Linear Full-information Feedback Approach to Closed-loop Stackelberg Strategy

Hongdan Li  
College of Electrical Engineering and Automation  
Shandong University of Science and Technology  
Qingdao, China  
lhd200908@163.com

Juanjuan Xu  
School of Control Science and Engineering  
Shandong University  
Jinan, China  
jnxujuanjuan@163.com

Huanshui Zhang  
College of Electrical Engineering and Automation  
Shandong University of Science and Technology  
Qingdao, China  
hszhang@sdu.edu.cn

Abstract—Stackelberg strategy has received great attention since the 1970s. The closed-loop solution still faces many difficulties though the open-loop problems have been well studied. An obvious fact for the difficulty is that the optimal closed-loop solution to the Stackelberg strategy is implicit nonlinear. By presenting a novel approach of full-information linear feedback strategy, substantial progress for the problem is made in this paper. We are the first to present explicit linear closed-loop Stackelberg controllers in terms of the current state and Riccati-like equations. Due to the elegance and simplicity of the solution, the proposed results will be extended to other game control problems with asymmetry status players. More importantly, the Q-learning algorithm for the Stackelberg strategy will become possible with the proposed optimal solution.

Index Terms—closed-loop Stackelberg strategy, full-information linear feedback, constrained maximum principle

I. INTRODUCTION

Stackelberg solution is of great significance in plenty of fields including economics, engineering, and biology [1], [9], [18], [19]. Specific examples are referred to government regulation of the macro economy [27] and transboundary pollution [14]. The salient feature of such problems is that at least two players are involved in a nonzero-sum game, well-known as Stackelberg game or leader-follower game, and one player named by the leader can enforce his strategy on the other player named by the follower. The asymmetry of players’ roles makes the derivation of Stackelberg solution extremely complex, which is completely different from the symmetric positions among players in the Nash game [17].

The dynamic Stackelberg solution was initially studied in the 1970s [8] and has achieved many fruitful results since then. For example, an implicit necessary and sufficient condition was given in [21] for the unique existence of the solution based on the operator technique. [10] proposed a sufficient existence condition by adopting a Lyapunov-type approach. In particular, the explicit Stackelberg solution is given for the LQ game in terms of three coupled and nonsymmetric Riccati equations. To avoid the solvability of the above nonstandard Riccati equations, [23] introduced three decoupled and symmetric Riccati equations to design the Stackelberg solution, which was later extended to solve the Stackelberg game with time delay [24]. The stochastic LQ leader-follower game with multiplicative noise was solved in [26]. It is noted that the Stackelberg solutions in the above-mentioned work are open-loop, that is, the information sets available to the players contain only initial value. This makes it easy to obtain the equilibrium conditions based on the maximum principle. However, it is not reasonable to assume that the players have access to only open-loop information in a dynamic game as has been pointed out in [2].

Another widely studied strategy is feedback strategy, i.e., in the feedback form of the state at the current time. Once the feedback structure of the Stackelberg strategy is known, the optimal feedback gain can be obtained by applying the matrix maximum principle or dynamic programming. For instance, [11] derived feedback Stackelberg strategies for the LQ two-player game by using dynamic programming. A multilevel feedback Stackelberg strategy was then studied in [12] for systems with $M$ players and sufficient conditions. The results were extended to more general stochastic case in [13], [15].

Compared with the open-loop and feedback strategies, the closed-loop strategy, whose information sets available to the players contain all the states from the initial time to the current time, is much more advantageous because the leader’s motivation for using the Stackelberg strategy is to reduce losses [16]. In this case, a pioneer work in [2] provided a clean characterization of the set of optimal closed-loop policies for the leader with an indirect method. And then [5] derived the team-optimal closed-loop Stackelberg strategies for a class of continuous-time two-player nonzero-sum differential games. [22] studied the team-optimal closed-loop Stackelberg strategies for systems with slow and fast modes.

On the other hand, a direct method for closed-loop Stackelberg strategy based on the maximum principle has been
investigated [4], [20]. However, the derivation of the closed-loop strategy by using the direct approach faces major challenges even in the linear quadratic game because the standard techniques such as dynamic programming are invalid in this case. The main reason lies in that the follower’s reaction can only be expressed implicitly for every announced strategy of the leader which further leads to a nonstandard optimization problem for the leader containing implicit constraints [3]. To overcome this difficulty, one way is to study the closed-loop memoryless Stackelberg strategy, that is, the information sets available to the players are restricted to consisting of only two parts: the state at the current time and the initial state. In this case, the leader faces a nonclassical optimization problem where the controlled system contains the derivative of the control for the state. And the derived controller is related to the backward state (co-state) which is an implicit solution and difficult to be calculated in practice [4], [20].

In this paper, we will further study the closed-loop Stackelberg strategy for an LQ leader-follower game by a new direct approach of linear full-information feedback. In particular, the strategy is in the linear closed-loop form with one-step memory. By applying the constrained maximum principle, the solvability of the leader-follower game is reduced to that of forward and backward difference equations. The main contribution is to derive the explicitly linear closed-loop Stackelberg strategy with one-step memory in terms of Riccati equations based on the model matrix parameters, which is very elegant like the standard LQ controller. The simple form will allow us to make an extension to other related problems and Q-learning algorithms in future work. A numerical example shows that the derived closed-loop strategy achieves better performance than the feedback strategy.

