $\{y_\tau\}_{\tau=\bar{t}+2}^T$ are determined by (2.8). Namely, for each $\nu = 1, 2, \cdots, N$ and $\bar{v}^\nu_{\bar{t}+1,T-1} \in \Pi^\nu_{\bar{t}+1,T-1}(x_{\bar{t}+1}, \bar{y}^\nu_{\bar{t}+1}, u^\star_{\bar{t}+1,T-1})$, we have

$$\begin{aligned} g^\nu_T(x_T, \bar{y}^\nu_T) + \sum_{\tau=\bar{t}+1}^{T-1} g^\nu_\tau(x_\tau, \bar{y}^\nu_\tau, u^\star_\tau, \bar{v}^\nu_\tau) \quad &\geq \quad g^\nu_T(x_T, y^\nu_T) + \sum_{\tau=\bar{t}+1}^{T-1} g^\nu_\tau(x_\tau, y^\nu_\tau, u^\star_\tau, v^{\nu\star}_\tau) \\ &= \quad \hat{V}^\nu_{T-\bar{t}-1}(x_{\bar{t}+1}, y_{\bar{t}+1}, u^\star_{\bar{t}+1}) = V^\nu_{T-\bar{t}-1}(x_{\bar{t}+1}, y_{\bar{t}+1}) \quad (2.9) \end{aligned}$$

for each $\nu = 1, 2, \cdots, N$, where $(x_{\bar{t}+1}, \bar{y}_{\bar{t}+1}) = (x_{\bar{t}+1}, y_{\bar{t}+1})$ and

$$x_{\tau+1} = F_\tau(x_\tau, u^\star_\tau), \quad \bar{y}^\nu_{\tau+1} = f^\nu_\tau(x_\tau, \bar{y}^\nu_\tau, u^\star_\tau, \bar{v}^\nu_\tau), \qquad \tau = \bar{t}+1, \cdots, T-1. \tag{2.10}$$

On the other hand, $(u^\star_{\bar{t}+1}(x_{\bar{t}+1}, y_{\bar{t}+1}), u^\star_{\bar{t}+2}(x_{\bar{t}+2}, y_{\bar{t}+2}), \cdots, u^\star_{T-1}(x_{T-1}, y_{T-1}))$ is an optimal decision of the leader and $(v^\star_{\bar{t}+1}(x_{\bar{t}+1}, y_{\bar{t}+1}), v^\star_{\bar{t}+2}(x_{\bar{t}+2}, y_{\bar{t}+2}), \cdots, v^\star_{T-1}(x_{T-1}, y_{T-1}))$ is the optimal response to the leader's decision, where $\{x_\tau\}_{\tau=\bar{t}+2}^T$ and $\{y_\tau\}_{\tau=\bar{t}+2}^T$ are determined by (2.8). Namely, for any decisions of the leader $(\hat{u}_{\bar{t}+1}, \hat{u}_{\bar{t}+2}, \cdots, \hat{u}_{T-1})$ and the corresponding response of the followers $\hat{v}^\nu_{\bar{t}+1,T-1} \in \Pi^\nu_{\bar{t}+1,T-1}(\hat{x}_{\bar{t}+1}, \hat{y}^\nu_{\bar{t}+1}, \hat{u}_{\bar{t}+1,T-1})$, $\nu = 1, 2, \cdots, N$, we have

$$\begin{aligned} G_T(\hat{x}_T, \hat{y}_T) + \sum_{\tau=\bar{t}+1}^{T-1} G_\tau(\hat{x}_\tau, \hat{y}_\tau, \hat{u}_\tau, \hat{v}_\tau) \quad &\geq \quad G_T(x_T, y_T) + \sum_{\tau=\bar{t}+1}^{T-1} G_\tau(x_\tau, y_\tau, u^\star_\tau, v^\star_\tau) \\ &= \quad V^0_{T-\bar{t}-1}(\hat{x}_{\bar{t}+1}, \hat{y}_{\bar{t}+1}), \end{aligned} \tag{2.11}$$

where $(x_{\bar{t}+1}, y_{\bar{t}+1}) = (\hat{x}_{\bar{t}+1}, \hat{y}_{\bar{t}+1})$ and

$$\hat{x}_{\tau+1} = F_\tau(\hat{x}_\tau, \hat{u}_\tau), \quad \hat{y}^\nu_{\tau+1} = f^\nu_\tau(\hat{x}_\tau, \hat{y}^\nu_\tau, \hat{u}_\tau, \hat{v}^\nu_\tau), \quad \tau = \bar{t}+1, \cdots, T-1. \tag{2.12}$$

Consider $t = \bar{t}$ and assume that $(u^\star_{\bar{t}}(x_{\bar{t}}, y_{\bar{t}}), v^\star_{\bar{t}}(x_{\bar{t}}, y_{\bar{t}}))$ is an optimal solution to the subproblem $P_{T-\bar{t}}(x_{\bar{t}}, y_{\bar{t}})$. Firstly, we show that $(v^\star_{\bar{t}}(x_{\bar{t}}, y_{\bar{t}}), v^\star_{\bar{t}+1}(x_{\bar{t}+1}, y_{\bar{t}+1}), \cdots, v^\star_{T-1}(x_{T-1}, y_{T-1}))$ is the optimal response of the followers to the leader's decisions

$$(u^\star_{\bar{t}}(x_{\bar{t}}, y_{\bar{t}}), u^\star_{\bar{t}+1}(x_{\bar{t}+1}, y_{\bar{t}+1}), \cdots, u^\star_{T-1}(x_{T-1}, y_{T-1})).$$

For the purpose of contradiction, suppose that there exist some $\nu \in \{1, 2, \cdots, N\}$ and

$$(\bar{v}^\nu_{\bar{t}}(x_{\bar{t}}, y_{\bar{t}}), \bar{v}^\nu_{\bar{t}+1}(x_{\bar{t}+1}, \bar{y}_{\bar{t}+1}), \cdots, \bar{v}^\nu_{T-1}(x_{T-1}, \bar{y}_{T-1}))$$

such that

$$g_T^\nu(x_T,\bar{y}_T^\nu)+\sum_{\tau=\bar{t}}^{T-1} g_\tau^\nu(x_\tau,\bar{y}_\tau^\nu,u_\tau^\star,\bar{v}_\tau^\nu) \;<\; g_T^\nu(x_T,y_T^\nu)+\sum_{\tau=\bar{t}}^{T-1} g_\tau^\nu(x_\tau,y_\tau^\nu,u_\tau^\star,v_\tau^{\nu\star}) \tag{2.13}$$

for $y_{\bar{t}}^\nu=\bar{y}_{\bar{t}}^\nu$, where $\{x_\tau\}_{\tau=\bar{t}+1}^T$ and $\{\bar{y}_\tau^\nu\}_{\tau=\bar{t}+1}^T$ are determined by (2.10). Moreover, we have

$$g_T^\nu(x_T,\bar{y}_T^\nu)+\sum_{\tau=\bar{t}}^{T-1} g_\tau^\nu(x_\tau,\bar{y}_\tau^\nu,u_\tau^\star,\bar{v}_\tau^\nu) \;\geq\; \hat{V}_{T-\bar{t}-1}^\nu(x_{\bar{t}+1},\bar{y}_{\bar{t}+1}^\nu,u_{\bar{t}+1}^\star)+g_{\bar{t}}^\nu(x_{\bar{t}},y_{\bar{t}}^\nu,u_{\bar{t}}^\star,\bar{v}_{\bar{t}}^\nu)$$

$$\geq\; \hat{V}_{T-\bar{t}}^\nu(x_{\bar{t}},y_{\bar{t}}^\nu,u_{\bar{t}}^\star) \;=\; V_{T-\bar{t}}^\nu(x_{\bar{t}},y_{\bar{t}}) \;=\; g_T^\nu(x_T,y_T^\nu)+\sum_{\tau=\bar{t}}^{T-1} g_\tau^\nu(x_\tau,y_\tau^\nu,u_\tau^\star,v_\tau^{\nu\star})$$

with $y_{\bar{t}}^\nu=\bar{y}_{\bar{t}}^\nu$, where the first inequality follows from (2.9) and the second inequality comes from Step 2 of Algorithm 2.2. However, this contradicts (2.13). Therefore,

$$(v_{\bar{t}}^\star(x_{\bar{t}},y_{\bar{t}}),v_{\bar{t}+1}^\star(x_{\bar{t}+1},y_{\bar{t}+1}),\cdots,v_{T-1}^\star(x_{T-1},y_{T-1}))$$

is the followers' optimal response to $(u_{\bar{t}}^\star(x_{\bar{t}},y_{\bar{t}}),u_{\bar{t}+1}^\star(x_{\bar{t}+1},y_{\bar{t}+1}),\cdots,u_{T-1}^\star(x_{T-1},y_{T-1}))$.

