Stochastic Adaptive Linear Quadratic Differential Games

Nian Liu and Lei Guo, *Fellow, IEEE*

***Abstract*— Game theory is playing more and more important roles in understanding complex systems and in investigating intelligent machines with various uncertainties. As a starting point, we consider the classical two-player zero-sum linear-quadratic stochastic differential games, but in contrast to most of the existing studies, the coefficient matrices of the systems are assumed to be unknown to both players, and consequently it is necessary to study adaptive strategies of the players, which may be termed as adaptive games and which has rarely been explored in the literature. In this paper, by introducing a suitable information structure for adaptive games, we will show that a theory can be established on adaptive strategies that are designed based on both the certainty equivalence principle and the diminishing excitation technique. Under almost the same physical structure conditions as those in the traditional known parameters case, it is shown that the closed-loop adaptive game systems will be globally stable and asymptotically reach the Nash equilibrium.**

***Index Terms*— Zero-sum games, uncertain parameters, adaptive strategy, least-squares, Nash equilibrium, stochastic differential games.**

I. Introduction

COMPLEX systems are currently at the research frontiers of many fields in scientific technology, such as economic and social systems, biology and environmental systems, physical and engineering systems, and artificial intelligent systems. It is quite common that the components or subsystems of complex systems have game-like relationships, and the theory of differential games appears to be a useful tool in modeling and analyzing conflicts in the context of dynamical systems.

The differential game theory was firstly introduced by Isaacs [1] in combat problems, and has been applied in many fields (see, e.g., [2]–[4]). A great deal of research effort has been devoted to the area in the past half a century and much progress has been made (see, e.g., [5]–[7]). In particular, the linear-quadratic differential games, which are described by linear systems and quadratic payoff functions, have attracted a lot of attention. Bernhard [8] gave necessary and sufficient conditions for the existence of a saddle point for deterministic two-player zero-sum differential games on a finite time interval. Starr and Ho [9] extended the zero-sum differential games to the general case, i.e., the players wish to minimize different performance criteria and they discussed three types of solutions. Differential game theory may also be used to study H^∞-optimal control problems [10], since this problem is actually a minimax optimization problem. However, in all the above-mentioned works, the parameters in the related mathematical models are assumed to be known to the players.

When the parameters in stochastic dynamical systems are unknown, there are a great deal of researches in the area of adaptive control and much progress has been made over the past half a century. A basic method in the design of adaptive control is called the certainty equivalence principle, which consists of two steps: firstly to use the observed information to get an estimate of the unknown parameters at each time instant, then to construct or update the controller by taking the estimate as "true" parameters at the same time. This design method is well-known to be quite powerful in dealing with dynamical systems with possible large uncertainties (see [11], [12]). However, since the closed-loop systems of adaptive control are usually described by a set of very complicated nonlinear stochastic dynamical equations, a rigorous theoretical investigation is well-known to be quite hard, even for linear uncertain stochastic systems. An example is the adaptive linear-quadratic-Gaussian control problem, where a key theoretical difficulty was how to guarantee the controllability of the online estimated model. This longstanding problem was reasonably resolved in the work [13] based on the self-convergence property of a class of weighted least squares and on a random regulation theory established in [14], which turn out to be the fundamental bases for solving the adaptive game problems in the current paper.

It goes without saying that uncertainties in the system structure, information and environment widely exist in dynamical games, and it is thus natural to consider adaptive game theory. To the best of our knowledge, only a little effort has been devoted to adaptive game theory, due to the complexity of the related theoretical investigation. For examples, Li and Guo [15] had considered a two-player zero-sum stochastic adaptive differential linear-quadratic game with state matrix to be known and stable. Yuan and Guo [16] investigated adaptive strategies for a zero-sum game described by an input-output stochastic model with known high gain parameters for both players. In a related but somewhat different context, reinforcement learning methods are also adopted to obtain the optimal strategies of players (see [17]–[22]). In [17]–[19], both players need to solve a least squares problem cooperatively and update their strategies respectively. In [20] and [21], a Stackelberg-like model is investigated, where the leader and follower update their strategies alternately with the leader first.

This paper was supported by the National Natural Science Foundational of China under Grant No.12288201.

The authors are with Institute of Systems Science, AMSS, Chinese Academy of Sciences, Beijing, 100190, China. (e-mails: liunian@amss.ac.cn, Lguo@amss.ac.cn).

The paper [20] needs to project the feedback gain of the leader to a known convex and compact set to ensure the stability of the system, and the paper [21] does not need such a projection but needs to assume that both players have an incentive to stabilize the system in the first place.

In this paper, we consider the problem of adaptive linear-quadratic zero-sum stochastic differential games, by using some of the powerful techniques developed in stochastic adaptive control but in a quite different and more complicated framework. First, a new information structure which describes a kind of complex situations involving both "competition and cooperation" will be introduced in the paper, where "cooperation" here implies that both players have a first priority to stabilize the game systems in the process of "competition". Second, a common adaptive parameter estimator will be designed and provided to both players with some guaranteed nice properties regardless of the players' strategies. Third, a theory will be established on adaptive strategies that are designed based on both the certainty equivalence principle and the diminishing excitation technique. To be specific, it will be shown that the closed-loop adaptive game systems will be globally stable and asymptotically reach the Nash equilibrium.

The remainder of the paper is organized as follows: In Section II, the design procedure of the adaptive strategies is provided and the main results on global stability, convergence of the estimate and Nash equilibrium of the closed-loop game systems are presented. Section III gives the proofs of above theorems and Section IV concludes this paper.

II. PROBLEM FORMULATION AND MAIN RESULTS

A. Problem Formulation

Consider the following basic stochastic linear quadratic zero-sum differential game:

$$dx(t) = (Ax(t) + B_1u_1(t) + B_2u_2(t))dt + Ddw(t), \quad (1)$$

where $x(t) \in \mathbb{R}^n$ is the state with the initial state $x(0) = x_0$, $u_i(t) \in \mathbb{R}^{m_i}$ is the strategy of Player i $(i = 1, 2)$, $(w(t), \mathscr{F}_t; t \geq 0)$ is a $\mathbb{R}^p$-valued standard Wiener process, and $A \in \mathbb{R}^{n\times n}, B_i \in \mathbb{R}^{n\times m_i}, D \in \mathbb{R}^{n\times p}$ are system matrices.

The payoff function is :

$$J(u_1, u_2) = \limsup_{T\to\infty} \frac{1}{T} \int_0^T (x^\mathsf{T}Qx + u_1^\mathsf{T}R_1u_1 - u_2^\mathsf{T}R_2u_2)dt, \quad (2)$$

where $Q = Q^\mathsf{T}, R_1 = R_1^\mathsf{T} > 0, R_2 = R_2^\mathsf{T} > 0$ are given weighting matrices. The objective of Player 1 is to minimize the payoff function, while Player 2 wants to maximize it.

It is well-known that the information structure and the timing of actions play a crucial role in games (see [7]). The basic information structure of the above zero-sum game is summarized as follows:

1) The system matrices $\{A, B_1, B_2, D\}$ are unknown to both players. But, the weighting matrices $\{Q, R_1, R_2\}$ and the state $x(t)$ are "common" knowledge at time instant t.