The remainder of the paper is organized as follows. Section 2 presents the studied problem. The main result is given in Section 3. A numerical example is given in Section 4. Some concluding remarks are given in the last section.

The following notations will be used throughout this paper: \( \mathbb{R}^n \) denotes the set of \( n \)-dimensional vectors; \( \forall \) denotes the transpose of \( x \); a symmetric matrix \( M > 0 \) \((\geq 0)\) means that \( M \) is strictly positive-definite (positive-semidefinite).

II. PROBLEM STATEMENT

Consider the following linear system

\[
x_{k+1} = Ax_k + Bu_k + B_2u_k,
\]

where \( x_k \in \mathbb{R}^n \) is a state and \( x_{-1} = 0 \). One of the players \( u_k^1 \in \mathbb{R}^{m_1} \) is as a follower, the other \( u_k^2 \in \mathbb{R}^{m_2} \) is a leader. \( A, B_1, B_2 \) are real matrices with compatible dimensions. Define the following two cost functionals

\[
J^f(N, s) = \sum_{k=s}^{N} [x_k'Qx_k + (u_k^1)'R_1u_k^1 + (u_k^2)'R_2u_k^2] + x_{N+1}'P_{N+1}x_{N+1},
\]

(1)

\[
J(N, s) = \sum_{k=s}^{N} [x_k'Qx_k + (u_k^1)'R_1u_k^1 + (u_k^2)'R_2u_k^2] + x_{N+1}'P_{N+1}x_{N+1},
\]

(2)

where the state weighting matrices \( Q, \bar{Q}, P_{N+1} \) and \( \bar{P}_{N+1} \) are semi-positive definite. The control weighting matrices \( R_i, i = 1, 2 \) are positive definite. In the following, we assume \( s = 0, i.e., the initial time s is 0.\)

According to the information structure used by the controller, it contains four pattern as following [6]:

1) an open-loop pattern if \( \eta_k = \{ x_0 \}; \)
2) a feedback information pattern if \( \eta_k = \{ x_k \}; \)
3) a closed-loop information with memoryless if \( \eta_k = \{ x_0, x_k \}; \)
4) a closed-loop information with memory if \( \eta_k = \{ x_0, x_1, \ldots, x_k \}. \)

In this paper, we are concerned with the closed-loop with memory, i.e., the forth case pattern. For the convenience of discussions, we consider the case of closed-loop with one-step memory, \( \eta_k = \{ x_{k-1}, x_k \}. \) It is noted that general case of information structure pattern \( \eta_k = \{ x_0, x_1, \ldots, x_k \} \) can be done by a similar discussion. Now we define the controller by linear information feedback as following:

\[
u_k^1 = K_kx_k + G_kx_{k-1},
\]

(3)

\[
u_k^2 = K_kx_k + G_kx_{k-1},
\]

(4)

where parameter matrices \( K_k, G_k \in \mathbb{R}^{m_1 \times n}, \bar{K}_k, \bar{G}_k \in \mathbb{R}^{m_2 \times n} \) with \( G_0 = G_0 = 0 \) are to be determined. The problem to be discussed in this paper is as follows.

Problem 1. Find the leader’s controller \( u^f \) as defined in (5) to minimize (3) under the condition that the follower constructs a controller \( u^l \) as defined in (4) minimizing (2) given \( u^f \).

Remark 1. This problem is called a linear closed-loop Stackelberg game control. It is different from the well-known state feedback strategy where the controller is assumed to be of current state feedback, i.e., \( u_k = K_kx_k \). The problem is also completely different from the open-loop problem addressed in the past decades where only initial state value is available.

III. MAIN RESULT

To solve Problem 1, we first review the following standard LQR with a constraint of current state-feedback. Consider the following linear system

\[
x_{k+1} = Ax_k + Bu_k,
\]

(6)

and define the cost functional as follows.

\[
J(N, 0) = \sum_{k=0}^{N} [x_k'Qx_k + u_k'R_ku_k] + x_{N+1}'P_{N+1}x_{N+1},
\]

(7)

where \( Q, R \) are semi-positive definite. Different from the standard LQR problem, we assume that the controller is with the following form

\[
u_k = K_kx_k.
\]

(8)

The constraint LQR problem is to find the controller with form of (8) to minimize (7). Now, we revisit the following result:

Lemma 1. Suppose the above constrained LQR problem admits a solution, then, the controller satisfies

\[
0 = Ru_k + B^T\lambda_k - \beta_k,
\]

(9)

where \( \lambda_k \) is as

\[
\begin{cases}
\lambda_{k-1} = Qx_k + A'\lambda_k + K_k'\beta_k, & 0 < k \leq N, \\
\lambda_N = P_{N+1}x_{N+1},
\end{cases}
\]

and \( \beta_k \) is arbitrary to be determined.
Remark 3. It is clear that the constraint maximum principle (MP) presented in Lemma 1 is different from the standard MP.