Now, we show that

$$\{(u_t^\star(x_t,y_t),u_{t+1}^\star(x_{t+1},y_{t+1}),\cdots,u_{T-1}^\star(x_{T-1},y_{T-1})),$$

$$(v_t^\star(x_t,y_t),v_{t+1}^\star(x_{t+1},y_{t+1}),\cdots,v_{T-1}^\star(x_{T-1},y_{T-1}))\}$$

is an optimal solution to the subproblem $P_{T-t}(x_t,y_t)$. If the results were false, there must exist the leader's decision $(\bar{u}_{\bar{t}},\bar{u}_{\bar{t}+1},\cdots,\bar{u}_{T-1})$ and the followers' response $(\bar{v}_{\bar{t}},\bar{v}_{\bar{t}+1},\cdots,\bar{v}_{T-1})$, along with the corresponding sequences $\{\bar{x}_\tau\}_{\tau=\bar{t}+1}^T$ and $\{\bar{y}_\tau\}_{\tau=\bar{t}+1}^T$ with $\bar{x}_{\bar{t}}=x_{\bar{t}}$ and $\bar{y}_{\bar{t}}=y_{\bar{t}}$, such that

$$G_T(\bar{x}_T,\bar{y}_T)+\sum_{\tau=\bar{t}}^{T-1} G_\tau(\bar{x}_\tau,\bar{y}_\tau,\bar{u}_\tau,\bar{v}_\tau) \;<\; G_T(x_T,y_T)+\sum_{\tau=\bar{t}}^{T-1} G_\tau(x_\tau,y_\tau,u_\tau^\star,v_\tau^\star). \tag{2.14}$$

However, we have

$$G_T(\bar{x}_T,\bar{y}_T)+\sum_{\tau=\bar{t}}^{T-1} G_\tau(\bar{x}_\tau,\bar{y}_\tau,\bar{u}_\tau,\bar{v}_\tau) \;\geq\; V_{T-\bar{t}-1}^0(\bar{x}_{\bar{t}+1},\bar{y}_{\bar{t}+1})+G_{\bar{t}}(x_{\bar{t}},y_{\bar{t}},\bar{u}_{\bar{t}},\bar{v}_{\bar{t}})$$

$$\geq\; V_{T-\bar{t}}^0(\bar{x}_{\bar{t}},\bar{y}_{\bar{t}}) \;=\; G_T(x_T,y_T)+\sum_{\tau=\bar{t}}^{T-1} G_\tau(x_\tau,y_\tau,u_\tau^\star,v_\tau^\star)$$

with $x_{\bar{t}}=\bar{x}_{\bar{t}}$ and $y_{\bar{t}}=\bar{y}_{\bar{t}}$, where the first inequality follows from (2.11) and the second inequality is induced from Step 2 of Algorithm 2.2. This contradicts (2.14). Therefore,

$$\{(u_{\bar{t}}^\star(x_{\bar{t}},y_{\bar{t}}),u_{\bar{t}+1}^\star(x_{\bar{t}+1},y_{\bar{t}+1}),\cdots,u_{T-1}^\star(x_{T-1},y_{T-1}),$$

$$(v_{\bar{t}}^\star(x_{\bar{t}},y_{\bar{t}}),v_{\bar{t}+1}^\star(x_{\bar{t}+1},y_{\bar{t}+1}),\cdots,v_{T-1}^\star(x_{T-1},y_{T-1}))\}$$

is optimal and, hence, the results hold for $\bar{t}$.

Consequently, the results hold for all $t=0,1,\cdots,T-1$ and the proof is complete. $\square$

In dynamic programming, the problem is decomposed into a sequence of minimization problems involving only decision variables of each stage, which are easier than the original problem. The following example illustrates the algorithm.

Example 1 Consider the following three-player problem with three stages. Among players, one is the leader and the other two are followers. The decision variables of the players are $u = (u_0, u_1, u_2), v^1 = (v_0^1, v_1^1, v_2^1)$ and $v^2 = (v_0^2, v_1^2, v_2^2)$, respectively. The state transition equations are given by

$$\begin{aligned} x_{t+1} &= x_t + u_t, \\ y_{t+1}^\nu &= y_t^\nu + v_t^\nu, \quad \nu = 1, 2; \quad t = 0, 1, 2. \end{aligned}$$

The cost functions are given by

$$\begin{aligned} G_3(x_3, y_3) &= 4x_3 + 3y_3^1 + 2y_3^2, \\ g_3^1(x_3, y_3) &= x_3 + 2y_3^1, \\ g_3^2(x_3, y_3) &= x_3 + 3y_3^2, \end{aligned}$$

and

$$\begin{aligned} G_t(x_t, y_t, u_t, v_t) &= (u_t)^2 + (v_t^1)^2 + (v_t^2)^2 + 2u_t x_t, \\ g_t^1(x_t, y_t^1, u_t, v_t^1) &= u_t + (v_t^1 - 1)^2, \\ g_t^2(x_t, y_t^2, u_t, v_t^2) &= (t-1)u_t v_t^2 + (v_t^2 + 1)^2 + 1, \end{aligned}$$

for $t = 0, 1, 2$. The initial states are $x_0 = 1, y_0^1 = 1, y_0^2 = 1$ and the admissible decisions are unrestricted, i.e., $\Pi_t^0(x_t) = R, \Pi_t^1(x_t, y_t^1, u_t) = R$ and $\Pi_t^2(x_t, y_t^1, u_t) = R$ for $t = 0, 1, 2$.

Let us apply Algorithm 2.2 to this example. Let

$$\begin{aligned} V_0^0(x_3, y_3) &= 4x_3 + 3y_3^1 + 2y_3^2, \\ V_0^1(x_3, y_3) &= x_3 + 2y_3^1, \\ V_0^2(x_3, y_3) &= x_3 + 3y_3^2. \end{aligned}$$

At the first step, the following problem, which corresponds to (2.6) with $t = 2$, is considered:

$$\begin{array}{ll} \min_{u_2} & 4(x_2 + u_2) + 3(y_2^1 + v_2^1) + 2(y_2^2 + v_2^2) + G_2 \\ \text{subject to} & v_2^1 \in \arg\min\{x_2 + u_2 + 2(y_2^1 + v_2^1) + g_2^1\} \\ & v_2^2 \in \arg\min\{x_2 + u_2 + 3(y_2^2 + v_2^2) + g_2^2\}. \end{array}$$

The solution to this problem is computed as $u_2 = -\frac{4}{5}x_2 - \frac{11}{5}, v_2^1 = 0, v_2^2 = -\frac{5+u_2}{2}$. Moreover, we have

$$\begin{aligned} \hat{V}_1^1(x_2, y_2, u_2) &= x_2 + 2u_2 + 2y_2^1 + 1, \\ \hat{V}_1^2(x_2, y_2, u_2) &= x_2 + u_2 + 3y_2^2 - (\tfrac{1}{2}u_2 + \tfrac{5}{2})^2 + 2, \\ V_1^0(x_2, y_2) &= 4x_2 + 3y_2^1 + 2y_2^2 - \tfrac{5}{4}(\tfrac{4}{5}x_2 + \tfrac{11}{5})^2 + \tfrac{5}{4}, \\ V_1^1(x_2, y_2) &= -\tfrac{3}{5}x_2 + 2y_2^1 - \tfrac{17}{5}, \\ V_1^2(x_2, y_2) &= \tfrac{1}{5}x_2 + 3y_2^2 - (\tfrac{2}{5}x_2 - \tfrac{7}{5})^2 - \tfrac{1}{5}. \end{aligned}$$

Then, the following problem, which corresponds to (2.6) with $t = 1$, is considered:

$$\begin{array}{ll} \min_{u_1} & 4x_1 + 4u_1 + 3y_1^1 + 3v_1^1 + 2y_1^2 + 2v_1^2 - \frac{5}{4}(\frac{4}{5}x_1 + \frac{4}{5}u_1 + \frac{11}{5})^2 + \frac{5}{4} + G_1 \\ \text{subject to} & v_1^1 \in \arg\min\{-\frac{3}{5}x_1 - \frac{3}{5}u_1 - \frac{17}{5} + 2v_1^1 + 2y_1^1 + g_1^1\} \\ & v_1^2 \in \arg\min\{-(\frac{2}{5}x_1 + \frac{2}{5}u_1 - \frac{7}{5})^2 + \frac{1}{5}x_1 + \frac{1}{5}u_1 - \frac{1}{5} + 3y_1^2 + 3v_1^2 + g_1^2\}. \end{array}$$

The solution to this problem is computed as $u_1 = -x_1 + 1, v_1^1 = 0, v_1^2 = -\frac{5}{2}$. Furthermore, we have

$$\begin{aligned}
&\hat{V}_2^1(x_1, y_1, u_1) = -\tfrac{3}{5}x_1 + \tfrac{2}{5}u_1 - \tfrac{12}{5} + 2y_1^1,\\
&\hat{V}_2^2(x_1, y_1, u_1) = -(\tfrac{2}{5}x_1 + \tfrac{2}{5}u_1 - \tfrac{7}{5})^2 + \tfrac{1}{5}x_1 + \tfrac{1}{5}u_1 + \tfrac{9}{5} + 3y_1^2 - \tfrac{25}{4},\\
&V_2^0(x_1, y_1) = 3y_1^1 + 2y_1^2 - x_1^2 - \tfrac{15}{4},\\
&V_2^1(x_1, y_1) = -x_1 + 2y_1^1 - 2,\\
&V_2^2(x_1, y_1) = 3y_1^2 - \tfrac{21}{4}.
\end{aligned}$$