2) The strategies $u_i(t), i = 1, 2$ are of adaptive patterns, i.e., $u_i(t)$ is adapted to $\{\mathscr{F}_t; t \geq 0\}$ where $\mathscr{F}_t$ is a known non-decreasing filtration containing the filtration $\mathscr{F}_t^x$ generated by the state process $x(t)$, i.e., $\mathscr{F}_t^x \triangleq \sigma(x(s); 0 \leq s \leq t)$.

3) Neither of the players knows its opponent's strategy, and we assume that a common adaptive parameter estimator can be provided to both players.

4) Both players are of rationality in the sense that they have a first priority in stabilizing the game system and know about some basic principles in improving their respective strategies.

Remark 1: In the information structure 2), the strategies are usually referred to as feedback patterns when $\mathscr{F}_t = \sigma\{x(t)\}$. The general concept of adaptive pattern introduced here will make the design more flexible to include, e.g., "exploration" signals in the adaptive strategies. Since neither of the players knows its opponent's strategy as assumed in the information structure 3), it is hard for the players to estimate their opponents' input matrices without the assistance of a common estimator. Moreover, the information structure 4) on securing the stability of the game systems by both players is similar to that assumed in the non-adaptive deterministic case in [6]. In this sense, the information structure stated above describes a kind of complex situations involving both "competition and cooperation" rather than the scenario of pure "competition", which exists widely in social and economic systems. In fact, the players who are involved in competition must exist in the same system, and the breakdown of the system will not be beneficial to any players in general.

Definition 1: A pair of strategies $(u_1(t), u_2(t))$ are said to be admissible if they are adapted to $\{\mathscr{F}_t; t \geq 0\}$ and under which the following properties hold for any initial state $x(0)$:

$$\limsup_{T\to\infty} \frac{1}{T} \int_0^T (|x(t)|^2 + |u_1(t)|^2 + |u_2(t)|^2)dt < \infty \text{ and}$$

$$\lim_{T\to\infty} \frac{|x(T)|^2}{T} = 0 \quad a.s.$$

Definition 2: For the zero-sum linear-quadratic differential game (1)-(2) with both players in the adaptive pattern, a pair of admissible strategies (u_1^*, u_2^*) are called to attain an adaptive Nash equilibrium if they satisfy

$$J(u_1^*, u_2) \leq J(u_1^*, u_2^*) \leq J(u_1, u_2^*),$$

for any admissible pairs (u_1^*, u_2) and (u_1, u_2^*).

It is well-known that if the algebraic Riccati equation (ARE)

$$A^\mathsf{T}P + PA + Q - PB_1R_1^{-1}B_1^\mathsf{T}P + PB_2R_2^{-1}B_2^\mathsf{T}P = 0 \quad (3)$$

admits a real symmetric solution P such that $A-(B_1R_1^{-1}B_1^\mathsf{T}-B_2R_2^{-1}B_2^\mathsf{T})P$ is stable, then the following pair of strategies constitute a feedback Nash equilibrium [6]:

$$u_i(t) = (-1)^iR_i^{-1}B_i^\mathsf{T}Px(t), i = 1, 2. \quad (4)$$

It is worth mentioning that such solution P is called the stabilizing solution to ARE (3) and it is unique.

Now, we introduce a class of matrices defined by

$$\mathscr{L} \triangleq \left\{ L \triangleq \begin{pmatrix} L_1 \\ L_2 \end{pmatrix} | A + B_1L_1 + B_2L_2 \text{ is stable} \right\},$$

and we need the following notations:

$$B = [B_1, B_2], \quad R = diag(R_1, -R_2), \quad (5)$$

$$G(s) = R + B^\mathsf{T}(-sI - A^\mathsf{T})^{-1}Q(sI - A)^{-1}B, \quad (6)$$

$$N_L(s) = I + L(sI - A - BL)^{-1}B, \tag{7}$$

$$\tilde{G}_L(s) = N_L^{\mathsf{H}}(s)G(s)N_L(s), \tag{8}$$

where A^{H} denotes the conjugate transpose of A.

Definition 3: A matrix function $G(s)$ is called antianalytic factorizable if there exist matrix functions $\Xi(s)$ and $\Omega(s)$ such that $G(j\omega) = \Xi(-j\omega)\Omega(j\omega)$, where $\Xi, \Xi^{-1}, \Omega, \Omega^{-1}$ are all proper rational matrix functions without poles on the imaginary axis.

In the spectral theory of linear-quadratic optimal control, the above factorization has been well discussed and more details may be found in [23] and [24].

Definition 4: Assume (A, B) is stabilizable. We will say that $G(s)$ is antianalytic prefactorizable (see [25]) if there exists $L \in \mathscr{L}$ such that $\tilde{G}_L(s)$ is antianalytic factorizable.

Proposition 1 [25]: The algebraic Riccati equation (3) has a stabilizing solution and R is nonsingular if and only if the pair (A, B) is stabilizable and the matrix function $G(s)$ is antianalytic prefactorizable.

Then, the following basic assumptions are made.

A1) the pair (A, B) is controllable.

A2) the matrix function $G(s)$ defined by (6) is antianalytic prefactorizable.

Remark 2: Assumption A1) and A2) ensure that the ARE (3) has a stabilizing solution. Unfortunately, the corresponding strategies (4) are not implementable, because the system matrices (A, B) are unknown to both players. To solve this problem, it is natural to resort to adaptive methods based on online estimation of the system matrices. The controllability assumption A1) also makes it possible to transform excitation properties from the input to the state signals, necessary for the convergence of the parameter estimation. These are the contents of the next subsections.

B. The WLS Estimation

To describe the estimation problem in the standard form, we first introduce the following notations:

$$\theta^{\mathsf{T}} = [A, B_1, B_2], \tag{9}$$

$$\varphi(t) = [x^{\mathsf{T}}(t), u_1^{\mathsf{T}}(t), u_2^{\mathsf{T}}(t)]^{\mathsf{T}}, \tag{10}$$

and rewrite the system (1) into the following compact form:

$$dx(t) = \theta^{\mathsf{T}}\varphi(t)dt + Ddw(t). \tag{11}$$

Now the continuous-time weighted least-square (WLS) estimates $(\theta(t), t \geq 0)$ are given by [13]

$$d\theta(t) = a(t)Q(t)\varphi(t)[dx^{\mathsf{T}}(t) - \varphi^{\mathsf{T}}(t)\theta(t)dt], \tag{12}$$

$$dQ(t) = -a(t)Q(t)\varphi(t)\varphi^{\mathsf{T}}(t)Q(t)dt, \tag{13}$$

where the initial conditions $Q(0) > 0$ and $\theta^{\mathsf{T}}(0) = [A(0), B_1(0), B_2(0)]$ are arbitrary deterministic values such that $(A(0), B(0))$ is controllable with $B(0) = [B_1(0), B_2(0)]$, $a(t) = 1/\log^{(1+\delta)} r(t), \delta > 0$ is a constant and

$$r(t) = \|Q^{-1}(0)\| + \int_0^t |\varphi(s)|^2 ds. \tag{14}$$

Lemma 1 [13]: Let $(\theta(t), t \geq 0)$ be defined by (12)-(13). Then the following properties are satisfied:

$$(1) \sup_{t\geq 0} \|Q^{-\frac{1}{2}}(t)\tilde{\theta}(t)\|^2 < \infty \quad a.s.$$

$$(2) \int_0^\infty a(t)\|\tilde{\theta}^{\mathsf{T}}(t)\varphi(t)\|^2 dt < \infty \quad a.s.$$

$$(3) \lim_{t\to\infty} \theta(t) = \bar{\theta} \quad a.s.$$

where $\tilde{\theta}(t) = \theta(t) - \theta$ and $\bar{\theta}$ is a finite random variable.

