2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Towards a Theory of Stochastic Adaptive Differential Games

Yan LI and Lei GUO

Abstract— **Complex systems with components or subsystems having game-like relationships are arguably the most complex ones. Much progress has been made in the traditional game theory over the past half a century, where the structure and the parameters are assumed to be known when the players make their decisions. However this is not the case in many practical situations where the players may have unknown parameters. To initiate a theoretical study of such problems, we consider in this paper a class of two-player zero-sum linear-quadratic stochastic differential games, assuming that the matrices associated with the strategies of the players are unknown to both players. By using the weighted least squares (WLS) estimation algorithms and a random regularization method, adaptive strategies will be constructed for both players. It is shown that both the adaptive strategies will converge to the optimal ones under some natural conditions on the true parameters of the system. To the best of our knowledge, this work seems to be the first to address adaptive stochastic differential game problems with rigorous convergence analysis.**

I. INTRODUCTION

Complex systems are currently at the research frontiers of many important scientific fields, ranging from biology to economy. Complex systems with components or subsystems having game-like relationships may be the most complicated ones to handle, and the theory of differential games appears to be a useful tool in modeling and analyzing conflicts in the context of dynamical systems. Differential games were motivated by combat problems [2], and have been applied in many disciplines, such as economics and management science, biology, ecology and sociology [19], [20]. A great deal of research effort has been devoted to this area in the past half century and much progress has been made [3]-[6]. The theory of differential games combines control theory and game theory [7] in some sense, which enriched the theory and applications of both [17], [26]. Among many different kinds of games, the linear-quadratic differential games, which are described by linear equations and quadratic payoff functions, have attracted much attention. Bernhard [4] gave necessary and sufficient conditions for the existence of a saddle point for deterministic two-player zero-sum differential games on the finite time interval. Starr and Ho [5] extended the zero-sum differential games to the nonzero-sum cases, i.e., the players wish to minimize different performance criteria and they discuss three types of solutions. Different state information patterns may give rise to different types of equilibria, which have been discussed extensively in [6]. Feedback Nash equilibria in the linear quadratic differential games on an infinite time horizon have been studied in [6], [25]. The existence of such equilibria is equivalent to the existence of solutions to a set of algebraic Riccati equations. Differential games are also related closely to optimal control problems. Basar [26] studied the H^∞-optimal control problems in the framework of dynamic game theory, since the original H^∞-optimal control problem is in fact a minimax optimization problem, and hence a zero-sum game. However, in all the above mentioned works, the game structure and the parameters are assumed to be known to all the players.

In control systems, when the structure and the parameters are unknown, a natural way is to use the online state information to estimate the unknowns, which are then used to update the controller. This is called adaptive control design, which is known to be a powerful tool in dealing with systems with large uncertainties. In the past half century, adaptive control systems have attracted much attention, see, e.g., [13]-[15]. Since adaptive control is a nonlinear feedback which performs identification and control simultaneously in the same feedback loop, a rigorous theoretical investigation is well-known to be complicated, even for linear stochastic systems. Notwithstanding, when facing with parameter uncertainties in linear stochastic discrete-time systems, a well-developed theory exists now (see, e.g., [9]-[11]). One of the most remarkable progress has been the establishment of a convergence theory of the well-known Least Square (LS) based self-tuning regulators, which was initiated by the seminar work of Åström and Wittenmark [27] and completed by the rigorous analysis in [9] and [12]. Another notable advance in linear stochastic adaptive control is the adaptive Linear-quadratic-Gaussian (LQG) control problem, where a key theoretical difficulty has been how to guarantee the controllability of the online estimated model. This longstanding problem was finally resolved reasonably in the works [10] and [11], which turn out to be the key bases for the adaptive game problems to be solved in the current paper. That may also explains why the natural adaptive two-player zero-sum game has not been solved earlier.

Similar to control systems, structural and parametric uncertainties may be faced with in the game systems. For example, in the pursuit-evasion game, the pursuer may not know the parameters or the structure of the evader and vice versa; in the economic market, if two firms compete for a fixed number of consumers by advertising, they may not know how their strategies will affect the market. What can be done in the game theoretical framework when facing with uncertainties? Of course, it is natural to consider adaptive

This work was supported by National Natural Science Foundation of China under Grant 60821091.

