

THE MAXIMUM PRINCIPLE FOR GLOBAL SOLUTIONS OF STOCHASTIC STACKELBERG DIFFERENTIAL GAMES*

ALAIN BENSOUSSAN[†], SHAOKUAN CHEN[‡], AND SURESH P. SETHI[‡]

Abstract. For stochastic Stackelberg differential games played by a leader and a follower, there are several solution concepts in terms of the players' information sets. In this paper we derive the maximum principle for the leader's global Stackelberg solution under the adapted closed-loop memoryless information structure, where the term *global* signifies the leader's domination over the entire game duration. As special cases, we study linear quadratic Stackelberg games under both adapted open-loop and adapted closed-loop memoryless information structures, as well as the resulting Riccati equations.

Key words. Stackelberg differential game, maximum principle, forward-backward stochastic differential equation, Riccati equation

AMS subject classifications. Primary, 93E20; Secondary, 65K10, 90C46

DOI. 10.1137/140958906

1. Game formulations and definitions of solutions. In 1934, H. von Stackelberg [35] introduced a concept of a hierarchical solution for markets where some firms have power of domination over others. This solution concept, now known as the Stackelberg equilibrium or the Stackelberg solution, involves players with asymmetric roles in the context of two-person nonzero-sum static games—one leading (called the leader) and the other following (called the follower). A Stackelberg game proceeds with the leader first announcing her action at the beginning of the game. With the knowledge of the leader's action, the follower chooses a response so as to optimize his own performance index. The leader, anticipating the follower's optimal response, will pick an action which optimizes her own performance index on the rational reaction curve of the follower. The leader's optimal action and the follower's rational response constitute a Stackelberg solution.

In this paper, we consider Stackelberg differential games in a stochastic context. To formulate this kind a game, we first introduce some notation. Let (Ω, \mathcal{F}, P) be a complete probability space on which is defined a d -dimensional standard Brownian motion $\{W(t), 0 \leq t \leq T\}$. We denote by $\{\mathcal{F}_t\}_{0 \leq t \leq T}$ the natural filtration generated by W and augmented by all the P -null sets in \mathcal{F} and by \mathcal{P} the predictable sub- σ -field of $\mathcal{B}([0, T]) \times \mathcal{F}$.

The state equation is described as the stochastic differential equation.

$$(1.1) \quad dx = f(t, x, u, v)dt + \sigma(t, x)dW(t), \quad x(0) = x_0,$$

where $x_0 \in \mathbb{R}^n$ is a given (deterministic) initial condition and $(u(\cdot), v(\cdot))$ are the decisions of the leader and the follower, with values in subsets U and V in some

*Received by the editors February 27, 2014; accepted for publication (in revised form) January 26, 2015; published electronically July 30, 2015.

<http://www.siam.org/journals/sicon/53-4/95890.html>

[†]Naveen Jindal School of Management, University of Texas at Dallas, Richardson, TX 75080–3021 (alain.bensoussan@utdallas.edu), and Department of Systems Engineering and Engineering Management, City University of Hong Kong, Hong Kong. The research of the first author was supported by the National Science Foundation, DMS 1303775, and by the Research Grant Council of the HKSAR, CityU 500113.

[‡]Naveen Jindal School of Management, University of Texas at Dallas, Richardson, TX 75080–3021 (sxc110120@utdallas.edu, sethi@udallas.edu).

appropriate Euclidean spaces, respectively. The cost functionals for the leader and the follower to minimize are described, respectively, as follows:

$$(1.2) \quad J_1(u, v) = \mathbb{E} \left\{ \int_0^T g_1(t, x, u, v) dt + G_1(x(T)) \right\}$$

and

$$(1.3) \quad J_2(u, v) = \mathbb{E} \left\{ \int_0^T g_2(t, x, u, v) dt + G_2(x(T)) \right\}.$$

In dynamic Stackelberg games, it becomes important to know the players' information sets at any given time as all the decisions are made based on available information sets. For continuous-time deterministic differential games on fixed duration $[0, T]$, Başar and Olsder [4] define the information structures η (see [4, Definition 5.6]) as follows:

- (i) open-loop (OL): $\eta(t) = \{x_0\}$, $t \in [0, T]$;
- (ii) feedback: $\eta(t) = \{x(t)\}$, $t \in [0, T]$;
- (iii) closed-loop memoryless (CLM): $\eta(t) = \{x_0, x(t)\}$, $t \in [0, T]$;
- (iv) closed-loop (CL): $\eta(t) = \{x(s), 0 \leq s \leq t\}$, $t \in [0, T]$.

When the state equation is driven by a Brownian motion as (1.1), it is natural to require that the players' decisions are adapted to the filtration $\{\mathcal{F}_t\}_{0 \leq t \leq T}$. Consequently, the above notions are extended to stochastic settings as follows:

- (i') adapted open-loop (AOL): $\eta(t) = \{x_0, \mathcal{F}_t\}$, $t \in [0, T]$;
- (ii') adapted feedback (AF): $\eta(t) = \{x(t), \mathcal{F}_t\}$, $t \in [0, T]$;
- (iii') adapted closed-loop memoryless (ACLM): $\eta(t) = \{x_0, x(t), \mathcal{F}_t\}$, $t \in [0, T]$;
- (iv') adapted closed-loop (ACL): $\eta(t) = \{x(s), 0 \leq s \leq t, \mathcal{F}_t\}$, $t \in [0, T]$.

With the exceptions of [37] and [23] that deal with the AOL solution of a stochastic Stackelberg game, there is scarce literature on the solution concepts under the other three information structures. Therefore we also undertake to elaborate the concepts (i')–(iv') in what follows. As we will see, the leader has only an instantaneous advantage over the follower in solutions under (ii) and (ii'), and these solutions are called feedback Stackelberg solutions. The leader dominates the follower over the entire duration of the game globally in solutions under the remaining structures, and these solutions are called global Stackelberg solutions. Since the paper is concerned with global solutions, we will omit the term “global” except when we are discussing feedback solutions. For example, a solution under (i') will be referred to simply as an AOL solution.

Now we articulate the ways in which Stackelberg games under different information structures are played as well as define the corresponding solution concepts.

Global solutions. The term “global” indicates that the leader has a global advantage over the follower as compared to the feedback solutions to be discussed later, where the leader has only an instantaneous advantage over the follower. The game is played in the following way. First, as the game begins at time 0, the leader announces her strategy over the whole planning horizon $[0, T]$ and commits to it. Then, with the knowledge of the leader's strategy over the entire horizon, the follower determines his response strategy over the entire horizon, also at time 0, to minimize his cost functional. Since the follower's optimal response can be determined by the leader, the leader can take it into account in finding and announcing her optimal strategy that minimizes her cost functional. One can see that the mechanism here resembles the original one stated by Stackelberg [35], except that now the players need to choose their strategies rather than single actions due to the dynamic nature of the problem.

A strategy is a mapping from the available information set to the action space, which specifies the rule for what action to take in every possible situation. While the realization of the state or the Brownian motion in the information set is uncertain, the strategies of the players are predetermined at the start of the game, and both players must commit to their respective strategies. In this sense, the actions of both players are actually contingent on the outcome of the elements in the information sets. According to different information structures (different elements in the information sets) on which players can base their decisions, there are three types of solution concepts for global Stackelberg games discussed below. One can also refer to Başar and Olsder [4], Dockner et al. [12], Long [19], Petit [32], and the references therein, for a discussion of the concept of a global Stackelberg solution when the underlying state is an ordinary differential equation.

AOL solutions. In a Stackelberg game under the AOL information pattern, the leader's information set is $\eta_t^L = \{x_0, \mathcal{F}_t\}$ and the follower's information set is $\eta_t^F = \{x_0, u(\cdot), \mathcal{F}_t\}$, where $u(\cdot)$ is the leader's strategy, since the follower makes his decision after the leader announces her whole strategy over $[0, T]$. Therefore, the value $u(t)$ of the leader's announced strategy at a future time t can only be contingent on the realization of the Brownian motion on $[0, t]$, i.e., measurable with respect to sigma-algebra \mathcal{F}_t . The value $v(t)$ of the follower's response strategy at a future time t depends on the leader's whole strategy $u(\cdot)$ and is also measurable with respect to \mathcal{F}_t . Then, the leader's strategy space and the follower's response strategy space are

$$\mathcal{U} = \{u \mid u : \{x_0\} \times \Omega \times [0, T] \rightarrow U \text{ is an } \mathcal{F}_t\text{-adapted process}\}$$

and

$$\mathcal{V} = \{v \mid v : \{x_0\} \times \Omega \times [0, T] \times \mathcal{U} \rightarrow V, \text{ and } v(\cdot, x_0, u) \text{ is an } \mathcal{F}_t\text{-adapted process for any } u \in \mathcal{U}\},$$

respectively. Since the follower aims at minimizing his cost functional $J_2(u, v)$ in accordance with the leader's strategy u on the whole duration of the game, his optimal response $v^* \in \mathcal{V}$ should satisfy

$$J_2(u, v^*(u)) \leq J_2(u, v(u)) \quad \forall (u, v) \in \mathcal{U} \times \mathcal{V}.$$

The leader, anticipating the follower's optimal response v^* , picks a policy u^* which optimizes her performance index on the rational reaction curve of the follower, i.e.,

$$J_1(u^*, v^*(u^*)) \leq J_1(u, v^*(u)) \quad \forall u \in \mathcal{U}.$$

The pair $(u^*, v^*(u^*))$ is an AOL solution of the Stackelberg game.

ACLM solutions. If the information structure is ACLM for the players, then the leader's and follower's information sets at time t are $\eta_t^L = \{x_0, x(t), \mathcal{F}_t\}$ and $\eta_t^F = \{x_0, x(t), u(\cdot), \mathcal{F}_t\}$, respectively. This means that both players can make the values of their decisions at time t contingent on additionally the current state information $x(t)$, when compared to the AOL information structure case. Therefore, the leader's and follower's strategy spaces are

$$\mathcal{U} = \{u \mid u : \{x_0\} \times \Omega \times [0, T] \times \mathbb{R}^n \rightarrow U \text{ is an } \mathcal{F}_t\text{-adapted random field}\}$$

and

$$\mathcal{V} = \{v \mid v : \{x_0\} \times \Omega \times [0, T] \times \mathbb{R}^n \times \mathcal{U} \rightarrow V, \text{ and } v(\cdot, x_0, u) \text{ is an } \mathcal{F}_t\text{-adapted random field for any } u \in \mathcal{U}\},$$

respectively. For each of the leader’s strategy $u \in \mathcal{U}$ announced at time 0, the follower aims to find his optimal response $v^* \in \mathcal{V}$ such that

$$J_2(u(\cdot), v^*(\cdot, u)) \leq J_2(u(\cdot), v(\cdot, u)) \quad \forall (u, v) \in \mathcal{U} \times \mathcal{V}.$$

Taking into account the follower’s optimal response, the leader should choose u^* such that

$$J_1(u^*(\cdot), v^*(\cdot, u^*)) \leq J_1(u(\cdot), v^*(\cdot, u)) \quad \forall u \in \mathcal{U}.$$

The pair $(u^*, v^*(u^*))$ is an ACLM solution of the Stackelberg game.