The final step is to consider the following problem, which corresponds to (2.6) with $t = 0$:

$$\begin{array}{ll}
\min_{u_0} & 3y_0^1 + 3v_0^1 + 2y_0^2 + 2v_0^2 - \frac{15}{4} - (x_0 + u_0)^2 + G_0\\
\text{subject to} & v_0^1 \in \arg\min\{2v_0^1 + 2y_0^1 - (x_0 + u_0 + 2) + g_0^1\}\\
& v_0^2 \in \arg\min\{3y_0^2 + 3v_0^2 - \frac{21}{4} + g_0^2\}.
\end{array}$$

The solution of this problem is computed as $u_0 = 3, v_0^1 = 0, v_0^2 = -\frac{5-u_0}{2}$. Furthermore, we have

$$\begin{aligned}
&\hat{V}_3^1(x_0, y_0, u_0) = 2y_0^1 - (x_0 + 1),\\
&\hat{V}_3^2(x_0, y_0, u_0) = 3y_0^2 - \tfrac{13}{4} - (\tfrac{1}{2}u_0 - \tfrac{5}{2})^2,\\
&V_3^0(x_0, y_0) = 3y_0^1 + 2y_0^2 - x_0^2 - \tfrac{19}{4},\\
&V_3^1(x_0, y_0) = 2y_0^1 - (x_0 + 1),\\
&V_3^2(x_0, y_0) = 3y_0^2 - \tfrac{17}{4}.
\end{aligned}$$

Therefore, the optimal decisions are $(u_0, v_0^1, v_0^2) = (3, 0, -1)$, $(u_1, v_1^1, v_1^2) = (-3, 0, -\frac{5}{2})$, $(u_2, v_2^1, v_2^2) = (-3, 0, -1)$, with the corresponding states $(x_1, y_1^1, y_1^2) = (4, 1, 0)$, $(x_2, y_2^1, y_2^2) = (1, 1, -\frac{5}{2})$ and $(x_3, y_3^1, y_3^2) = (-2, 1, -\frac{7}{2})$. The optimal values to the leader and the followers are $-\frac{3}{4}$ and 0, $-\frac{5}{4}$, respectively.

In many practical situations, it is not possible to obtain an optimal solution analytically, and one has to resort to numerical execution of the dynamic programming algorithm possibly through discretization of the state and decision spaces [7].

3 DBOP Structure with Dependent Followers under Feedback Information

In the previous section, we have assumed that the followers act independently. In practice, when a follower makes decisions, he/she often has to take into account the strategies of the other followers and behaves in a noncooperative manner. In such a case, the lower level problems comprise a Nash game among the followers. Therefore, the whole problem is a feedback Stackelberg game with Nash game constraints, which is stated as follows: Given the initial state (x_0, y_0),

$$\begin{array}{ll}
\text{minimize} & J^0(x_0, y_0, u, v)\\
\text{subject to} & \bar{V}^\nu(x_0, y_0, u, v^{-\nu}) = \bar{J}^\nu(x_0, y_0, u, v)\\
& u \in \Pi^0(x_0), v^\nu \in \Pi^\nu(x_0, y_0^\nu, u), \quad \nu = 1, 2, \cdots, N,
\end{array} \tag{3.1}$$

where

$$
\begin{aligned}
v^{-\nu} &:= (v^1, v^2, \cdots, v^{\nu-1}, v^{\nu+1}, \cdots, v^N),\\
J^0(x_0, y_0, u, v) &:= G_T(x_T, y_T) + \sum_{t=0}^{T-1} G_t(x_t, y_t, u_t, v_t),\\
\bar{J}^\nu(x_0, y_0, u, v) &:= \bar{g}_T^\nu(x_T, y_T) + \sum_{t=0}^{T-1} \bar{g}_t^\nu(x_t, y_t, u_t, v_t),
\end{aligned}
$$

and $\bar{V}^\nu(x_0, y_0, u, v^{-\nu})$ are the optimal value functions for followers $\nu = 1, 2, \cdots, N$ defined by

$$
\bar{V}^\nu(x_0, y_0, u, v^{-\nu}) := \min\{\bar{J}^\nu(x_0, y_0, u, v) | (1.1), (1.2) \text{ and } v^\nu \in \Pi^\nu(x_0, y_0^\nu, u)\}.
$$

Therefore, the constraints in (3.1) imply that

$$
v^\nu \in \arg\min\{\bar{J}^\nu(x_0, y_0, u, v) | (1.1), (1.2) \text{ and } v^\nu \in \Pi^\nu(x_0, y_0^\nu, u)\},
$$

for all $\nu = 1, 2, \cdots, N$. Namely, v is a Nash equilibrium (NE) in the lower level problem for a given u. In DBOP (3.1) with dependent followers, the feasible region of a follower is independent of the other followers but the cost functions are dependent in the lower level problem.

Recall that, for the lower level problems in (3.1), $(v_0^\star, v_1^\star, \cdots, v_{T-1}^\star)$ is called a Nash equilibrium corresponding to (x_0, y_0, u) if and only if, for given (x_0, y_0, u),

$$
\bar{J}^\nu(x_0, y_0, u, v^\star) \le \bar{J}^\nu(x_0, y_0, u, (v^\nu, v^{-\nu\star})) \tag{3.2}
$$

for any $v^\nu \in \Pi^\nu(x_0, y_0^\nu, u)$ and all $\nu = 1, 2, \cdots, N$.

We will develop a dynamic programming algorithm for DBOP (3.1) with dependent followers under feedback information structure. The following result, which is an extension of Theorem 2.1, shows the principle of optimality for this problem. Some related notation are given below.

$$
\begin{aligned}
& v_t^{-\nu} := (v_t^1, \cdots, v_t^{\nu-1}, v_t^{\nu+1}, \cdots, v_t^N), \quad v_{t,T-1}^{-\nu} := (v_t^{-\nu}, \cdots, v_{T-1}^{-\nu}),\\
& v_{0,t}^{-\nu} := (v_0^{-\nu}, \cdots, v_t^{-\nu}), \quad v_{0,t}^\nu := (v_0^\nu, \cdots, v_t^\nu), \quad v_t = (v_t^\nu, v_t^{-\nu}),\\
& v_{t,T-1} = (v_{t,T-1}^\nu, v_{t,T-1}^{-\nu}), \quad y_t = (y_t^\nu, y_t^{-\nu}),\\
& J_{T-t}^0(x_t, y_t, u_{t,T-1}, v_{t,T-1}) := G_T(x_T, y_T) + \sum_{\tau=t}^{T-1} G_\tau(x_\tau, y_\tau, u_\tau, v_\tau),\\
& \bar{J}_{T-t}^\nu(x_t, y_t, u_{t,T-1}, v_{t,T-1}) := \bar{g}_T^\nu(x_T, y_T) + \sum_{\tau=t}^{T-1} \bar{g}_\tau^\nu(x_\tau, y_\tau, u_\tau, v_\tau),
\end{aligned}
$$

$$
\begin{aligned}
\bar{V}_{T-t}^\nu(x_t, y_t, u_{t,T-1}, v_{t,T-1}^{-\nu}) \quad := \quad & \min\{\bar{J}_{T-t}^\nu(x_t, y_t, u_{t,T-1}, v_{t,T-1}) : \\
& v_{t,T-1}^\nu \in \Pi_{t,T-1}^\nu(x_t, y_t^\nu, u_{t,T-1})\}.
\end{aligned} \tag{3.3}
$$

For $t = 0, 1, \cdots, T-1$, consider the subproblem

$$
\begin{array}{ll}
\text{minimize} & J_{T-t}^0(x_t, y_t, u_{t,T-1}, v_{t,T-1})\\
\text{subject to} & \bar{V}_{T-t}^\nu(x_t, y_t, u_{t,T-1}, v_{t,T-1}^{-\nu}) = \bar{J}_{T-t}^\nu(x_t, y_t, u_{t,T-1}, v_{t,T-1}),\\
& u_{t,T-1} \in \Pi_{t,T-1}^0(x_t),\\
& v_{t,T-1}^\nu \in \Pi_{t,T-1}^\nu(x_t, y_t^\nu, u_{t,T-1}),\\
& \nu = 1, 2, \cdots, N,
\end{array} \tag{3.4}
$$

where the initial state (x_t, y_t) is given, and $\{x_\tau\}_{\tau=t}^T$, $\{y_\tau\}_{\tau=t}^T$ are determined by (1.1) and (1.2). This problem is referred to as $\bar{P}_{T-t}(x_t, y_t)$. Moreover, for the subproblem $\bar{P}_{T-t}(x_t, y_t)$, $v_{t,T-1}^\star = (v_t^\star, v_{t+1}^\star, \cdots, v_{T-1}^\star)$ is called a Nash equilibrium corresponding to $(x_t, y_t, u_{t,T-1})$ if and only if, for given $(x_t, y_t, u_{t,T-1})$,

$$\bar{J}_{T-t}^\nu(x_t, y_t, u_{t,T-1}, v_{t,T-1}^\star) \;\leq\; \bar{J}_{T-t}^\nu(x_t, y_t, u_{t,T-1}, (v_{t,T-1}^\nu, v_{t,T-1}^{-\nu\star})) \tag{3.5}$$

for any $v_{t,T-1}^\nu \in \Pi_{t,T-1}^\nu(x_t, y_t^\nu, u_{t,T-1})$ and all $\nu = 1, 2, \cdots, N$. All the results in this section are based on the NE formulations (3.2) and (3.5), which will be assumed to have a unique solution throughout.