The authors are with the Institute of Systems Science, AMSS, Chinese Academy of Sciences, Beijing 100190, China. liyan@amss.ac.cn, lguo@amss.ac.cn

game theory. As a starting point, we will consider the two-player zero-sum stochastic differential linear-quadratic games. The dynamics of the system is described by the following stochastic differential equation:

$$\mathrm{d}X(t) = (AX(t) + B_1U_1(t) + B_2U_2(t))\mathrm{d}t + D\mathrm{d}W(t),$$

where U_1 is the strategy of Player 1 and U_2 is the strategy of Player 2. As is well-known, there are relatively complete theories in the case where the parameters of the game and the exact state information are known to both players. Much research effort has been devoted to the linear-quadratic differential games in which the system state is affected by external disturbances. Bagchi and Olsder [23] solved this problem by converting it into an optimization problem in an infinite-dimensional state space. Leondes and Mons [22] gave computable strategies which were shown to be closely related to a Kalman filter. Recently, Mu and Guo in [33] have considered the identification problem between a human and a machine based on the generic 2-player-2-action game including the prisoner dilemma game. Vrabie and Lewis [24] give a method for finding online the Nash equilibrium solution by a continuous-time adaptive dynamic programming procedure that uses the idea of integral reinforcement learning, where the matrix A, which is part of the dynamic of the system, need not be known. However, to the best of our knowledge, few have considered the problem where the unknown parameters exist in stochastic differential games. In this paper, we will deal with this kind of problem. We assume that both players only know the parameters of the system, i.e., Player 1 does not know B_1 and B_2, neither does Player 2. Although the problem formulation as well as the theoretical difficulties are different with those in adaptive control, it turns out that the theory and techniques developed for stochastic adaptive LQG control are quite useful (see, [10], [11]). The adaptive strategies, which are designed via an adaptive Riccati equation with parameter estimates given by weighted least-squares algorithms regularized by a random method, are shown to be asymptotically optimal.

The remainder of the paper is organized as follows. In Section II, we give the problem formulation and the main results. Section III gives the proof of Theorem 1. Section IV will conclude the paper.

II. PROBLEM FORMULATION AND MAIN RESULTS

A. Problem Formulation

Consider the standard two-player zero-sum stochastic differential game. The system is described by the following differential equation

$$\mathrm{d}X(t) = (AX(t) + B_1U_1(t) + B_2U_2(t))\mathrm{d}t + D\mathrm{d}W(t), \quad (1)$$

where $X(t) \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times n}$, $B_1 \in \mathbb{R}^{n\times m_1}$, $B_2 \in \mathbb{R}^{n\times m_2}$, $D \in \mathbb{R}^{n\times p}$. We will define A as system matrix, B_1 as action matrix for Player 1, and B_2 as action matrix for Player 2. $U_1(t) \in \mathbb{R}^{m_1}$ is the strategy of Player 1, $U_2(t) \in \mathbb{R}^{m_2}$ is the strategy of Player 2, $(W(t), \mathscr{F}_t; t \geq 0)$ is an $\mathbb{R}^p$-valued standard Wiener process. The random variables are defined on a fixed complete probability space and the filtration $(\mathscr{F}_t, t \geq 0)$ is defined on this space. It is assumed that A is known to both players, while B_1 and B_2 are unknown to both.

The payoff function is:

$$J = \limsup_{T\to\infty} \frac{1}{T} \int_0^T \big(X^T(t)QX(t) + U_1^T(t)R_1U_1(t) - U_2^T(t)R_2U_2(t)\big)\mathrm{d}t, \quad (2)$$

where $Q = Q^T \geq 0$, $R_1 = R_1^T > 0$, $R_2 = R_2^T > 0$, are known to both players. Player 1 aims to minimize the payoff function, while Player 2 maximize it. We may denote the payoff function J by $J(U_1(t), U_2(t))$, since it depends on the choices of $U_1(t)$ and $U_2(t)$.

If we want to find the solution to the game, we should first know the information that the players have about the game when they make their decisions. Now we will introduce the following definition [6] to explicit this problem clearly.

Definition 1: Let $\eta^i(t) = \{X(s), 0 \leq s \leq \varepsilon_t^i\}$, where $0 \leq \varepsilon_t^i \leq t$, $i = 1,2$, $\eta^i(t)$ determines the state information gained by Player i at time t, and ε_t^i denotes the last time of Player i gaining his information, so Player i can only make strategy depending on $\eta^i(t)$. We say Player i's *information pattern* is
open-loop pattern: if $\eta^i(t) = \{X(0)\}$ for $i = 1,2$;
feedback pattern: if $\eta^i(t) = \{X(t)\}$ for $i = 1,2$.

In this paper, we will pay attention to the feedback pattern only. Then the equilibrium of the above game is defined as the following.

Definition 2: For the zero-sum linear-quadratic differential game with both players of the feedback pattern, a pair of strategies (U_1^0, U_2^0) constitutes a *feedback Nash equilibrium* if it satisfies

$$J(U_1^0, U_2) \leq J(U_1^0, U_2^0) \leq J(U_1, U_2^0),$$

for all U_1, U_2 in the feedback pattern.