C. Regularization

To construct the adaptive version of the ARE (3) by using the estimates $(A(t), B(t))$ with guaranteed solvability, one needs at least the stabilizability of the estimates $(A(t), B(t))$ which may not be provided by the above WLS algorithm. To solve this problem, We resort to the regularization method introduced in [14] to modify the WLS estimates to ensure their uniform controllablility. We first introduce the following definition in [13]:

Definition 5: A family of system matrices $(A(t), B(t); t \geq 0)$ is said to be uniformly controllable if there is a constant $c > 0$ such that

$$\sum_{i=0}^{n-1} A^i(t)B(t)B^{\mathsf{T}}(t)A^{i\mathsf{T}}(t) \geq cI,$$

for all $t \in [0, \infty)$, where $A(t) \in \mathbb{R}^{n\times n}, B(t) \in \mathbb{R}^{n\times m}$.

By Lemma 1(3), it is known that $\theta(t)$ converges to a certain random matrix $\bar{\theta}$ which may not be the true parameter matrix θ and naturally, the controllability of the estimate models may not be guaranteed. To solve this problem, we observe that by Lemma 1(1), the matrix sequence $\{Q^{-\frac{1}{2}}(t)(\theta - \theta(t)),\ t \geq 0\}$ is bounded. In other words, there exists a bounded random sequence $\{\beta^*(t), t \geq 0\}$ such that

$$\theta = \theta(t) - Q^{\frac{1}{2}}(t)\beta^*(t).$$

Note that (A, B) is controllable, this inspires the following modification for getting controllable estimated models:

$$\theta(t, \beta(t)) = \theta(t) - Q^{\frac{1}{2}}(t)\beta(t),$$

where $\beta(t) \in \mathbb{R}^{(n+m_1+m_2)\times n}$ is a sequence of bounded matrices to be defined shortly. For simplicity, we denote

$$\theta^{\mathsf{T}}(t, \beta(t)) = [\bar{A}(t), \bar{B}(t)], \quad \bar{B}(t) = [\bar{B}_1(t), \bar{B}_2(t)].$$

To guarantee the uniform controllability of $(\bar{A}(t), \bar{B}(t))$, we need only to select the sequence $\{\beta(t), t \geq 0\}$ to guarantee the uniform positivity of $Y(t)$ defined by

$$Y(t, \beta(t)) = \det\Big(\sum_{i=0}^{n-1} \bar{A}^i(t)\bar{B}(t)\bar{B}^{\mathsf{T}}(t)\bar{A}^{i\mathsf{T}}(t)\Big).$$

For this purpose, we proceed to choose a suitable process $\{\beta(t), t \geq 0\}$ to prevent $Y(t, \beta(t))$ from being close to zero. We adopt a method inspired by that in random optimization [13]. Let $\{\eta_k \in \mathbb{R}^{(n+m_1+m_2)\times n}, k \in \mathbb{N}\}$ be a sequence of independent random variables which are uniformly distributed in the unit ball for a norm of the matrices and are also

independent of $(w(t), t \geq 0)$. The procedure of choosing $\beta(k)$ is recursively given by the following:

$$\beta(0) = 0,$$
$$\beta(k) = \begin{cases} \eta_k, & \text{if } Y(k, \eta_k) \geq (1+\gamma) Y(k, \beta(k-1)) \\ \beta(k-1), & \text{otherwise,} \end{cases} \tag{15}$$

where $\gamma \in (0, \sqrt{2}-1)$ is a fixed constant. Thus, a sequence of regularized estimates $(\bar{\theta}_k, k \in \mathbb{N})$ can be defined by

$$\bar{\theta}_k = \theta(k) - Q^{\frac{1}{2}}(k)\beta(k). \tag{16}$$

Finally, the continuous-time estimates used for the design of adaptive strategies can be defined piecewise as follows:

$$\hat{\theta}(t) = \bar{\theta}_k, \tag{17}$$

for any $t \in (k, k+1]$ and for all $k \in \mathbb{N}$.

The following lemma shows that the regularized estimates $(\hat{\theta}(t), t \geq 0)$ defined above do indeed ensure the uniform controllability of the estimated models, while keeping the nice properties of the WLS.

Lemma 2: Let A1) be satisfied for the game system (1)-(2). Then the family of regularized WLS estimates $(\hat{\theta}(t), t \geq 0)$ defined by (12)-(17) has the following properties.

(1) Self-convergence, that is, $\hat{\theta}(t)$ converges a.s. to a finite random matrix as $t \to \infty$.

(2) The family $(A(t), B(t); t \geq 0)$ is uniformly controllable where $[A(t), B(t)] = \hat{\theta}^{\mathsf{T}}(t)$.

(3) Semi-consistency, that is,

$$\int_0^t |(\hat{\theta}(s) - \theta)^{\mathsf{T}} \varphi(s)|^2 ds = o(r(t)) + O(1),$$

where $r(t)$ is defined in (14).

The proof is similar to Lemma 2 of [13].

D. The Main Result

For simplicity, we rewrite the estimates given by (17) as

$$\hat{\theta}^{\mathsf{T}}(t) = [A(t), B_1(t), B_2(t)].$$

For any $k \in \mathbb{N}$, it is well-known that the following ARE:

$$A^{\mathsf{T}}(k)P(k) + P(k)A(k) + Q - P(k)S(k)P(k) = 0 \tag{18}$$

has at most one Hermitian matrix solution $P(k)$ such that

$$A_{cl}\big(P(k)\big) \triangleq A(k) - S(k)P(k) \tag{19}$$

is stable (see [28]), where

$$S(k) = B_1(k)R_1^{-1}B_1^{\mathsf{T}}(k) - B_2(k)R_2^{-1}B_2^{\mathsf{T}}(k).$$

We can rewrite such $P(k)$ as

$$P(k) = P_1(k) + P_2(k)j, \tag{20}$$

where $P_1(k)$ is a real symmetric matrix, $P_2(k)$ is a real skew-symmetric matrix and j is the imaginary unit with $j^2 = -1$.

Then, we construct the desired strategy pair by considering two cases separately.

