Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Adaptive State Feedback Nash Strategies for Linear Quadratic Discrete-Time Games

Dan Shen * and **Jose B. Cruz, Jr.** **

* *Intelligent Automation Inc., Rockville, MD 20858 USA (email: dshen@i-a-i.com).*
** *The Ohio State University, Columbus, OH 43210 USA(email: jbcruz@ieee.org).*

Abstract: A substantial effort has been devoted to various adaptive techniques of systems. Most of these concepts work in the control domain, where every system only has one controller. Yet, for the multi-controller counterpart — dynamic games, adaptations are usually considered from a perspective of systems, for an example, evolutionary games. In this paper, we propose a new adaptive approach for linear quadratic discrete-time games with scalar inputs and state feedback Nash strategies. We consider the effort of adaptation under a Fictitious Play (FP) framework with learning algorithms derived from conventional adaptive control methods. Convergence to Nash strategies is proved with the condition that there exists a unique state feedback strategy, which implies that the associated coupled discrete-time algebraic Riccati equations (DAREs) have a unique positive semi-definite solution. The requirement of Persistency of Excitation (PE) is satisfied by proper reference signals to be tracked.

Keywords: Adaptive Control; Adaptive Linear Quadratic Games; Adaptive State Feedback Nash Strategies; Riccati Equations; Fictitious Play; Persistency of Excitation.

1. INTRODUCTION

In general, every closed-loop controller is "adaptive" in the sense that it modifies its output in response to changes of system states. However, an "adaptive controller" can adapt not only its output, but its underlying control strategy as well. Adaptive controllers can outperform their fixed-parameter counterparts in terms of efficiency. They can often eliminate errors faster and with fewer fluctuations. In Åström and Wittenmark [1995], an adaptive controller is defined as "a controller with adjustable parameters and a mechanism for adjusting the parameters". Usually, an adaptive control system has two loops: one is the information feedback between the process and the controller; the other is for the parameter adjustment.

In the control domain, where every system only has one controller, a range of techniques, such as adaptive dual control methods Alster and Belanger [1974], Bar-Shalom and Tse [1974], normalization or dead zones Peterson and Narendra [1982], Kreisselmeier and Anderson [1986], model reference adaptive control Morse [1980], Narendra et al. [1980], Goodwin et al. [1980], and self-tuning regulators Åström and Wittenmark [1973], are used to design the adaptive schemes. While in the game domain, there are two concepts related to the adaptation design. The first one is called evolutionary game theory (EGT) Smith [1982] that originated as an application of the mathematical theory of games to biological contexts. Arising from the realization that frequency dependent fitness introduces a strategic aspect to evolution, EGT is the application of population genetics-inspired models of change in gene frequency in populations to game theory. It differs from classical game theory by focusing on the dynamics of strategy change more than the properties of strategy equilibria. The second one is Fictitious Play (FP) Brown [1951], Fudenberg and Levine [1998]. In it, each player presumes that its opponents are playing stable (possibly mixed) strategies. Each player starts with some initial beliefs and chooses a best response to those beliefs as a strategy in this round. Then, after observing their opponents' actions, the players update their beliefs according to some learning rule (e.g. reinforcement learning or Bayes' rule). The process is then repeated.

In this paper, we present a new adaptive approach for linear quadratic games with state feedback Nash strategies. We consider the effort of adaptation under a FP framework with learning algorithms derived from conventional adaptive control methods. When the parameters of the objective functions are not shared among the decision makers, an on-line adaptive scheme is provided for each player to estimate the actual control gain used by the other player. The convergence to Nash strategies is proved with the condition that the associated coupled DAREs have a unique positive definite solution. The requirement of Persistency of Excitation (PE) is satisfied by proper reference signals to be tracked.

The remainder of the paper is organized as follows. The state feedback Nash strategies for linear quadratic games are reviewed in Section 2. Then in Section 3 (one-side adaptation) and Section 4 (two-side adaptation), we design

⋆ This work was supported in part by the Defense Advanced Research Project Agency (DARPA) under Contract F33615-01-C3151 issued by the AFRL/VAK, and in part by The Ohio State University College of Engineering.

adaptive laws for linear quadratic games, and provide the proof of the convergence to Nash strategies given that the associated coupled discrete-time algebraic Riccati equations (DAREs) have a unique semi-positive definite solution. Numerical systems are simulated in Section 5 to illustrate the proposed adaptation designs for linear quadratic games with state feedback Nash strategies. Finally, conclusions are drawn in Section 6.

2. STATE FEEDBACK NASH STRATEGIES FOR LINEAR QUADRATIC GAMES

Let us consider two-person infinite-horizon linear quadratic games

$$x_{k+1} = Ax_k + B^1 u_k^1 + B^2 u_k^2 \quad (1)$$

with the cost functions

$$J^1 = \sum_{k=0}^{+\infty} \left(x_k^\top Q^1 x_k + u_k^{1\top} R^{11} u_k^1 + u_k^{2\top} R^{12} u_k^2 \right) \quad (2)$$

$$J^2 = \sum_{k=0}^{+\infty} \left(x_k^\top Q^2 x_k + u_k^{1\top} R^{21} u_k^1 + u_k^{2\top} R^{22} u_k^2 \right) \quad (3)$$

where $x_k \in \mathcal{R}^n$, $u_k^i \in \mathcal{R}^{m_i}$, $Q^i > 0$, $R^{ij} > 0$, (A, B^i) is stabilizable and (A, C^i) is observable (where $C^{i\top} C^i = Q^i$) for $i = 1, 2$, $k = 1, 2 \cdots$. We assume that both players have perfect information structures, with which both players know the exact system dynamics $x_{k+1} = Ax_k + B^1 u_k^1 + B^2 u_k^2$ and measure exact system states x_k. It is well known (see Başar and Olsder [January, 1999]) that the Nash strategy pair is specified by