When the parameters are known, the feedback Nash equilibrium for the above game is expressed as

$$U_1(t) = -R_1^{-1}B_1^T LX(t) \quad (3)$$

$$U_2(t) = R_2^{-1}B_2^T LX(t), \quad (4)$$

where L is the symmetric solution of the following algebraic Riccati equation (5), which makes $A - (B_1R_1^{-1}B_1^T - B_2R_2^{-1}B_2^T)L$ stable

$$LA + A^TL + Q - L(B_1R_1^{-1}B_1^T - B_2R_2^{-1}B_2^T)L = 0. \quad (5)$$

This Riccati equation seems to be similar to the standard Riccati equation associated with the standard LQG problem in control theory:

$$LA + A^TL + Q - LML = 0,$$

where the key difference is that M is positive definite in the standard LQG problem, while $B_1R_1^{-1}B_1^T - B_2R_2^{-1}B_2^T$ is indefinite in the zero-sum linear-quadratic game, which makes the analysis in the current paper more complicated.

Under the feedback pattern, the pair of the above strategies is the unique Nash equilibrium solution for the game [6].

In order to give a sufficient condition for the existence of (5), we will first introduce a matrix function given by

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

where $B = [B_1, B_2], R = \begin{bmatrix} R_1 & \\ & -R_2 \end{bmatrix}$.

If the following assumptions made on the true parameters are satisfied, (see, e.g., [16])

$A1$) the system matrix A is stable, and the pair $(A, [B_1, B_2])$ is controllable;

$A2$) the matrix function $G(s)$ is antianalytic perfactorizable;

then the Riccati equation (5) has a positive definite solution L which makes $A - (B_1R_1^{-1}B_1^T - B_2R_2^{-1}B_2^T)L$ stable.

Remark 1: Consider the following relaxation of A1):

$A1)'$ The pair $(A, [B_1, B_2])$ is stabilizable.

It has been shown in [16] that Assumptions $A1)'$ and $A2)$ are equivalent to the property that the Riccati equation (5) has a solution that makes $A - (B_1R_1^{-1}B_1^T - B_2R_2^{-1}B_2^T)L$ stable.

Next we introduce a class of matrices defined by

$$\mathscr{F}(A, B_1, B_2) \triangleq \{ \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} \triangleq F \mid A + B_1F_1 + B_2F_2 \text{ exponentially stable} \}.$$

Definition 3: Assume $(A, [B_1, B_2])$ is stabilizable. We will say that $G(s)$ is *antianalytic perfactorizable* (see. e.g., [16]) if there exists $F \in \mathscr{F}(A, B_1, B_2)$ such that $\widetilde{G}(s)$ is antianalytic facotorizable,
where

$$\begin{aligned} \widetilde{G}(s) = R &+ B^T(-sI - \widetilde{A}^T)^{-1}F^TR + RF(sI - \widetilde{A})^{-1}B \\ &+ B^T(-sI - \widetilde{A}^T)^{-1}(Q + F^TRF)(sI - \widetilde{A})^{-1} \end{aligned}$$

where $\widetilde{A} = A + BF,\ B = [B_1, B_2], R = \begin{bmatrix} R_1 & \\ & -R_2 \end{bmatrix},\ F = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} \in \mathscr{F}(A, B_1, B_2)$. For $\widetilde{G}(s)$ is associated to F, we can denote it by $\widetilde{G}_F(s)$.

Remark 2: A matrix $\widetilde{G}(s)$ is *antianalytic facotorizable*, if $\widetilde{G}(s)$ and its inverse can be factorized into two proper rational matrix functions [16].

The following lemma can be found in [16], which is helpful for us to determine whether the estimated parameters satisfy $A2)$ or not.

Lemma 1: Assume the pair $(A, [B_1, B_2])$ is stabilizable. If there exists one matrix $F \in \mathscr{F}(A, B_1, B_2)$ such that the function $\widetilde{G}(s)$ associated to F is antianalytic facotorizable, then so does for any matrix $F \in \mathscr{F}(A, B_1, B_2)$.