ACL solutions. If the information structure is ACL for the players, then the leader’s and follower’s information sets at time t are $\eta_t^L = \{x(s), 0 \leq s \leq t, \mathcal{F}_t\}$ and $\eta_t^F = \{x(s), 0 \leq s \leq t, u(\cdot), \mathcal{F}_t\}$, respectively. This means that both players can make the values of their decisions at time t contingent on the information history of the state $x(s), 0 \leq s \leq t$. Then, the leader’s and the follower’s strategy spaces are

$$\mathcal{U} = \{u \mid u(t, \mathbf{x}_t) \in U \text{ is } \mathcal{F}_t\text{-measurable for any } \mathbf{x}_t \in (\mathbb{R}^n)^{[0,t]} \quad \forall t \in [0, T]\}$$

and

$$\mathcal{V} = \{v \mid v(t, \mathbf{x}_t, u) \in V \text{ is } \mathcal{F}_t\text{-measurable for any } \mathbf{x}_t \in (\mathbb{R}^n)^{[0,t]} \text{ and } u \in \mathcal{U} \quad \forall t \in [0, T]\},$$

respectively. Similar to the AOL and ACLM cases, a pair $(u^*, v^*(u^*)) \in \mathcal{U} \times \mathcal{V}$ is an ACL solution of the Stackelberg game if

$$J_2(u(\cdot), v^*(\cdot, u)) \leq J_2(u(\cdot), v(\cdot, u)) \quad \forall (u, v) \in \mathcal{U} \times \mathcal{V},$$

and

$$J_1(u^*(\cdot), v^*(\cdot, u^*)) \leq J_1(u(\cdot), v^*(\cdot, u)) \quad \forall u \in \mathcal{U}.$$

When both players determine their strategies at time 0, the evolution of the state is then determined, and the state $x(\cdot)$ is a process adapted to \mathcal{F}_t . Then, at time t , both players will know the value of the state $x(t)$ at time t (see, e.g., p. 53 in [8]). However, they cannot adjust their actions at time t based on the realized value of the state $x(t)$, since in the global games the players fix their strategies over $[0, T]$ at time 0 and must commit to those strategies at that time. If both players, on the other hand, determine their actions based on the realized $x(t)$ at each t , then the follower only observes the leader’s instant action at each time t other than her whole strategy over $[0, T]$ and makes an instant response. This different mechanism leads to the so-called feedback Stackelberg solutions discussed in the following.

Remark 1.1. With the knowledge of an announced leader’s strategy over $[0, T]$ under the AOL or ACLM information structure, the follower is confronted with an optimal control problem regardless of the information structure (AOL, ACLM, or ACL) under which he plays the game. Therefore, from the follower’s perspective, his optimal response strategies under the AOL, ACLM, and ACL structures collapse in one, in the sense that they have the same value at each time t . On the other hand, for a given leader’s strategy under the ACL information structure, the follower’s problem is no longer an optimal control problem in the standard sense. Therefore, his optimal response strategy will in general depend on the information history of the state. As

a result, the follower's optimal response strategies under the AOL, ACLM, and ACL information structures are different in general.

Remark 1.2. The leader's global solutions under the AOL, ACLM, and ACL information structures are generally not the same due to the leader-follower discipline in the game framework. This is a critical departure from the well-known result in the optimal control problem setting that the optimal AOL, ACLM, and ACL controls coincide with each other, as illustrated for the follower's problem in Remark 1.1. The reason the leader's optimal AOL and ACLM solutions are not identical in the present game setting is as follows: The realization of the Brownian motion on $[0, t]$ alone is not enough for the leader to infer the state $x(t)$ when she makes her decision at time 0 in the AOL information structure, since the follower makes his decision after the leader announces her whole strategy over $[0, T]$ and the evolution of the state also depends on the follower's strategy. This does not mean that the process $x(\cdot)$ is not adapted to \mathcal{F}_t ; it only means that the evolution of the state is not available to the leader when she first makes her decision at time 0. One can refer to Simaan and Cruz [34] to see the difference between the leader's optimal OL and CLM (corresponding to AOL and ACLM in the deterministic case) strategies. Another example is from section 5 in this paper, where we treat a linear quadratic (LQ) Stackelberg game under the AOL information structure and an equivalent problem of the LQ game under the ACLM information structure. It can be seen that the leader has one more decision variable in the latter case when compared to the AOL case.

Feedback solutions. A Stackelberg game under the AF information structure differs from those under the above-mentioned three information structures not only in the information structure itself, but also in the way the game is played.

AF solutions. In a Stackelberg game under the AF information structure, the leader's information set is $\eta_t^L = \{x(t), \mathcal{F}_t\}$ and the follower's information set is $\eta_t^F = \{x(t), u(t), \mathcal{F}_t\}$, where $u(t)$ is the leader's action at time t . The way the game is played is as follows: In each instant t , the leader takes an action based on $\eta_t^L = \{\mathcal{F}_t, x(t)\}$, before the follower makes his decision; after observing the leader's instant action at time t , rather than her whole strategy over $[0, T]$, the follower makes an instant response based on $\eta_t^F = \{x(t), u(t), \mathcal{F}_t\}$. Therefore, the leader's and follower's strategy spaces are

$$\mathcal{U} = \{u \mid u : \Omega \times [0, T] \times \mathbb{R}^n \rightarrow U \text{ is an } \mathcal{F}_t\text{-adapted random field}\}$$

and

$$\mathcal{V} = \{v \mid v : \Omega \times [0, T] \times \mathbb{R}^n \times U \rightarrow V \text{ is an } \mathcal{F}_t\text{-adapted random field}\},$$

respectively. This game requires the players to find menus $(u, v) \in \mathcal{U} \times \mathcal{V}$ so that any one deviating from the respective menus will be worse off. Therefore, an AF solution is a pair of strategies $(u^*, v^*) \in \mathcal{U} \times \mathcal{V}$ such that

$$\begin{aligned} J_1(u^*(\cdot), v^*(\cdot, u^*(\cdot))) &\leq J_1(u(\cdot), v^*(\cdot, u(\cdot))) \quad \forall u \in \mathcal{U}, \\ J_2(u^*(\cdot), v^*(\cdot, u^*(\cdot))) &\leq J_2(u^*(\cdot), v(\cdot, u^*(\cdot))) \quad \forall v \in \mathcal{V}, \end{aligned}$$

where we should stress that $u(\cdot)$, $v(\cdot, u(\cdot))$ evaluated at any (t, y) are $u(t, y)$, $v(t, y, u(t, y))$, respectively. It can also be shown that in the Markovian case, one can find deterministic menus $u(t, x)$ and $v(t, x, u(t, x))$ that are independent of the sigma algebra \mathcal{F}_t .

In the above definitions, we can see that the inequalities of the follower in the AOL, ACLM, and ACL solutions must hold for all $(u, v) \in \mathcal{U} \times \mathcal{V}$, whereas the inequality in the AF solution must hold only for all $v \in \mathcal{V}$ with $u^*(\cdot)$ appearing on both sides. This means that the AF Stackelberg solution is obtained as an equilibrium, whereas the AOL, ACLM, and ACL solutions are derived from two optimization problems defined over the entire game duration $[0, T]$ and solved in sequence—first one by the follower and the second one by the leader.

Remark 1.3. We can see from the above definitions that the leader is better off in a Stackelberg game under the ACLM information structure than the AF information structure. Indeed, if (u^*, v^*) is an AF solution and the leader chooses u^* in a Stackelberg game under the ACLM information structure, her cost is the same as that in the Stackelberg game under the AF information structure since the follower's optimal responses are the same in both games.

In this paper, we will focus on Stackelberg games under the AOL and ACLM information structures. The approach employed here is that of the maximum principle for controlled stochastic differential equations (SDEs) and forward-backward stochastic differential equations (FBSDEs). To be precise, we view the follower's Hamiltonian system as a state equation of the leader and derive the necessary conditions for the leader's global Stackelberg solutions under the AOL and ACLM information structures. In the next section we review the relevant literature, discuss how our results are related to it, and describe our contributions.

2. Literature review. Here we review the literature on Stackelberg games under the information structures specified in (i) and (i'), (iii), and (iii'). For other types of games, one can refer to [1], [2], [3], [5], [6], [24], [26], [27], and [34].

2.1. Stackelberg differential games under OL and AOL information structures. Deterministic Stackelberg differential games under the OL information structure can be tackled by utilizing the standard results of optimal control theory. One can refer to the monograph [4] for details. For those under the AOL information structure, since the follower's adjoint equation turns out to be a backward stochastic differential equation (BSDE), the leader will be confronted with a control problem (as we will see in section 3) where the state equation is an FBSDE. Yong [37] considers an LQ game under the AOL structure in a very general setting. The related Riccati equations for the follower and for the leader are derived *sequentially*, and the sufficient conditions for their solvability are provided in the special case of deterministic coefficients. Øksendal, Sandal, and Uboe [23] consider a general stochastic Stackelberg differential game with delayed information, prove the maximum principle, and apply it to continuous-time newsvendor problems.

In subsection 5.1, we investigate the LQ case where the weight matrices in the cost functionals are positive definite and controls do not appear in the diffusion term of the state equation. The novelty here is that we derive the related stochastic Riccati equation for the follower and the leader *simultaneously* by having the follower's Hamiltonian system as the leader's state equation. Moreover, we identify easily verifiable conditions under which the stochastic Riccati equation has a unique solution (Proposition 5.2).

2.2. Stackelberg differential games under CLM and ACLM information structures. The CLM or ACLM information pattern possesses the feature that the dependence of the strategy on the history of the state is restricted to the initial and the current states. As a result, the follower's problem can be solved by the

maximum principle in optimal control theory. However, the leader will encounter, after incorporating the follower's adjoint variable as an augmented state, a nonclassical optimal control problem where the controlled system contains the derivative of the control with respect to the state. For the CLM case, Papavassilopoulos and Cruz [25] provide two approaches to get the necessary conditions satisfied by the leader's optimal strategy. One is to directly apply the variational techniques to the state system with mixed-boundary conditions (the adjoint equation of the follower with a terminal condition). The other is to establish an equivalent relationship between the resulting nonclassical control problem and a classical control problem, which shows that the optimal strategy could be found in the space of affine functions. The phenomenon of time inconsistency is also analyzed by the authors.

The problem with the ACL information structure is mentioned but not addressed in [37]. In this paper we address the ACLM information case and show that the maximum principle for the leader's ACLM solution of a stochastic Stackelberg game is closely related to the theory of controlled FBSDEs. As far as we know, not much is available on ACL solutions of stochastic Stackelberg differential games in the literature, and they are yet to be studied.

The remainder of this paper is organized as follows. In section 3, we present the maximum principle for a Stackelberg game under the AOL information pattern, which is well known (see, e.g., [33] and [39]) and serves as a basis for the study of the ACLM case. In section 4, we focus on Stackelberg games under the ACLM information pattern and derive the maximum principle for the leader's optimal strategy. As examples, LQ Stackelberg games under the AOL and ACLM information structures are considered and compared in section 5. In the former case, we show the existence and uniqueness of the solution to the associated stochastic Riccati equation under some assumptions. In the latter case, we merely derive the associated Riccati equation which consists of FBSDE, and leave the issue of the existence of its solution for future work due to the quadratic and irregular feature. We conclude the paper in section 6.

3. Notation and preliminaries. We first introduce some notation. For two vectors x and y in \mathbb{R}^n , $\langle x, y \rangle$ means the inner product $\sum_{i=1}^n x_i y_i$. Throughout the paper, all the vectors are column vectors. The gradient of a scalar function f is $\frac{\partial f}{\partial x} = (\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n})^\top$, while the gradient of a vector function $f = (f_1, \dots, f_m)^\top$ is the matrix

$$\frac{\partial f}{\partial x} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}.$$

We introduce two spaces of adapted processes to be used in the definition of the solution to an FBSDE:

$$\begin{aligned} \mathcal{S}^2(0, T; \mathbb{R}^n) &:= \{ \psi \mid \psi : \Omega \times [0, T] \rightarrow \mathbb{R}^n \text{ is a continuous } \mathcal{F}_t\text{-adapted process such that} \\ &\quad \mathbb{E} \sup_{0 \leq t \leq T} |\psi(t)|^2 < \infty \}, \\ \mathcal{M}^2(0, T; \mathbb{R}^n) &:= \left\{ \psi \mid \psi : \Omega \times [0, T] \rightarrow \mathbb{R}^n \text{ is an } \mathcal{F}_t\text{-adapted process such that} \right. \\ &\quad \left. \mathbb{E} \int_0^T |\psi(t)|^2 dt < \infty \right\}. \end{aligned}$$

The above two spaces will be simply written as \mathcal{S}^2 and \mathcal{M}^2 , respectively, if no confusion arises.