Theorem 3.1 *Let $(u_0^\star, u_1^\star, \cdots, u_{T-1}^\star)$ and $(v_0^\star, v_1^\star, \cdots, v_{T-1}^\star)$ constitute an optimal policy for DBOP (3.1) under feedback information structure with corresponding trajectories $(x_0^\star,\ x_1^\star, \cdots, x_T^\star)$ and $(y_0^\star, y_1^\star, \cdots,\ y_T^\star)$ such that $(v_0^\star, v_1^\star, \cdots, v_{T-1}^\star)$ comprises the solution to NE (3.2) among the followers. Consider the subproblem $\bar{P}_{T-t}(x_t^\star, y_t^\star)$ for each $t = 1, 2, \cdots, T-1$. Then, the truncated policy*

$$\{(u_t^\star, u_{t+1}^\star, \cdots, u_{T-1}^\star), (v_t^\star, v_{t+1}^\star, \cdots, v_{T-1}^\star)\}$$

is optimal for the subproblem $\bar{P}_{T-t}(x_t^\star, y_t^\star)$, where $(v_t^\star, v_{t+1}^\star, \cdots, v_{T-1}^\star)$ comprises the solution to NE (3.5) among the followers.

Proof First, we show that $(v_t^\star, v_{t+1}^\star, \cdots, v_{T-1}^\star)$ is the solution to NE (3.5) corresponding to the leader's decision $(u_t^\star, u_{t+1}^\star, \cdots, u_{T-1}^\star)$. If it were not a solution to NE (3.5) of the followers, then there would exist some $\nu \in \{1, 2, \cdots, N\}$ and

$$(\hat{v}_t^\nu, \hat{v}_{t+1}^\nu, \cdots, \hat{v}_{T-1}^\nu) \in \Pi_{t,T-1}^\nu(x_t^\star, \hat{y}_t^\nu, u_{t,T-1}^\star),$$

with the corresponding sequences $\{x_\tau^\star\}_{\tau=t}^T$ and $\{\hat{y}_\tau^\nu\}_{\tau=t}^T$, such that

$$\bar{J}_{T-t}^\nu(x_t^\star, (\hat{y}_t^\nu, y_t^{-\nu\star}), u_{t,T-1}^\star, (\hat{v}_{t,T-1}^\nu, v_{t,T-1}^{-\nu\star})) < \bar{J}_{T-t}^\nu(x_t^\star, y_t^\star, u_{t,T-1}^\star, v_{t,T-1}^\star)$$

with $\hat{y}_t^\nu = y_t^{\nu\star}$. Moreover, we have

$$(v_{0,t-1}^{\nu\star}, \hat{v}_{t,T-1}^\nu) \in \Pi_{0,T-1}^\nu(x_0^\star, y_0^{\nu\star}, u_{0,T-1}^\star)$$

with $\hat{y}_t^\nu = y_t^{\nu\star}$. Therefore, we obtain

$$\begin{aligned}
&\bar{J}^\nu(x_0, y_0, u_{0,T-1}^\star, (v_{0,t-1}^\star, (\hat{v}_{t,T-1}^\nu, v_{t,T-1}^{-\nu\star}))) \\
&= \; \bar{J}_{T-t}^\nu(x_t^\star, (\hat{y}_t^\nu, y_t^{-\nu\star}), u_{t,T-1}^\star, (\hat{v}_{t,T-1}^\nu, v_{t,T-1}^{-\nu\star})) + \sum_{\tau=0}^{t-1} \bar{g}_\tau^\nu(x_\tau^\star, y_\tau^\star, u_\tau^\star, v_\tau^\star) \\
&< \; \bar{J}_{T-t}^\nu(x_t^\star, y_t^\star, u_{t,T-1}^\star, v_{t,T-1}^\star) + \sum_{\tau=0}^{t-1} \bar{g}_\tau^\nu(x_\tau^\star, y_\tau^\star, u_\tau^\star, v_\tau^\star) \;=\; \bar{J}^\nu(x_0, y_0, u_{0,T-1}^\star, v_{0,T-1}^\star).
\end{aligned}$$

This indicates that $(v_0^\star, v_1^\star, \cdots, v_{T-1}^\star)$ is not a solution to NE (3.2) corresponding to $(u_0^\star, u_1^\star, \cdots, u_{T-1}^\star)$, which contradicts the hypothesis of this theorem. Consequently, $(v_t^\star, v_{t+1}^\star, \cdots, v_{T-1}^\star)$ is the solution to NE (3.5) corresponding to the leader's decision $(u_t^\star, u_{t+1}^\star, \cdots, u_{T-1}^\star)$.

To prove the theorem by contradiction, suppose that $\{(u_t^\star, u_{t+1}^\star, \cdots, u_{T-1}^\star), (v_t^\star,\ v_{t+1}^\star, \cdots, v_{T-1}^\star)\}$ is not an optimal solution to the subproblem $\bar{P}_{T-t}(x_t^\star, y_t^\star)$, where $(v_t^\star,\ v_{t+1}^\star, \cdots,\ v_{T-1}^\star)$ is the solution to NE (3.5). Then, there must exist another policy $\{(\bar{u}_t, \bar{u}_{t+1}, \cdots, \bar{u}_{T-1}), (\bar{v}_t, \bar{v}_{t+1}, \cdots,\ \bar{v}_{T-1})\}$

such that $(\bar{v}_t, \bar{v}_{t+1}, \cdots, \bar{v}_{T-1})$ is the solution to NE (3.5) corresponding to the leader's decision $(\bar{u}_t, \bar{u}_{t+1}, \cdots, \bar{u}_{T-1})$ and

$$G_T(\bar{x}_T, \bar{y}_T) + \sum_{\tau=t}^{T-1} G_\tau(\bar{x}_\tau, \bar{y}_\tau, \bar{u}_\tau, \bar{v}_\tau) \ < \ G_T(x_T^\star, y_T^\star) + \sum_{\tau=t}^{T-1} G_\tau(x_\tau^\star, y_\tau^\star, u_\tau^\star, v_\tau^\star), \tag{3.6}$$

where $\{(x_t^\star, x_{t+1}^\star, \cdots, x_T^\star), (y_t^\star, y_{t+1}^\star, \cdots, y_T^\star)\}$ and $\{(\bar{x}_t, \bar{x}_{t+1}, \cdots, \bar{x}_T), (\bar{y}_t, \bar{y}_{t+1}, \cdots, \bar{y}_T)\}$ are the sequences of the leader's and the followers' states generated by the corresponding decisions, with the initial conditions $x_t = x_t^\star = \bar{x}_t$ and $y_t = y_t^\star = \bar{y}_t$, respectively.

Note that the decisions $\{(u_t^\star, u_{t+1}^\star, \cdots, u_{T-1}^\star), (v_t^\star, v_{t+1}^\star, \cdots, v_{T-1}^\star)\}$ and $\{(\bar{u}_t, \bar{u}_{t+1}, \cdots, \bar{u}_{T-1}), (\bar{v}_t, \bar{v}_{t+1}, \cdots, \bar{v}_{T-1})\}$ all satisfy the constraints

$$(v_t^{\nu\star}, \cdots, v_{T-1}^{\nu\star}) \in \Pi_{t,T-1}^{\nu}(x_t^\star, y_t^{\nu\star}, u_{t,T-1}^\star),$$

$$(\bar{v}_t^{\nu}, \cdots, \bar{v}_{T-1}^{\nu}) \in \Pi_{t,T-1}^{\nu}(\bar{x}_t, \bar{y}_t^{\nu}, \bar{u}_{t,T-1})$$

for $\nu = 1, 2, \cdots, N$, where $x_t^\star = \bar{x}_t$ and $y_t^\star = \bar{y}_t$. Accordingly, we have

$$(v_{0,t-1}^{\nu\star}, v_{t,T-1}^{\nu\star}) \in \Pi_{0,T-1}^{\nu}(x_0^\star, y_0^{\nu\star}, (u_{0,t-1}^\star, u_{t,T-1}^\star)),$$

$$(v_{0,t-1}^{\nu\star}, \bar{v}_{t,T-1}^{\nu}) \in \Pi_{0,T-1}^{\nu}(x_0^\star, y_0^{\nu\star}, (u_{0,t-1}^\star, \bar{u}_{t,T-1}))$$

for $\nu = 1, 2, \cdots, N$, where $x_t^\star = \bar{x}_t$ and $y_t^\star = \bar{y}_t$. In a similar manner to the proof of Theorem 2.1, we can show that $(v_{0,t-1}^\star, \bar{v}_{t,T-1})$ is an NE corresponding to the decision $(u_{0,t-1}^\star, \bar{u}_{t,T-1})$. Furthermore, from (3.6), we have

$$\begin{aligned} & G_T(\bar{x}_T, \bar{y}_T) + \sum_{\tau=t}^{T-1} G_\tau(\bar{x}_\tau, \bar{y}_\tau, \bar{u}_\tau, \bar{v}_\tau) + \sum_{\tau=0}^{t-1} G_\tau(x_\tau^\star, y_\tau^\star, u_\tau^\star, v_\tau^\star) \\ < \ & G_T(x_T^\star, y_T^\star) + \sum_{\tau=t}^{T-1} G_\tau(x_\tau^\star, y_\tau^\star, u_\tau^\star, v_\tau^\star) + \sum_{\tau=0}^{t-1} G_\tau(x_\tau^\star, y_\tau^\star, u_\tau^\star, v_\tau^\star). \end{aligned} \tag{3.7}$$

This contradicts the optimality of the policy $\{(u_0^\star, u_1^\star, \cdots, u_{T-1}^\star), (v_0^\star, v_1^\star, \cdots, v_{T-1}^\star)\}$ in DBOP (3.1). The proof is complete. $\Box$

Based on Theorem 3.1, a dynamic programming algorithm for DBOP with dependent followers under feedback information structure is presented as follows.