Remark 3: Since it is assumed that the true parameters satisfy $A1)$ and $A2)$, the existence of the positive definite solution to the corresponding Riccati equation (5) can be ensured. However, since B_1 and B_2 are unknown to both players, we need to estimate B_1 and B_2 by the online state information first, and then to replace B_1 and B_2 in (5) by its estimate denoted by $\hat{B}_1(t)$ and $\hat{B}_2(t)$. Now it is sufficient to check whether the pair $\left(A, \hat{B}_1(t), \hat{B}_2(t)\right)$ satisfies Assumptions $A1)$ and $A2)$. To this end, we note first that the random regularized method introduced in [10] can be used to ensure the controllability of $(A, [\hat{B}_1(t), \hat{B}_2(t)])$. Then we need only to verify $A2)$ for $\left(A, \hat{B}_1(t), \hat{B}_2(t)\right)$. Notice that the true parameters $\left(A, B_1, B_2\right)$ are assumed to satisfy $A2)$, then it can be concluded from Definition 3 that there exists at least one $F \in \mathscr{F}(A, B_1, B_2)$ such that the associated $\widetilde{G}_F(s)$ is antianalytic facotorizable. By the controllability of $(A, [B_1, B_2])$, we can know that there exists the special form of $F' = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ such that $F' \in \mathscr{F}(A, B_1, B_2)$. Then by Lemma 1, $\widetilde{G}_{F'}(s)$ is antianalytic facotorizable. So $\hat{G}(s)$ corresponding to $\left(A, \hat{B}_1(t), \hat{B}_2(t)\right)$ is antianalytic perfactorizable by Definition 3. Above all, it only needs to check that the pair $\left(A, [\hat{B}_2(t), \hat{B}_2](t)\right)$ is controllable in order to ensure the associated Riccati equation is solvable. If we denote the solution by $L(t)$, then the adaptive strategy can be designed by replacing L in (3) and (4) by $L(t)$, and (B_1, B_2) in (3)-(4) by $(\hat{B}_1(t), \hat{B}_2(t))$.

B. WLS Estimation

Because both players do not know the parameters B_1 and B_2, they need to estimate the parameters first in order to construct adaptive strategies. For simplicity, we assume that the two players use a common estimator, just like there is an independent agency providing parameter estimation or prediction for them.

Remark 4: It is also worth considering the case where the two players estimate the parameters independently, which will be considered elsewhere.

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

$$\theta^T = [B_1, B_2]$$

and

$$\varphi(t) = \begin{bmatrix} U_1(t) \\ U_2(t) \end{bmatrix},$$

so (1) can be rewritten as

$$\mathrm{d}X(t) = \left(AX(t) + \theta^T\varphi(t)\right)\mathrm{d}t + D\mathrm{d}W(t).$$

Now the continuous-time WLS estimates, $\left(\theta(t), t \geq 0\right)$, are given by

$$\mathrm{d}\theta(t) = a(t)P(t)\varphi(t)[\mathrm{d}X^T(t) - X^T(t)A^T - \varphi^T(t)\theta(t)\mathrm{d}t], \tag{6}$$

$$\mathrm{d}P(t) = -a(t)P(t)\varphi(t)\varphi^T(t)P(t)\mathrm{d}t, \tag{7}$$

where $P(0)>0$, $B_1(0)$ and $B_2(0)$ are arbitrary deterministic matrices such that the pair $\big(A,[B_1(0),B_2(0)]\big)$ is controllable,

$$a(t)=\frac{1}{f\big(r(t)\big)} \tag{8}$$

$$r(t)=\| P^{-1}(0) \|+\int_0^t U_1^T(s)U_1(s)+U_2^T(s)U_2(s)\mathrm{d}s \tag{9}$$

and $f\in\mathbb{F}$ with

$$\mathbb{F}=\{f|f:\mathbb{R}_+\to\mathbb{R}_+, f \text{ is slowly increasing and } \int_c^\infty \frac{\mathrm{d}x}{xf(x)}<\infty \text{ for some } c\geq 0\}, \tag{10}$$

where a function is called slowly increasing if it is increasing and satisfies $f\geq 1$. (see, e.g., [10])

Remark 5: It can be concluded from [10] that if $f\in\mathbb{F}$, $f(x)=o(\log x)$. So $a(t)$ can be chosen as $\log\log(r(t))$.

The following lemma is helpful for us to get the convergence result in this paper, which can be found in [11].

Lemma 2: Let $(\theta(t),t\geq 0)$ satisfy (6) and (7) . Then the following properties are satisfied:

$$1)\sup_{t\geq 0}|P^{-1}(t)\widetilde{\theta}(t)|^2<\infty \quad \text{a.s. ;}$$
$$2)\int_0^\infty a(t)|\widetilde{\theta}^T(t)\varphi(t)|^2\mathrm{d}t<\infty \quad \text{a.s.;}$$
$$3)\lim_{t\to\infty}\theta(t)=\overline{\theta} \quad \text{a.s.;}$$

for $i=1,2$, where $\widetilde{\theta}(t)=\theta(t)-\theta$ and $\theta^T(t)=[B_1(t),B_2(t)]$.

Although the WLS algorithms is self-convergent [10], the controllability of $\Big(A,[B_1(t),B_2(t)]\Big)$ is not apparent. Fortunately, the random regularization method introduced in [10] can be used to overcome this difficulty. The details are given in the next section.