$$u_k^{1*} := \arg\min_{u_k^1} J^1 = \gamma^1(x_k) = L^1 x_k \quad (4)$$

$$u_k^{2*} := \arg\min_{u_k^2} J^2 = \gamma^2(x_k) = L^2 x_k \quad (5)$$

where L^1 and L^2 are defined, respectively, by

$$L^1 = -[R^{11} + B^{1\top} K^1 B^1]^{-1} B^{1\top} K^1 [A + B^2 L^2] \quad (6)$$

$$L^2 = -[R^{22} + B^{2\top} K^2 B^2]^{-1} B^{2\top} K^2 [A + B^1 L^1] \quad (7)$$

K^1 and K^2 are specified in the following coupled DAREs, that, by assumption, have a unique positive semi-definite solution (in general, uniqueness and positive definiteness of solutions is not necessary to obtain a Nash equilibrium, see Başar and Olsder [January, 1999] Prop. 6.3. But in this paper, we need them to prove convergence of our designed adaptive laws).

$$\begin{aligned} K^1 =& Q^1 + L^{1\top} R^{11} L^1 + L^{2\top} R^{12} L^2 \\ &+ (A + B^1 L^1 + B^2 L^2)^\top K^1 (A + B^1 L^1 + B^2 L^2) \quad (8) \\ K^2 =& Q^2 + L^{1\top} R^{21} L^1 + L^{2\top} R^{22} L^2 \\ &+ (A + B^1 L^1 + B^2 L^2)^\top K^2 (A + B^1 L^1 + B^2 L^2) \quad (9) \end{aligned}$$

It is straightforward to extend state feedback Nash strategies to M-player cases.

3. ONE-SIDE ADAPTATION SCHEME

In the one-side adaptation scheme, only one of the two players has perfect information of the cost functions of both players, i.e., the player knows exact Q^1, Q^2, R^{11}, R^{12}, R^{21}, R^{22}, while the other player only has access to its cost function and does not know the parameters of the cost function of its opponent. Without loss of generality, we assume that player 1 knows exact Q^i , R^{ij}, A, B^i while player 2 has access to Q^2, R^{21}, R^{22}, A, B^i. We also assume that player 1 will apply its state feedback Nash strategies calculated based on the information of system dynamics, system states and cost functions of both players. Player 2 knows in advance that its opponent will implement state feedback Nash strategies. By following the concept of Fictitious Play (PE), player 2 will estimate the control gain L^1 first, and then calculate/estimate its own best response control gain L^2.

Assumption 1. In our proposed adaptive schemes including one-side adaptation here and two-side one in Section 4, we assume that u^i is scalar.

3.1 Adaptation Design

In the one-side adaptation scheme, player 1, who has perfect information structure, will apply its real state feedback Nash strategies $u^{1*} = L^1 x_k$, so we follow the conventional indirect adaptive control design method Tao [2003]. Consider the system defined in (1) with fixed controller (4) for player 1, let $\hat{L}_k^1$ be an estimate of the control gain L^1 of player 1, from the point view of player 2. The block diagram of indirect adaptive control system is shown in Fig. 1.

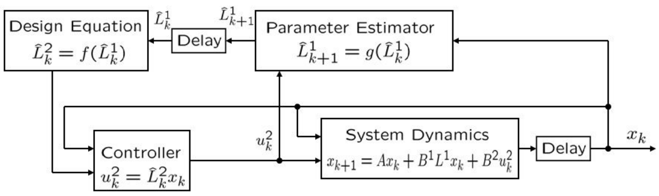

Fig. 1. Indirect adaptive control design for one-side adaptation scheme

First, we have

$$\hat{x}_{k+1} = Ax_k + B^1 \hat{L}_k^1 x_k + B^2 u_k^2 \quad (10)$$

Since $B^{1\top} B^1$ is a scalar, then from (1) and (10), we have

$$[B^{1\top} B^1]^{-1} B^{1\top} (\hat{x}_{k+1} - x_{k+1}) = (\hat{L}_k^1 - L^1) x_k \quad (11)$$

We introduce the estimation error $e_k = [B^{1\top} B^1]^{-1} B^{1\top} (\hat{x}_k - x_k)$, ($e_k$ is a scalar) then

$$e_k = \tilde{L}_{k-1}^1 \phi_k \quad (12)$$

where $\tilde{L}_{k-1}^1 = \hat{L}_{k-1}^1 - L^1$ and $\phi_k = \frac{1}{z}[x]_k = x_{k-1}$. (Note, $\frac{1}{z}[x]_k$ denotes the output of the system with transfer function $\frac{1}{z}$ and input x_k.)

Choose the adaptive law for $\hat{L}_k^1$ as

$$\hat{L}_k^1 = \hat{L}_{k-1}^1 - \frac{e_k \phi_k^\top \Gamma}{m_k^2} \quad (13)$$

where $0 < \Gamma = \Gamma^\top < 2I_n$ is a gain matrix, and

$$m_k = \sqrt{\kappa + \phi_k^\top \phi_k}, \quad \kappa > 0$$

Remark 1. Given the persistence of excitation (PE) Åström and Wittenmark [1995], (which will be satisfied and discussed in Section 3.2), we can easily prove convergence by following the convergence proof of Normalized Gradient algorithm (see Tao [2003],page 115-116).

Next, we need to specify a design function or mapping, as shown in Fig. 1, to calculate $\hat{L}_k^2$ for each for $\hat{L}_k^1$. Following the concept of FP, we use best response strategies as the required design function.

