第 35 卷第 5 期
2018 年 5 月

控 制 理 论 与 应 用
Control Theory & Applications

Vol. 35 No. 5
May 2018

DOI: 10.7641/CTA.2018.80039

多人非合作随机自适应博弈

胡浩洋, 郭　雷†

(中国科学院 系统控制重点实验室; 中国科学院 数学与系统科学研究院, 北京 100190)

摘要: 本文考虑系数未知的离散时间线性随机系统多人非合作的自适应博弈问题, 每个参与者运用最小二乘算法和“必然等价原则”来设计博弈策略组合, 目的是自适应优化自身的一步超前收益函数. 本文证明此自适应策略组合使得闭环系统全局稳定, 并且在一定意义下是该博弈问题的渐近纳什均衡解.

关键词: 线性随机系统; 自适应博弈; 最小二乘法; 全局稳定性; 渐近纳什均衡

引用格式: 胡浩洋, 郭雷. 多人非合作随机自适应博弈. 控制理论与应用, 2018, 35(5): 637 – 643

中图分类号: TP273　　**文献标识码**: A

Non-cooperative stochastic adaptive multi-player games

HU Hao-yang, GUO Lei†

(The Key Laboratory of Systems and Control; Academy of Mathematics and Systems Science,
Chinese Academy of Sciences, Beijing 100190, China)

Abstract: In this paper, we consider non-cooperative stochastic adaptive multi-player games described by linear discrete-time stochastic systems with unknown parameters. The least-squares algorithm together with the certainty equivalence principle is used by each player in designing the strategy for optimizing its own one-step-ahead payoff function. It will be shown that the resulting adaptive strategy profile can make the closed-loop system globally stable and at the same time, the profile converges to an asymptotic Nash equilibrium in some sense.

Key words: linear stochastic system; adaptive games; least-squares algorithm; globally stable; asymptotic Nash equilibrium

Citation: HU Haoyang, GUO Lei. Non-cooperative stochastic adaptive multi-player games. *Control Theory & Applications*, 2018, 35(5): 637 – 643

1 Introduction

In the areas of the control and game, much progress have been made in both theory and application over the past half century. Game theory and control theory have been developed in parallel to a large extent[1–7], although they are related in many aspects (e.g. [8]). With the investigation of complex adaptive systems, the combination of the control theory and the game theory has received increasing attention in recent years (e.g. [9–12]). A classical situation is the differential games[3, 13] which were motivated by combat problems, and have been applied in many disciplines, such as sociology, biology, economics, management science and power systems (e.g. [14–17]). However, in the classical game model, almost all of the existing studies assume that the model parameters are known to the players, which are unrealistic in many practical situations, since there always exist uncertainties in modelling real world systems[18], and the uncertainties may even change from time to time. When the structure and the parameters are unknown, a nature way in control systems design is to use the online measurement information to estimate the unknowns, which are then used to construct or update the controller. This is usually called adaptive control design[4–7], which is known to be a powerful tool in dealing with systems with large structure uncertainties. Since the adaptive control is a typically nonlinear feedback which performs identification and control simultaneously in the same feedback loop, a rigorous theoretical investigation for the closed-loop adaptive control systems is well-known to be complicated, even if the open-loop control systems are linear. Therefore, when we deal with game problems using adaptive control approaches, a central theoretical problem is how to establish the convergence of the adaptive strategy profile by overcoming difficulties arising from nonlinearities. Li and Guo[11–12] have studied adaptive game problems described by continuous-time state space stochastic mod-

Received 2018−01−13; revised 2018−03−23.
†Corresponding author. E-mail: lguo@amss.ac.cn
Recommended by Associate Editor MEI Sheng-wei.
Supported by the National Natural Science Foundation of China (11688101).

els with full state information. Moreover, Yuan and Guo[19] have investigated related problems for a zero-sum game described by an input-output stochastic model with stringent assumptions on the system parameters.