C. Regularization

We will introduce the following definition first which can be found in [11].

Definition 4: A family of linear system models $(A(t),B(t),A(t)\in\mathbb{R}^{n\times n},B(t)\in\mathbb{R}^{n\times m},t\geq 0)$ is said to be *uniformly controllable* if there is a constant $c>0$ such that

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

for all $t\in[0,\infty)$.

From Lemma 2, it is known that $\theta(t)$ converges to some random variable. But the estimates may not satisfy the desired controllability property, so it is necessary to have them modified suitably. By Lemma 2, we have

$$\| \theta-\theta(t) \|=O\Big(\| P(t) \|\Big).$$

So we can modify the estimates by the following way:

$$\theta(t,\beta)=\theta(t)-P^{1/2}(t)\beta \tag{11}$$

where $\beta\in\mathscr{M}(m_1+m_2,n)$ denotes the family of $(m_1+m_2)\times n$ real matrices, and we can denote that

$$\theta^T(t,\beta)=[B_1(t,\beta),B_2(t,\beta)] \tag{12}$$

To guarantee the uniform controllability of $\big(A,[B_1(t,\beta),B_2(t,\beta)]\big)$, it is only needed the uniform positivity of $F(t,\beta)$ [10] where

$$F(t,\beta)=\det\Big(\sum_{k=0}^{n-1}A^k[B_1(t,\beta),B_2(t,\beta)]\begin{bmatrix}B_1^T(t,\beta)\\ B_2^T(t,\beta)\end{bmatrix}A^{kT}\Big) \tag{13}$$

Note that if $\beta=P^{-1/2}(t)\begin{bmatrix}B_1^T(t,\beta)-B_1\\ B_2^T(t,\beta)-B_2\end{bmatrix}$, then $B_1(t,\beta)$ and $B_2(t,\beta)$ would be respectively the true system parameters B_1 and B_2. Then $\big(A,[\overline{B}_1(t,\beta),B_2(t,\beta)]\big)$ would be certainly uniformly controllability. But this ideal condition can not be obtained, so the following mechanism is used, which can be found in [10].

Let $(\eta_k,k\in\mathbb{N})$ be independent sequences of i.i.d. $\mathscr{M}(m_1+m_2,n)$-valued random variables that are independent of $(W(t),t\geq 0)$. Let η_k be uniformly distributed on the unit ball of a norm of the matrices. The procedure of choosing β is given by the following way, which can be found in [10]:

$$\beta_0=0$$
$$\beta_k=\begin{cases}\eta_k, & \text{if } F(k,\eta_k)\geq(1+\gamma)F(k,\beta_{k-1})\\ \beta_{k-1}, & \text{otherwise}\end{cases} \tag{14}$$

where $\gamma\in(0,\sqrt{2}-1)$ is fixed and the sequences of the regularized parameters $[\bar{B}_1(k),\bar{B}_2(k)]$ are given by

$$\begin{bmatrix}\bar{B}_1^T(k)\\ \bar{B}_2^T(k)\end{bmatrix}=\begin{bmatrix}B_1^T(k)\\ B_2^T(k)\end{bmatrix}-P^{1/2}(k)\beta_k \tag{15}$$

The following piecewise constant functions induced by (15) are used for the adaptive games:

$$\hat{B}_1(t)=\bar{B}_1(k) \tag{16}$$
$$\hat{B}_2(t)=\bar{B}_2(k) \tag{17}$$

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

D. Main Results

From the following Lemma 3, it is known that $\big(A,[\hat{B}_1(t),\hat{B}_2(t)]\big)$ is uniformly controllable with respect to t. Then, as is explained in Remark 3, the following algebraic Riccati equation will have a real stable positive solution for each $t\in[0,\infty)$:

$$A^TL(t)+L(t)A+Q$$
$$-L_1(t)\Big(\hat{B}_1(t)R_1^{-1}\hat{B}_1^T(t)-\hat{B}_2(t)R_2^{-1}\hat{B}_2^T(t)\Big)L(t)=0 \tag{18}$$

Then Player 1 can use the strategy given by

$$U_1(t)=-R_1^{-1}\hat{B}_1^T(t)L(t)X(t) \tag{19}$$

while the strategy for Player 2 is given by

$$U_2(t)=R_2^{-1}\hat{B}_2^T(t)L(t)X(t). \tag{20}$$

The first theorem shows that if the two players use the strategies (19) and (20), the solution of (1) will be stable in the average sense.