For the AOL information structure, the admissible strategy spaces for the leader and the follower are denoted by

$$\mathcal{U} = \left\{ u \mid u : \Omega \times [0, T] \rightarrow U \text{ is } \mathcal{F}_t\text{-adapted and } \mathbb{E} \int_0^T |u(t)|^2 dt < +\infty \right\},$$

$$\mathcal{V} = \left\{ v \mid v : \Omega \times [0, T] \times \mathcal{U} \rightarrow V, v(\cdot, u) \text{ is } \mathcal{F}_t\text{-adapted,} \right.$$

$$\left. \text{and } \mathbb{E} \int_0^T |v(t, u)|^2 dt < +\infty \text{ for } u \in \mathcal{U} \right\},$$

where U and V are subsets of \mathbb{R}^{m_1} and \mathbb{R}^{m_2} , respectively. We omit the dependence of the strategies on the initial state x_0 , which is fixed and is commonly known by both players.

We assume that the coefficients f and σ in the state equation (1.1) and g_i and G_i , $i = 1, 2$, in the cost functionals (1.2) and (1.3) are specified as follows:

$$f : \Omega \times [0, T] \times \mathbb{R}^n \times \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \rightarrow \mathbb{R}^n \text{ and } \sigma : \Omega \times [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^{n \times d}$$

are $\mathcal{P} \times \mathcal{B}(\mathbb{R}^{n+m_1+m_2})/\mathcal{B}(\mathbb{R}^n)$ and $\mathcal{P} \times \mathcal{B}(\mathbb{R}^n)/\mathcal{B}(\mathbb{R}^{n \times d})$ measurable, respectively, and

$$g_i : \Omega \times [0, T] \times \mathbb{R}^n \times U \times V \rightarrow \mathbb{R} \text{ and } G_i : \Omega \times \mathbb{R}^n \rightarrow \mathbb{R}$$

are $\mathcal{P} \times \mathcal{B}(\mathbb{R}^n) \times \mathcal{B}(U) \times \mathcal{B}(V)/\mathcal{B}(\mathbb{R})$ and $\mathcal{F}_T \times \mathcal{B}(\mathbb{R}^n)/\mathcal{B}(\mathbb{R})$ measurable, respectively. For $\psi(t, x, u, v) = f(t, x, u, v)$, $\sigma(t, x)$, $g_1(t, x, u, v)$, $g_2(t, x, u, v)$, $G_1(x)$, $G_2(x)$, we assume throughout the paper that ψ and its first and second derivatives are uniformly Lipschitz with respect to (x, u, v) and $\psi(\cdot, x, u, v) \in \mathcal{M}^2$, for $(x, u, v) \in \mathbb{R}^n \times \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$.

For completeness of this paper, we state the formulation of a general Stackelberg game under the AOL information pattern and the corresponding maximum principle. From the definition in section 1, given the leader's strategy $u \in \mathcal{U}$, the follower faces the stochastic control problem

$$\min_{v \in \mathcal{V}} J_2(u, v) = \mathbb{E} \left\{ \int_0^T g_2(t, x, u, v) dt + G_2(x(T)) \right\}$$

subject to

$$dx = f(t, x, u, v)dt + \sigma(t, x)dW(t), \quad x(0) = x_0.$$

Suppose there exists a solution $v^*(u) \in \mathcal{V}$ to the above problem for each $u \in \mathcal{U}$. If we define

$$H_2(t, x, u, v, p_2, q_2) := \langle p_2, f(t, x, u, v) \rangle + \langle q_2, \sigma(t, x) \rangle + g_2(t, x, u, v),$$

then the maximum principle (see, e.g., pp. 4131–4133 in [39]) states that there exists a pair of adapted processes $(p_2, q_2) \in \mathcal{S}^2 \times \mathcal{M}^2$ such that

$$(3.1) \quad \begin{cases} dx = f(t, x, u, v^*)dt + \sigma(t, x)dW(t), \\ -dp_2 = \left\{ \left(\frac{\partial f}{\partial x} \right)^\top (t, x, u, v^*)p_2 + \left(\frac{\partial \sigma}{\partial x} \right)^\top (t, x)q_2 \right. \\ \quad \left. + \frac{\partial g_2}{\partial x}(t, x, u, v^*) \right\} dt - q_2 dW(t), \\ x(0) = x_0, \quad p_2(T) = \frac{\partial G_2}{\partial x}(x(T)), \\ v^*(t, u) = \arg \min_{v \in V} H_2(t, x(t), u(t), v, p_2(t), q_2(t)). \end{cases}$$

We should stress that in the last equation, in addition to the pointwise dependence of $v^*(t, u)$ on the current value $u(t)$, the functional dependence of v^* on the entire strategy u is implicit and reflected by the adjoint variable term p_2 in H_2 . We assume that by the last equation in (3.1), the function $v = v^*(t, x, u, p_2)$ is uniquely defined and is uniformly Lipschitz continuous with respect to (x, u, p_2) and continuously differentiable. After substituting $v = v^*(t, x, u, p_2)$ into the follower’s maximum principle, we get the following control problem faced by the leader:

$$\min_{u \in \mathcal{U}} J_1(u) = \mathbb{E} \left\{ \int_0^T g_1(t, x, u, v^*(t, x, u, p_2))dt + G_1(x(T)) \right\}$$

subject to

$$(3.2) \quad \begin{cases} dx = f(t, x, u, v^*(t, x, u, p_2))dt + \sigma(t, x)dW(t), \\ -dp_2 = \left\{ \left(\frac{\partial f}{\partial x} \right)^\top (t, x, u, v^*(t, x, u, p_2))p_2 + \left(\frac{\partial \sigma}{\partial x} \right)^\top (t, x)q_2 \right. \\ \quad \left. + \frac{\partial g_2}{\partial x}(t, x, u, v^*(t, x, u, p_2)) \right\} dt - q_2 dW(t), \\ x(0) = x_0, \quad p_2(T) = \frac{\partial G_2}{\partial x}(x(T)). \end{cases}$$

We assume that

$$F_2(t, u, x, p_2, q_2) := \left(\frac{\partial f}{\partial x} \right)^\top (t, x, u, v^*(t, x, u, p_2))p_2 + \left(\frac{\partial \sigma}{\partial x} \right)^\top (t, x)q_2 + \frac{\partial g_2}{\partial x}(t, x, u, v^*(t, x, u, p_2))$$

is uniformly Lipschitz with respect to (u, x, p_2, q_2) and $F_2(\cdot, u, x, p_2, q_2) \in \mathcal{M}^2$. We further assume the following monotone conditions (see [31]): there exists an $n \times n$ full-rank matrix M such that

$$\begin{aligned} \langle \Lambda(t, u, \lambda) - \Lambda(t, u, \bar{\lambda}), \lambda - \bar{\lambda} \rangle &\leq -\beta_1 |M\hat{x}|^2 - \beta_2 (|M^\top \hat{p}_2|^2 + |M^\top \hat{q}_2|^2), \\ \left\langle \frac{\partial G_2}{\partial x}(x) - \frac{\partial G_2}{\partial x}(\bar{x}), M(x - \bar{x}) \right\rangle &\geq \mu |M\hat{x}|^2, \end{aligned}$$

where $\beta_1, \beta_2,$ and μ are nonnegative constants with $\beta_1 + \beta_2 > 0, \mu + \beta_2 > 0, \lambda = (x, p_2, q_2), \bar{\lambda} = (\bar{x}, \bar{p}_2, \bar{q}_2), \hat{x} = x - \bar{x}, \hat{p}_2 = p_2 - \bar{p}_2, \hat{q}_2 = q_2 - \bar{q}_2,$ and

$$\Lambda(t, u, \lambda) = \begin{pmatrix} M^\top F_2 \\ Mf \\ M\sigma \end{pmatrix} (t, u, \lambda).$$

The above conditions, first proposed by Peng and Wu [31], guarantee that the system (3.2) has a unique solution and thereby make the leader’s problem well-posed. In fact, since the celebrated work of Pardoux and Peng [28], nonlinear BSDEs and FBSDEs have been extensively studied in the last two decades. One can, e.g., refer to [14], [20], [21], [29], [31], [38] and the references therein for the developed theory as well as related applications.

Although not motivated initially by stochastic Stackelberg differential games, the maximum principle for controlled systems of FBSDEs provides an answer for solving the leader’s problem in the AOL case. One can refer to Shi and Wu [33] for the case when the forward diffusion coefficient does not involve the control variable and to Yong [38, 39] for a general case. Since we will establish the maximum principle for the leader’s ACLM solution on the basis of that for controlled systems of FBSDEs, we now state the latter formally.

PROPOSITION 3.1. *Let the Lipschitz and monotone assumptions on the coefficients hold. Suppose u^* is an optimal strategy for the leader. If we denote*

$$(3.3) \quad \begin{aligned} & H_1(t, u, x, y, p_1, p_2, q_1, q_2) \\ & = \langle p_1, f(t, x, u, v^*(t, x, u, p_2)) \rangle + \langle q_1, \sigma(t, x) \rangle + g_1(t, x, u, v^*(t, x, u, p_2)) \\ & \quad - \langle y, F_2(t, u, x, p_2, q_2) \rangle, \end{aligned}$$

then there exists a triple of adapted processes $(y, p_1, q_1) \in \mathcal{S}^2 \times \mathcal{S}^2 \times \mathcal{M}^2$ such that

$$(3.4) \quad u^*(t) = \arg \min_{u \in U} H_1(t, u, x(t), y(t), p_1(t), p_2(t), q_1(t), q_2(t))$$

and

$$(3.5) \quad \left\{ \begin{aligned} & dy = -\frac{\partial H_1}{\partial p_2} dt - \frac{\partial H_1}{\partial q_2} dW(t), \\ & = -\left\{ \left(\frac{\partial f}{\partial v} \frac{\partial v^*}{\partial p_2} \right)^\top p_1 - \frac{\partial f}{\partial x} y - \sum_{i=1}^n y_i \left(\frac{\partial v^*}{\partial p_2} \right)^\top \frac{\partial}{\partial v} \left(\frac{\partial f}{\partial x_i} \right)^\top p_2 \right. \\ & \quad \left. - \left(\frac{\partial^2 g_2}{\partial x \partial v} \frac{\partial v^*}{\partial p_2} \right)^\top y + \left(\frac{\partial v^*}{\partial p_2} \right)^\top \frac{\partial g_1}{\partial v} \right\} dt - \frac{\partial \sigma}{\partial x} y dW(t), \\ & dp_1 = -\frac{\partial H_1}{\partial x} dt + q_1 dW(t) \\ & = -\left\{ \frac{\partial f}{\partial x} + \frac{\partial f}{\partial v} \frac{\partial v^*}{\partial x} + \left(\frac{\partial \sigma}{\partial x} \right)^\top q_1 + \frac{\partial g_1}{\partial x} + \left(\frac{\partial v^*}{\partial x} \right)^\top \frac{\partial g_1}{\partial v} \right. \\ & \quad \left. - \sum_i y_i \left[\frac{\partial}{\partial x} \left(\frac{\partial f}{\partial x_i} \right)^\top + \left(\frac{\partial v^*}{\partial x} \right)^\top \frac{\partial}{\partial v} \left(\frac{\partial f}{\partial x_i} \right)^\top \right] p_2 \right. \\ & \quad \left. - \sum_i y_i \frac{\partial}{\partial x} \left(\frac{\partial \sigma}{\partial x_i} \right)^\top q_2 - \left(\frac{\partial^2 g_2}{\partial x^2} + \frac{\partial^2 g_2}{\partial x \partial v} \frac{\partial v^*}{\partial x} \right)^\top y \right\} dt + q_1 dW(t), \\ & y(0) = 0, \quad p_1(T) = -\frac{\partial^2 G_2}{\partial x^2}(x(T))y(T) + \frac{\partial G_1}{\partial x}(x(T)). \end{aligned} \right.$$

4. Stackelberg games under ACLM information structure. The difficulty in studying Stackelberg games under the CL or ACL information structure arises from the fact that the reaction of the follower may not be determined explicitly if the leader’s strategy depends on the whole history of the state. However, if the leader’s strategy is restricted to be memoryless, i.e., if only the current state, rather than the whole history of the state, is involved in the strategy, then a way to solve the problem under the CLM information pattern is given by Papavassilopoulos and Cruz [25].