Algorithm 3.2 *(DP Algorithm for DBOP with Dependent Followers)*
Step 1: Set $t := T$ and let for each $(x_T, y_T) \in \mathcal{X} \times \mathcal{Y}$

$$V_0^0(x_T, y_T) := G_T(x_T, y_T), \quad \check{V}_0^{\nu}(x_T, y_T) := \bar{g}_T^{\nu}(x_T, y_T) \quad \nu = 1, 2, \cdots, N.$$

Step 2: Set $t := t-1$ and solve the following problem for each $(x_t, y_t) \in \mathcal{X} \times \mathcal{Y}$:

$$\begin{aligned} & \min\nolimits_{u_t \in \Pi_t^0(x_t)} \{G_t(x_t, y_t, u_t, v_t(x_t, y_t, u_t)) \\ & + V_{T-t-1}^0(F_t(x_t, u_t), f_t(x_t, y_t, u_t, v_t(x_t, y_t, u_t)))\}, \end{aligned} \tag{3.8}$$

where $v_t(x_t, y_t, u_t)$ comprises the solution to NE (3.5) associated with the lower level problems

$$\min\nolimits_{v_t^{\nu} \in \Pi_t^{\nu}(x_t, y_t^{\nu}, u_t, v_t^{-\nu})} \quad \bar{g}_t^{\nu}(x_t, y_t, u_t, v_t) + \check{V}_{T-t-1}^{\nu}(F_t(x_t, u_t), f_t(x_t, y_t, u_t, v_t)) \tag{3.9}$$

for $\nu = 1, 2, \cdots, N$. Let $V^0_{T-t}(x_t, y_t)$ and $\hat{V}^{\nu}_{T-t}(x_t, y^{\nu}_t, u_t, v^{-\nu}_t)$ denote the optimal values of problems (3.8)–(3.9), respectively. Let $u^{\star}_t(x_t, y_t)$ be an optimal solution of (3.8) and $v^{\star}_t(x_t, y_t) := v_t(x_t, y_t, u^{\star}_t(x_t, y_t))$. Define $\check{V}^{\nu}_{T-t}(x_t, y_t), \nu = 1, 2, \cdots, N$, by

$$\check{V}^{\nu}_{T-t}(x_t, y_t) := \hat{V}^{\nu}_{T-t}(x_t, y^{\nu}_t, u^{\star}_t(x_t, y_t), v^{-\nu\star}_t(x_t, y_t)).$$

Step 3: If $t = 0$, a solution of DBOP (3.1) is obtained and stop. Otherwise, go to Step 2.

In Step 2, the lower level problems constitute a single stage Nash game with given (x_t, y_t, u_t). The next theorem shows that Algorithm 3.2 produces an optimal solution to DBOP (1.3) for any initial state $(x_0, y_0) \in \mathcal{X} \times \mathcal{Y}$. For convenience, we denote

$$\tilde{V}^{\nu}_{T-t-1}(x_t, y_t, u_t, v_t) := \check{V}^{\nu}_{T-t-1}(F_t(x_t, u_t), f_t(x_t, y_t, u_t, v_t)).$$

Theorem 3.3 *Under feedback information structure, for each $t = 0, 1, \cdots, T$, and $(x_t, y_t) \in \mathcal{X} \times \mathcal{Y}$, let $\{u^{\star}_t(x_t, y_t), v^{\star}_t(x_t, y_t)\}$ be an optimal solution to (3.8)–(3.9). Then, for any t and (x_t, y_t), an optimal solution to the subproblem $\bar{P}_{T-t}(x_t, y_t)$ is given by*

$$\{(u^{\star}_t(x_t, y_t), u^{\star}_{t+1}(x_{t+1}, y_{t+1}), \cdots, u^{\star}_{T-1}(x_{T-1}, y_{T-1})),$$

$$(v^{\star}_t(x_t, y_t), v^{\star}_{t+1}(x_{t+1}, y_{t+1}), \cdots, v^{\star}_{T-1}(x_{T-1}, y_{T-1}))\},$$

where $\{x_\tau\}^T_{\tau=t+1}$ and $\{y_\tau\}^T_{\tau=t+1}$ are determined by (2.8). In particular,

$$\{(u^{\star}_0(x_0, y_0), u^{\star}_1(x_1, y_1), \cdots, u^{\star}_{T-1}(x_{T-1}, y_{T-1})),$$

$$(v^{\star}_0(x_0, y_0), v^{\star}_1(x_1, y_1), \cdots, v^{\star}_{T-1}(x_{T-1}, y_{T-1}))\}$$

is an optimal solution to DBOP (3.1).

Proof We prove by induction in t. Apparently, the results hold for $t = T - 1$. Assuming the conclusions are correct for all $t = T - 1, \cdots, \bar{t} + 1$, we show that the statement also holds for $t = \bar{t}$.

Since the results hold when $t = \bar{t} + 1$, it follows that, for any $(x_{\bar{t}+1}, y_{\bar{t}+1}) \in \mathcal{X} \times \mathcal{Y}$,

$$\{(u^{\star}_{\bar{t}+1}(x_{\bar{t}+1}, y_{\bar{t}+1}), u^{\star}_{\bar{t}+2}(x_{\bar{t}+2}, y_{\bar{t}+2}), \cdots, u^{\star}_{T-1}(x_{T-1}, y_{T-1})),$$

$$(v^{\star}_{\bar{t}+1}(x_{\bar{t}+1}, y_{\bar{t}+1}), v^{\star}_{\bar{t}+2}(x_{\bar{t}+2}, y_{\bar{t}+2}), \cdots, v^{\star}_{T-1}(x_{T-1}, y_{T-1}))\}$$

is an optimal solution to the subproblem $\bar{P}_{T-\bar{t}-1}(x_{\bar{t}+1}, y_{\bar{t}+1})$. On one hand,

$$(v^{\star}_{\bar{t}+1}(x_{\bar{t}+1}, y_{\bar{t}+1}), v^{\star}_{\bar{t}+2}(x_{\bar{t}+2}, y_{\bar{t}+2}), \cdots, v^{\star}_{T-1}(x_{T-1}, y_{T-1}))$$

is the solution to NE (3.5) corresponding to the leader's decision

$$(u^{\star}_{\bar{t}+1}(x_{\bar{t}+1}, y_{\bar{t}+1}), u^{\star}_{\bar{t}+2}(x_{\bar{t}+2}, y_{\bar{t}+2}), \cdots, u^{\star}_{T-1}(x_{T-1}, y_{T-1})),$$

where $\{x_\tau\}^T_{\tau=\bar{t}+2}$ and $\{y_\tau\}^T_{\tau=\bar{t}+2}$ are determined by (2.8). Namely, for any $\bar{v}^{\nu}_{\bar{t}+1,T-1} \in \Pi^{\nu}_{\bar{t}+1,T-1}$ $(x_{\bar{t}+1}, \bar{y}^{\nu}_{\bar{t}+1}, u^{\star}_{\bar{t}+1,T-1})$, with the corresponding sequence $\{\bar{y}^{\nu}_\tau\}^{T-1}_{\tau=\bar{t}+1}$, we have

$$\begin{aligned} &\bar{J}^{\nu}_{T-\bar{t}-1}(x_{\bar{t}+1}, (\bar{y}^{\nu}_{\bar{t}+1}, y^{-\nu}_{\bar{t}+1}), u^{\star}_{\bar{t}+1,T-1}, (\bar{v}^{\nu}_{\bar{t}+1,T-1}, v^{-\nu\star}_{\bar{t}+1,T-1})) \\ \geq\; &\bar{J}^{\nu}_{T-\bar{t}-1}(x_{\bar{t}+1}, y_{\bar{t}+1}, u^{\star}_{\bar{t}+1,T-1}, v^{\star}_{\bar{t}+1,T-1}) \end{aligned} \tag{3.10}$$

and

$$\bar{J}^{\nu}_{T-\bar{t}-1}(x_{\bar{t}+1}, y_{\bar{t}+1}, u^{\star}_{\bar{t}+1,T-1}, v^{\star}_{\bar{t}+1,T-1}) \;=\; \check{V}^{\nu}_{T-\bar{t}-1}(x_{\bar{t}+1}, y_{\bar{t}+1}) \tag{3.11}$$

for each $\nu = 1, 2, \cdots, N$, where $\bar{y}^{\nu}_{\bar{t}+1} = y^{\nu}_{\bar{t}+1}$.