Theorem 1: The solution $(X(t), t \geq 0)$ of (1) with the adaptive strategies of the players given by (19) and (20) is stable in the sense that

$$\limsup_{T \longrightarrow \infty} \frac{1}{T} \int_0^T |X(s)|^2 \mathrm{d}s < \infty \text{ a.s.} \tag{21}$$

The details will be given in the next section.

To obtain the optimal strategy, it is necessary to obtain the strong consistency for the estimates $(\hat{B}_1(t), t \geq 0)$ and $(\hat{B}_2(t), t \geq 0)$. By the method in [10], diminishing excitations are added to the strategies (19) and (20) respectively given by

$$U_1^*(t) = -R_1^{-1}\hat{B}_1 L(k)X(t) + \gamma_k[V(t) - V(k)] \tag{22}$$

$$U_2^*(t) = R_2^{-1}\hat{B}_2 L(k)X(t) + \gamma'_k[V'(t) - V'(k)] \tag{23}$$

for $t \in (k, k+1]$ and $k \in \mathbb{N}$ where $U_1^*(0) \in \mathbb{R}^{m_1}$, $U_2^*(0) \in \mathbb{R}^{m_2}$ are arbitrary deterministic vectors. γ_k and γ'_k can be any sequences satisfying the following:

$$\frac{1}{k}\sum_{i=1}^{k} \gamma_i^2 = o(1), \ \log^l k = o(\sum_{i=1}^{k} \gamma_i^2) \text{ for any } l \geq 1,$$

$$\frac{1}{k}\sum_{i=1}^{k} \gamma_i'^2 = o(1), \ \log^l k = o(\sum_{i=1}^{k} \gamma_i'^2) \text{ for any } l \geq 1,$$

$(V(t), t \geq 0)$ and $V'(t)$ are chosen as sequences of independent standard Wiener Processes that are independent of $\big(W(t), t \geq 0\big)$ and $(\eta_k, k \in \mathbb{N})$.

The following theorem shows that the estimated parameters converge to the true ones. The proof is similar to that in [11], which will be omitted in this paper.

Theorem 2: Let $(\hat{B}_1(t), t \geq 0)$ and $(\hat{B}_2(t), t \geq 0)$ be given by (15) and the players use the strategies (22) and (23) respectively in the system (1). If $A1)$ and $A2)$ are satisfied, then

$$\lim_{t \to \infty} \hat{B}_1(t) = B_1 \quad \text{a.s.} \tag{24}$$

$$\lim_{t \to \infty} \hat{B}_2(t) = B_2 \quad \text{a.s.} \tag{25}$$

Since the parameters converge to the true parameters, the optimality of the strategies given by (22) and (23) can be proved easily.

Theorem 3: If $A1)$ and $A2)$ are satisfied for the differential game (1) and (2), then the strategies given by (22) and (23) are optimal, that is

$$\limsup_{t \to \infty} \frac{1}{t} \int_0^t [X^T(t)QX(t) + U_1^{*T}(t)R_1U_1^*(t) + U_2^{*T}(t)R_2U_2^*(t)]\mathrm{d}t = tr(D^T L D) \tag{26}$$

where L is the solution of (5).

III. PROOF OF THEOREM 1

We need the following results on WLS, which can be found in [11].

Lemma 3: Let $A1)$ and $A2)$ be satisfied for the game (1) (2). Then for any admissible strategies $(U_1(t), U_2(t); t \geq 0)$, the family of regularized WLS estimates $(\hat{B}_i(t), t \geq 0, i = 1,2)$ defined by (14)-(15) have the following properties.

1) Self-convergence, that is, $\hat{B}_i(t)$ converges a.s. to some finite random matrix as $t \to \infty$ for $i = 1,2$.

2) The family $\big(A, [\hat{B}_1(t), \hat{B}_2(t)]\big)$ is uniformly controllable.

3) Semiconsistency, that is, as $t \to \infty$,

$$\int_0^t |(\hat{B}_i(s) - B_i)U_i(s)|^2 \mathrm{d}s = o(r(t)) + O(1) \text{ a.s.}$$

for $i = 1,2$.

Proof of Theorem 1

By Lemma 3 there are random matrices $B_1(\infty)$ and $B_2(\infty)$ such that

$$\lim_{t \to \infty} \hat{B}_1(t) = B_1(\infty) \text{ a.s.}$$

$$\lim_{t \to \infty} \hat{B}_2(t) = B_2(\infty) \text{ a.s.}$$

It can be concluded from Lemma (3) that $\big(A, [B_1(\infty), B_2(\infty)]\big)$ are controllable a.s.. Then form Remark 3, it is easily to know that 18 has a positive solution which makes $A - \Big(\hat{B}_1(t)R_1^{-1}\hat{B}_1^T(t) - \hat{B}_2(t)R_2^{-1}\hat{B}_2^T(t)\Big)L(t)$ stable.