From (7) and (9), we have a best response strategy for player 2 given the estimated control gain $\hat{L}_k^1$ of player 1,

$$\hat{L}_k^2 = -[R^{22} + B^{2\top}\hat{K}_k^2 B^2]^{-1} B^{2\top}\hat{K}_k^2[A + B^1\hat{L}_k^1] \quad (14)$$

$$\hat{K}_k^2 = Q^2 + \hat{L}_k^{1\top} R^{21}\hat{L}_k^1 + \hat{L}_k^{2\top} R^{22}\hat{L}_k^2 + (A + B^1\hat{L}_k^1 + B^2\hat{L}_k^2)^\top \hat{K}_k^2 (A + B^1\hat{L}_k^1 + B^2\hat{L}_k^2) \quad (15)$$

Fortunately, as shown in the following Lemma 1, we can calculate the control gains via a standard Riccati equation of a transformed system.

Lemma 1. The $\hat{K}_k^2$ specified in (14) and (15) is the solution to the discrete-time algebraic Riccati equation (DARE)

$$\bar{A}^\top X\bar{A} - X - \bar{A}^\top X\bar{B}[\bar{B}^\top X\bar{B} + \bar{R}]^{-1}\bar{B}^\top X\bar{A} + \bar{Q} = 0$$

where $\bar{A} = A + B^1\hat{L}_k^1$, $\bar{B} = B^2$, $\bar{R} = R^{22}$, and $\bar{Q} = Q^2 + \hat{L}_k^{1\top} R^{21}\hat{L}_k^1$. And with the regularity condition specified below (3), i.e., $Q^i > 0$, $R^{ij} > 0$, (A, B^i) is stabilizable, (A, C^i) is observable (where $C^{i\top} C^i = Q^i$), there exists a positive semi-definite solution to the above discrete time algebraic Riccati equation.

Proof: By matrix manipulation, we can rewrite (15) as

$$\begin{aligned}
\hat{K}_k^2 &= \bar{Q} + \hat{L}_k^{2\top}\bar{R}\hat{L}_k^2 + (\bar{A} + \bar{B}\hat{L}_k^2)^\top \hat{K}_k^2(\bar{A} + \bar{B}\hat{L}_k^2) \\
&= \bar{Q} + \bar{A}^\top\hat{K}_k^2\bar{A} + \hat{L}_k^{2\top}[\bar{R} + \bar{B}^\top\hat{K}_k^2\bar{B}]\hat{L}_k^2 \\
&\quad + \bar{A}^\top\hat{K}_k^2\bar{B}\hat{L}_k^2 + \hat{L}_k^{2\top}\bar{B}^\top\hat{K}_k^2\bar{A} \\
&= \bar{Q} + \bar{A}^\top\hat{K}_k^2\bar{A} + \bar{A}^\top\hat{K}_k^2\bar{B}[\bar{R} + \bar{B}^\top\hat{K}_k^2\bar{B}]^{-1}\bar{B}^\top\hat{K}_k^2\bar{A} \\
&\quad - \bar{A}^\top\hat{K}_k^2\bar{B}[\bar{R} + \bar{B}^\top\hat{K}_k^2\bar{B}]^{-1}\bar{B}^\top\hat{K}_k^2\bar{A} \\
&\quad - \bar{A}^\top\hat{K}_k^2\bar{B}[\bar{R} + \bar{B}^\top\hat{K}_k^2\bar{B}]^{-1}\bar{B}^\top\hat{K}_k^2\bar{A} \\
&= \bar{Q} + \bar{A}^\top\hat{K}_k^2\bar{A} \\
&\quad - \bar{A}^\top\hat{K}_k^2\bar{B}[\bar{R} + \bar{B}^\top\hat{K}_k^2\bar{B}]^{-1}\bar{B}^\top\hat{K}_k^2\bar{A} \qquad (16)
\end{aligned}$$

In the above, the equality in the third line follows (14). Wonham [1968] has proved that the necessary and sufficient conditions for the existence of positive semi-definite solution to the discrete time algebraic Riccati equation (16) are $(\bar{A}, \bar{B})$ is stabilizable and $(\bar{A}, \bar{C})$ is observable (where $\bar{C}^\top\bar{C} = \bar{Q}$).

Since (A, B^2) is stabilizable, there exists a $\bar{L}$ such that $(A + B^2\bar{L})$ is Hurwitz, that is it has all its eigenvalues inside the unit circle. Since $B^{2\top}B^2$ is a scalar, there exists a $\bar{\bar{L}}$, such that

$$A + B^2\bar{L} = \bar{A} + B^2\bar{\bar{L}}$$

Actually, $\bar{\bar{L}} = (B^{2\top}B^2)^{-1}B^{2\top}(A + B^2\bar{L} - \bar{A})$. Then $(\bar{A}, \bar{B})$ is stabilizable (note that $\bar{B} = B^2$).

Similarly, since (A, C^2) is detectable (where $C^{2\top}C^2 = Q^2$), we can always find a $\bar{p}$ such that $(A^\top + C^{2\top}\bar{p})$ is Hurwitz. From the condition $Q^2 > 0$, we have $\bar{Q} = Q^2 + \hat{L}_k^{1\top}R^{21}\hat{L}_k^1 > 0$. Then $\bar{C}^\top > 0$ where $\bar{C}^\top\bar{C} = \bar{Q}$. So there exist a $\bar{\bar{p}}$ such that

$$A^\top + C^{2\top}\bar{p} = \bar{A}^\top + \bar{C}^\top\bar{\bar{p}}$$

Actually $\bar{\bar{p}} = (\bar{C}^\top)^{-1}(A^\top + C^{2\top}\bar{p} - \bar{A}^\top)$. Then $(\bar{A}^\top + \bar{C}^\top\bar{\bar{p}})$ is Hurwitz, too. Thus, $(\bar{A}, \bar{C})$ is observable. Applying Wonham's result, Lemma 1 follows. This completes our proof. ■

Remark 2. We can easily obtain $\hat{K}_k^2$ by using MATLAB command `dare`$(\bar{A}, \bar{B}, \bar{Q}, \bar{R})$.