On the other hand, $(u^{\star}_{\bar{t}+1}(x_{\bar{t}+1}, y_{\bar{t}+1}), u^{\star}_{\bar{t}+2}(x_{\bar{t}+2}, y_{\bar{t}+2}), \cdots, u^{\star}_{T-1}(x_{T-1}, y_{T-1}))$ is an optimal decision of the leader and $(v^{\star}_{\bar{t}+1}(x_{\bar{t}+1}, y_{\bar{t}+1}), v^{\star}_{\bar{t}+2}(x_{\bar{t}+2}, y_{\bar{t}+2}), \cdots, v^{\star}_{T-1}(x_{T-1}, y_{T-1}))$ is the solution to NE (3.5) corresponding to the leader's decision. Namely, for any decisions of the leader $(\hat{u}_{\bar{t}+1}, \hat{u}_{\bar{t}+2}, \cdots, \hat{u}_{T-1})$ and the corresponding response of the followers $(\hat{v}_{\bar{t}+1}, \hat{v}_{\bar{t}+2}, \cdots, \hat{v}_{T-1})$, we have

$$\begin{aligned} G_T(\hat{x}_T, \hat{y}_T) + \sum_{\tau=\bar{t}+1}^{T-1} G_\tau(\hat{x}_\tau, \hat{y}_\tau, \hat{u}_\tau, \hat{v}_\tau) \;&\geq\; G_T(x_T, y_T) + \sum_{\tau=\bar{t}+1}^{T-1} G_\tau(x_\tau, y_\tau, u^{\star}_\tau, v^{\star}_\tau) \\ &=\; V^{0}_{T-\bar{t}-1}(x_{\bar{t}+1}, y_{\bar{t}+1}) \end{aligned} \tag{3.12}$$

with $(x_{\bar{t}+1}, y_{\bar{t}+1}) = (\hat{x}_{\bar{t}+1}, \hat{y}_{\bar{t}+1})$, where $\{\hat{x}_\tau\}_{\tau=\bar{t}+2}^{T}$ and $\{\hat{y}_\tau\}_{\tau=\bar{t}+2}^{T}$ are determined by (2.12).

Consider $t = \bar{t}$ and assume that $(u^{\star}_{\bar{t}}(x_{\bar{t}}, y_{\bar{t}}), v^{\star}_{\bar{t}}(x_{\bar{t}}, y_{\bar{t}}))$ is optimal to (3.8)–(3.9). First, we show that $(v^{\star}_{\bar{t}}(x_{\bar{t}}, y_{\bar{t}}), v^{\star}_{\bar{t}+1}(x_{\bar{t}+1}, y_{\bar{t}+1}), \cdots, v^{\star}_{T-1}(x_{T-1}, y_{T-1}))$ is the solution to NE (3.5) corresponding to the leader's decision $(u^{\star}_{\bar{t}}(x_{\bar{t}}, y_{\bar{t}}), u^{\star}_{\bar{t}+1}(x_{\bar{t}+1}, y_{\bar{t}+1}), \cdots, u^{\star}_{T-1}(x_{T-1}, y_{T-1}))$. For the purpose of contradiction, suppose that there exist some $\nu \in \{1, 2, \cdots, N\}$ and $\bar{v}^{\nu}_{\bar{t},T-1} \in \Pi^{\nu}_{\bar{t},T-1}(x_{\bar{t}}, \bar{y}^{\nu}_{\bar{t}}, u^{\star}_{\bar{t},T-1})$ with the corresponding sequence $\{\bar{y}^{\nu}_\tau\}_{\tau=\bar{t}+1}^{T}$ such that

$$\bar{J}^{\nu}_{T-\bar{t}}(x_{\bar{t}}, y_{\bar{t}}, u^{\star}_{\bar{t},T-1}, (\bar{v}^{\nu}_{\bar{t},T-1}, v^{-\nu\star}_{\bar{t},T-1})) \;<\; \bar{J}^{\nu}_{T-\bar{t}}(x_{\bar{t}}, y_{\bar{t}}, u^{\star}_{\bar{t},T-1}, v^{\star}_{\bar{t},T-1}) \tag{3.13}$$

with $\bar{y}^{\nu}_{\bar{t}} = y^{\nu}_{\bar{t}}$. Consider two decisions $\{u^{\star}_{\bar{t},T-1}, ((\bar{v}^{\nu}_{\bar{t}}, v^{-\nu\star}_{\bar{t}}), (\bar{v}^{\nu}_{\bar{t}+1,T-1}, v^{-\nu\star}_{\bar{t}+1,T-1}))\}$ and $\{u^{\star}_{\bar{t},T-1}, ((\bar{v}^{\nu}_{\bar{t}}, v^{-\nu\star}_{\bar{t}}), v^{\star}_{\bar{t}+1,T-1})\}$ with the initial state $(x_{\bar{t}}, y_{\bar{t}})$. Correspondingly, there exist two state trajectories $\{(x_{\bar{t}}, y_{\bar{t}}), (x_{\bar{t}+1}, (\bar{y}^{\nu}_{\bar{t}+1}, y^{-\nu}_{\bar{t}+1})), \cdots, (x_T, (\bar{y}^{\nu}_T, y^{-\nu}_T))\}$ and $\{(x_{\bar{t}}, y_{\bar{t}}), (x_{\bar{t}+1}, \bar{y}_{\bar{t}+1}), \cdots, (x_T, \bar{y}_T)\}$, respectively. Note that $\bar{y}^{\nu}_{\bar{t}+1} = y^{\nu}_{\bar{t}+1} = f^{\nu}_{\bar{t}}(x_{\bar{t}}, y^{\nu}_{\bar{t}}, u_{\bar{t}}, v^{\nu}_{\bar{t}})$ and hence we have $(\bar{y}^{\nu}_{\bar{t}+1}, y^{-\nu}_{\bar{t}+1}) = y_{\bar{t}+1}$. From (3.10) and Step 2 of Algorithm 3.2, we therefore have

$$\begin{aligned} &\bar{J}^{\nu}_{T-\bar{t}}(x_{\bar{t}}, y_{\bar{t}}, u^{\star}_{\bar{t},T-1}, (\bar{v}^{\nu}_{\bar{t},T-1}, v^{-\nu\star}_{\bar{t},T-1})) \\ &= \bar{g}^{\nu}_{\bar{t}}(x_{\bar{t}}, y_{\bar{t}}, u^{\star}_{\bar{t}}, (\bar{v}^{\nu}_{\bar{t}}, v^{-\nu\star}_{\bar{t}})) + \bar{J}^{\nu}_{T-\bar{t}-1}(x_{\bar{t}+1}, (\bar{y}^{\nu}_{\bar{t}+1}, y^{-\nu}_{\bar{t}+1}), u^{\star}_{\bar{t}+1,T-1}, (\bar{v}^{\nu}_{\bar{t}+1,T-1}, v^{-\nu\star}_{\bar{t}+1,T-1})) \\ &\geq \bar{g}^{\nu}_{\bar{t}}(x_{\bar{t}}, y_{\bar{t}}, u^{\star}_{\bar{t}}, (\bar{v}^{\nu}_{\bar{t}}, v^{-\nu\star}_{\bar{t}})) + \bar{J}^{\nu}_{T-\bar{t}-1}(x_{\bar{t}+1}, y_{\bar{t}+1}, u^{\star}_{\bar{t}+1,T-1}, v^{\star}_{\bar{t}+1,T-1}) \\ &= \bar{g}^{\nu}_{\bar{t}}(x_{\bar{t}}, y_{\bar{t}}, u^{\star}_{\bar{t}}, (\bar{v}^{\nu}_{\bar{t}}, v^{-\nu\star}_{\bar{t}})) + \check{V}^{\nu}_{T-\bar{t}-1}(x_{\bar{t}+1}, y_{\bar{t}+1}) \\ &= \bar{g}^{\nu}_{\bar{t}}(x_{\bar{t}}, y_{\bar{t}}, u^{\star}_{\bar{t}}, (\bar{v}^{\nu}_{\bar{t}}, v^{-\nu\star}_{\bar{t}})) + \tilde{V}^{\nu}_{T-\bar{t}-1}(x_{\bar{t}}, y_{\bar{t}}, u^{\star}_{\bar{t}}, (\bar{v}^{\nu}_{\bar{t}}, v^{-\nu\star}_{\bar{t}})) \\ &\geq \bar{g}^{\nu}_{\bar{t}}(x_{\bar{t}}, y_{\bar{t}}, u^{\star}_{\bar{t}}, v^{\star}_{\bar{t}}) + \tilde{V}^{\nu}_{T-\bar{t}-1}(x_{\bar{t}}, y_{\bar{t}}, u^{\star}_{\bar{t}}, v^{\star}_{\bar{t}}) = \bar{J}^{\nu}_{T-\bar{t}}(x_{\bar{t}}, y_{\bar{t}}, u^{\star}_{\bar{t},T-1}, v^{\star}_{\bar{t},T-1}), \end{aligned}$$

where $v^{\star}_{\bar{t}} = v^{\star}_{\bar{t}}(x_{\bar{t}}, y_{\bar{t}})$ and $v^{\nu\star}_{\bar{t}} = v^{\nu\star}_{\bar{t}}(x_{\bar{t}}, y^{\nu}_{\bar{t}})$, for all $\nu = 1, 2, \cdots, N$. This contradicts (3.13). Therefore, $(v^{\star}_{\bar{t}}(x_{\bar{t}}, y_{\bar{t}}), v^{\star}_{\bar{t}+1}(x_{\bar{t}+1}, y_{\bar{t}+1}), \cdots, v^{\star}_{T-1}(x_{T-1}, y_{T-1}))$ is the solution to NE (3.5) corresponding to the decision $(u^{\star}_{\bar{t}}(x_{\bar{t}}, y_{\bar{t}}), u^{\star}_{\bar{t}+1}(x_{\bar{t}+1}, y_{\bar{t}+1}), \cdots, u^{\star}_{T-1}(x_{T-1}, y_{T-1}))$ and

$$\bar{J}^{\nu}_{T-\bar{t}}(x_{\bar{t}}, y_{\bar{t}}, u^{\star}_{\bar{t},T-1}, v^{\star}_{\bar{t},T-1}) \;=\; \check{V}^{\nu}_{T-\bar{t}}(x_{\bar{t}}, y_{\bar{t}}), \tag{3.14}$$

for any $\nu = 1, 2, \cdots, N$.