By the continuity property of the solution to the Riccati equation (see in [16]),

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

where $L(\infty)$ is the solution of the static Riccati equation obtained by taking $t \to \infty$ in (18).

Let us denote

$$\Phi(t) = A - \Big(\hat{B}_1(t)R_1^{-1}\hat{B}_1^T(t) - \hat{B}_2(t)R_2^{-1}\hat{B}_2^T(t)\Big)L(t)$$

$$\Phi(\infty) = A - \Big(\hat{B}_1(\infty)R_1^{-1}\hat{B}_1^T(\infty) - \hat{B}_2(\infty)R_2^{-1}\hat{B}_2^T(\infty)\Big)L(\infty)$$

Then we can get

$$\lim_{t \to \infty} \Phi(t) = \Phi(\infty).$$

Since $\Phi(t)$ is stable and converges to a stable matrix, there exists some bounded positive definite matrices $\bar{P}(t)$, such that

$$\Phi(t)\bar{P}(t) + \bar{P}(t)\Phi(t) = -I.$$

Now, note that the system can be expressed as:

$$\begin{aligned} \mathrm{d}X(t) &= \Big(AX(t) + B_1U_1(t) + B_2U_2(t)\Big)\mathrm{d}t + D\mathrm{d}W(t) \\ &= \Big(\Phi(t) + \delta(t)\Big)\mathrm{d}t + D\mathrm{d}W(t) \end{aligned}$$

where

$$\delta(t) = \widetilde{B}_1(t)U_1(t) + \widetilde{B}_2(t)U_2(t).$$

Now, let us analyze the term $\delta(t)$ first,

$$|\delta(t)|^2 \leq 2\Big(|\widetilde{B}_1(t)U_1(t)|^2 + |\widetilde{B}_2(t)U_2(t)|^2\Big)$$

where $|\cdot|^2$ denotes the Euclidean norm of a vector.

From Lemma 3, it is known that

$$\int_0^t |\delta(t)|^2 \mathrm{d}t = o\big(r(T)\big) + O(1).$$

Then, applying Itŏ's formula to $\langle \bar{P}(t)X(t),X(t)\rangle$, and noting that $\bar{P}(t)$ is actually constant in any interval $t \in (k,k+1], k \in \mathbb{N}$, it follows that

$$\begin{aligned}
&\mathrm{d}\langle \bar{P}(t)X(t),X(t)\rangle \\
=&2\langle \bar{P}(t)X(t),\Phi(t)X(t)+\delta(t)\rangle \\
&+\mathrm{tr}(\bar{P}(t)DD^T)\mathrm{d}+2\langle \bar{P}(t)X(t),D\mathrm{d}W(t)\rangle. \qquad (27)
\end{aligned}$$

Furthermore, by Lemma 12.3 of [13] it follows that

$$\left|\int_0^T \langle X(t),\bar{P}(t)D\mathrm{d}W(t)\rangle\right| = O\left(\left[\int_0^T |X(t)|^2\mathrm{d}t\right]^{1/2+\varepsilon}\right)$$

for each $\varepsilon \in (0,1/2)$.

By integrating (27), we can get that

$$\begin{aligned}
&\langle \bar{P}(T)X(T),X(T)\rangle - \langle \bar{P}(0)X(0),X(0)\rangle + \int_0^T |X(t)|\mathrm{d}t \\
\le& o\big(r(T)\big) + O(1) + O\left(\left[\int_0^T |X(t)|^2\mathrm{d}t\right]^{\frac{1}{2}+\varepsilon}\right) \\
&+\int_0^T \mathrm{tr}[\bar{P}(t)DD^T]\mathrm{d}t
\end{aligned}$$

So we can easily get that

$$(1+o(1))\int_0^T |X(t)|\mathrm{d}t \le O(1)+O(T).$$

Hence, Theorem 1 is true.

IV. CONCLUSIONS

In this paper, we have proposed and analyzed a class of adaptive stochastic differential games, specifically, linear quadratic two-player zero-sum stochastic differential games with unknown parameters. As a starting point, we have considered the case where both players share the information of the system matrix, but have unknown action parameters. It is demonstrated that the optimality of the payoff function can be obtained by adaptive strategies. However, many problems remains to be done in this direction. For example, how to design and analyze the adaptive strategies when all the parameters are unknown to the players? What will happen if the player is heterogeneous in the sense that both players use different estimation algorithms with strategies updated in different time scales? These remains further investigation.

REFERENCES

[1] John von Neumann and Oskar Morgenstern, Theory of Games and Economic Behavior, *Princeton University Press*, Princeton, 1944.