3.2 Persistency of Excitation

We propose a reference signal tracking method to satisfy the excitation conditions (see Åström and Wittenmark [1995], page 63-67). The cost functions of decision makers need to be modified as follows:

$$J^1 = \sum_{k=0}^{+\infty}\left[(x_k - r_k)^\top Q^1(x_k - r_k) + R^{11}(u_k^1 - v_k^1)^2 + R^{12}(u_k^2 - v_k^2)^2\right] \quad (17)$$

$$J^2 = \sum_{k=0}^{+\infty}\left[(x_k - r_k)^\top Q^2(x_k - r_k) + R^{21}(u_k^1 - v_k^1)^2 + R^{22}(u_k^2 - v_k^2)^2\right] \quad (18)$$

To let the problem be well-defined (the optimal costs for both player are bounded), we require that $B^1v_k^1 + B^2v_k^2 = r_{k+1} - Ar_k$. (If the cost functions are bounded, then $k \to +\infty \Rightarrow x_k \to r_k$ and $u_k^i \to v_k^i \Rightarrow B^1v_k^1 + B^2v_k^2 = r_{k+1} - Ar_k$.)

Denoting $\bar{x}_k = x_k - r_k$ and $\bar{u}_k^i = u_k^i - v_k^i$, we can rewrite system state equation (1) and cost functions (17)-(18) as

$$\bar{x}_{k+1} = A\bar{x}_k + B^1\bar{u}_k^1 + B^2\bar{u}_k^2$$

$$J^1 = \sum_{k=0}^{+\infty}\left(\bar{x}_k^\top Q^1\bar{x}_k + \bar{u}_k^{1\top}R^{11}\bar{u}_k^1 + \bar{u}_k^{2\top}R^{12}\bar{u}_k^2\right)$$

$$J^2 = \sum_{k=0}^{+\infty}\left(\bar{x}_k^\top Q^2\bar{x}_k + \bar{u}_k^{1\top}R^{21}\bar{u}_k^1 + \bar{u}_k^{2\top}R^{22}\bar{u}_k^2\right)$$

It is a standard simultaneous linear quadratic game problem with Nash strategies

$$\bar{u}_k^{1*} := \arg\min_{\bar{u}_k^1} J^1 = \gamma^1(\bar{x}_k) = L^1\bar{x}_k$$

$$\bar{u}_k^{2*} := \arg\min_{\bar{u}_k^2} J^2 = \gamma^2(\bar{x}_k) = L^2\bar{x}_k$$

where L^1 and L^2 are defined in (6)-(7). Then

$$u_k^{i*} = L^i\bar{x}_k + v_k^i = L^ix_k + v_k^i - L^ir_k = L^ix_k + p_k^1$$

Assumption 2. In this scenario, we assume that player 2 only needs to estimate the control gain L^1. Thus the excitation condition of the adaptive law in (13) is satisfied by choosing a proper reference input signal r_k.

4. TWO-SIDE ADAPTATION SCHEME

For the same linear quadratic infinite horizon simultaneous game defined in (1)-(9), we provide adaptation design, parameter convergence, and persistence of excitation for the two-side adaptation scheme, in which each player needs to estimate the control gain of its opponent. We assume that each player has access to its own cost function as well as the system dynamics, and does not know the parameters of the cost function of its opponent. i.e., player 1 knows Q^1 , R^{12}, R^{11}, A, B^1 and B^2 while player 2 has access to Q^2, R^{21}, R^{22}, A, B^1 and B^2. We also assume that each player knows in advance that its opponent will implement state feedback Nash strategies. So, player 2 (or player 1) will estimate the control gain L^1 (or L^2) first, and then calculate/estimate its own best response control gain L^2 (or L^1) by using FP concept.

4.1 Adaptation Design

Consider the system defined in (1), let $\hat{L}_k^{12}$ and $\hat{L}_k^{21}$ be, respectively, the estimate of the control gain L^2 and L^1 from the point view of player 1 and player 2. Player 1 then calculates the associated estimate $\hat{L}_k^{11}$ of its own control gain L^1 from $\hat{L}_k^{12}$. Similarly, player 2 calculates the associated estimate $\hat{L}_k^{22}$ of its own control gain L^2 from $\hat{L}_k^{21}$. The block diagram of indirect adaptive control system is shown in Fig. 2.