Then, we show that

$$\{(u^{\star}_t(x_t, y_t), u^{\star}_{t+1}(x_{t+1}, y_{t+1}), \cdots, u^{\star}_{T-1}(x_{T-1}, y_{T-1})),$$

$$(v^{\star}_t(x_t, y_t), v^{\star}_{t+1}(x_{t+1}, y_{t+1}), \cdots, v^{\star}_{T-1}(x_{T-1}, y_{T-1}))\}$$

is an optimal solution to the subproblem $\bar{P}_{T-\bar{t}}(x_t, y_t)$. If this were false, there must exist the leader's decision $(\bar{u}_{\bar{t}}, \bar{u}_{\bar{t}+1}, \cdots, \bar{u}_{T-1})$ together with the corresponding solution $(\bar{v}_{\bar{t}}, \bar{v}_{\bar{t}+1}, \cdots, \bar{v}_{T-1})$ to NE (3.5) of the followers such that

$$G_T(\bar{x}_T, \bar{y}_T) + \sum_{\tau=\bar{t}}^{T-1} G_\tau(\bar{x}_\tau, \bar{y}_\tau, \bar{u}_\tau, \bar{v}_\tau) \ < \ G_T(x_T, y_T) + \sum_{\tau=\bar{t}}^{T-1} G_\tau(x_\tau, y_\tau, u^\star_\tau, v^\star_\tau), \tag{3.15}$$

where the sequences $\{\bar{x}_\tau\}_{\tau=\bar{t}+1}^{T}$ and $\{\bar{y}_\tau\}_{\tau=\bar{t}+1}^{T}$ are generated according to $\{\bar{u}_\tau\}_{\tau=\bar{t}}^{T-1}$ and $\{\bar{v}_\tau\}_{\tau=\bar{t}}^{T-1}$, respectively, with $x_{\bar{t}} = \bar{x}_{\bar{t}}$ and $y_{\bar{t}} = \bar{y}_{\bar{t}}$. However, from Algorithm 3.2 and (3.12), we have

$$\begin{aligned} & G_T(\bar{x}_T, \bar{y}_T) + \sum_{\tau=\bar{t}}^{T-1} G_\tau(\bar{x}_\tau, \bar{y}_\tau, \bar{u}_\tau, \bar{v}_\tau) \ \geq \ V^0_{T-\bar{t}-1}(\bar{x}_{\bar{t}+1}, \bar{y}_{\bar{t}+1}) + G_{\bar{t}}(x_{\bar{t}}, y_{\bar{t}}, \bar{u}_{\bar{t}}, \bar{v}_{\bar{t}}) \\ & \geq \ V^0_{T-\bar{t}}(\bar{x}_{\bar{t}}, \bar{y}_{\bar{t}}) \ = \ G_T(x_T, y_T) + \sum_{\tau=\bar{t}}^{T-1} G_\tau(x_\tau, y_\tau, u^\star_\tau, v^\star_\tau), \end{aligned}$$

which is in contradiction with (3.15). Therefore,

$$\begin{aligned} \{ & (u^\star_{\bar{t}}(x_{\bar{t}}, y_{\bar{t}}), u^\star_{\bar{t}+1}(x_{\bar{t}+1}, y_{\bar{t}+1}), \cdots, u^\star_{T-1}(x_{T-1}, y_{T-1}), \\ & (v^\star_{\bar{t}}(x_{\bar{t}}, y_{\bar{t}}), v^\star_{\bar{t}+1}(x_{\bar{t}+1}, y_{\bar{t}+1}), \cdots, v^\star_{T-1}(x_{T-1}, y_{T-1}))\} \end{aligned}$$

is optimal for $\bar{P}_{T-\bar{t}}(x_{\bar{t}}, y_{\bar{t}})$ and hence the results hold for $\bar{t}$.

Consequently, the results hold for all $t = 0, 1, \cdots, T-1$ and the proof is complete. □

The following example illustrates Algorithm 3.2.

Example 2 Consider the following three-player problem with three stages. Among players, one is the leader and the other two are followers. The decision variables of the players are $u = (u_0, u_1, u_2), v^1 = (v^1_0, v^1_1, v^1_2)$ and $v^2 = (v^2_0, v^2_1, v^2_2)$, respectively. The state transition equations are given by

$$\begin{aligned} x_{t+1} &= x_t + u_t - 2v^1_t + v^2_t, \\ y^\nu_{t+1} &= y^\nu_t + 2v^\nu_t, \qquad\qquad \nu = 1, 2; \quad t = 0, 1, 2. \end{aligned}$$

The cost functions are given by

$$\begin{aligned} G_3(x_3, y_3) &= 4x_3 + 3y^1_3 + 2y^2_3, \\ g^1_3(x_3, y_3) &= x_3 + 2y^1_3 - 10y^2_3, \\ g^2_3(x_3, y_3) &= 2x_3 + 3y^2_3, \end{aligned}$$

and

$$\begin{aligned} G_t(x_t, y_t, u_t, v_t) &= (u_t)^2 + (v^1_t)^2 - (v^2_t)^2 + 2u_t x_t + x^2_t, \\ g^1_t(x_t, y_t, u_t, v_t) &= -15u_t + (v^1_t - 1)^2 - 2v^1_t v^2_t + (v^2_t)^2, \\ g^2_t(x_t, y_t, u_t, v_t) &= 2u_t v^2_t + (v^1_t + 1)^2 + (v^2_t + 1)^2, \end{aligned}$$

for $t = 0, 1, 2$. The initial states are $x_0 = 1, y^1_0 = 1, y^2_0 = 1$ and the admissible decisions are unrestricted, i.e., $\Pi^0_t(x_t) = R, \Pi^1_t(x_t, y^1_t, u_t) = R$ and $\Pi^2_t(x_t, y^1_t, u_t) = R$ for $t = 0, 1, 2$.

Let us employ Algorithm 3.2 to solve the above example. Let

$$\begin{aligned} V^0_0(x_3, y_3) &= 4x_3 + 3y^1_3 + 2y^2_3, \\ \check{V}^1_0(x_3, y_3) &= x_3 + 2y^1_3 - 10y^2_3, \\ \check{V}^2_0(x_3, y_3) &= 2x_3 + 3y^2_3. \end{aligned}$$

At the first step, the following problem, which corresponds to (3.8) with $t = 2$, is considered:

$$\begin{array}{ll} \min_{u_2} & 4(x_2 + u_2 - 2v_2^1 + v_2^2) + 3(y_2^1 + 2v_2^1) + 2(y_2^2 + 2v_2^2) + G_2 \\ \text{subject to} & v_2^1 \in \arg\min\{(x_2 + u_2 - 2v_2^1 + v_2^2) + 2(y_2^1 + 2v_2^1) - 10(y_2^2 + 2v_2^2) + g_2^1\}, \\ & v_2^2 \in \arg\min\{2(x_2 + u_2 - 2v_2^1 + v_2^2) + 3(y_2^2 + 2v_2^2) + g_2^2\}. \end{array}$$

The unique Nash equilibrium for the lower level problem is given by $(v_2^1, v_2^2) = (-(5+u_2), -(5+u_2))$. Accordingly, the solution is computed as $u_2 = 1 - x_2, v_2^1 = v_2^2, v_2^2 = -5 - u_2$. Furthermore, we have

$$\begin{aligned} &\hat{V}_1^1(x_2, y_2, u_2) = x_2 + 2y_2^1 - 10y_2^2 + 5u_2 + 96, \\ &\hat{V}_1^2(x_2, y_2, u_2) = 2x_2 + 3y_2^2 + 12 + 4u_2, \\ &V_1^0(x_2, y_2) = 6x_2 + 3y_2^1 + 2y_2^2 - 31, \\ &\check{V}_1^1(x_2, y_2) = 101 - 4x_2 + 2y_2^1 - 10y_2^2, \\ &\check{V}_1^2(x_2, y_2) = -2x_2 + 3y_2^2 + 16. \end{aligned}$$

Then, the following problem, which corresponds to (3.8) with $t = 1$, is considered:

$$\begin{array}{ll} \min_{u_1} & 6x_1 + 6u_1 + 3y_1^1 - 6v_1^1 + 2y_1^2 + 10v_1^2 - 31 + G_1 \\ \text{subject to} & v_1^1 \in \arg\min\{101 - 4x_1 + 2y_1^1 - 10y_1^2 - 4u_1 + 12v_1^1 - 24v_1^2 + g_1^1\}, \\ & v_1^2 \in \arg\min\{-2(x_1 + u_1 - 2v_1^1 + v_1^2) + 3(y_1^2 + 2v_1^2) + 16 + g_1^2\}. \end{array}$$