In this section, we treat stochastic Stackelberg differential games under the ACLM information structure and thus generalize the CLM case studied in [25]. It is noteworthy to mention that in contrast to the AOL case in the preceding section, the leader in the ACLM case ends up with an optimal control problem in which the state equation consists of an SDE and a BSDE, with the feature that both the control u and its derivative $\frac{\partial u}{\partial x}$ are involved in the controlled system. This gives rise to a difficult nonstandard control problem. Nevertheless, armed with the developments in the study of controlled systems of FBSDEs (see, e.g., [33], [38], [39]) and the technique of equivalent transformation in [25], we are able to prove the necessary conditions for the leader’s ACLM solution.

We first introduce the admissible strategy spaces for the leader and the follower:

$$\mathcal{U} := \left\{ u | u : \Omega \times [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^{m_1} \text{ is } \mathcal{F}_t\text{-adapted for any } x \in \mathbb{R}^n, u(t, x) \right. \\ \left. \text{is continuously differentiable in } x \text{ for any } (\omega, t) \in \Omega \times [0, T], \text{ and } \left| \frac{\partial u}{\partial x} \right| \leq K \right\},$$

$$\mathcal{V} := \{ v | v : \Omega \times [0, T] \times \mathbb{R}^n \times \mathcal{U} \rightarrow \mathbb{R}^{m_2} \text{ is } \mathcal{F}_t\text{-adapted for any } x \in \mathbb{R}^n \text{ and } u \in \mathcal{U} \}.$$

Then, given the leader’s strategy $u \in \mathcal{U}$, the follower’s optimal response strategy $v^* \in \mathcal{V}$ is a solution of the following classical optimal control problem:

$$(4.1) \quad \min_{v \in \mathcal{V}} J_2 = \mathbb{E} \left\{ \int_0^T g_2(t, x, u(t, x), v) dt + G_2(x(T)) \right\}$$

subject to

$$(4.2) \quad dx = f(t, x, u(t, x), v) dt + \sigma(t, x) dW(t), \quad x(0) = x_0.$$

Similar to the AOL case, the maximum principle yields that there exists a pair of adapted processes $(p_2, q_2) \in \mathcal{S}^2 \times \mathcal{M}^2$ such that

$$(4.3) \quad v^*(t, x(t), u(\cdot)) = \arg \min_{v \in \mathcal{V}} \{ \langle p_2(t), f(t, x(t), u(t, x(t)), v) \rangle + \langle q_2, \sigma(t, x) \rangle \\ + g_2(t, x(t), u(t, x(t)), v) \}$$

and

$$(4.4) \quad \begin{cases} dp_2 = - \left[\left(\frac{\partial f}{\partial x} + \frac{\partial f}{\partial u} \frac{\partial u}{\partial x} \right)^\top p_2 + \left(\frac{\partial \sigma}{\partial x} \right)^\top q_2 \right. \\ \left. + \frac{\partial g_2}{\partial x} + \left(\frac{\partial u}{\partial x} \right)^\top \frac{\partial g_2}{\partial u} \right] dt + q_2 dW(t), \\ p_2(T) = \frac{\partial G_2}{\partial x}(x(T)), \end{cases}$$

where $x(\cdot)$ is the solution of (4.2) with the policies u and v^* . Suppose that for any leader's strategy $u \in \mathcal{U}$, there exists a unique strategy v^* for the follower that minimizes his cost functional J_2 and that (4.3) yields $v^* = \varphi(t, x, u, p_2)$. Then, taking into account the follower's optimal response, the leader will be confronted with solving the optimal control problem:

$$(4.5) \quad \min_{u \in \mathcal{U}} J_1 = \mathbb{E} \left\{ \int_0^T g_1(t, x, u(t, x), \varphi(t, x, u(t, x), p_2)) dt + G_1(x(T)) \right\}$$

subject to

$$(4.6) \quad \begin{cases} dx = f(t, x, u(t, x), \varphi(t, x, u(t, x), p_2)) dt + \sigma(t, x) dW(t), \\ dp_2 = - \left[\left(\frac{\partial f}{\partial x} + \frac{\partial f}{\partial u} \frac{\partial u}{\partial x} \right)^\top p_2 + \left(\frac{\partial \sigma}{\partial x} \right)^\top q_2 \right. \\ \quad \left. + \frac{\partial g_2}{\partial x} + \left(\frac{\partial u}{\partial x} \right)^\top \frac{\partial g_2}{\partial u} \right] dt + q_2 dW(t), \\ x(0) = x_0, \quad p_2(T) = \frac{\partial G_2}{\partial x}(x(T)). \end{cases}$$

The solvability of (4.6) is due to our assumption that the follower has a unique optimal response strategy v^* for every strategy $u \in \mathcal{U}$ of the leader. Here we further assume that the leader's problem is well-posed, i.e., for each $u \in \mathcal{U}$, there exists a unique triple $(x, p_2, q_2) \in \mathcal{S}^2 \times \mathcal{S}^2 \times \mathcal{M}^2$ that solves FBSDE (4.6). As mentioned earlier, the presence of the derivative $\frac{\partial u}{\partial x}$ of the strategy u in (4.6) results in a nonstandard optimal control problem for the leader. We now introduce a standard stochastic optimal control problem and establish the equivalence between the two in the sense that they have the same optimal trajectory and cost.

Consider the following optimization problem of a controlled FBSDE:

$$(4.7) \quad \min_{u_1, u_2} J(u_1, u_2) = \mathbb{E} \left\{ \int_0^T g_1(t, x, u_2 x + u_1, \varphi(t, x, u_2 x + u_1, p_2)) dt + G_1(x(T)) \right\}$$

subject to

$$(4.8) \quad \begin{cases} dx = f(t, x, u_2 x + u_1, \varphi(t, x, u_2 x + u_1, p_2)) dt + \sigma(t, x) dW(t), \\ dp_2 = - \left[\left(\frac{\partial f}{\partial x} + \frac{\partial f}{\partial u} u_2 \right)^\top p_2 + \left(\frac{\partial \sigma}{\partial x} \right)^\top q_2 \right. \\ \quad \left. + \frac{\partial g_2}{\partial x} + (u_2)^\top \frac{\partial g_2}{\partial u} \right] dt + q_2 dW(t), \\ x(0) = x_0, \quad p_2(T) = \frac{\partial G_2}{\partial x}(x(T)), \end{cases}$$

where u_1 and u_2 are adapted decision processes with values in \mathbb{R}^{m_1} and the ball $B_K(\mathbb{R}^{m_1 \times n})$ with radius K in $\mathbb{R}^{m_1 \times n}$, respectively. Again we assume that for any $(u_1, u_2) \in \mathbb{R}^{m_1} \times B_K(\mathbb{R}^{m_1 \times n})$,

$$\tilde{F}_2(t, u_1, u_2, x, p_2, q_2) := \left(\frac{\partial f}{\partial x} + \frac{\partial f}{\partial u} u_2 \right)^\top p_2 + \left(\frac{\partial \sigma}{\partial x} \right)^\top q_2 + \frac{\partial g_2}{\partial x} + (u_2)^\top \frac{\partial g_2}{\partial u}$$

is uniformly Lipschitz in (x, p_2, q_2) and f, σ, \tilde{F}_2 , and $\frac{\partial G_2}{\partial x}$ satisfy the monotone condition in section 3 (the LQ case in the next section satisfies these assumptions). Therefore, the problem (4.7)–(4.8) is well-posed and the maximum principle can be referred to Proposition 3.1.

THEOREM 4.1. *Suppose that the above Lipschitz and monotone conditions hold for the coefficients of the problem (4.7)–(4.8). If $u^* \in \mathcal{U}$ is a solution to the leader’s problem (4.5)–(4.6) with the corresponding state trajectory (x^*, p_2^*, q_2^*) , then there exists a triple $(y, p_1, q_1) \in \mathcal{S}^2 \times \mathcal{S}^2 \times \mathcal{M}^2$ such that*

$$(4.9) \quad \left(u^*(t, x^*(t)) - \frac{\partial u^*}{\partial x}(t, x^*(t))x^*(t), \frac{\partial u^*}{\partial x}(t, x^*(t)) \right) \\ = \arg_{(u_1, u_2)} \min H_1(t, u_1, u_2, x^*(t), y(t), p_1(t), p_2^*(t), q_1(t), q_2^*(t))$$

and

$$(4.10) \quad \begin{cases} dy = -\frac{\partial H_1}{\partial p_2} dt - \frac{\partial H_1}{\partial q_2} dW(t), \\ dp_1 = -\frac{\partial H_1}{\partial x} dt + q_1 dW(t), \\ y(0) = 0, \quad p_1(T) = -\frac{\partial^2 G_2}{\partial x^2}(x^*(T))y(T) + \frac{\partial G_1}{\partial x}(x^*(T)). \end{cases}$$

Here

$$(4.11) \quad H_1(t, u_1, u_2, x, y, p_1, p_2, q_1, q_2) \\ = \langle p_1, f(t, x, u_2x + u_1, \varphi(t, x, u_2x + u_1, p_2)) \rangle + \langle q_1, \sigma(t, x) \rangle - \langle y, \tilde{F}_2(t, u_1, u_2, x, p_2, q_2) \rangle \\ + g_1(t, x, u_2x + u_1, \varphi(t, x, u_2x + u_1, p_2)),$$

and $\frac{\partial H_1}{\partial p_2}, \frac{\partial H_1}{\partial q_2}$, and $\frac{\partial H_1}{\partial x}$ in (4.10) are evaluated at the point

$$\left(t, \left\{ u^*(t, x^*(t)) - \frac{\partial u^*}{\partial x}(t, x^*(t))x^*(t) \right\}, \left\{ \frac{\partial u^*}{\partial x}(t, x^*(t)) \right\}, \right. \\ \left. x^*(t), y(t), p_1(t), p_2^*(t), q_1(t), q_2^*(t) \right).$$

Proof. It is obvious that if J_1^* and J^* denote the optimal costs associated with problems (4.5)–(4.6) and (4.7)–(4.8), respectively, then $J_1^* \leq J^*$. On the other hand, from the solution u^* of problem (4.5)–(4.6), we can construct a pair of control processes (u_1^*, u_2^*) for problem (4.7)–(4.8) as follows:

$$(4.12) \quad u_1^*(t) := u^*(t, x^*(t)) - \frac{\partial u^*}{\partial x}(t, x^*(t))x^*(t), \\ u_2^*(t) := \frac{\partial u^*}{\partial x}(t, x^*(t)).$$

With these controls, the FBSDE (4.8) has the same solution as that of (4.6) with the optimal strategy u^* and, moreover, along the trajectory $x^*(t)$,

$$u_2^*(t)x^*(t) + u_1^*(t) = u^*(t, x^*(t)).$$

Therefore, $J_1^* = J^*$ and the above constructed (u_1^*, u_2^*) is an optimal control for problem (4.7)–(4.8), leading to the same state trajectory (x^*, p_2^*, q_2^*) .