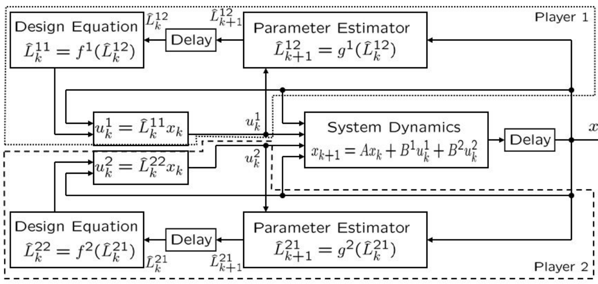

Fig. 2. Indirect adaptive control design for two-side adaptation scheme

First we have

$$\hat{x}_{k+1}^1 = Ax_k + B^1 \hat{L}_k^{11} x_k + B^2 \hat{L}_k^{12} x_k \tag{19}$$

$$\hat{x}_{k+1}^2 = Ax_k + B^1 \hat{L}_k^{21} x_k + B^2 \hat{L}_k^{22} x_k \tag{20}$$

$$x_{k+1} = Ax_k + B^1 \hat{L}_k^{11} x_k + B^2 \hat{L}_k^{22} x_k \tag{21}$$

where x_k is the real system state, $\hat{x}_k^1$ and $\hat{x}_k^2$ are the estimates of system state for player 1 and player 2, respectively. Since $B^{1\top} B^1$ and $B^{2\top} B^2$ are non-zero scalars, we obtain, from (19)-(21),

$$[B^{2\top} B^2]^{-1} B^{2\top} (\hat{x}_{k+1}^1 - x_{k+1}) = (\hat{L}_k^{12} - \hat{L}_k^{22}) x_k \tag{22}$$

$$[B^{1\top} B^1]^{-1} B^{1\top} (\hat{x}_{k+1}^2 - x_{k+1}) = (\hat{L}_k^{21} - \hat{L}_k^{11}) x_k \tag{23}$$

We introduce the estimation errors $e_k^1 = [B^{2\top} B^2]^{-1} B^{2\top} (\hat{x}_k^1 - x_k)$ and $e_k^2 = [B^{1\top} B^1]^{-1} B^{1\top} (\hat{x}_k^2 - x_k)$. ($e_k^i$ is a scalar) for player 1 and player 2, respectively, then

$$e_k^1 = \tilde{L}_{k-1}^2 \phi_k \tag{24}$$

$$e_k^2 = \tilde{L}_{k-1}^1 \phi_k \tag{25}$$

where $\tilde{L}_{k-1}^2 = \hat{L}_{k-1}^{12} - \hat{L}_{k-1}^{22}$, $\tilde{L}_{k-1}^1 = \hat{L}_{k-1}^{21} - \hat{L}_{k-1}^{11}$ and $\phi_k = \frac{1}{z}[x]_k = x_{k-1}$. (Note, $\frac{1}{z}[x]_k$ denotes the output of the system with transfer function $\frac{1}{z}$ and input x_k.)

Choose the adaptive law for $\hat{L}_k^{12}$ and $\hat{L}_k^{21}$ as

$$\hat{L}_k^{12} = \hat{L}_{k-1}^{12} - \frac{e_k^1 \phi_k^\top}{(m_k^1)^2} \Gamma_1 \tag{26}$$

$$\hat{L}_k^{21} = \hat{L}_{k-1}^{21} - \frac{e_k^2 \phi_k^\top}{(m_k^2)^2} \Gamma_2 \tag{27}$$

where $m_k^1 = \sqrt{\kappa_1 + \phi_k^\top \phi_k}$, $m_k^2 = \sqrt{\kappa_2 + \phi_k^\top \phi_k}$, $\kappa_1 > 0$, $\kappa_2 > 0$, $0 < \Gamma_1 = \Gamma_1^\top \leq I$, and $0 < \Gamma_2 = \Gamma_2^\top \leq I$. Convergence is proved and discussed in Section 4.2.

As shown in Fig. 2, for any $\hat{L}_k^{12}$ (or $\hat{L}_k^{21}$), a design function or mapping will be used to calculate the associated $\hat{L}_k^{11}$ (or $\hat{L}_k^{22}$). From (6)-(9), we have a best response strategy for player 1 (or player 2) given the estimated control gain $\hat{L}_k^{12}$ (or $\hat{L}_k^{21}$) of player 2 (or player 1) at time k,

$$\hat{L}_k^{11} = -[R^{11} + B^{1\top} \hat{K}_k^1 B^1]^{-1} B^{1\top} \hat{K}_k^1 [A + B^2 \hat{L}_k^{12}] \tag{28}$$

$$\hat{L}_k^{22} = -[R^{22} + B^{2\top} \hat{K}_k^2 B^2]^{-1} B^{2\top} \hat{K}_k^2 [A + B^1 \hat{L}_k^{21}] \tag{29}$$

where

$$\begin{aligned}\hat{K}_k^1 = {} & Q^1 + \hat{L}_k^{11\top} R^{11} \hat{L}_k^{11} + \hat{L}_k^{12\top} R^{12} \hat{L}_k^{12} \\ & +(A + B^1 \hat{L}_k^{11} + B^2 \hat{L}_k^{12})^\top \hat{K}_k^1 (A + B^1 \hat{L}_k^{11} + B^2 \hat{L}_k^{12}) \\ \hat{K}_k^2 = {} & Q^2 + \hat{L}_k^{21\top} R^{21} \hat{L}_k^{21} + \hat{L}_k^{22\top} R^{22} \hat{L}_k^{22} \\ & +(A + B^1 \hat{L}_k^{21} + B^2 \hat{L}_k^{22})^\top \hat{K}_k^2 (A + B^1 \hat{L}_k^{21} + B^2 \hat{L}_k^{22})\end{aligned}$$

By Lemma 1, we can prove that $\hat{K}^1$ and $\hat{K}^2$ satisfy the following discrete time algebraic Riccati equations, respectively,

$$\begin{aligned}\bar{A}_1^\top \hat{K}_k^1 \bar{A}_1 - \hat{K}_k^1 - \bar{A}_1^\top \hat{K}_k^1 B^1 [B^{1\top} \hat{K}_k^1 B^1 & \\ +R^{11}]^{-1} B^{1\top} \hat{K}_k^1 \bar{A}_1 + \bar{Q}^1 = 0 & \quad (30) \\ \bar{A}_2^\top \hat{K}_k^2 \bar{A}_2 - \hat{K}_k^2 - \bar{A}_2^\top \hat{K}_k^2 B^2 [B^{2\top} \hat{K}_k^2 B^2 & \\ +R^{22}]^{-1} B^{2\top} \hat{K}_k^2 \bar{A}_2 + \bar{Q}^2 = 0 & \quad (31)\end{aligned}$$

where $\bar{A}_1 = A + B^2 \hat{L}_k^{12}$, $\bar{Q}^1 = Q^1 + \hat{L}_k^{12\top} R^{12} \hat{L}_k^{12}$, $\bar{A}_2 = A + B^1 \hat{L}_k^{21}$, and $\bar{Q}^2 = Q^2 + \hat{L}_k^{21\top} R^{21} \hat{L}_k^{21}$.