[2] R. Isaacs, Differential games I, II, III, IV, *RAND Corporation Research Memorandum* RM-1391, 1399, 1411, 1468, 1954-1956.

[3] Y. C. Ho, A. E. Bryson JR. and S. Baron, Differential Games and Optimal Pursuit-Evasion Strategies, *IEEE Trans. Automat. Contr.*, vol. 10, 1965, pp. 385-389.

[4] 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.

[5] 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.

[6] T. Basar and G. J. Olsder, Dynamic Noncooperative Game Theory, *Society for Industrial and Applied Mathematics*, Philadelphia, 1999.

[7] Y. C. Ho, Differential Games, Dynamic Optimization, and Generalized Control Theory, *Journal of Optimization Theory and Applications*, vol. 6, no. 3, 1979, pp. 179-209.

[8] J. C. Engwerda , LQ Dynamic Optimization and Differential Games, *Chichester*, Wiley, 2005.

[9] L. Guo and H. F. Chen, The Astrom-Wittenmark Self-tuning Regulator Revisited and ELS Based Adaptive Trackers, *IEEE Trans. Automat. Contr.*, Vol. 36, No. 7, 1991, pp. 802-812.

[10] 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.

[11] 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.

[12] L. Guo, Convergence and Logatithm Laws of Self-Tuning Regulators, *Automatica*, vol. 31, no. 3, 1995, pp. 435-450.

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

[14] M. Krstic, I. Kanellakopoulos, and P. V. Kokotovic, Nonlinear and Adaptive Control Design, *Wiley*, New York, 1995.

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

[16] V. Ionescu and M.Weiss, Continuous and Discrete-Time Riccati Theory: A Popov-Function Approach, *Linear Algebra Appl.*, vol. 193, 1993, pp. 173-209.

[17] David W. K. Yeung and Leon A. Petrosyan, Cooperative Stochastic Differential Games, *Spinger*, New York, 2006.

[18] K. J. Åström and B. Wittenmark, On Self-Tuning Regulators, *Automatica*, vol. 9, 1973, pp. 195-199.

[19] A. Bagchi, Stackelberg Differential Games in Economic Models, *Spinger*, Berlin, 1984.

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

[21] Y. C. Ho, On the Minimax Principle and Zero-Sum Stochsatic Differential Games, *Journal of Optimization Theory and Applications*, vol. 13, no. 3, 1974, pp. 343-361.

[22] C. T. Leondes and B. Mons, Differential Games with Noise-Corrupted Measurements, *Journal of Optimization Theory and Applications*, vol. 28, no. 2, 1979, pp. 233-251.

[23] C. Bagchi and G. J. Olsder, Linear-Quadratic Stochasitc Pursuit-Evasion Games, *Appl. Math. Optim.*, vol. 7, no. 1, 1981, pp. 95-123.

[24] D. Vrabie and F. L. Lewis, Adaptive Dynamic Programming for Online Solution of A Zero-sum Differential Game, *J. Control Theory Appl.*, vol. 9, no. 3, 2011, pp. 353-360.

[25] J. C. Engwerds, W. A. van den Broek, and J. M. Schumacher, Feedback Nash Equilibria in Uncertain Infinite Time Horizon Differential Games, *Proceedings CD of the fourteenth International Symposium of Mathematical Theory of Networks and Systems MTNS 2000*, Perpignan, France,2000, CD Rom 1-6.

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

[27] K. J. Åström and B. Wittenmark, On Self-Tuning Regulators, *Automatica*, vol. 9, 1973, pp. 185-199.

[28] L. Ljung, Analysis of Recursive Stochastic Algorithms, *IEEE Trans. Automat. Contr.*, vol. 22, no.1, 1977 pp. 551-575.

[29] I. Liptser and D. Luenberger, Differential Games with Imperfect State Information, *IEEE Trans. Automat. Contr.*, vol. AC-14, no. 1, 1969, pp.29-38.

[30] I. B. Rhodes and D. G. Luenberger, Stochastic Differential Games with Constrained State Estimators, *IEEE Trans. Automat. Contr.*, vol. AC-14, no. 1, 1969, pp.29-38.

[31] L. Guo, Further Results on Least Squares Based Adaptive Minimum Variance Control, *SIAM J. Contr. Optimization*, vol. 32, no. 1, 1969, pp. 187-212.

[32] T. E. Duncan and B. Pasik-Duncan, Adaptive Control of Continuous Time Linear Stochastic Systems, *Math. Contr., Signals, Syst.*, vol. 3, 1990, pp. 45-60.

[33] Y. Mu and L. Guo, Optimization and Identification in Nonequilibrium Dynamical Games, *Proceedings of the Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference, December*, 2009, Shanghai, pp. 5750-5755.