Case (i): If the ARE (18) has a Hermitian matrix solution $P(k)$ such that both $A_{cl}\big(P(k)\big)$ and $A_{cl}\big(P_1(k)\big)$ are stable, then Player i $(i = 1, 2)$ can use the following strategies respectively:

$$u_i(t) = (-1)^i R_i^{-1} B_i^{\mathsf{T}}(k) P_1(k) x(t), \text{ for } t \in (k, k+1]. \tag{21}$$

Case (ii): If the ARE (18) does not admit any Hermitian matrix solution $P(k)$ such that both $A_{cl}\big(P(k)\big)$ and $A_{cl}\big(P_1(k)\big)$ are stable, then we choose the following strategy pair

$$(u_1^{\mathsf{T}}(t), u_2^{\mathsf{T}}(t))^{\mathsf{T}} = -B^{\mathsf{T}}(k)W_k^{-1}x(t), \text{ for } t \in (k, k+1], \tag{22}$$

where $W_k = \int_0^1 e^{-A(k)\tau} B(k) (e^{-A(k)\tau} B(k))^{\mathsf{T}} d\tau$ and $B(k) = [B_1(k), B_2(k)]$.

Remark 3: The strategy pair (21) is designed by the certainty equivalence principle, and the strategy pair (22) is designed to stabilize the system (1) whenever the ARE (18) does not have the desired solution. In the following section, one may see that the estimates will converge to the true parameters, and so the strategy pair $(u_1(t), u_2(t))$ will not take the form (22) when t is large enough.

By the well-known Fel'dbaum dual principle in optimal control, the optimal strategy should achieve a good balance between control and estimation for uncertain systems. A similar philosophy used in reinforcement learning is the trade-off between "exploitation and exploration". Inspired by this and following the ideas in [14], some diminishing excitation or exploration signals that are helpful for estimation but not essentially influence the control, are incorporated into the adaptive strategies of both players, i.e., for $t \in (k, k+1]$,

$$u_i^*(t) = L_i(k)x(t) + \gamma_k^{(i)}(v_i(t) - v_i(k)) \text{ or}$$
$$du_i^*(t) = L_i(k)dx(t) + \gamma_k^{(i)} dv_i(t), i = 1, 2 \tag{23}$$

where $[L_1(t), L_2(t)] = [-R_1^{-1}B_1^{\mathsf{T}}(k)P_1(k), R_2^{-1}B_2^{\mathsf{T}}(k)P_1(k)]$ in Case (i); $[L_1^{\mathsf{T}}(t), L_2^{\mathsf{T}}(t)]^{\mathsf{T}} = -B^{\mathsf{T}}(t)W_k^{-1}$ in Case (ii). The sequences $\{\gamma_k^{(i)}, k \in \mathbb{N}\}$, $i = 1, 2$ can be any sequences satisfying the following:

$$\lim_{k\to\infty} \gamma_k^{(1)} = \lim_{k\to\infty} \gamma_k^{(2)} = 0,$$
$$\min\{\gamma_k^{(1)}, \gamma_k^{(2)}\} \geq \gamma_k = (\frac{\log k}{\sqrt{k}})^{\frac{1}{2}}, \text{ with } \gamma_0 = 0.$$

For simplicity, we choose $\gamma_k^{(1)} = \gamma_k^{(2)} = \gamma_k$ in the rest of the paper. The processes $(v_1(t), t \geq 0)$ and $(v_2(t), t \geq 0)$ are chosen as sequences of independent standard Wiener processes that are independent of $(w(t), t \geq 0)$ and $(\eta_k, k \in \mathbb{N})$.

Now, we have the following theorems and the proofs will be given in the next section.

Theorem 1: Let A1) be satisfied. Then, under the adaptive strategies (23) of the players, the following properties hold:

1) The system (1) is globally stable in the sense that for any initial state $x(0)$,

$$\limsup_{T\to\infty} \frac{1}{T} \int_0^T |x(t)|^2 dt < \infty \quad a.s. \tag{24}$$

2) The parameter estimates $\hat{\theta}(t)$ adopted by the players are strongly consistent, i.e.,

$$\lim_{t\to\infty} \hat{\theta}(t) = \theta \quad a.s. \tag{25}$$

where θ is the true system parameter defined by (9).

Theorem 2: For the stochastic game system (1) with the payoff function (2), if A1) and A2) are satisfied, then the above adaptive strategies (23) constitute an adaptive Nash equilibrium. Moreover, we have

$$J(u_1^*, u_2^*) = tr(D^{\mathsf{T}} P D) \quad a.s. \tag{26}$$

where P is the stabilizing solution to the ARE (3).

III. Proof

We first present the following lemmas:

Lemma 3 [26]: If the system $\dot{x} = Ax + Bu$ is controllable, then under the control $u = -B^{\mathsf{T}} W^{-1}(0, T_0)x$, the closed-loop system is stable, where $W(0, T_0) = \int_0^{T_0} e^{-At} B (e^{-At} B)^{\mathsf{T}} dt$ and $T_0 > 0$ is any given constant.

Lemma 4 [13]: For the processes $(v_i(t), t \geq 0), i = 1, 2$ and the sequences $\{\gamma_k, k \in \mathbb{N}\}$ defined in (23), we have

$$\limsup_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \int_k^{k+1} \gamma_k^2 |v_i(t) - v_i(k)|^2 dt = 0 \quad a.s.$$

Lemma 5 [28]: With each triple $(A, S, Q) \in (\mathbb{C}^{n\times n})^3$ satisfying $S = S^{\mathsf{H}}$ and $Q = Q^{\mathsf{H}}$ to be piecewise continuous and locally bounded matrix functions $A, S, Q : \mathbb{D} \to \mathbb{C}^{n\times n}$ on some interval $\mathbb{D} \in \mathbb{R}$, we associate the matrix function

$$E = \begin{pmatrix} Q & A^{\mathsf{H}} \\ A & -S \end{pmatrix} : \mathbb{D} \to \mathbb{C}^{2n\times 2n},$$

and the ARE

$$A^{\mathsf{H}} P + PA + Q - PSP = 0, \tag{27}$$

where A^{H} denotes the conjugate transpose of A. Assume that for some $E = E_0$ there is a stabilizing solution P_0 for which (27) is fulfilled. Then there exists $r(E_0) > 0$ such that for E ranging $||E - E_0|| < r(E_0)$, there is a unique analytic function $E \mapsto P(E)$ such that $A - SP(E)$ is stable and $P(E)$ is a Hermitian matrix solution to the equation (27) satisfying $P(E_0) = P_0$.