From Proposition 3.1 we can then conclude that if (u_1^*, u_2^*) is a solution for problem (4.7)–(4.8), there exists a triple $(y, p_1, q_1) \in \mathcal{S}^2 \times \mathcal{S}^2 \times \mathcal{M}^2$ such that

$$(4.13) \quad \begin{aligned} & (u_1^*, u_2^*) \\ & = \arg_{(u_1, u_2)} \min H_1(t, u_1, u_2, x(t), y(t), p_1(t), p_2(t), q_1(t), q_2(t)) \end{aligned}$$

and

$$(4.14) \quad \begin{cases} dy = -\frac{\partial H_1}{\partial p_2} dt - \frac{\partial H_1}{\partial q_2} dW(t), \\ dp_1 = -\frac{\partial H_1}{\partial x} dt + q_1 dW(t), \\ y(0) = 0, \quad p_1(T) = -\frac{\partial^2 G_2}{\partial x^2}(x(T))y(T) + \frac{\partial G_1}{\partial x}(x(T)), \end{cases}$$

where (x, p_2, q_2) is the solution of the state equation (4.8) with control (u_1^*, u_2^*) , and $\frac{\partial H_1}{\partial p_2}$, $\frac{\partial H_1}{\partial q_2}$, and $\frac{\partial H_1}{\partial x}$ in (4.14) are evaluated at the point

$$(t, u_1^*(t), u_2^*(t), x(t), y(t), p_1(t), p_2(t), q_1(t), q_2(t)).$$

From the above arguments we know that

$$\left\{ \left(u^*(t, x^*(t)) - \frac{\partial u^*}{\partial x}(t, x^*(t))x^*(t), \frac{\partial u^*}{\partial x}(t, x^*(t)) \right) \right\}_{t \in [0, T]}$$

is an optimal control for problem (4.7)–(4.8) if $u^* \in \mathcal{U}$ is a solution for problem (4.5)–(4.6) with the corresponding forward state x^* . Therefore, we can finally obtain the maximum principle for problem (4.5)–(4.6) of the leader by substituting

$$\left\{ \left(u^*(t, x^*(t)) - \frac{\partial u^*}{\partial x}(t, x^*(t))x^*(t), \frac{\partial u^*}{\partial x}(t, x^*(t)) \right) \right\}_{t \in [0, T]}$$

into the necessary conditions satisfied by the optimal control for problem (4.7)–(4.8). \square

Remark 4.1. If u is independent of x , we conclude by comparison with the arguments in section 3 that the ACLM Stackelberg solution is reduced to the AOL Stackelberg solution, and thus the maximum principle for both cases coincides.

5. LQ Stackelberg games. In this section we consider LQ Stackelberg games under the AOL and ACLM information structures. In the former case, Yong [37] derives stochastic Riccati equations for the follower and the leader sequentially in a setting in which the coefficients are random, the diffusion term of the state equation contains controls, and the weight matrices in the cost functionals are not necessarily positive definite. To be more precise, first the follower obtains his Riccati equation for any given strategy of the leader. Then the leader solves her problem involving a system of FBSDE, whose coefficients depend on the solution of the follower’s Riccati equation. Finally, a further analysis of the state feedback representation of the leader’s optimal strategy gives the leader’s Riccati equation. Under some assumptions, the solvability of the leader’s Riccati equation in the case of deterministic coefficients is also discussed.

In contrast, we consider the follower's Hamiltonian system as the leader's controlled state equation, and hence the state feedback representation of the AOL Stackelberg solution can be obtained simultaneously for the leader and the follower. As a result, the corresponding Riccati equation is different from that in [37]. Moreover, by means of a linear transformation to a standard stochastic Riccati equation, we also prove that under some assumptions there exists a unique solution to the Riccati equation with stochastic coefficients. For an LQ Stackelberg game under the ACLM structure, we will see that the follower's Hamiltonian system is no longer linear, and that prevents us from getting a Riccati equation if we proceed the same way as in the AOL case. Instead, we assume that the forward variable y is linear with respect to the original state x and derive an FBSDE that plays the same role as the Riccati equation in the AOL case.

Throughout this section, we make the following assumptions on the coefficients:

$$(5.1) \quad \begin{aligned} &A, B_i, C, Q_i, R_i, G_i \text{ are adapted bounded matrices,} \\ &Q_i, R_i, G_i \text{ are symmetric and nonnegative,} \\ &\text{and } R_i \text{ is uniformly positive, } i = 1, 2. \end{aligned}$$

5.1. The AOL case. The state equation and the cost functionals are given, separately, as follows:

$$(5.2) \quad dx = (Ax + B_1u + B_2v)dt + CxdW(t), \quad x(0) = x_0,$$

$$(5.3) \quad J_1(u, v) = \frac{1}{2} \mathbb{E} \left\{ \int_0^T (\langle Q_1x, x \rangle + \langle R_1u, u \rangle) dt + \langle G_1x(T), x(T) \rangle \right\},$$

$$(5.4) \quad J_2(u, v) = \frac{1}{2} \mathbb{E} \left\{ \int_0^T (\langle Q_2x, x \rangle + \langle R_2v, v \rangle) dt + \langle G_2x(T), x(T) \rangle \right\}.$$

Given the leader's strategy $u \in \mathcal{U}$, it is well known that the follower's problem, i.e.,

$$\min_{v \in \mathcal{V}} J_2(u, v) = \frac{1}{2} \mathbb{E} \left\{ \int_0^T (\langle Q_2x, x \rangle + \langle R_2v, v \rangle) dt + \langle G_2x(T), x(T) \rangle \right\}$$

subject to

$$dx = (Ax + B_1u + B_2v)dt + CxdW(t), \quad x(0) = x_0,$$

is a standard LQ optimal control problem, and its unique solution is

$$v^*(t, u(\cdot)) = -R_2^{-1}B_2^\top p_2(t),$$

where p_2 is the first part of the solution $(p_2, q_2) \in \mathcal{S}^2 \times \mathcal{M}^2$ to the adjoint equation

$$(5.5) \quad -dp_2 = (A^\top p_2 + C^\top q_2 + Q_2x)dt - q_2dW(t), \quad p_2(T) = G_2x(T).$$

Then, the leader's problem is

$$\min_{u \in \mathcal{U}} J_1(u) = \frac{1}{2} \mathbb{E} \left\{ \int_0^T (\langle Q_1x, x \rangle + \langle R_1u, u \rangle) dt + \langle G_1x(T), x(T) \rangle \right\}$$

subject to (the Hamiltonian system of the follower)

$$(5.6) \quad \begin{cases} dx = (Ax + B_1u - B_2R_2^{-1}B_2^\top p_2)dt + CxdW(t), \\ -dp_2 = (A^\top p_2 + C^\top q_2 + Q_2x)dt - q_2dW(t), \\ x(0) = x_0, p_2(T) = G_2x(T). \end{cases}$$

The leader’s problem is well-posed since for every $u \in \mathcal{U}$, the coefficients of the system (5.6) satisfy the monotonicity condition proposed by Peng and Wu [31]. Moreover, by similar arguments as the proof of Theorem 2.2 in Tang [36], we can get the estimate

$$(5.7) \quad \mathbb{E} \sup_{0 \leq t \leq T} |p_2(t)|^2 + \mathbb{E} \sup_{0 \leq t \leq T} |x(t)|^2 + \mathbb{E} \int_0^T |q_2(t)|^2 dt \leq L \left(|x_0|^2 + \mathbb{E} \int_0^T |u(t)|^2 dt \right),$$

where L is a positive constant. With this estimate, we can adopt the relevant arguments for standard LQ optimal control problems in [22] to show that the leader’s objective functional $J_1(u)$ is convex in u , $\lim_{\|u\| \rightarrow \infty} J_1(u) = \infty$, and $J_1(u)$ is Fréchet differentiable over \mathcal{U} with the representation

$$(5.8) \quad \langle J'_1(u), w \rangle = \mathbb{E} \left\{ \int_0^T (\langle Q_1(t)x(t; x_0, u), x(t; 0, w) \rangle + \langle R_1(t)u(t), w(t) \rangle) dt + \langle G_1x(T; x_0, u), x(T; 0, w) \rangle \right\}.$$

Here we use $x(\cdot; x_0, u)$ to represent the solution of (5.6) with the initial state $x(0) = x_0$ and control $u \in \mathcal{U}$. From Proposition 2.1.2 in [13], we can conclude that the leader has a unique optimal strategy $u^* \in \mathcal{U}$, which satisfies $J'_1(u^*) = 0$. Now we use the dual representation to characterize the optimal strategy u^* .

THEOREM 5.1. *Let assumption (5.1) hold. For each $u \in \mathcal{U}$, there exists a unique solution $(x, y, p_1, q_1, p_2, q_2)$ of the FBSDE*

$$(5.9) \quad \begin{cases} dx = (Ax + B_1u - B_2R_2^{-1}B_2^\top p_2)dt + CxdW(t), \\ -dp_2 = (A^\top p_2 + C^\top q_2 + Q_2x)dt - q_2dW(t), \\ dy = (Ay + B_2R_2^{-1}B_2^\top p_1)dt + CydW(t), \\ -dp_1 = (A^\top p_1 + C^\top q_1 - Q_2y + Q_1x)dt - q_1dW(t), \\ x(0) = x_0, y(0) = 0, p_1(T) = -G_2y(T) + G_1x(T), p_2(T) = G_2x(T). \end{cases}$$

Furthermore, the necessary and sufficient condition for u to be the leader’s optimal strategy is

$$u(t) = -R_1^{-1}B_1p_1(t).$$

Proof. It can be seen that the FBSDEs consisting of (x, p_2, q_2) and (y, p_1, q_1) are two decoupled systems. Therefore, for given $u \in \mathcal{U}$, we can first get the unique solution (x, p_2, q_2) to the FBSDE

$$(5.10) \quad \begin{cases} dx = (Ax + B_1u - B_2R_2^{-1}B_2^\top p_2)dt + CxdW(t), \\ -dp_2 = (A^\top p_2 + C^\top q_2 + Q_2x)dt - q_2dW(t), \\ x(0) = x_0, p_2(T) = G_2x(T). \end{cases}$$

Let $\tilde{y} := -y$. Then the FBSDE consisting of (y, p_1, q_1) in (5.9) can be converted into

$$(5.11) \quad \begin{cases} d\tilde{y} = (A\tilde{y} - B_2R_2^{-1}B_2^\top p_1)dt + C\tilde{y}dW(t), \\ -dp_1 = (A^\top p_1 + C^\top q_1 + Q_2\tilde{y} + Q_1x)dt - q_1dW(t), \\ \tilde{y}(0) = 0, \quad p_1(T) = G_2\tilde{y}(T) + G_1x(T). \end{cases}$$

The coefficients in the above system also satisfy the monotonicity condition in [31]. So there exists a unique solution of (5.11), which also implies the existence and uniqueness of the solution $(x, y, p_1, q_1, p_2, q_2)$ of the FBSDE (5.9). The necessity part comes directly from the maximum principle (3.4) and (3.5). Now we prove the sufficiency part. Denote

$$(x(\cdot; x_0, u), y(\cdot; x_0, u), p_1(\cdot; x_0, u), q_1(\cdot; x_0, u), p_2(\cdot; x_0, u), q_2(\cdot; x_0, u))$$

and

$$(x(\cdot; 0, w), y(\cdot; 0, w), p_1(\cdot; 0, w), q_1(\cdot; 0, w), p_2(\cdot; 0, w), q_2(\cdot; 0, w))$$

as the solutions to the FBSDEs (5.9) with the initial states and controls as (x_0, u) and $(0, w)$, respectively. Using Itô's formula to compute

$$\langle p_1(t; x_0, u), x(t; 0, w) \rangle + \langle p_2(t; 0, w), y(t; x_0, u) \rangle$$

and taking the expectation, we get

$$(5.12) \quad \begin{aligned} \langle J'_1(u), w \rangle &= \mathbb{E} \langle G_1x(T; x_0, u), x(T; 0, w) \rangle \\ &\quad + \mathbb{E} \int_0^T \langle Q_1(t)x(t; x_0, u), x(t; 0, w) \rangle + \langle R_1(t)u(t), w(t) \rangle dt \\ &= \mathbb{E} \int_0^T \langle R_1(t)u(t) + B_1^\top(t)p_1(t; x_0, u), w(t) \rangle dt. \end{aligned}$$