Next, we derive the controller using Lemma 1. Firstly, we consider $k = N$. From (9) with $k = N$, we have that

$$0 = R_{N} + B'P_{N+1}(Ax_N + B_{N}) - \beta_N = (R + B'P_{N+1}B)u_N + B'P_{N+1}Ax_N - \beta_N.$$  

Suppose that $R + B'P_{N+1}B$ is positive definite, it yields

$$u_N = -(R + B'P_{N+1}B)^{-1}B'P_{N+1}Ax_N + (R + B'P_{N+1}B)^{-1}\beta_N.$$  

Consider the linear form $u_N = K_N x_N$, combining (11), we get

$$[K_N + (R + B'P_{N+1}B)^{-1}B'P_{N+1}A]x_N = -(R + B'P_{N+1}B)^{-1}\beta_N = 0.$$  

Due to the fact that the above relationship holds for any $x_N$, it deduces that

$$K_N = -(R + B'P_{N+1}B)^{-1}B'P_{N+1}A, \quad \beta_N = 0,$$  

i.e.,

$$u_N = -(R + B'P_{N+1}B)^{-1}B'P_{N+1}Ax_N.$$  

Based on this, from (9), we have

$$\lambda_{N-1} = (Q + A'P_{N+1}A - A'P_{N+1}B(R + B'P_{N+1}B)^{-1}B'P_{N+1}A)x_N.$$  

Following the line of $k = N$, by mathematical induction, we can deduce that the optimal solution is

$$u_k = -(R + B'P_{k+1}B)^{-1}B'P_{k+1}Ax_k.$$  

where $P_k = Q + A'P_{k+1}A - A'P_{k+1}B(R + B'P_{k+1}B)^{-1}B'P_{k+1}A$ with terminal value $P_{N+1}$. Moreover, the relationship between state and co-state is $\lambda_{k-1} = P_k x_k$.

Remark 2. To this end, we firstly define the following Riccati equation

$$P_k = Q + K_k' R_k K_k + (A + B_k \bar{K}_k)' P_{k+1} (A + B_k \bar{K}_k) + (A + B_k \bar{K}_k)' S_{k+1} (A + B_k \bar{K}_k) + (A + B_k \bar{K}_k)' G_{k+1} \Delta_{k+1} G_{k+1}' (A + B_k \bar{K}_k)' M_k' M_k (A + B_k \bar{K}_k)$$  

where $\Delta_k = M_k' M_k - B_k' S_{k+1} B_k$ and $G_{k+1} = K_{k+1} R_{k+1}$. 

and

$$J^f(N, 0) = \sum_{h=0}^{N} \left[ x_h' (Q + K_h' R_h K_h) x_h + (u_h')' R_h u_h' + 2x_h' K_h' R_h K_h x_h - x_{h+1}' G_h' R_h G_h x_{h+1} + x_{N+1}' P_{N+1} x_{N+1} \right]. $$  

Similar to the line for Lemma 1, we have the following Lemma.

Lemma 2. Suppose the follower’s optimization problem admits a solution, then the controller $u_k^f$ satisfies

$$0 = R_k u_k^f + B_k' \lambda_k^f - \beta_k^f,$$  

where $\lambda_k^f$ is as follows:

$$\lambda_{k-1} = (Q + K_k' R_k K_k + (A + B_k \bar{K}_k)' P_{k+1} (A + B_k \bar{K}_k) + (A + B_k \bar{K}_k)' S_{k+1} (A + B_k \bar{K}_k) + (A + B_k \bar{K}_k)' G_{k+1} \Delta_{k+1} G_{k+1}' (A + B_k \bar{K}_k)' M_k' M_k (A + B_k \bar{K}_k)$$  

with $\bar{G}_0 = \bar{G}_{N+1} = 0$. And $\beta_k^f$ is arbitrary to be determined.
**Proof.** From (1) and the controller’s linear form (4)-(5), it is not hard to derive that \( x_k = A(k, 0)x_0 \) and \( x_{k-1} = A(k-1, 0)x_0 \), where \( A(k, i) = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} A_i + A(k, i + 2) B_{i+1} \) with \( A(k, k) = I \), \( A(k, k-1) = A_{k-1} \) and \( A_i = A + B_k K_i + B_2 K_i \). \( B_i = B_1 G_i + B_2 G_i \), \( i = 0, 1, \ldots, k - 1 \). Hence, equation (31) is rewritten as follows.

\[
[W_k A(k, 0) + V_k A(k-1, 0)]x_0 + b_k = 0. \tag{32}
\]

In view of the arbitrariness of the initial state \( x_0 \), obviously, it deduces that \( b_k = 0 \).

Accordingly, we give the main result of the follower’s optimization problem.

**Theorem 1.** The follower’s optimal controller is

\[
 u_k^f = - (\Gamma_k^f)^{-1} (M_k^B + M_k^B I_k + B_1^I S_{k+1}) x_k - (\Gamma_k^f)^{-1} M_k^B G_k x_{k-1} \tag{33}
\]

and the co-state \( \lambda_{k-1}^f \) is given as follows.

\[
 \lambda_{k-1}^f = P_k x_k + S_k x_{k-1}. \tag{34}
\]

Moreover, the follower’s optimal cost value is

\[
 J^f(N, 0) = x_0^T P_0 x_0. \tag{35}
\]