The unique Nash equilibrium for the lower level problem is given by $(v_1^1, v_1^2) = (-8 - u_1, -3 - u_1)$. Consequently, the solution is computed as $u_1 = -6 - x_1, v_1^1 = -5 + v_1^2, v_1^2 = -3 - u_1$. Moreover, we have

$$\begin{aligned} &\hat{V}_2^1(x_1, y_1, u_1) = 119 - 4x_1 + 2y_1^1 - 10y_1^2 - 5u_1, \\ &\hat{V}_2^2(x_1, y_1, u_1) = -2x_1 + 2u_1 + 3y_1^2 + 25, \\ &V_2^0(x_1, y_1) = -6x_1 + 3y_1^1 + 2y_1^2 + 6, \\ &\check{V}_2^1(x_1, y_1) = 149 + x_1 + 2y_1^1 - 10y_1^2, \\ &\check{V}_2^2(x_1, y_1) = -4x_1 + 3y_1^2 + 13. \end{aligned}$$

The final step is to consider the following problem, which corresponds to (3.8) with $t = 0$:

$$\begin{array}{ll} \min_{u_0} & -6x_0 + 3y_0^1 + 2y_0^2 - 6u_0 + 18v_0^1 - 2v_0^2 + 6 + G_0 \\ \text{subject to} & v_0^1 \in \arg\min\{149 + x_0 + 2y_0^1 - 10y_0^2 + u_0 + 2v_0^1 - 19v_0^2 + g_0^1\} \\ & v_0^2 \in \arg\min\{13 - 4x_0 - 4u_0 + 8v_0^1 + 2v_0^2 + 3y_0^2 + g_0^2\}. \end{array}$$

The unique Nash equilibrium for the lower level problem is given by $(v_0^1, v_0^2) = (-u_0 - 2, -u_0 - 2)$. Therefore, the solution is computed as $u_0 = -x_0 + 11, v_0^1 = v_0^2, v_0^2 = -u_0 - 2$. Moreover, we have

$$\begin{aligned} &\hat{V}_3^1(x_0, y_0, u_0) = 188 + x_0 + 2y_0^1 - 10y_0^2 + 5u_0, \\ &\hat{V}_3^2(x_0, y_0, u_0) = -5 - 4x_0 - 14u_0 + 3y_0^2, \\ &V_3^0(x_0, y_0) = -6x_0 + 3y_0^1 + 2y_0^2 + x_0^2 - (x_0 - 11)^2 - 26, \\ &\check{V}_3^1(x_0, y_0) = 243 - 4x_0 + 2y_0^1 - 10y_0^2, \\ &\check{V}_3^2(x_0, y_0) = -159 + 10x_0 + 3y_0^2. \end{aligned}$$

Thus, the optimal decisions are $(u_0, v_0^1, v_0^2) = (10, -12, -12)$, $(u_1, v_1^1, v_1^2) = (-29, 21, 26)$, $(u_2, v_2^1, v_2^2) = (23, -28, -28)$, with the corresponding states $(x_1, y_1^1, y_1^2) = (23, -23, -23)$, $(x_2, y_2^1, y_2^2) = (-22, 19, 29)$ and $(x_3, y_3^1, y_3^2) = (29, -37, -27)$. The optimal values to the leader and the two followers are -126 and 231, -146, respectively.

4 Concluding Remarks

A backward dynamic programming algorithm for DBOP under feedback information structure is put forward in Section 2. It is further extended to the case of dependent followers in Section 3. Since various decision-making problems can be modelled as discrete time dynamic Stackelberg games, the results obtained in this paper are expected to be useful in practice. Nevertheless, we have to point out some limits of the dynamic programming approach. As is well known, the major difficulty in dynamic programming algorithms consists in the curse of dimensionality. Also, the proposed dynamic programming algorithms may not be applied to the case where the state transition equations of a follower depend on other followers' decision. Another difficulty lies in the fact that, in practice, it is not easy to find a global optimal solution to each subproblem. Moreover, in Algorithm 2.2 and Algorithm 3.2, we assume that there exists a unique solution for the lower level problem at each stage. If there exist non-unique responses of the followers, the problem will become intractable. Therefore, the practical applicability of the proposed approach may be somewhat limited. It is an interesting and important subject to develop efficient algorithms for solving more general DBOPs.

Acknowledgment The authors are grateful to an anonymous referee for helpful comments and suggestions on an earlier version of the paper.

References

[1] S. Ambler and A. Paquet: Recursive methods for computing equilibria of general equilibrium dynamic Stackelberg games, Economic Modelling, 14 (1997), pp. 155–173.

[2] G. Anandalingam and T.L. Friesz (eds.): Hierarchical Optimization, Annals of Operations Research, Volume 34, J.C Baltzer AG Scientific Publishing Company, Switzerland, 1992.

[3] J.F. Bard: Optimality conditions for the bilevel programming problem, Naval Research Logistics Quarterly, 31 (1984), pp. 13–26.

[4] T. Başar and G.J. Olsder: Dynamic Noncooperative Game Theory, Second Edition, Academic Press, New York, NY, 1995.

[5] T. Başar and H. Selbuz: Closed-loop Stackelberg strategies with applications in the optimal control of multilevel systems, IEEE Transactions on Automatic Control, 24 (1979), pp. 166–179.

[6] T. Başar and G.J. Olsder: Team-optimal closed-loop Stackelberg strategies in hierarchical control problems, Automatica, 16 (1980), pp. 409–414.

[7] D.P. Bertsekas: Dynamic Programming and Optimal Control, Volume 1, Athena Scientific, Belmont, MA, 1995.

[8] D. Castanon and M. Athans: On stochastic dynamic stackelberg strategies, Automatica, 12 (1976), pp. 177–183.

[9] B. Chen and P. A. Zadrozny: An anticipative feedback solution for the infinite-horizon, linear-quadratic, dynamic, Stackelberg game, Journal of Economic Dynamics and Control, 26 (2002), pp. 1397–1416.

[10] C.I. Chen and J.B. Cruz: Stackelberg solution for two person games with biased information patterns, IEEE Transactions on Automatic Control, 6 (1972), pp. 791–798.

[11] S. Dempe: Foundations of Bilevel Programming, Kluwer Academic Publishers, Dordrechts, The Netherlands, 2002.

[12] G.E. Fruchter and P.R. Messinger: Optimal management of fringe entry over time, Journal of Economic Dynamics and Control, 28 (2003), pp. 445–466.

[13] M. Fukushima, Z.Q. Luo and J.S. Pang: A globally convergent sequential quadratic programming algorithm for mathematical programs with linear complementarity constraints, Computational Optimization and Applications, 10 (1998), pp. 5–34.

[14] M. Fukushima and P. Tseng: An implementable active-set algorithm for computing a B-stationary point of the mathematical program with linear complementarity constraints, SIAM Journal on Optimization, 12 (2002), pp. 724–739.

[15] M. Li, J.B. Cruz, Jr. and M.A. Simaan: An approach to discrete-time incentive feedback Stackelberg games, IEEE Transactions on Systems, Man and Cybernetics, Part A: System and Humans, 32 (2002), pp. 472–481.

[16] G.H. Lin and M. Fukushima: A new relaxation method for mathematical programs with complementarity constraints, Journal of Optimization Theory and Applications, 118 (2003), pp. 81–116.

[17] Z.Q. Luo, J.S. Pang and D. Ralph: Mathematical Programs with Equilibrium Constraints, Cambridge University Press, Cambridge, UK, 1996.

[18] J.S. Pang and M. Fukushima: Complementarity constraint qualifications and simplified B-stationarity conditions for mathematical programs with equilibrium constraints, Computational Optimization and Applicatins, 13 (1999), pp. 111–136.

[19] H. Scheel and S. Scholtes: Mathematical programs with complementarity constraints: Stationarity, optimality and sensitivity, Mathematics of Operation Research, 25 (2000), pp. 1–22.

[20] K. Shimizu, Y. Ishizuka and J. F. Bard: Nondifferentiable and Two Level Mathematical Programming, Kluwer Academic Publishers, Boston, MA, 1997.

[21] M. Simaan and J.B. Cruz, Jr.: On the Stackelberg strategy in nonzero-sum games, Journal of Optimization Theory and Applications, 11 (1973), pp. 533–555.

[22] M. Simaan and J.B. Cruz, Jr.: Additional aspects of the Stackelberg strategy in nonzero-sum games, Journal of Optimization Theory and Applications, 11 (1973), pp. 613–626.

[23] M. Simaan and J.B. Cruz, Jr.: A Stackelberg solution for games with many players, IEEE Transactions on Automatic Control, 18 (1973), pp. 322–324.

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

[25] J.J. Ye: Necessary conditions for bilevel dynamic optimization problems, SIAM Journal on Control and Optimization, 33 (1995), pp. 1208–1223.

[26] J.J. Ye: Optimal strategies for bilevel dynamic problems, SIAM Journal on Control and Optimization, 35 (1997), pp. 512–531.