Remark 3. We may obtain $\hat{K}_k^1$ and $\hat{K}_k^2$ by MATLAB command `dare(`$\bar{A}_1, B^1, \bar{Q}^1, R^{11}$`)` and `dare(`$\bar{A}_2, B^2, \bar{Q}^2, R^{22}$`)`, respectively.

4.2 Parameter Convergence and Persistency of Excitation

To satisfy the excitation conditions (see Åström and Wittenmark [1995], page 63-67), we propose a reference signal tracking method, in which the cost functions of decision makers need to be modified as specified in (17)-(18). As mentioned in the one-side adaptation scheme (see Section 3.2), we also require that $B^1 v_k^1 + B^2 v_k^2 = r_{k+1} - Ar_k$. (If the cost functions are bounded, then $k \to +\infty \Rightarrow x_k \to r_k$ and $u_k^i \to v_k^i \Rightarrow B^1 v_k^1 + B^2 v_k^2 = r_{k+1} - Ar_k$.)

It is already verified in Section 3.2 that under the perfect information structure the Nash strategies are

$$\bar{u}_k^{1*} := \arg\min_{\bar{u}_k^1} J^1 = \gamma^1(\bar{x}_k) = L^1 \bar{x}_k = L^1 x_k + p_k^1$$

$$\bar{u}_k^{2*} := \arg\min_{\bar{u}_k^2} J^2 = \gamma^2(\bar{x}_k) = L^2 \bar{x}_k = L^2 x_k + p_k^2$$

where L^1 and L^2 are defined in (6)-(7).

Assumption 3. In this scenario, we assume that each player only needs to estimate the control gain of its opponent. Thus the excitation conditions of the adaptive laws in (26)-(27) are satisfied by choosing a proper reference input signal r_k.

From Lemma 1, we know that for any $\hat{L}_k^{12}$ (or $\hat{L}_k^{21}$), we can obtain the unique positive semi-define solution $\hat{K}_k^1$ (or $\hat{K}_k^2$), then we can calculate $\hat{L}_k^{11}$ (or $\hat{L}_k^{22}$) by using equation (28) or (29). Let us define two functions as follows:

$$f^1 : \hat{L}_k^{12} \longmapsto \hat{L}_k^{11} \tag{32}$$

$$f^2 : \hat{L}_k^{21} \longmapsto \hat{L}_k^{22} \tag{33}$$

In general, the functions operate on row vectors.

Definition 1. Similar to the "reaction curve" concept of game theory, we define two properties as

$$\text{Property I:} \begin{cases} \|f^1(f^2(a)) - L^1\|_2 \le \|a - L^1\|_2 \\ \|f^2(f^1(b)) - L^2\|_2 \le \|b - L^2\|_2 \end{cases}$$

$$\text{Property II:} \begin{cases} \|a_1 - L^1\|_2 \le \|a_2 - L^1\|_2 \quad \Rightarrow \\ \qquad \|f^2(a_1) - L^2\|_2 \le \|f^2(a_2) - L^2\|_2 \\ \|b_1 - L^2\|_2 \le \|b_2 - L^2\|_2 \quad \Rightarrow \\ \qquad \|f^1(b_1) - L^1\|_2 \le \|f^1(b_2) - L^1\|_2 \end{cases}$$

where $\|a\|_2$ is the standard Euclidean norm. For row vector a, $\|a\|_2 = \sqrt{aa^\top}$. The equalities in Property I hold only when $a = L^1$ and $b = L^2$.

For the scalar case, the reaction-curve-like relation between f^1 and f^2 is shown in Fig. 3.

Remark 4. Property I means that each player will use its reaction chance to improve its strategy in the sense that the distance from the current control gain to the optimal control gain is reduced at each reaction time.

Remark 5. Property II means that the distance (from the current control gain to the optimal control gain) relation between the two players is positive.

Theorem 1. For the system defined in (1) with objective functions in (17)-(18), if the associated coupled discrete time algebraic Riccati equation specified in (6)-(9) has a unique positive semi-definite solution, then the two-side adaptation with adaptive laws in (26)-(27) and design equations in (28)-(29) will guarantee convergence, i.e.,

$$\lim_{k\to+\infty} \hat{L}_k^{ij} = L^j$$

where $i = 1, 2$ and $j = 1, 2$.