**Proof.** We adopt the backward iteration method to find the optimal controller. When \( k = N \), given (17), we get

\[
 0 = (R_1 + B_1^I P_{N+1} B_1) u_N^f + B_1^I P_{N+1} (A + B_2 K_N) x_N + B_1^I P_{N+1} B_2 G_N x_{N-1} - \beta_N^f, \tag{36}
\]

i.e.,

\[
 u_N^f = -(\Gamma_N^N)^{-1} (M_N^B + M_N^B I_N + B_1^I S_{N+1}) x_N - (\Gamma_N^N)^{-1} M_N^B G_N x_{N-1} + (\Gamma_N^N)^{-1} \beta_N^f. \tag{37}
\]

From the linear closed-loop form, i.e., \( u_N^f = K_N x_N + G_N x_{N-1} \), we derive that

\[
 [K_N + (\Gamma_N^N)^{-1} (M_N^B + M_N^B I_N + B_1^I S_{N+1})] x_N + [G_N + (\Gamma_N^N)^{-1} M_N^B G_N] x_{N-1} - (\Gamma_N^N)^{-1} \beta_N^f = 0. \tag{38}
\]

From Lemma 3 and (38), it yields that \( (\Gamma_N^N)^{-1} \beta_N^f = 0 \). Further, we obtain that \( \beta_N^f = 0 \) due to the invertibility of \( (\Gamma_N^N)^{-1} \). That is,

\[
 [K_N + (\Gamma_N^N)^{-1} (M_N^B + M_N^B I_N + B_1^I S_{N+1})] x_N + [G_N + (\Gamma_N^N)^{-1} M_N^B G_N] x_{N-1} = 0.
\]

Therefore, the optimal controller is given by

\[
 u_N^f = K_N x_N + G_N x_{N-1} = -(\Gamma_N^N)^{-1} (M_N^B + M_N^B I_N + B_1^I S_{N+1}) x_N - (\Gamma_N^N)^{-1} M_N^B G_N x_{N-1}. \tag{39}
\]

For finding the relationship between co-state and state, from (18), we have

\[
 \lambda_{N-1}^f = [Q + \bar{K}_N R_N R_2 K_N + \bar{K}_N R_2 R_2 G_N x_{N-1} + (A + B_2 K_N)^T P_{N+1} x_N + B_1^I G_N x_{N-1} + B_2 G_N x_{N-1}] - B_1 (\Gamma_{N-1}^N)^{-1} (M_{N-1}^B + M_{N-1}^B I_{N-1} + B_1^I S_{N+1}) x_{N-1} + (\Gamma_{N-1}^N)^{-1} \beta_{N-1}^f \tag{40}
\]

For any \( n \) with \( 0 \leq n \leq N \), we assume that \( \Gamma_n^f > 0 \) and the follower’s optimal controller \( u_k^f \) and co-state \( \lambda_{k-1}^f \) are respectively given as in (33) and (34) for \( k > n \). Now we show \( \lambda_n^f \) and \( \lambda_{k-1}^f \) have the same expression as (33) and (34) for \( k = n \), respectively. For \( u_n^f \), from (17), we have

\[
 0 = [R_1 + B_1^I P_{n+1} B_1] u_n^f + [(M_n^B + M_n^B I_n + B_1^I S_{n+1}) x_n + M_n^B G_n x_{n-1} - \beta_n^f].
\]

It yields that

\[
 u_n^f = -(\Gamma_n^N)^{-1} (M_n^B + M_n^B I_n + B_1^I S_{n+1}) x_n - (\Gamma_n^N)^{-1} M_n^B G_n x_{n-1} + (\Gamma_n^N)^{-1} \beta_n^f,
\]

combining the linear form \( u_n^f = K_n x_n + G_n x_{n-1} \), hence, we have that

\[
 [K_n + (\Gamma_n^N)^{-1} (M_n^B + M_n^B I_n + B_1^I S_{n+1})] x_n + [G_n + (\Gamma_n^N)^{-1} M_n^B G_n] x_{n-1} - (\Gamma_n^N)^{-1} \beta_n^f = 0. \tag{41}
\]

From Lemma 3 and (41), it yields that \( (\Gamma_n^N)^{-1} \beta_n^f = 0 \). Moreover, it deduces that \( \beta_n^f = 0 \) for the invertibility of \( (\Gamma_n^N)^{-1} \). Then, we have that

\[
 [K_n + (\Gamma_n^N)^{-1} (M_n^B + M_n^B I_n + B_1^I S_{n+1})] x_n + [G_n + (\Gamma_n^N)^{-1} M_n^B G_n] x_{n-1} = 0.
\]

That is, the optimal controller is given by

\[
 u_n^f = K_n x_n + G_n x_{n-1} = -(\Gamma_n^N)^{-1} (M_n^B + M_n^B I_n + B_1^I S_{n+1}) x_n - (\Gamma_n^N)^{-1} M_n^B G_n x_{n-1}. \tag{42}
\]

Thus, by induction it is easy to know that the controller \( u_k^f \) for any \( 0 \leq k \leq N \) has the form as (33).