Lemma 6 [29]: Let the processes $(x(t) \in \mathbb{R}^n, t \geq 0)$ *and* $(V(t) \in \mathbb{R}^{n\times n}, t \geq 0)$ *satisfy*

$$\limsup_{T\to\infty} \frac{1}{T} \int_0^T |x(t)|^2 dt < \infty \quad a.s. \tag{28}$$

and

$$\lim_{t\to\infty} V(t) = 0 \quad a.s. \tag{29}$$

then, the following equality is true:

$$\lim_{t\to\infty} \frac{1}{T} \int_0^T x^{\mathsf{T}}(t) V(t) x(t) dt = 0 \quad a.s. \tag{30}$$

A. Proof of Theorem 1:

By Lemma 2 1) and 2), there are random matrices $A(\infty)$ and $B_i(\infty)$ such that

$$\lim_{t\to\infty} A(t) = A(\infty) \quad a.s.$$

$$\lim_{t\to\infty} B_i(t) = B_i(\infty), i = 1, 2 \quad a.s.$$

and that $(A(\infty), B(\infty))$ is controllable where

$$B(\infty) = [B_1(\infty), B_2(\infty)].$$

For simplicity of the remaining descriptions, we denote

$$\Phi(t) = A(t) + B_1(t)L_1(t) + B_2(t)L_2(t). \tag{31}$$

where $L_1(t)$ and $L_2(t)$ are defined in (23).

We now proceed to verify that $\Phi(t)$ is uniformly stable and convergent. By the definition (23) and Lemma 3, we only need to consider the case where t is sufficiently large. For this, we consider the following two cases separately:

Case (1): If the ARE

$$A^{\mathsf{T}}(\infty)P(\infty) + P(\infty)A(\infty) - P(\infty)B_1(\infty)R_1^{-1}B_1^{\mathsf{T}}(\infty)P(\infty)$$
$$+ Q + P(\infty)B_2(\infty)R_2^{-1}B_2^{\mathsf{T}}(\infty)P(\infty) = 0 \tag{32}$$

has a Hermitian matrix solution $P(\infty)$ such that both $A_{cl}\big(P(\infty)\big)$ and $A_{cl}\big(P_1(\infty)\big)$ defined by (19)-(20) are stable, then by Lemma 5, there exist solutions $P(k)$ defined in (18) for all large enough k such that

$$\lim_{k\to\infty} P(k) = P(\infty) \quad a.s.$$

and that both $A_{cl}\big(P(k)\big)$ and $A_{cl}\big(P_1(k)\big)$ (by continuity) are stable, which means that when $k \in \mathbb{N}$ is large enough, $L_1(t)$ and $L_2(t)$ will take the following form for $t \in (k, k+1]$:

$$L_1(t) = -R_1^{-1}B_1^{\mathsf{T}}(k)P_1(k), \quad L_2(t) = R_2^{-1}B_2^{\mathsf{T}}(k)P_1(k).$$

Hence, $\Phi(t)$ is uniformly stable and convergent.

Case (2): In the case where the ARE (32) does not admit any Hermitian matrix solution $P(\infty)$ such that both $A_{cl}\big(P(\infty)\big)$ and $A_{cl}\big(P_1(\infty)\big)$ are stable, by the definition (23) and Lemma 5, $L_1(t)$ and $L_2(t)$ will take the following form for all sufficiently large k:

$$[L_1^{\mathsf{T}}(t), L_2^{\mathsf{T}}(t)]^{\mathsf{T}} = -B^{\mathsf{T}}(k)W_k^{-1}, \text{ for } t \in (k, k+1].$$

Then, by Lemma 3, we can easily see that $\Phi(t)$ is uniformly stable and convergent.

To summarize, we know that $\Phi(t)$ is uniformly stable and converges to a stable matrix a.s. Hence, by Lyapunov equation, there exist some uniformly bounded positive definite matrices $K(t)$ such that

$$\Phi^{\mathsf{T}}(t)K(t) + K(t)\Phi(t) = -I. \tag{33}$$

Next, we proceed to verify that

$$\sum_{k=0}^{N} |x(k)|^2 = O(N) + o(r(N)) \quad a.s. \tag{34}$$

where $r(t)$ is defined in (14). Note that for $t \in (k, k+1]$ and $k \in \mathbb{N}$, under adaptive strategies (23) the system (1) will be

$$\begin{aligned} dx(t) &= (Ax(t) + B_1u_1^*(t) + B_2u_2^*(t))dt + Ddw(t) \qquad (35)\\ &= (\Phi(t)x(t) + \delta(t) + \gamma_k(v(t) - v(k)))dt + Ddw(t), \end{aligned}$$

where $\delta(t) = (\theta - \hat{\theta}(t))^{\mathsf{T}}\varphi(t)$ and $v(t) = B_1v_1(t) + B_2v_2(t)$.

Then, it follows that

$$\begin{aligned} x(k+1) &= e^{\Phi(k)}x(k) + \int_k^{k+1} e^{(k+1-t)\Phi(k)}Ddw(t)\\ &+ \int_k^{k+1} e^{(k+1-t)\Phi(k)}(\delta(t) + \gamma_k(v(t) - v(k)))dt. \end{aligned}$$

Since $\Phi(k)$ is uniformly stable and convergent, by Cauchy-Schwarz inequality, it is easy to get

$$\begin{aligned} |x(k+1)|^2 &\le m|x(k)|^2 + m_1(\int_k^{k+1} e^{(k+1-t)\Phi(k)}Ddw(t))^2\\ &+m_2(\int_k^{k+1} |\delta(t)|^2dt + \int_k^{k+1} \gamma_k^2|v(t) - v(k)|^2dt), \end{aligned}$$

where $0 < m < 1$ and $m_1, m_2 > 0$ are some fixed constants. Then, it is easy to see that

$$\begin{aligned} &\frac{1}{N}(1-m)\sum_{k=1}^{N}|x(k)|^2\\ &= O(\frac{1}{N}\sum_{k=1}^{N}(\int_k^{k+1} e^{(k+1-t)\Phi(k)}Ddw(t))^2)+\\ &O(\frac{1}{N}\int_1^N |\delta(t)|^2dt) + O(\frac{1}{N}\sum_{k=1}^{N}\int_k^{k+1}\gamma_k^2|v(t) - v(k)|^2dt)\\ &= O(1) + o(\frac{1}{N}r(N)), \end{aligned}$$

where the first part can use Lemma 1 (Etemadi) of 5.2 in [27], the second part is the direct result of Lemma 2(3) and the third part can be estimated by the consequence of Lemma 4.

Finally, we proceed to prove Theorem 1. Applying the Ito's formula to $\langle K(t)x(t), x(t)\rangle$ where $\langle\cdot,\cdot\rangle$ represents the inner product, and noting that $K(t)$ defined in (33) is actually constant in any interval $t \in (i, i+1]$ and any $i \in \mathbb{N}$, it follows that

$$\begin{aligned} d\langle K(t)x(t), x(t)\rangle &= tr(K(t)DD^{\mathsf{T}})dt + 2\langle K(t)x(t), Ddw(t)\rangle\\ &+2\langle K(t)x(t), \Phi(t)x(t) + \delta(t) + \gamma_i(v(t) - v(i))\rangle dt, \end{aligned}$$

which in conjunction with equation (33) gives

$$\begin{aligned} d\langle K(t)x(t), x(t)\rangle + |x(t)|^2dt &= tr(K(t)DD^{\mathsf{T}})dt +\\ 2\langle K(t)x(t), Ddw(t)\rangle &+ 2\langle K(t)x(t), \delta(t) + \gamma_i(v(t) - v(i))\rangle dt. \end{aligned} \qquad (36)$$

For the second part on the right-hand-side, by the boundedness of $K(t)$, it follows from Lemma 12.3 of [11] that

$$|\int_0^t \langle K(t)x(t), Ddw(t)\rangle| = O\big([\int_0^t |x(t)|^2dt]^{\frac{1}{2}+\epsilon}\big), \qquad (37)$$

for any $\epsilon \in (0, 1/2)$.