Obviously $u = -R_1^{-1}B_1^\top p_1$ makes $J'_1(u)$ equal to zero, so it is an optimal strategy for the leader. \square

From the uniqueness of the optimal strategy, we also know that the FBSDE

$$(5.13) \quad \begin{cases} dx = (Ax - B_1R_1^{-1}B_1^\top p_1 - B_2R_2^{-1}B_2^\top p_2)dt + CxdW(t), \\ -dp_2 = (A^\top p_2 + C^\top q_2 + Q_2x)dt - q_2dW(t), \\ dy = (Ay + B_2R_2^{-1}B_2^\top p_1)dt + CydW(t), \\ -dp_1 = (A^\top p_1 + C^\top q_1 - Q_2y + Q_1x)dt - q_1dW(t), \\ x(0) = x_0, \quad y(0) = 0, \quad p_1(T) = -G_2y(T) + G_1x(T), \quad p_2(T) = G_2x(T), \end{cases}$$

has a unique solution $(x, y, p_1, q_1, p_2, q_2)$. Moreover, the AOL Stackelberg solution (u^*, v^*) can be written as

$$(5.14) \quad u^* = -R_1^{-1}B_1^\top p_1, \quad v^*(u^*) = -R_2^{-1}B_2^\top p_2.$$

In what follows, we derive the feedback representation of the Stackelberg solution (u^*, v^*) in terms of the state (x, y) . We denote

$$\hat{x} = \begin{pmatrix} x \\ y \end{pmatrix}, \hat{p} = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}, \hat{q} = \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}$$

and

$$\hat{A} = \begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix}, \hat{B} = \begin{pmatrix} B_1 R_1^{-1} B_1^\top & B_2 R_2^{-1} B_2^\top \\ -B_2 R_2^{-1} B_2^\top & 0 \end{pmatrix}, \hat{C} = \begin{pmatrix} C & 0 \\ 0 & C \end{pmatrix},$$

$$\hat{Q} = \begin{pmatrix} Q_1 & -Q_2 \\ Q_2 & 0 \end{pmatrix}, \hat{G} = \begin{pmatrix} G_1 & -G_2 \\ G_2 & 0 \end{pmatrix}.$$

Then the FBSDE (5.13) can be rewritten as

$$(5.15) \quad \begin{cases} d\hat{x} = (\hat{A}\hat{x} - \hat{B}\hat{p})dt + \hat{C}\hat{x}dW(t), \\ d\hat{p} = -(\hat{A}^\top \hat{p} + \hat{C}^\top \hat{q} + \hat{Q}\hat{x})dt + \hat{q}dW(t), \\ \hat{x}(0) = 0, \hat{p}(T) = \hat{G}\hat{x}(T). \end{cases}$$

Suppose there is a matrix-valued process K such that

$$(5.16) \quad \hat{p} = K\hat{x},$$

where K has the stochastic differential form

$$(5.17) \quad dK = Mdt + LdW(t).$$

By applying Itô's formula to $K\hat{x}$, we get

$$(5.18) \quad \begin{aligned} & M\hat{x}dt + L\hat{x}dW(t) + K(\hat{A}\hat{x} - \hat{B}K\hat{x})dt + K\hat{C}\hat{x}dW(t) + L\hat{C}\hat{x}dt \\ &= d\hat{p}(t) \\ &= -(\hat{A}^\top K\hat{x} + \hat{C}^\top \hat{q} + \hat{Q}\hat{x})dt + \hat{q}dW(t). \end{aligned}$$

Comparing the diffusion terms in (5.18), we have

$$(5.19) \quad \hat{q} = L\hat{x} + K\hat{C}\hat{x}.$$

We then substitute this expression into (5.18), compare the drift terms, and obtain

$$(5.20) \quad \begin{aligned} & M\hat{x} + K(\hat{A}\hat{x} - \hat{B}K\hat{x}(t)) + L\hat{C}\hat{x} \\ &= -\hat{A}^\top K\hat{x} - \hat{C}^\top (L\hat{x} + K\hat{C}\hat{x}) - \hat{Q}\hat{x}, \end{aligned}$$

which yields

$$M = -K\hat{A} - \hat{A}^\top K + K\hat{B}K - L\hat{C} - \hat{C}^\top L - \hat{C}^\top K\hat{C} - \hat{Q}.$$

Therefore, we get the Riccati equation

$$(5.21) \quad \begin{cases} dK = -(K\hat{A} + \hat{A}^\top K - K\hat{B}K + L\hat{C} + \hat{C}^\top L + \hat{C}^\top K\hat{C} + \hat{Q})dt + LdW(t), \\ K(T) = \hat{G}. \end{cases}$$

The difference between the Riccati equation (5.21) and the standard one from stochastic LQ problems without control in the diffusion terms (see, e.g., [30]) is that \hat{B} , \hat{Q} , and \hat{G} here are not symmetric matrices. For $n = 1$ and under some appropriate assumptions on the coefficient matrices, we show in the following proposition that the Riccati

equation (5.21) can be connected to a standard one through a linear transformation of FBSDE (5.15).

PROPOSITION 5.2. *Suppose that $n = 1$ and $\frac{Q_2}{Q_1} = \frac{G_2}{G_1}$. Then the Riccati equation (5.21) has a unique solution.*

Proof. We first set

$$\frac{Q_2}{Q_1} = \frac{G_2}{G_1} = \alpha, \quad \frac{B_2 R_2^{-1} B_2^\top}{B_1 R_1^{-1} B_1^\top} = \beta,$$

and make the transformation $\hat{x} = \tilde{x}$, $\hat{p} = \Phi \tilde{p}$, $\hat{q} = \Phi \tilde{q}$ with

$$\Phi = \begin{pmatrix} 1 & -2\beta \\ 2\alpha & 1 \end{pmatrix}.$$

Then the FBSDE (5.15) can be converted into

$$(5.22) \quad \begin{cases} d\tilde{x} = (\tilde{A}\tilde{x} - \tilde{B}\tilde{p})dt + \tilde{C}\tilde{x}dW(t), \\ d\tilde{p} = -(\tilde{A}^\top \tilde{p} + \tilde{C}^\top \tilde{q} + \tilde{Q}\tilde{x})dt + \tilde{q}dW(t), \\ \tilde{x}(0) = 0, \quad \tilde{p}(T) = \tilde{G}\tilde{x}(T), \end{cases}$$

where

$$\tilde{A} = \hat{A}, \quad \tilde{C} = \hat{C},$$

$$\tilde{B} = \begin{pmatrix} B_1 R_1^{-1} B_1^\top + 2\alpha B_2 R_2^{-1} B_2^\top & -B_2 R_2^{-1} B_2^\top \\ -B_2 R_2^{-1} B_2^\top & 2\beta B_2 R_2^{-1} B_2^\top \end{pmatrix},$$

$$\tilde{Q} = \frac{1}{4\alpha\beta} \begin{pmatrix} Q_1 + 2\beta Q_2 & -Q_2 \\ -Q_2 & 2\alpha Q_2 \end{pmatrix},$$

$$\tilde{G} = \begin{pmatrix} G_1 + 2\beta G_2 & -G_2 \\ -G_2 & 2\alpha G_2 \end{pmatrix}.$$

Now the matrices \tilde{B} , \tilde{Q} , and \tilde{G} are symmetric and positive definite. Suppose $\tilde{p} = \tilde{K}\tilde{x}$ and

$$d\tilde{K} = \tilde{K}_1 dt + \tilde{L}dW(t).$$

With the same procedure used to derive the Riccati equation (5.21), we can get the following standard Riccati equation for (\tilde{K}, \tilde{L}) :

$$(5.23) \quad \begin{cases} d\tilde{K} = -(\tilde{K}\tilde{A} + \tilde{A}^\top \tilde{K} - \tilde{K}\tilde{B}\tilde{K} + \tilde{L}\tilde{C} + \tilde{C}^\top \tilde{L} + \tilde{C}^\top \tilde{K}\tilde{C} + \tilde{Q})dt + \tilde{L}dW(t), \\ \tilde{K}(T) = \tilde{G}. \end{cases}$$

According to the results in [7], [30], or the more general case in [36], we know that the Riccati equation (5.23) has a unique solution (\tilde{K}, \tilde{L}) and

$$\tilde{p} = \tilde{K}\tilde{x}, \quad \tilde{q} = (\tilde{L} + \tilde{K}\tilde{C})\tilde{x}.$$

Consequently,

$$(5.24) \quad \begin{aligned} \hat{p} &= \Phi\tilde{p} = \Phi\tilde{K}\tilde{x} = \Phi\tilde{K}\hat{x}, \\ \hat{q} &= \Phi\tilde{q} = \Phi(\tilde{L} + \tilde{K}\tilde{C})\tilde{x} = \Phi(\tilde{L} + \tilde{K}\tilde{C})\hat{x}. \end{aligned}$$

By comparing (5.24) with (5.16) and (5.19), we finally get

$$K = \Phi\tilde{K}, \quad L = \Phi\tilde{L}.$$

We conclude from (5.14) that the AOL Stackelberg solution (u^*, v^*) has a feedback representation in terms of the state (x, y) . \square

5.2. The ACLM case. From the equivalent construction in section 3 we can see that the leader’s strategy u and its derivative $\frac{\partial u}{\partial x}$ are quite independent. If there is no restriction on the derivative $\frac{\partial u}{\partial x}$ (equivalently, no restriction on the control u_2 in the standard problem (4.7)–(4.8)), the Hamiltonian H (see (5.30)) for the standard problem could go to $-\infty$. One way to avoid this issue is to add a penalty term on $\frac{\partial u}{\partial x}$ in the leader’s cost functional (see [25]) so that H is convex with respect to $(u, \frac{\partial u}{\partial x})$. Another way is to impose a priori bounds on $\frac{\partial u}{\partial x}$ to keep H finite. In this subsection, we adopt the latter approach by assuming the derivative $\frac{\partial u}{\partial x}$ to be bounded. Since the derivative appears as a part of the coefficient in the adjoint equation (5.25), its boundedness implies the well-posedness of the leader’s problem when affine strategies are adopted. For simplicity, we consider a one-dimensional LQ game, with the state equation and the cost functionals of the two players, separately, as follows:

$$dx = [Ax + B_1u + B_2v]dt + CxdW(t), \quad x(0) = x_0$$

and

$$\begin{aligned} J_1 &= \frac{1}{2}\mathbb{E}\left\{ \int_0^T (Q_1x^2 + R_1u^2)dt + G_1x^2(T) \right\}, \\ J_2 &= \frac{1}{2}\mathbb{E}\left\{ \int_0^T (Q_2x^2 + R_2v^2)dt + G_2x^2(T) \right\}. \end{aligned}$$