Next, we consider the expression of \( \lambda_n^f \). In view of (15) and (18), it follows

\[
 \lambda_{n-1}^f = [Q + \bar{K}_N^f R_N R_2 K_N + \bar{K}_N^f R_2 R_2 G_N x_{n-1} + (A + B_2 K_N)^T P_{n+1} x_n + B_1^I G_N x_{n-1} + B_2 G_N x_{n-1}] - B_1 (\Gamma_{n-1}^f)^{-1} (M_{n-1}^B + M_{n-1}^B I_{n-1} + B_1^I S_{n+1}) x_{n-1} + (\Gamma_{n-1}^f)^{-1} \beta_{n-1}^f \tag{43}
\]

For finding the relationship between co-state and state, from (18), we have

\[
 \lambda_{n-1}^f = [Q + \bar{K}_N^f R_N R_2 K_N + \bar{K}_N^f R_2 R_2 G_N x_{n-1} + (A + B_2 K_N)^T P_{n+1} x_n + B_1^I G_N x_{n-1} + B_2 G_N x_{n-1}] - B_1 (\Gamma_{n-1}^f)^{-1} (M_{n-1}^B + M_{n-1}^B I_{n-1} + B_1^I S_{n+1}) x_{n-1} + (\Gamma_{n-1}^f)^{-1} \beta_{n-1}^f \tag{43}
\]

The above second equation is based on the hypothetical of \( \lambda_n^f \). Thus, by induction it is easy to know that the co-state \( \lambda_k^f \) for any \( 0 \leq k \leq N \) has the form as (34).

For finding the follower’s optimal cost value, define \( V_k^f = x_k^T P_k x_k + 2 x_k^T S_k x_{k-1} \), we get
\[ V'_k - V_{k+1} = x'_k [P_k - A' \cdot P_{k+1} \cdot A - 2(M^B_k' \cdot K_k - 2A' \cdot S_{k+1} - K'_k \cdot B_k P_{k+1} \cdot B_k \cdot K_k - 2K'_k B_k S_{k+1}) \cdot x_k + 2x'_k \cdot (S_k - (M^B_k') \cdot \bar{G}_k - K'_k B_k P_{k+1} \cdot B_k \cdot G_k x_k - 2x'_k(M^B_k' \cdot x'_k) - (u'_k) \cdot (u'_k) \cdot (M^B_k' \cdot x'_k - 2u'_k) \cdot (M^B_k' \cdot G_k x_k - 2u'_k) \cdot (M^B_k' \cdot G_k x_k - 2u'_k) \cdot (M^B_k' \cdot G_k x_k - 2u'_k) \cdot (M^B_k' \cdot G_k x_k - 2u'_k) \cdot (M^B_k' \cdot G_k x_k - 2u'_k)] \]

According to Lemma 1, we can obtain the following result.

**Lemma 4.** Suppose the leader’s optimization problem admits a solution, then, the leader’s controller \( u^L_k \) satisfies

\[
0 = (M^B_k) (\Gamma^G_k)^{-1} R_k (\Gamma^G_k)^{-1} (M^B_k + B'_k S'_{k+1}) \cdot x_k + (R^*_k + (M^B_k') \cdot (\Gamma^G_k)^{-1}) R_k (\Gamma^G_k)^{-1} x_k + B_2 - B_1 (\Gamma^G_k)^{-1} M^B_k' \cdot \lambda_k - \beta_k, \tag{47}
\]

where \( \lambda_k \) is as

\[
\lambda'_k = \left[ Q + (M^B_k + B'_k S'_{k+1}) \right] R_k \cdot \left[ (M^B_k + B'_k S'_{k+1}) \right] x_k + (M^B_k + B'_k S'_{k+1}) \cdot x_k + (M^B_k + B'_k S'_{k+1}) \cdot x_k + (M^B_k + B'_k S'_{k+1}) \cdot x_k + (M^B_k + B'_k S'_{k+1}) \cdot x_k \cdot \lambda_k + \bar{K}_k \cdot \beta_k + \bar{G}_{k+1} \cdot \lambda_{k+1}, \tag{48}
\]

\[
\beta_k = P_{k+1} x_{N+1} \tag{49}
\]

with \( \bar{G}_0 = \bar{G}_{N+1} = 0. \) And \( \beta_k \) is arbitrary to be determined.

**Proof.** It is immediately obtained by similar lines of discussion for Lemma 2. \( \square \)

Next, we shall solve the controller \( u^L_k \) from Lemma 4. To this end, we firstly define the following Riccati equation

\[
\dot{P}_k = Q + (M^B_k + B'_k S'_{k+1}) \cdot (\Gamma^G_k)^{-1} R_k \cdot (\Gamma^G_k)^{-1} (M^B_k + B'_k S'_{k+1}) \cdot x_k + (M^B_k + B'_k S'_{k+1}) \cdot x_k + (M^B_k + B'_k S'_{k+1}) \cdot x_k + (M^B_k + B'_k S'_{k+1}) \cdot x_k \cdot \bar{K}_k \cdot \lambda_k + \bar{K}_k \cdot \beta_k + \bar{G}_{k+1} \cdot \lambda_{k+1}, \tag{50}
\]

\[
\lambda'_k = \left[ Q + (M^B_k + B'_k S'_{k+1}) \right] R_k \cdot \left[ (M^B_k + B'_k S'_{k+1}) \right] x_k + (M^B_k + B'_k S'_{k+1}) \cdot x_k + (M^B_k + B'_k S'_{k+1}) \cdot x_k + (M^B_k + B'_k S'_{k+1}) \cdot x_k + (M^B_k + B'_k S'_{k+1}) \cdot x_k \cdot \lambda_k + \bar{K}_k \cdot \beta_k + \bar{G}_{k+1} \cdot \lambda_{k+1}, \tag{51}
\]

\[
M^B_k = (M^B_k') (\Gamma^G_k)^{-1} R_k (\Gamma^G_k)^{-1} (M^B_k + B'_k S'_{k+1}) \cdot x_k + (M^B_k + B'_k S'_{k+1}) \cdot x_k + (M^B_k + B'_k S'_{k+1}) \cdot x_k + (M^B_k + B'_k S'_{k+1}) \cdot x_k \cdot \bar{K}_k \cdot \lambda_k + \bar{K}_k \cdot \beta_k + \bar{G}_{k+1} \cdot \lambda_{k+1}, \tag{52}
\]