By Lemma 2(3), it follows that

$$\int_0^t |\delta(s)|^2ds = o(r(t)) + O(1). \qquad (38)$$

Then, integrating the equation (36) over the interval $(0, T)$, and using (37)-(38), Lemma 4 and the Cauchy-Schwarz inequality, it follows that

$$\begin{aligned} &\sum_{i=0}^{\lfloor T\rfloor-1}\big(\langle K(i)x(i+1), x(i+1)\rangle - \langle K(i)x(i), x(i)\rangle\big)\\ &+\langle K(\lfloor T\rfloor)x(T), x(T)\rangle - \langle K(\lfloor T\rfloor)x(\lfloor T\rfloor), x(\lfloor T\rfloor)\rangle\\ &+\int_0^T |x(t)|^2dt = \int_0^T tr(K(t)DD^{\mathsf{T}})dt+\\ &O\big((\int_0^T |x(t)|^2dt)^{\frac{1}{2}+\epsilon}\big) + (\int_0^T |x(t)|^2dt)^{\frac{1}{2}}\cdot o(T^{\frac{1}{2}})+\\ &(\int_0^T |x(t)|^2dt)^{\frac{1}{2}}(o(r(T)) + O(1))^{\frac{1}{2}}, \end{aligned} \qquad (39)$$

where $\lfloor T\rfloor$ denotes the integer part of T.

Since $K(t)$ is uniformly bounded, we have

$$\begin{aligned} &\sum_{i=0}^{\lfloor T\rfloor-1}\big(\langle K(i)x(i+1), x(i+1)\rangle - \langle K(i)x(i), x(i)\rangle\big)\\ &+\langle K(\lfloor T\rfloor)x(T), x(T)\rangle - \langle K(\lfloor T\rfloor)x(\lfloor T\rfloor), x(\lfloor T\rfloor)\rangle\\ &= O(\sum_{i=0}^{\lfloor T\rfloor}|x(i)|^2) + \langle K(\lfloor T\rfloor)x(T), x(T)\rangle. \end{aligned} \qquad (40)$$

By Lemma 4 and the convergence of $[L_1(t), L_2(t)]$, it follows that

$$r(T) = \|P^{-1}(0)\| + \int_0^T |\varphi(s)|^2ds = O(\int_0^T |x(t)|^2dt). \qquad (41)$$

Finally, by (34) and (40)-(41), the equality (39) will be

$$\begin{aligned} &\langle K(\lfloor T\rfloor)x(T), x(T)\rangle + \int_0^T |x(t)|^2dt\\ &= O(T) + o(\int_0^T |x(t)|^2dt) + \int_0^T tr(K(t)DD^{\mathsf{T}})dt, \end{aligned} \qquad (42)$$

which implies the desired result of Theorem 1 1).

In order to prove the strong consistency of the WLS estimates, we need to verify the excitation condition on $\phi(t)$ needed for the convergence $\tilde{\theta}(t) \to 0$.

By Lemma 1(1), it follows that

$$\|\theta(t) - \theta\|^2 \le \|Q(t)\|\|Q^{-\frac{1}{2}}(t)(\theta(t) - \theta)\|^2 = O(\|Q(t)\|).$$

From (16)-(17), we need only to verify that $Q(k) \to 0$.

By the definition (13), it is easy to see that

$$\begin{aligned} Q(k) &= \Big(Q^{-1}(0) + \int_0^k a(s)\varphi(s)\varphi^{\mathsf{T}}(s)ds\Big)^{-1}\\ &\le \Big(Q^{-1}(0) + a(k)\int_0^k \varphi(s)\varphi^{\mathsf{T}}(s)ds\Big)^{-1}\\ &\le \Big(Q^{-1}(0) + \frac{M}{\log^{1+\delta}k}\int_0^k \varphi(s)\varphi^{\mathsf{T}}(s)ds\Big)^{-1}\\ &= O\left(\frac{\log^{1+\delta}k}{\lambda_{min}(\int_0^k \varphi(s)\varphi^{\mathsf{T}}(s)ds)}\right) \end{aligned} \qquad (43)$$

where we have used the fact that $r(k) = O(k)$ (see (41) and Theorem 1 1)), $M > 0$ is a constant and $\lambda_{min}(\cdot)$ denotes the minimum eigenvalue.

Next, we only need to verify that there exists some $c > 0$ such that for all sufficiently large k,

$$\lambda_{min}\Big(\int_0^k \varphi(s)\varphi^{\mathsf{T}}(s)ds\Big) \geq c\sqrt{k}. \tag{44}$$

Since (A, B) is controllable and $[L_1(t), L_2(t)]$ is convergent, the desired excitation condition (44) can be proved using the similar arguments as in the proof of Theorem 2 [13] (see (61) in [13]), details will not be repeated here. Hence, the proof is completed.

B. Proof of Theorem 2:

First, we verify that the adaptive strategy pair $(u_1^*(t), u_2^*(t))$ is admissible. By Theorem 1 1), we only need to verify that

$$\lim_{T\to\infty} \frac{|x(T)|^2}{T} = 0 \quad a.s. \tag{45}$$

By Theorem 1 2), it follows that

$$\lim_{t\to\infty} A(t) = A \quad a.s.$$
$$\lim_{t\to\infty} B_i(t) = B_i, i = 1, 2 \quad a.s.$$

Under Assumptions A1) and A2), the ARE (3) admits a unique stabilizing solution P (see Proposition 1). By Lemma 5, we know that when k is large enough, there a.s. exists a Hermitian matrix solution $P(k)$ to the ARE (18) such that $A_{cl}\big(P(k)\big)$ is stable. Consequently, for the real part and imaginary part of such $P(k)$, we have

$$\lim_{k\to\infty} P_1(k) = P_1(\infty) = P \text{ and } \lim_{k\to\infty} P_2(k) = 0 \quad a.s. \tag{46}$$

Hence, for any sample point w, there exists some sufficiently large $T_w > 0$ such that $\Phi(t) \triangleq A + B_1L_1(t) + B_2L_2(t)$ is stable for any $t \geq T_w$, where $[L_1(t), L_2(t)]$ is defined in (23).

For the system (1) with adaptive strategies (23), we have

$$\begin{aligned} dx(t) &= (Ax(t) + B_1u_1^*(t) + B_2u_2^*(t))dt + Ddw(t) \quad (47)\\ &= \big(\Phi(t)x(t) + \gamma_{\lfloor t\rfloor}(v(t) - v(\lfloor t\rfloor))\big)dt + Ddw(t), \end{aligned}$$

where $v(t) = B_1v_1(t) + B_2v_2(t)$ and $\lfloor t\rfloor$ denotes the integer part of t. Integrating the equation (47) over interval $[T_w, t]$, it follows that

$$\begin{aligned} x(t) = \Psi(t, T_w)x(T_w) &+ \int_{T_w}^t \Psi(t, \tau)\gamma_{\lfloor\tau\rfloor}(v(\tau) - v(\lfloor\tau\rfloor))d\tau\\ &+ \int_{T_w}^t \Psi(t, \tau)Ddw(\tau), \end{aligned}$$

where $d\Psi(t, s) = \Phi(t)\Psi(t, s)dt$ with $\Psi(s, s) = I_n$, for $s \geq 0$.