5. NUMERICAL SIMULATIONS

Consider the system defined in (1) - (3) with

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}, Q^1 = \begin{bmatrix} 1.5 & 0 \\ 0 & 1.5 \end{bmatrix}, Q^2 = \begin{bmatrix} 1.6 & 0 \\ 0 & 2.1 \end{bmatrix}$$

$B^1 = [1.3, 1.7]^\top$, $B^2 = [1.5, 1.6]^\top$, $R^{11} = 0.4$, $R^{12} = 1.0$, $R^{21} = 1.0$, and $R^{22} = 0.5$. We can calculate K^1, K^2, L^1 and L^2 from the coupled discrete-time algebraic Riccati equation specified in (6)-(9),

$$K^1 = \begin{bmatrix} 2.3609 & 1.0957 \\ 1.0957 & 3.4812 \end{bmatrix}, K^2 = \begin{bmatrix} 2.5639 & 1.2059 \\ 1.2059 & 4.2965 \end{bmatrix}$$

$L^1 = [-0.4992, -1.1682]$ and $L^2 = [-0.5215, -1.1295]$. We first carry out the two-side adaptation based on the adaptive laws (26)-(27) for the system with cost functions in (17)-(18). The parameters of adaptive laws are chosen

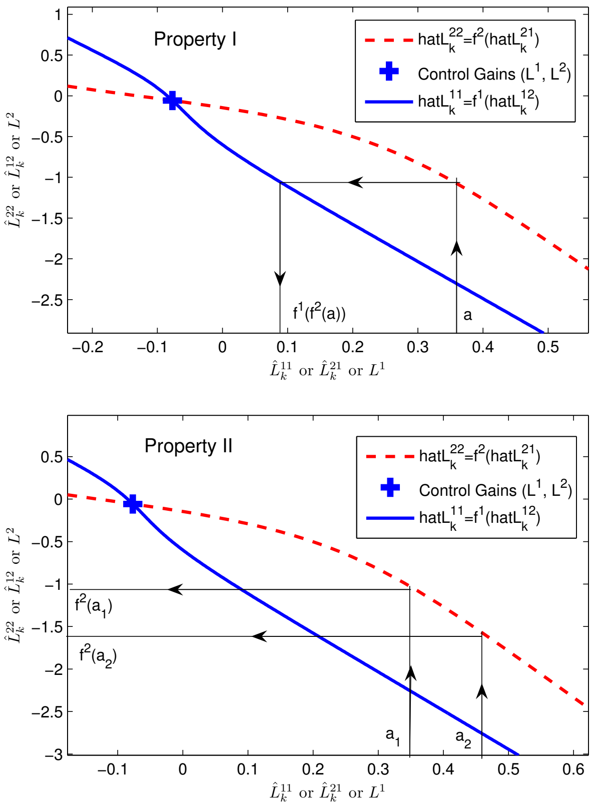

Fig. 3. Reaction-curve-like relation between function $\hat{L}_k^{11} = f^1(\hat{L}_k^{12})$ and $\hat{L}_k^{22} = f^2(\hat{L}_k^{21})$. Since we assume that there exist a unique positive semi-definite solution to the coupled discrete time algebraic Riccati equations in (6)-(9), there is only one cross point (marked by a green cross in this plot), which is composed of the real control gains L^2 and L^1 of Player 2 and Player 1, respectively.

as $\Gamma_1 = 0.8I$, $\Gamma_2 = 0.9I$, $\kappa_1 = 1.1$ and $\kappa_2 = 1.7$. The reference signals are generated by

$$r_k = \begin{bmatrix} 1.7\sin(3.5k) \\ 3.5\cos(0.9k^2 + 0.4k + 0.1) \end{bmatrix}$$

As shown in Fig. 4, we can see that when $k \longrightarrow +\infty$,

$$\hat{L}_k^{12} \longrightarrow L^2, \quad \hat{L}_k^{22} \longrightarrow L^2, \quad \hat{L}_k^{21} \longrightarrow L^1,$$
$$\hat{L}_k^{11} \longrightarrow L^1, \quad \hat{x}_k^2 \longrightarrow x_k, \quad \text{and } \hat{x}_k^2 \longrightarrow x_k.$$

We can verify the relations specified in (??)-(??) from Fig. 5, which shows that $\max(\|\hat{L}_k^{11} - L^1\|_2, \|\hat{L}_k^{21} - L^1\|_2)$ and $\max(\|\hat{L}_k^{22} - L^2\|_2, \|\hat{L}_k^{12} - L^2\|_2)$ are non-increasing sequences.

6. CONCLUSIONS

In this paper, we focused on the adaptivity analysis of the infinite-horizon perfect-state two-player linear quadratic games under a FP framework, in which both players will apply Nash strategies based on the available information set. For each player, we designed an indirect adaptive method to estimate the control gain of its opponent, using the concept of Fictitious Plays. With stabilizability and observability conditions, we proved that given the estimated control gain, each player can always calculate