Now, we give the results of the leader’s optimization problem. It should be pointed out that the discussion of leader’s optimization problem is based on the solvability of the follower’s optimization problem.

**Theorem 2.** The leader’s optimal linear strategy is\( u^L_k = - (\Gamma^G_k)^{-1} M^B_k \cdot x_k, \tag{53} \)

and the relationship between state and co-state is

\[
\lambda_{k+1} = P_{k+1} x_{N+1}. \tag{54}
\]

Moreover, the leader’s optimal cost value is given by

\[
J^L_k (N, 0) = x'_k P_{k+1} x_{N+1}. \tag{55}
\]

**Proof.** Using the backward iteration method, from (47) with \( k = N \), we obtain

\[
0 = (M^B_N) (\Gamma^G_N)^{-1} R_N (\Gamma^G_N)^{-1} M^B_N x_N + \left[ R^*_N + (M^B_N') (\Gamma^G_N)^{-1} R_N (\Gamma^G_N)^{-1} x_N + B_2 - B_1 (\Gamma^G_N)^{-1} M^B_N' \cdot \lambda_N - \beta_N \right.
\]

\[
\lambda_N = P_{N+1} x_{N+1}, \tag{56}
\]

i.e.,

\[
u^L_N = - (\Gamma^G_N)^{-1} M^B_N x_N + (\Gamma^G_N)^{-1} \beta_N. \tag{56}
\]
From the linear closed-loop form $\bar{K}_N x_N + \bar{G}_N x_{N-1}$, we derive that

$$[\bar{K}_N + (\Gamma_N^{-1} - \Gamma_N^{-1}) \mathcal{M}_N] x_N + \bar{G}_N x_{N-1} - (\Gamma_N^{-1})^{-1} \beta_N = 0. \tag{57}$$

From Lemma 3 and (57), it yields that $(\Gamma_N^{-1})^{-1} \beta_N = 0$. Moreover, it deduces that $\beta_N = 0$ for the invertibility of $(\Gamma_N^{-1})^{-1}$. Thus, we get

$$[\bar{K}_N + (\Gamma_N^{-1} - \Gamma_N^{-1}) \mathcal{M}_N] x_N + \bar{G}_N x_{N-1} - (\Gamma_N^{-1})^{-1} \beta_N = 0.$$

That is, the optimal controller is given by

$$u_N^* = \bar{K}_N x_N + \bar{G}_N x_{N-1} - (\Gamma_N^{-1})^{-1} \mathcal{M}_N x_N. \tag{58}$$

For $\lambda^N_{N-1}$, from (48), we have

$$\lambda^N_{N-1} = (\bar{Q} + (\Gamma_N')^{-1} \bar{R}_1(\Gamma_N')^{-1} (\mathcal{M}_N') x_N + (\mathcal{M}_N') (\Gamma_N')^{-1} x_N + (\bar{A} - (\Gamma_N')^{-1} (\mathcal{M}_N') x_N + (\mathcal{M}_N') (\Gamma_N')^{-1} x_N) P_{N+1} \mathcal{M}_N x_N. \tag{59}$$

For any $n$ with $0 \leq n \leq N$, we assume that $\Gamma_n' > 0$ and the optimal follower’s controller $u_n$ and co-state $\lambda^N_n$ are respectively given as in (53) and (54) for $k > n$. Next, we show that $u_k$ and $\lambda^N_k$ have the same form as in (53) and (54) for $k = n$, respectively. For $u_n'$, from (47), we have that

$$0 = (\mathcal{M}_n') (\Gamma_n')^{-1} \bar{R}_1(\Gamma_n')^{-1} (\mathcal{M}_n' + \mathcal{B}_n' s_{n+1}' ) x_n + (\bar{R}_2 + (\mathcal{M}_n') (\Gamma_n')^{-1} x_n + (\bar{A} - (\Gamma_n')^{-1} (\mathcal{M}_n') x_n + (\mathcal{M}_n') (\Gamma_n')^{-1} x_n) P_{n+1} (\mathcal{B}_n' + (\mathcal{M}_n') (\Gamma_n')^{-1} x_n) u_n = P_n x_n. \tag{60}$$

The second equality holds from equations (50). Hence, taking summation from $k = 0$ to $k = N$, we have that $u_k$ in (53), the optimal cost values is $J^* (N, 0) = x_0^T P_0 x_0$. Now the proof of Theorem 2 is completed. \qed

**Remark 5.** It is interesting to know that the presented optimal leader’s strategy is in fact a current state feedback. Based on the definition of the leader’s strategy in (5), it is known that

$$u_k = \bar{K}_k x_k + \bar{G}_k x_{k-1} - (\Gamma_k^{-1})^{-1} \mathcal{M}_k x_k.$$

Clearly, it can not determine $\bar{K}_k$ and $\bar{G}_k$ uniquely from the above identity because $x_k$ is dependent on $x_{k-1}$. So further optimization will be made to solve $\bar{K}_k$ and $\bar{G}_k$, which are necessary to design both the follower’s and leader’s strategy.