Since $\Phi(t)$ is stable and convergent for any $t \geq T_w$, there exist $\alpha, \beta > 0$ such that $\|\Psi(t, T_w)\| \leq \beta e^{-\alpha t}$ for $t \geq T_w$ (see Theorem 2.4.1 [30]). Therefore, by the Cauchy-Schwarz inequality, it is easy to get

$$\begin{aligned} |x(t)|^2 = O\big(|\Psi(t, T_w)x(T_w)|^2\big) &+ O\Big(\Big|\int_{T_w}^t \Psi(t, \tau)Ddw(\tau)\Big|^2\Big)\\ &+ O\Big(\int_{T_w}^t \gamma_{\lfloor\tau\rfloor}^2|v(\tau) - v(\lfloor\tau\rfloor)|^2d\tau\Big) = o(t), \end{aligned} \tag{48}$$

where the second part can be estimated by Lemma 12.3 of [11] and the third part is a direct consequence of Lemma 4. Hence, we can see that (48) implies the desired result (45).

Next, we proceed to show that

$$J(u_1^*, u_2^*) = tr(D^{\mathsf{T}}PD). \tag{49}$$

From (47), applying the Ito's formula to $\langle Px(t), x(t)\rangle$, where P is the stabilizing solution to the ARE (3), it follows that

$$\begin{aligned} d\langle Px(t), x(t)\rangle &= 2\langle Px(t), \Phi(t)x(t) + \gamma_{\lfloor t\rfloor}(v(t) - v(\lfloor t\rfloor))\rangle dt\\ &+ tr(PDD^{\mathsf{T}})dt + 2\langle Px(t), Ddw(t)\rangle, \end{aligned} \tag{50}$$

by integrating (50) over the interval $[0, T]$, we have

$$\begin{aligned} &\langle Px(T), x(T)\rangle - \langle Px(0), x(0)\rangle =\\ &2\int_0^T \langle Px(t), \Phi(t)x(t)\rangle dt + 2\int_0^T \langle Px(t), \gamma_{\lfloor t\rfloor}(v(t) - v(\lfloor t\rfloor))\rangle dt\\ &+ Ttr(PDD^{\mathsf{T}}) + 2\int_0^T \langle Px(t), Ddw(t)\rangle. \end{aligned} \tag{51}$$

We now analyze the right-hand-side of (51) term by term. First, by the Cauchy-Schwarz inequality, it follows that

$$\begin{aligned} &\Big(\int_0^T \langle Px(t), \gamma_{\lfloor t\rfloor}(v(t) - v(\lfloor t\rfloor))\rangle dt\Big)^2\\ &\leq \Big(\int_0^T |Px(t)|^2dt\Big)\Big(\int_0^T \gamma_{\lfloor t\rfloor}^2|v(t) - v(\lfloor t\rfloor)|^2dt\Big). \end{aligned} \tag{52}$$

Similar to (38), it follows that for any $\epsilon \in (0, 1/2)$,

$$\Big|\int_0^t \langle Px(t), Ddw(t)\rangle\Big| = O\Big(\Big(\int_0^t |x(t)|^2dt\Big)^{\frac{1}{2}+\epsilon}\Big). \tag{53}$$

By (52)-(53), Theorem 1 1) and Lemma 4, (51) implies that

$$\limsup_{T\to\infty} \frac{1}{T}\int_0^T \langle -2Px(t), \Phi(t)x(t)\rangle dt = tr(PDD^{\mathsf{T}}). \tag{54}$$

Now, let us denote $V = Q + PB_1R_1^{-1}B_1^{\mathsf{T}}P - PB_2R_2^{-1}B_2^{\mathsf{T}}P$. By Lemma 6 and the ARE (3), it follows that

$$\begin{aligned} tr(PDD^{\mathsf{T}}) &= \limsup_{T\to\infty} \frac{1}{T}\int_0^T \langle -2Px(t), \Phi(t)x(t)\rangle dt\\ &= \limsup_{T\to\infty} \frac{1}{T}\int_0^T \langle -2Px(t), \Phi(\infty)x(t)\rangle dt\\ &= \limsup_{T\to\infty} \frac{1}{T}\int_0^T \langle Vx(t), x(t)\rangle dt. \end{aligned} \tag{55}$$

Therefore, we have

$$\begin{aligned} &J(u_1^*, u_2^*)\\ &= \limsup_{T\to\infty} \frac{1}{T}\int_0^T \big(x^{\mathsf{T}}Qx + u_1^{*\mathsf{T}}R_1u_1^* - u_2^{*\mathsf{T}}R_2u_2^*\big)dt\\ &= \limsup_{T\to\infty} \frac{1}{T}\int_0^T \langle Vx(t), x(t)\rangle dt = tr(D^{\mathsf{T}}PD), \end{aligned} \tag{56}$$

where Theorem 1 1), Lemma 4 and Lemma 6 are used.

It remains to prove that (u_1^*, u_2^*) constitutes a Nash equilibrium. Because of symmetry, we only prove $J(u_1^*, u_2^*) \leq J(u_1, u_2^*)$ for any admissible strategy pair (u_1, u_2^*).

For the system (1) with any admissible strategy pair (u_1, u_2^*), it follows that

$$\begin{aligned} dx(t) &= (Ax(t) + B_1u_1(t) + B_2u_2^*(t))dt + Ddw(t) \\ &= \big(\tilde{A}(t)x(t) + B_1u_1(t) + \gamma_{\lfloor t \rfloor}B_2(v_2(t) - v_2(\lfloor t \rfloor))\big)dt \\ &\quad + Ddw(t), \end{aligned} \tag{57}$$

where $\tilde{A}(t) = A + B_2L_2(t)$. Applying the Ito's formula to $\langle Px(t), x(t)\rangle$, it follows that

$$d\langle Px(t), x(t)\rangle = tr(PDD^{\mathsf{T}})dt + 2\langle Px(t), Ddw(t)\rangle +$$

$$2\langle Px(t), \tilde{A}(t)x(t) + B_1u_1(t) + \gamma_{\lfloor t \rfloor}B_2(v_2(t) - v_2(\lfloor t \rfloor))\rangle dt.$$

Furthermore, repeating the similar method as in the proof of (54), it is easy to see that

$$\begin{aligned} &\limsup_{T\to\infty} \frac{1}{T}\int_0^T \langle -2Px(t), \tilde{A}(t)x(t) + B_1u_1(t)\rangle dt \\ &\qquad = tr(PDD^{\mathsf{T}}). \end{aligned}$$