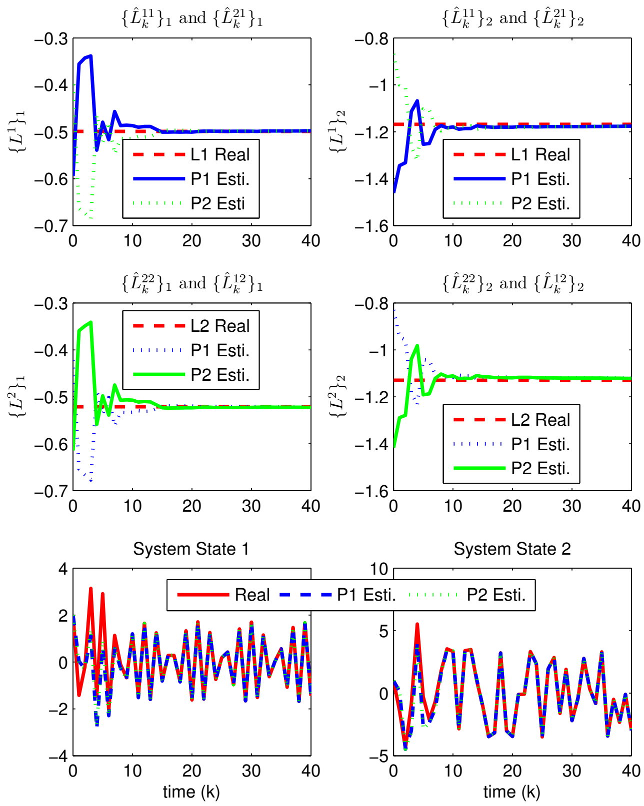

Fig. 4. $\hat{L}^{ij}$, L^j, x_k and $\hat{x}_k^i$ in two-side adaptation for a simultaneous game: reference signal tracking case. $\{\hat{L}_k^{ij}\}_s$ is the s^{th} element of $\hat{L}_k^{ij}$. The reference signal is $r_k = [1.7\sin(3.5k), \;\; 3.5\cos(0.9k^2+0.4k+0.1)]^\top$. We can see the convergence of $\hat{L}_k^{ij}$ to L^j.

his best response from the associated discrete-time algebraic Riccati equations. Convergence was proved and the excitation conditions were satisfied as well. Finally, numerical examples were simulated to illustrate our adaptive approaches for the two-side adaptive schemes.

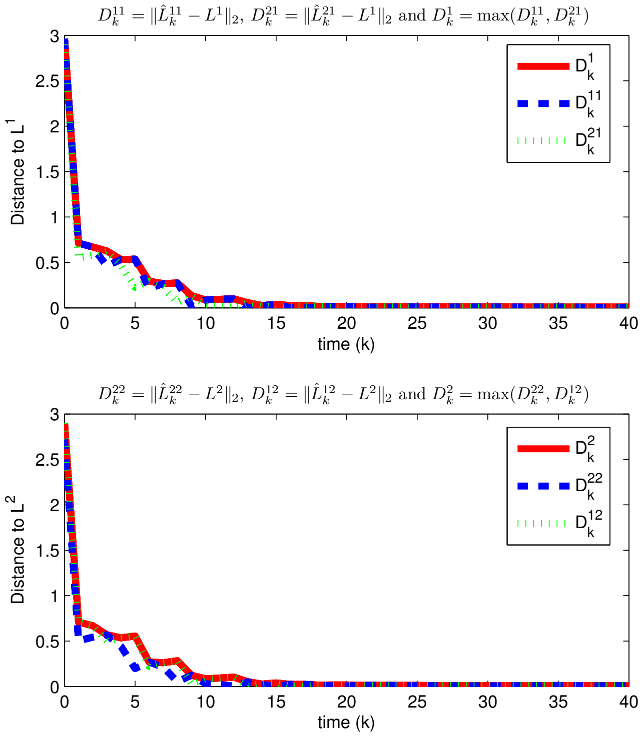

Fig. 5. As proved in Theorem 1, $\max(\|\hat{L}_k^{11} - L^1\|_2, \|\hat{L}_k^{21} - L^1\|_2)$ and $\max(\|\hat{L}_k^{22} - L^2\|_2, \|\hat{L}_k^{12} - L^2\|_2)$ are theoretically non-increasing sequences. In our simulation, due to the calculation accuracy, we found several increasing points: $D_{11}^1 = 0.0861$, $D_{12}^1 = 0.0944$, and $D_{11}^2 = 0.0838$, $D_{12}^2 = 0.0931$.

REFERENCES

J. Alster and P. R. Belanger. A technique of dual adaptive control. *Automatica*, 10:627–634, 1974.

K. J. Åström and B. Wittenmark. *Adaptive Control.* Addison-Wesley Series in Electrical Engineering: Control Engineering. Addison-Wesley, second edition, 1995.

K. J. Åström and B. Wittenmark. On self-tuning regulators. *Automatica*, 9:185–199, 1973.

T. Başar and G. J. Olsder. *Dynamic Noncooperative Game Theory.* SIAM Series in Classics in Applied Mathematics. Philadelphia, second edition, January, 1999.

Y. Bar-Shalom and E. Tse. Dual effect, certainty equivalence and separation in stochastic control. *IEEE Trans. Automat. Contr.*, AC-19:494–500, 1974.

G. W. Brown. Iterative solutions of games by fictitious play. In T. C. Koopmans, editor, *Activity Analysis of Production and Allocation.* Wiley, New York, 1951.

D. Fudenberg and D. K. Levine. *The Theory of Learning in Games.* MIT Press, Cambridge, 1998.

G. C. Goodwin, P. J. Ramadge, and P. E. Caines. Discrete time multivariable control. *IEEE Trans. Automat. Contr.*, 25:449–456, 1980.

G. Kreisselmeier and B. Anderson. Robust model reference adaptive control. *IEEE Trans. Automat. Contr.*, 31: 127–133, 1986.

A. S. Morse. Global stability of parameter-adaptive control systems. *IEEE Trans. Automat. Contr.*, 25:433–439, 1980.

K. S. Narendra, Y. H. Lin, and L. S. Valavani. Stable adaptive controller design, part ii: Proof of stability. *IEEE Trans. Automat. Contr.*, 25:440–448, 1980.

B. B. Peterson and K. S. Narendra. Bounded error adaptive control. *IEEE Trans. Automat. Contr.*, 27(6): 1161–1168, 1982.

John Maynard Smith. *Evolution and the Theory of Games.* Cambridge University Press, 1982.

G. Tao. *Adaptive Control Design and Analysis.* Adaptive and Learning Systems for Signal Processing, Communications and Control Series. Wiley-Interscience, Hoboken, N.J., 2003.

W. M. Wonham. On a matrix Riccati eqution of stochastic control. *SIAM Journal on Control*, 6(4):681–697, 1968.