**C. The Determination of $\bar{K}_k$ and $\bar{G}_k$ by Twice Optimization**

It can be seen from Theorem 1 and 2 that the design of controller and the associated performance are based on the gain matrices $\bar{K}_k$ and $\bar{G}_k$ which are to be determined. In the following, we shall derive the gain matrices by a method of twice optimization for the performance of the follower, which will lead to a better performance.

In view of (2), and following the discussion in Part A, it is easy to know that $J^* (N, s) = x_0^T P_0 x_0$, where $P_0$ is as (24). Note that $P_n$ contains $\bar{K}$ in quadratic form, so it is one possible way to determine $\bar{K}$ by taking minimization of $J^* (N, s)$, that is,

$$\bar{K}_k = \arg \min_{\bar{K}} J^* (N, s).$$
With the above twice optimization, we have the following results.

**Theorem 3.** The gain matrices $\bar{K}_k$ and $\bar{G}_k$ are given as follows

\[
\bar{K}_k = -\Delta_k^{-1}M_k^f, \\
\bar{G}_k = -\Delta_k^{-1}M_k^f + \Delta_k^{-1}M_k^n[A - B_1(G_k^{-1})^{-1}]M_{k-1}^f - B_2(G_k^{-1})^{-1}M_{k-1}. 
\]

where the parameters are as in (24)-(30) and (50)-(52), and

\[
M_k^f = M_k^f + B_k^1S_{k+1}^f - M_k^B(\Gamma_k^f)^{-1}M_k^f. 
\]

**Proof.** Firstly, for the follower’s optimal controller $u_k^f$, from the result of Theorem 1 and 2, we can further calculate it as below.

\[
u_k^f = -\Gamma_k^f M_k^f \bar{K}_k + B_k^1S_{k+1}^f x_k - \Gamma_k^f M_k^f G_k x_{k-1} = -(\Gamma_k^f)^{-1}(M_k^f + B_k^1S_{k+1}^f)x_k + (\Gamma_k^f)^{-1}M_k^B(\Gamma_k^f)^{-1}M_k^f x_k = -(\Gamma_k^f)^{-1}M_k^f x_k. 
\]

In this case, combining (1) and (53), it is easy to deduce that

\[
x_k = [A - B_1(\Gamma_k^f)^{-1}]^{-1}M_k^f x_k - B_2(\Gamma_k^{-1})^{-1}M_k^f x_k. 
\]

In this basis, we find $\bar{G}_k$ in the sequel. By the linear constraint (5) and (67), we get

\[
\bar{G}_k x_{k-1} = u_k^f - \bar{K}_k x_k = -\Gamma_k^f M_k^f x_k - \bar{K}_k x_k = -[\Gamma_k^f M_k^f + \bar{K}_k][A - B_1(\Gamma_k^f)^{-1}]^{-1}M_k^f x_k - B_2(\Gamma_k^{-1})^{-1}M_k^f x_k. 
\]

Due to the fact that the above relationship holds for any $x_{k-1}$, thus, we have

\[
\bar{G}_k = -[\Gamma_k^f M_k^f + \bar{K}_k][A - B_1(\Gamma_k^f)^{-1}]^{-1}M_k^f x_k - B_2(\Gamma_k^{-1})^{-1}M_k^f x_k. 
\]

It is clear that it cannot solve $\bar{K}_k$, $\bar{G}_k$ uniquely from equation (68). Therefore, we need to further determine $K_k$ in the next step. To this end, using the twice optimization stated in the above, it is derived that

\[
0 = \frac{\partial P_k}{\partial K_k} = R_k \bar{K}_k + B_k^1P_{k+1}(A + B_2\bar{K}_k) + B_k^1S_{k+1} = -(M_k^B)^{(\Gamma_k^f)^{-1}}(M_k^f + M_k^B \bar{K}_k + B_k^1S_{k+1}) = R_k + B_k^1P_{k+1}B_2(M_k^B)^{(\Gamma_k^f)^{-1}}M_k^f \bar{K}_k + B_k^1P_{k+1}A + B_k^1S_{k+1} - (M_k^B)^{(\Gamma_k^f)^{-1}}(M_k^f + B_k^1S_{k+1}) = \Delta_k \bar{K}_k + \Delta_k \bar{G}_k. 
\]

From (69), when $\Delta_k$ is invertible, it yields

\[
\bar{K}_k = -\Delta_k^{-1}M_k^f. 
\]

Thus, combining (68), $\bar{G}_k$ is determined as in (64). Now, we have determined the gain matrices $K_k$ and $G_k$. The proof is completed. 

Based on the above discussion, we give the following main result.