The admissible strategy spaces from which the leader and the follower choose their strategies are given, separately, by

$$\mathcal{U} := \left\{ u|u : \Omega \times [0, T] \times \mathbb{R} \rightarrow \mathbb{R} \text{ is } \mathcal{F}_t\text{-adapted for any } x \in \mathbb{R}, u(t, x) \text{ is continuously differentiable in } x \text{ for any } (\omega, t) \in \Omega \times [0, T], \text{ and the derivative } \left| \frac{\partial u}{\partial x} \right| \leq K \text{ for some postive constant } K \right\},$$

$$\mathcal{V} := \{v|v : \Omega \times [0, T] \times \mathbb{R} \times \mathcal{U} \rightarrow \mathbb{R} \text{ is } \mathcal{F}_t\text{-adapted for any } x \in \mathbb{R} \text{ and } u \in \mathcal{U}\}.$$

Suppose for every leader’s strategy $u \in \mathcal{U}$, the follower has a unique optimal response $v^* \in \mathcal{V}$. From (4.3), we know that

$$v^*(u) = -R_2^{-1}B_2p_2,$$

where p_2 satisfies

$$(5.25) \quad dp_2 = - \left[\left(A + B_1 \frac{\partial u}{\partial x} \right) p_2 + Cq_2 + Q_2x \right] dt + q_2 dW(t), \quad p_2(T) = G_2x(T).$$

Therefore, the leader’s problem is

$$(5.26) \quad \min_{u \in \mathcal{U}} J_1 = \frac{1}{2} \mathbb{E} \left\{ \int_0^T (Q_1x^2 + R_1u^2) dt + G_1x^2(T) \right\}$$

subject to

$$(5.27) \quad \begin{cases} dx = [Ax + B_1u(t, x) - R_2^{-1}B_2^2p_2] dt + CxdW(t), \\ dp_2 = - \left[\left(A + B_1 \frac{\partial u}{\partial x} \right) p_2 + Cq_2 + Q_2x \right] dt + q_2 dW(t), \\ x(0) = x_0, \quad p_2(T) = G_2x(T). \end{cases}$$

Suppose the leader’s problem is well-posed. According to the discussions in the proof of Theorem 4.1, we know that the leader will lose nothing if she chooses her strategy among affine functions

$$u(t, x) = u_2(t)x + u_1(t),$$

with u_1 and u_2 being adapted processes and $|u_2| \leq K$. Then, the leader’s equivalent problem can be written as follows:

$$(5.28) \quad \min_{u_1, u_2} J_1 = \frac{1}{2} \mathbb{E} \left\{ \int_0^T [Q_1x^2 + R_1(u_2x + u_1)^2] dt + G_1x^2(T) \right\}$$

subject to

$$(5.29) \quad \begin{cases} dx = [(A + B_1u_2)x + B_1u_1 - R_2^{-1}B_2^2p_2] dt + CxdW(t), \\ dp_2 = -[(A + B_1u_2)p_2 + Cq_2 + Q_2x] dt + q_2 dW(t), \\ x(0) = x_0, \quad p_2(T) = G_2x(T). \end{cases}$$

For each pair (u_1, u_2) , the monotonicity condition guarantees the existence and uniqueness of the solution of (5.29). Therefore, the leader’s problem with strategies restricted to be of affine form is well-posed. In what follows, we use the maximum principle to get the Hamiltonian system and the related Riccati equation for the leader’s problem (5.28)–(5.29). Denote

$$(5.30) \quad \begin{aligned} & H_1(t, u_1, u_2, x, y, p_1, p_2, q_1, q_2) \\ &= p_1[(A + B_1u_2)x + B_1u_1 - R_2^{-1}B_2^2p_2] + Cxq_1 \\ &\quad - y[(A + B_1u_2)p_2 + Cq_2 + Q_2x] + \frac{1}{2}[Q_1x^2 + R_1(u_2x + u_1)^2]. \end{aligned}$$

To obtain (u_1^*, u_2^*) that minimizes $H_1(t, u_1, u_2, x, y, p_1, p_2, q_1, q_2)$, we first fix u_2 and minimize H_1 with respect to u_1 . This procedure yields

$$(5.31) \quad u_1^* = -u_2x - R_1^{-1}B_1p_1.$$

By substituting (5.31) into the expression (5.30) of H , we can see that the only term containing u_2 is $-B_1yp_2u_2$. Therefore, the optimal u_2^* is given as

$$(5.32) \quad u_2^* = \text{bang}(K, -K; \Delta) := \begin{cases} -K & \text{if } \Delta > 0, \\ K & \text{if } \Delta < 0, \\ \text{arbitrary} & \text{if } \Delta = 0, \end{cases}$$

where $\Delta := -B_1yp_2$. To find a candidate for the optimal pair (u_1^*, u_2^*) , we set

$$u_2^* = \text{bang}(K, -K; \Delta) = \text{sgn}(B_1yp_2)K = \text{sgn}(y)\text{sgn}(B_1p_2)K = \text{sgn}(p_2)\text{sgn}(B_1y)K,$$

where the sgn function is defined as

$$\text{sgn}(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -1 & \text{if } x < 0. \end{cases}$$

From (5.31), we get

$$(5.33) \quad u_1^* = -\text{bang}(K, -K; \Delta)x - R_1^{-1}B_1p_1.$$

If (u_1^*, u_2^*) is a solution of the leader’s problem (5.28)–(5.29), then by the maximum principle (Theorem 4.1), there exist adapted processes y , p_1 , and q_1 such that

$$(5.34) \quad \begin{cases} dx = [(A + B_1u_2^*)x + B_1u_1^* - R_2^{-1}B_2^2p_2]dt + CxdW(t), \\ dy = [(A + B_1u_2^*)y + R_2^{-1}B_2^2p_1]dt + CydW(t), \\ dp_1 = -[(A + B_1u_2^*)p_1 + Cq_1 - Q_2y + Q_1x + R_1u_2^*(u_2^*x + u_1^*)]dt + q_1dW(t), \\ dp_2 = -[(A + B_1u_2^*)p_2 + Cq_2 + Q_2x]dt + q_2dW(t), \\ x(0) = x_0, y(0) = 0, p_1(T) = -G_2y(T) + G_1x(T), p_2(T) = G_2x(T), \\ u_1^* = -\text{bang}(K, -K; \Delta)x - R_1^{-1}B_1p_1, u_2^* = \text{bang}(K, -K; \Delta). \end{cases}$$

Like in the AOL case, we proceed by expressing the optimal strategy (u_1^*, u_2^*) in a nonanticipating way by means of the state feedback representation. By substituting the expressions of u_1^* and u_2^* into the FBSDE (5.34), we get

$$(5.35) \quad \begin{cases} dx = [Ax - R_1^{-1}B_1^2p_1 - R_2^{-1}B_2^2p_2]dt + CxdW(t), \\ dy = [(A + B_1\text{bang}(K, -K; \Delta))y + R_2^{-1}B_2^2p_1]dt + CydW(t) \\ \quad = [Ay + \text{sgn}(p_2)K|B_1y| + R_2^{-1}B_2^2p_1]dt + CydW(t), \\ dp_1 = -[Ap_1 + Cq_1 - Q_2y + Q_1x]dt + q_1dW(t), \\ dp_2 = -[(A + B_1\text{bang}(K, -K; \Delta))p_2 + Cq_2 + Q_2x]dt + q_2dW(t) \\ \quad = -[Ap_2 + \text{sgn}(y)K|B_1p_2| + Cq_2 + Q_2x]dt + q_2dW(t), \\ x(0) = x_0, y(0) = 0, p_1(T) = -G_2y(T) + G_1x(T), p_2(T) = G_2x(T). \end{cases}$$

In contrast to the FBSDE (5.13) in the AOL case, the presence of the additional nonlinear term $\text{bang}(K, -K; \Delta)$ in the FBSDE (5.35) makes it a nonlinear system. Moreover, the Lipschitz continuity assumption usually made for the coefficients in the literature does not hold here. Therefore, the existence and uniqueness of the solution to (5.35), as far as we know, is still not available. On the other hand, if we still view

(x, y) as the “state” and represent (p_1, p_2) in terms of (x, y) as in the AOL case, we are not able to derive a Riccati equation which is independent of the state. Instead, we only see x as the state and suppose

$$(5.36) \quad y(t) = \xi(t)x(t), \quad p_1(t) = \eta(t)x(t), \quad p_2(t) = \zeta(t)x(t)$$

and

$$(5.37) \quad \begin{aligned} d\xi(t) &= \xi_1(t)dt + \xi_2(t)dW(t), \\ d\eta(t) &= \eta_1(t)dt + \eta_2(t)dW(t), \\ d\zeta(t) &= \zeta_1(t)dt + \zeta_2(t)dW(t). \end{aligned}$$

By Itô’s formula and (5.36), we obtain

$$(5.38) \quad \begin{aligned} dy &= \xi dx + x d\xi + Cx\xi_2 dt \\ &= \xi[Ax - R_1^{-1}B_1^2 p_1 - R_2^{-1}B_2^2 p_2]dt + C\xi x dW(t) \\ &\quad + \xi_1 x dt + \xi_2 x dW(t) + C\xi_2 x dt \\ &= \{[A - R_1^{-1}B_1^2 \eta(t) - R_2^{-1}B_2^2 \zeta] \xi \\ &\quad + \xi_1 + C\xi_2\} x dt + [C\xi + \xi_2] x dW(t). \end{aligned}$$

On the other hand,

$$(5.39) \quad \begin{aligned} dy &= [(A + B_1 \text{bang}(K, -K; \Delta))y + R_2^{-1}B_2^2 p_1]dt + Cy dW(t) \\ &= [(A + B_1 \text{bang}(K, -K; \tilde{\Delta}))\xi + R_2^{-1}B_2^2 \eta]x dt + C\xi x dW(t), \end{aligned}$$

where

$$\tilde{\Delta} := -B_1 \xi(t) \zeta(t).$$

By comparing (5.38) and (5.39), we have

$$\begin{aligned} \xi_2 &= 0, \\ \xi_1 &= [R_1^{-1}B_1^2 \eta + R_2^{-1}B_2^2 \zeta + B_1 \text{bang}(K, -K; \tilde{\Delta})] \xi + R_2^{-1}B_2^2 \eta. \end{aligned}$$

By applying Itô’s formula to p_1 and p_2 , and proceeding in the same way as above, we can get

$$\begin{cases} \eta_1 = [R_1^{-1}B_1^2 \eta + R_2^{-1}B_2^2 \zeta - 2A - C^2] \eta + Q_2 \xi - 2C \eta_2 - Q_1, \\ \zeta_1 = [R_1^{-1}B_1^2 \eta + R_2^{-1}B_2^2 \zeta - 2A - C^2 - B_1 \text{bang}(K, -K; \tilde{\Delta})] \zeta - 2C \zeta_2 - Q_2. \end{cases}$$

Therefore, we derive the related Riccati equation for problem (5.28)–(5.29):

$$\left\{ \begin{aligned} d\xi &= \{[R_1^{-1}B_1^2 \eta + R_2^{-1}B_2^2 \zeta + B_1 \text{bang}(K, -K; \tilde{\Delta})] \xi + R_2^{-1}B_2^2 \eta\} dt \\ &\quad = \{[R_1^{-1}B_1^2 \eta + R_2^{-1}B_2^2 \zeta] \xi + \text{sgn}(\zeta) |B_1 \xi| + R_2^{-1}B_2^2 \eta\} dt, \\ d\eta &= \{[R_1^{-1}B_1^2 \eta + R_2^{-1}B_2^2 \zeta - 2A - C^2] \eta + Q_2 \xi - 2C \eta_2 \\ &\quad - Q_1\} dt + \eta_2 dW(t), \\ d\zeta &= \{[R_1^{-1}B_1^2 \eta + R_2^{-1}B_2^2 \zeta - 2A - C^2 - B_1 \text{bang}(K, -K; \tilde{\Delta})] \zeta \\ &\quad - 2C \zeta_2 - Q_2\} dt + \zeta_2 dW(t) \\ &= \{[R_1^{-1}B_1^2 \eta + R_2^{-1}B_2^2 \zeta - 2A - C^2] \zeta - \text{sgn}(\xi) |B_1 \zeta| \\ &\quad - 2C \zeta_2 - Q_2\} dt + \zeta_2 dW(t), \\ \xi(0) &= 0, \quad \eta(T) = -G_2 \xi(T) + G_1, \quad \zeta(T) = G_2. \end{aligned} \right.$$

Suppose $(\xi, \eta, \zeta, \eta_2, \zeta_2)$ is a solution of the above FBSDE and x^* solves the linear SDE

$$dx = [A - R_1^{-1}B_1^2\eta - R_2^{-1}B_2^2\zeta]xdt + CxdW(t), \quad x(0) = x_0.$$

Then, we can use Itô's formula to verify that

$$\begin{aligned} y(t) &:= \xi(t)x^*(t), \quad p_1(t) := \eta(t)x^*(t), \quad p_2(t) := \zeta(t)x^*(t), \\ q_1(t) &:= [C\eta(t) + \eta_2(t)]x^*(t), \quad q_2(t) := [C\zeta(t) + \zeta_2(t)]x^*(t), \end{aligned}$$

together with x^* , solve the leader's Hamiltonian system (5.35). Therefore,

$$u(t, x) = \text{bang}(K, -K; \tilde{\Delta})x - \text{bang}(K, -K; \tilde{\Delta})x^*(t) - R_1^{-1}B_1\eta(t)x^*(t),$$

where $\tilde{\Delta} = -B_1\xi(t)\zeta(t)$, is a candidate for the leader's optimal strategy.