Finally, it is not hard to verify that

$$\begin{aligned} &J(u_1, u_2^*) = \limsup_{T\to\infty} \frac{1}{T}\int_0^T \big(x^{\mathsf{T}}Qx + u_1^{\mathsf{T}}R_1u_1 - u_2^{*\mathsf{T}}R_2u_2^*\big)dt \\ &= \limsup_{T\to\infty} \frac{1}{T}\int_0^T \big(\langle -2Px(t), (A + B_2R_2^{-1}B_2^{\mathsf{T}}P)x(t)\rangle \\ &\quad + \langle R_1(u_1(t) + R_1^{-1}B_1^{\mathsf{T}}Px(t)), u_1(t) + R_1^{-1}B_1^{\mathsf{T}}Px(t)\rangle \\ &\quad + \langle -2Px(t), B_1u_1(t)\rangle\big)dt \\ &\geq \limsup_{T\to\infty} \frac{1}{T}\int_0^T \langle -2Px(t), \tilde{A}(t)x(t) + B_1u_1(t)\rangle dt \\ &= tr(PDD^{\mathsf{T}}) = J(u_1^*, u_2^*). \end{aligned}$$

Hence, Theorem 2 is true.

IV. Conclusions

In this paper, we have established an adaptive theory on linear quadratic zero-sum stochastic differential games when the players know neither the system parameters nor their opponents' strategies. This has been a longstanding problem, partly because the simpler adaptive linear quadratic stochastic control problem is already complicated enough. To study adaptive game problems, one need to introduce a suitable information structure for the game in the face of system uncertainties, that will inevitably lead to more complicated framework and problems than those in the traditional adaptive control. Under the information structure introduced in this paper, we have established a theory on global stability and asymptotic performance for adaptive strategies that are designed based on the well-known philosophy of "exploitation and exploration". However, many interesting problems still remain to be investigated in this direction. For examples, how to design and analyze the adaptive strategies when the system parameters are time-varying and unknown to the players? What will happen if the players are heterogeneous in the sense that different players may have asymmetric information? How to regulate the adaptive Nash equilibrium if there is a global regulator over the two players?

References

[1] R. Isaacs, *Differential games I, II, III, IV*, RAND Corporation Research Memorandum, 1954-1956.

[2] A. Bagchi, *Stackelberg Differential Games in Economic Models*, Springer, Berlin, 1984.

[3] J. M. Smith, *Evolution and the Theory of Games*, Cambridge University Press, 1982.

[4] D. Fudenberg and J. Tirole, *Game Theory*, Princeton University Press, 2012.

[5] Y. Ho, A. Bryson and S. Baron, "Differential games and optimal pursuit-evasion strategies," *IEEE Trans. Automat. Contr.*, vol. 10, no. 4, 1965, pp. 385-389.

[6] J. Engwerda, *LQ Dynamic Optimization and Differential Games*, John Wiley and Sons, 2005.

[7] T. Basar and G. J. Olsder, *Dynamic Noncooperative Game Theory*, SIAM, 1999.

[8] P. Bernhard, "Linear-quadratic, two-person, zero-sum differential games: Necessary and sufficient conditions," *Journal of Optimization Theory and Applications*, vol. 27, no. 1, 1979, pp. 51-69.

[9] A. W. Starr and Y. C. Ho, "Nonzero-sum differential games," *Journal of Optimization Theory and Applications*, vol. 3, no. 3, 1969, pp. 184-206.

[10] T. Basar and P. Bernhard, *H_∞ Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach*, MA: Birkhäuser, 1991.

[11] H. F. Chen and L. Guo, *Identification and Stochastic Adaptive Control*, Boston, MA: Birkhäuser, 1991.

[12] K. J. Aström and B. Wittenmark, *Adaptive Control*, Dover Publications, Mineola, N.Y., 2008.

[13] T. E. Duncan, L. Guo and B. Pasik-Duncan, "Adaptive continuous-time linear quadratic Gaussian control," *IEEE Trans. Automat. Contr.*, vol. 44, no. 9, 1999, pp. 1653-1662.

[14] L. Guo, "Self-convergence of weighted least-squares with applications to stochastic adaptive control," *IEEE Trans. Automat.Contr.*, vol. 41, no. 1, 1996, pp. 79-89.

[15] Y. Li and L. Guo, "Towards a theory of stochastic adaptive differential games," *2011 50th IEEE Conference on Decision and Control and European Control Conference*, 2011, pp. 5041-5046.

[16] S. Yuan and L. Guo, "Stochastic adaptive dynamical games," *Scientia Sinica Mathematica*, vol. 46, no. 10, 2016, pp. 1367-1382.

[17] S.A.A. Rizvi and Z. Lin, "Output feedback adaptive dynamic programming for linear differential zero-sum games," *Automatica*, vol. 122, 2020, 109272.

[18] S.A.A. Rizvi and Z. Lin, "Output feedback Q-learning for discrete-time linear zero-sum games with application to the H-infinity control," *Automatica*, vol. 95, 2018, pp. 213-221.

[19] H. Li, D. Liu and D. Wang, "Integral reinforcement learning for linear continuous-time zero-sum games with completely unknown dynamics," *IEEE Transactions on Automation Science And Engineering*, vol. 11, no. 3, 2014, pp. 706-714.

[20] K. Zhang, Z. Yang and T. Başar, "Policy optimization provably converges to Nash equilibria in zero-sum linear quadratic games," *arXiv preprint*, arXiv:1906.00729, 2021.

[21] J. Bu, L. J. Ratliff and M. Mesbahi, "Global convergence of policy gradient for sequential zero-sum linear quadratic dynamic games," *arXiv preprint*, arXiv:1911.04672, 2019.

[22] A. Ozdaglar, M. O. Sayin and K. Zhang, "Independent learning in stochastic games," *arXiv preprint*, arXiv:2111.11743, 2021.

[23] M. Rabindranathan, "On the inversion of Toeplitz operators," *J. Math. Mech.*, vol. 19, 1969, pp. 195-206.

[24] E. A. Jonckheere and L. M. Silverman, "Spectral theory of the linear-quadratic optimal control problem: analytic factorization of rational matrix-valued functions," *SIAM J. Control Optim.*, vol. 19, no. 2, 1981, pp. 262-281.

[25] V. Ionescu and M.Weiss, "Continuous and discrete-time Riccati theory: A Popov-function approach," *Linear Algebra Appl.*, vol. 193, 1993, pp. 173-209.

[26] A. Bacciotti, *Stability and Control of Linear Systems*, Springer-Verlag, 2019.

[27] Y. Chow and H. Teicher, *Probability Theory: Independence, Interchangeability, Martingales*, Springer-Verlag, 2008.

[28] H. Abou-Kandil, et al., *Matrix Riccati Equations in Control and Systems Theory*, Birkhäuser Basel, 2003.

[29] T. E. Duncan and B. Pasik-Duncan, "Adaptive control of continuous time linear stochastic systems," *Math. Control Signal Systems*, vol. 3, 1990, pp. 45–60.

[30] L. Guo, *Time-Varing Stochastic Systems: Stability and Adaptive Theory*, Science Press, Beijing, 2020.