**Theorem 4.** The optimal linear closed-loop Stackelberg strategy for Problem 1 is given by

\[
u_k^f = -\Gamma_k^f M_k^f x_k, \quad u_k^* = -(\Gamma_k^f)^{-1}M_k^f x_k, 
\]

where $\Gamma_k^f, M_k^f, \Gamma_k^f$ and $M_k^f$ are as in (28), (65), (51) and (52). Moreover, the follower’s and leader’s optimal cost value are given by

\[
J_k^*(N, 0) = x_0^* P_0 x_0^*, \quad J_k^*(N, 0) = x_0^* P_0 x_0^*, 
\]

where $P$ and $P$ are the solution of Riccati equations (24)-(30) and (50)-(52).

**Remark 6.** Theorem 4 shows that the linear closed-loop Stackelberg strategy has an elegant form, i.e., current state feedback. However, it is completely different from the classical feedback strategy though they have similar form of controller. In fact, the closed-loop controller has better performance which is verified in simulation examples in next section. Secondly, the obtained linear closed-loop optimal strategy is different from that in [2] where the team optimization approach is adopted to solve the closed-loop Stackelberg game with one-step memory and an additional implicit equation is required to be solved. Thirdly, it is different from maximum principle based approach [4]. More importantly, under the presented controller structure, the Q-learning algorithm becomes possible, which will be given in our future work.

**IV. SIMULATION**

In this section, we take a numerical example to show that the derived closed-loop strategy achieves better performance than feedback strategy.

We consider the above Stackelberg problem with the following coefficients: $A = 0.7$, $B_1 = 1$, $B_2 = -1$, $Q = 2$, $R_1 = 2$, $R_2 = 10$, $Q = 1$, $R_1 = 4$, $R_2 = 2$, $N = 5$. Given the terminal values $P(N + 1) = P(N + 1) = 2$ and an initial value $x_0 = 10$. Next, we will give the results of feedback case and liner closed-loop case.

First, for the standard feedback case, i.e., $u_k^f = K_k x_k$, $u_k^f = \bar{K}_k x_k$, by calculating, the optimal strategies are as follows (TABLE I). And in this case, the optimal cost values are given by $J_k^* = 323.3451, J_k^* = 145.5862$.

On the other hand, for the linear closed-loop case, i.e., $u_k^f = K_k x_k + G_k x_{k-1}$, $u_k^f = \bar{K}_k x_k + G_k x_{k-1}$, by deriving the gain matrices, it can be written as a form of a current state feedback, and the results are calculated as follows (TABLE II). where the results of $K_k$ and $G_k$ are given: It should be pointed out that the gain matrices $K_k$, $G_k$ given in TABLE III satisfy the relationship $K_k x_k + G_k x_{k-1} = K_k x_k$, where $K_k$ is the corresponding gain matrix in the form of state feedback in TABLE II. And the optimal cost values of linear closed-loop case are $J_k^*(N, 0) = 250.9362, J_k^*(N, 0) = 142.2957$.

Comparing the cost values of the feedback case and the linear closed-loop case, it is not hard to note that the feedback result may not always minimize the cost values of Stackelberg problem. Actually, it is rational due to that more information is obtained by players for the linear closed-loop case.
TABLE I
THE FEEDBACK RESULTS

<table>
<thead>
<tr>
<th>Step k</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>u_k^1</td>
<td>-0.2313x_0</td>
<td>-0.2313x_1</td>
<td>-0.2313x_2</td>
<td>-0.2313x_3</td>
<td>-0.2298x_4</td>
<td>-0.2x_5</td>
</tr>
<tr>
<td>u_k^2</td>
<td>0.3256x_0</td>
<td>0.3256x_1</td>
<td>0.3256x_2</td>
<td>0.3256x_3</td>
<td>0.3201x_4</td>
<td>0.3x_5</td>
</tr>
</tbody>
</table>

TABLE II
THE LINEAR CLOSED-LOOP RESULTS

<table>
<thead>
<tr>
<th>Step k</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>u_k^1</td>
<td>-0.2214x_0</td>
<td>-0.2214x_1</td>
<td>-0.2214x_2</td>
<td>-0.2214x_3</td>
<td>-0.2202x_4</td>
<td>-0.2x_5</td>
</tr>
<tr>
<td>u_k^2</td>
<td>0.3021x_0</td>
<td>0.3021x_1</td>
<td>0.3021x_2</td>
<td>0.3021x_3</td>
<td>0.2998x_4</td>
<td>0.3x_5</td>
</tr>
</tbody>
</table>

TABLE III
VALUES OF $\hat{K}_k$ AND $\hat{G}_k$

<table>
<thead>
<tr>
<th>Step k</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{K}_k$</td>
<td>0.0701</td>
<td>0.0701</td>
<td>0.0701</td>
<td>0.0701</td>
<td>0.0694</td>
<td>0.0636</td>
</tr>
<tr>
<td>$\hat{G}_k$</td>
<td>0.0409</td>
<td>0.0409</td>
<td>0.0409</td>
<td>0.0409</td>
<td>0.0407</td>
<td>0.0426</td>
</tr>
</tbody>
</table>

V. CONCLUSION

This paper solves the linear closed-loop Stackelberg control by proposing a novel approach of linear full-information feedback. Different from the earlier works, it is the first time to derive an explicit closed-loop controller based on Riccati equations with a direct derivation structure. Due to the elegance of the results, it will find significant applications for game control of players with asymmetric positions and the Q-learning for Stackelberg control will become possible which have never been given in literatures.

REFERENCES