6. Conclusion. In this paper we study the solutions of stochastic Stackelberg differential games under two information structures: adapted open-loop and adapted closed-loop memoryless patterns. Based on the maximum principle for the former kind of game, we derive the necessary conditions for the leader's optimal strategy in the latter game. For a comparison, we further show the differences by applying the results to the special case of LQ stochastic Stackelberg differential games and derive the related Riccati equations for the two cases. Under appropriate conditions, we prove the existence and uniqueness of the solution to the Riccati equation in the adapted open-loop case and leave the issue of the Riccati equation in the adapted closed-loop memoryless case as a topic for future research.

As for applications, Stackelberg differential games under the feedback and adapted feedback information structures have been employed in treating supply chain management, marketing channel management, and economics problems in [9], [10], [12], [15], [16], and [17]. One can also read [18] for a recent application of the feedback Stackelberg game in a discrete-time supply chain setting. It would be interesting to characterize global Stackelberg solutions for some of these problems and compare them with the associated feedback Stackelberg solutions.

Another possible research direction is to extend the various solution concepts to Stackelberg differential games where the follower has private information and/or hidden action, as is the case of the so-called principle-agent or contract design problems in the economics literature. For some recent work on the contract theory with moral hazard in the continuous-time framework, see Cvitanić and Zhang [11] and the references therein.

Acknowledgments. The authors are grateful for the comments and suggestions from Kateřina Staňková, the associate editor, and the referees.

REFERENCES

[1] T. BAŞAR, *A new approach for derivation of closed-loop Stackelberg strategies*, in Proceedings of the IEEE 17th Conference on Decision and Control, IEEE Computer Society Press, Los Alamitos, CA, 1979, pp. 1113–1118.
 [2] T. BAŞAR AND A. HAURIE, *Feedback equilibria in differential games with structural and modal uncertainties*, in Advances in Large Scale Systems, J. B. Cruz, Jr., ed., JAI Press, Greenwich, CT, 1984, pp. 163–201.
 [3] T. BAŞAR AND G. J. OLSDER, *Team-optimal closed-loop Stackelberg strategies in hierarchical control problems*, Automatica, 16 (1980), pp. 409–414.
 [4] T. BAŞAR AND G. J. OLSDER, *Dynamic Noncooperative Game Theory*, Classics Appl. Math., SIAM, Philadelphia, 1999.

- [5] T. BAŞAR AND H. SELBUZ, *Closed-loop Stackelberg strategies with applications in the optimal control of multilevel systems*, IEEE Trans. Automat. Control, 24 (1979), pp. 166–179.
- [6] A. BENSOUSSAN, S. CHEN, AND S. P. SETHI, *Feedback Stackelberg solutions of infinite-horizon stochastic differential games*, in Models and Methods in Economics and Management Science, Internat. Ser. Oper. Res. Management Sci. 198, F. El Ouardighi and K. Kogan, eds., Springer, New York, 2014, pp. 3–15.
- [7] J. M. BISMUT, *Contrôle des systèmes linéaires quadratiques: applications de l'intégrale stochastique*, in Séminaire de Probabilités XII, C. Dellacherie, P. A. Meyer, and M. Weil, eds., Lecture Notes in Math. 649, Springer-Verlag, Berlin, 1978, pp. 180–264.
- [8] T. BJÖRK, *Arbitrage Theory in Continuous Time*, Oxford University Press, New York, 1998.
- [9] A. CHUTANI AND S. P. SETHI, *Optimal advertising and pricing in a dynamic durable goods supply chain*, J. Optim. Theory Appl., 154 (2012), pp. 615–643.
- [10] A. CHUTANI AND S. P. SETHI, *Cooperative advertising in a dynamic retail market oligopoly*, Dyn. Games Appl., 2 (2012), pp. 347–375.
- [11] J. CVITANIĆ AND J. ZHANG, *Contract Theory in Continuous Time Models*, Springer, Berlin, 2012.
- [12] E. DOCKNER, S. JØGENSEN, N. V. LONG, AND G. SORGER, *Differential Games in Economics and Management Science*, Cambridge University Press, Cambridge, UK, 2000.
- [13] I. EKELAND AND R. TÉMAM, *Convex Analysis and Variational Problems*, North-Holland, Amsterdam, 1976.
- [14] N. EL KAROUI, S. PENG, AND M. C. QUENEZ, *Backward stochastic differential equations in finance*, Math. Finance, 7 (1997), pp. 1–71.
- [15] X. HE, A. PRASAD, AND S. P. SETHI, *Cooperative advertising and pricing in a dynamic stochastic supply chain: Feedback Stackelberg strategies*, Production Operations Management, 18 (2009), pp. 78–94.
- [16] X. HE, A. PRASAD, S. P. SETHI, AND G. J. GUTIERREZ, *A survey of Stackelberg differential game models in supply and marketing channels*, J. Systems Sci. Systems Eng., 16 (2007), pp. 385–413.
- [17] K. KOGAN AND C. S. TAPIERO, *Supply Chain Games: Operations Management and Risk Valuation*, Springer, New York, 2007.
- [18] T. LI, S. P. SETHI, AND X. HE, *Dynamic Pricing, Production, and Channel Coordination with Stochastic Learning*, Prod. Oper. Manag., 24 (2015), pp. 857–882.
- [19] N. V. LONG, *A Survey of Dynamic Games in Economics*, World Scientific, Singapore, 2010.
- [20] J. MA, P. PROTTER, AND J. YONG, *Solving forward-backward stochastic differential equations explicitly—a four step scheme*, Probab. Theory Related Fields, 98 (1994), pp. 339–359.
- [21] J. MA AND J. YONG, *Forward-Backward Stochastic Differential Equations and Their Applications*, Lecture Notes in Math. 1702, Springer-Verlag, Berlin, 1999.
- [22] Q. MENG, *General Linear Quadratic Optimal Stochastic Control Problem Driven by a Brownian Motion and a Poisson Random Martingale Measure with Random Coefficients*, arXiv:1102.3295v2, 2011.
- [23] B. ØKSENDAL, L. SANDAL, AND J. UBOE, *Stochastic Stackelberg equilibria with applications to time dependent newsvendor models*, J. Econom. Dynam. Control, 37 (2013), pp. 1284–1299.
- [24] G. P. PAPAVALLOPOULOS, *Leader-Follower and Nash Strategies with State Information*, Ph.D. thesis, University of Illinois at Urbana-Champaign, 1979.
- [25] G. P. PAPAVALLOPOULOS AND J. B. CRUZ, JR., *Nonclassical control problems and Stackelberg games*, IEEE Trans. Automat. Control, 24 (1979), pp. 155–166.
- [26] G. P. PAPAVALLOPOULOS AND J. B. CRUZ, JR., *Sufficient conditions for Stackelberg and Nash strategies with memory*, in Proceedings of the Conference on Systems Engineering for Power: Organizational Forms for Large Scale Systems, Vol. II, U.S. Department of Energy, Washington, DC, 1979, pp. 2.59–2.69.
- [27] G. P. PAPAVALLOPOULOS AND J. B. CRUZ, JR., *Sufficient conditions for Stackelberg and Nash strategies with memory*, J. Optim. Theory Appl., 31 (1980), pp. 233–260.
- [28] E. PARDOUX AND S. PENG, *Adapted solution of a backward stochastic differential equation*, Systems Control Lett., 14 (1990), pp. 55–61.
- [29] E. PARDOUX AND S. TANG, *Forward-backward stochastic differential equations and quasilinear parabolic PDEs*, Probab. Theory Related Fields, 114 (1999), pp. 123–150.
- [30] S. PENG, *Stochastic Hamilton-Jacobi-Bellman equations*, SIAM J. Control Optim., 30 (1992), pp. 284–304.
- [31] S. PENG AND Z. WU, *Fully coupled forward-backward stochastic differential equations and applications to optimal control*, SIAM J. Control Optim., 37 (1999), pp. 825–843.

- [32] M. L. PETIT, *Control Theory and Dynamic Games in Economic Policy Analysis*, Cambridge University Press, Cambridge, UK, 2009.
- [33] J. SHI AND Z. WU, *The maximum principle for fully coupled forward-backward stochastic control system*, *Acta Automat. Sinica*, 32 (2006), pp. 161–169.
- [34] M. SIMAAN AND J. B. CRUZ, JR., *Additional aspects of the Stackelberg strategy in nonzero-sum games*, *J. Optim. Theory Appl.*, 11 (1973), pp. 613–626.
- [35] H. VON STACKELBERG, *Marktform und Gleichgewicht*, Springer-Verlag, Wien New York, 1934 (in German); *Market Structure and Equilibrium*, Springer-Verlag, Berlin Heidelberg, 2011 (in English).
- [36] S. TANG, *General linear quadratic optimal stochastic control problems with random coefficients: Linear stochastic Hamilton systems and backward stochastic Riccati equations*, *SIAM J. Control Optim.*, 42 (2003), pp. 53–75.
- [37] J. YONG, *A leader-follower stochastic linear quadratic differential game*, *SIAM J. Control Optim.*, 41 (2002), pp. 1015–1041.
- [38] J. YONG, *Forward-backward stochastic differential equations with mixed initial-terminal conditions*, *Trans. Amer. Math. Soc.*, 362 (2010), pp. 1047–1096.
- [39] J. YONG, *Optimality variational principle for controlled forward-backward stochastic differential equations with mixed initial-terminal conditions*, *SIAM J. Control Optim.*, 48 (2010), pp. 4119–4156.
- [40] J. YONG AND X. ZHOU, *Stochastic Controls: Hamiltonian Systems and HJB Equations*, Springer, New York, 1999.



TREASURES
@UT Dallas

Naveen Jindal School of Management

The Maximum Principle for Global Solutions of Stochastic Stackelberg Differential Games

©2015 Society for Industrial and Applied Mathematics

Citation:

Bensoussan, Alain, Shaokuan Chen, and Suresh P. Sethi. 2015. "The Maximum Principle for Global Solutions of Stochastic Stackelberg Differential Games." *SIAM Journal on Control and Optimization* 53(4), doi:10.1137/140958906.

This document is being made freely available by the Eugene McDermott Library of The University of Texas at Dallas with permission from the copyright owner. All rights are reserved under United States copyright law unless specified otherwise.