In this paper, we will extend the work of Yuan and Guo[19] and Hu and Guo[20] to general multi-player non-cooperative stochastic games. We assume that the parameters are unknown to the players in the system model, and attempt to use the ideas and methods of adaptive control to deal with the unknown parameters and to design a strategy profile for the players, aiming at minimizing the respective payoff functions. We will prove that the adaptive strategy profile can make the system globally stable, and at the same time, the profile is an asymptotic Nash equilibrium solution to our game problem.

The remainder of the paper is organized as follows: The problem formulation and the main results will be presented in the next section, the stability and performance analysis will be conducted in Sections 3, and finally some concluding remarks will be given in the last section.

2 The main results

Consider the following linear stochastic input-output model with r players:

$$A(z)y_{t+1} = B_1(z)u_{1t} + \cdots + B_r(z)u_{rt} + w_{t+1}, \tag{1}$$

where $\{y_t\}$ and $\{w_t\}$ are, respectively, the system output and noise process, $\{u_{it}\}(i = 1, \cdots, r)$ are the controls or strategies of the players (without loss of generality, we assume $y_t = w_t = u_{it} = 0,\ \forall\, t < 0,\ i = 1, \cdots, r$), and $A(z)$, $B_i(z)(i = 1, \cdots, r)$ are polynomials in the backward-shift operator z:

$$\begin{cases} A(z) = 1 + a_1 z + \cdots + a_p z^p, & p \geqslant 0, \\ B_1(z) = b_{11} + b_{12} z + \cdots + b_{1q_1} z^{q_1 - 1}, & q_1 \geqslant 1, \\ \quad \vdots \\ B_r(z) = b_{r1} + b_{r2} z + \cdots + b_{rq_r} z^{q_r - 1}, & q_r \geqslant 1, \end{cases} \tag{2}$$

where a_i, b_{jk} are coefficients with non-zero leading coefficients $b_{11}, \cdots, b_{r1}$ and $p, q_1, \cdots, q_r$ are upper bounds on the true orders. We assume that $\{w_t\}$ satisfies the following condition:

A1) The noise sequence $\{w_t, \mathscr{F}_t\}$ is a martingale difference sequence (where $\mathscr{F}_t$ is a sequence of nondecreasing σ-algebras) with conditional variance σ^2, i.e.

$$\mathrm{E}[w_{t+1}^2 | \mathscr{F}_t] = \sigma^2 > 0,\ \text{a.s.}. \tag{3}$$

Also, assume that there exists a constant $\kappa > 2$ such that

$$\sup_t \mathrm{E}[|w_{t+1}|^\kappa | \mathscr{F}_t] < \infty,\ \text{a.s.}. \tag{4}$$

For convenience of subsequent discussions, we also assume that there is a nondecreasing positive deterministic sequence $\{d_t\}$ such that

$$w_t^2 = \mathrm{O}(d_t),\ \text{a.s.},\ d_{t+1} = \mathrm{O}(d_t). \tag{5}$$

It is easy to prove[21] that under Condition A1), $\{d_t\}$ can be taken as

$$d_t = t^\delta,\ \forall\, \delta \in (\frac{2}{\kappa}, 1), \tag{6}$$

where κ is given by (4).

The objective of this paper is to design the players strategies $u_{1t}, \cdots, u_{rt}$ based on the past measurements $\{y_0, \cdots, y_{t-1}, u_{10}, \cdots, u_{1(t-1)}, \cdots, u_{r0}, \cdots, u_{r(t-1)}\}$ to minimize the following payoff functions respectively:

$$\begin{cases} J_1[u_1(t)\ \cdots\ u_r(t)] = \limsup\limits_{n\to\infty} \dfrac{1}{n} \sum\limits_{t=0}^{n-1} J_{1t}, \\ \quad \vdots \\ J_r[u_1(t)\ \cdots\ u_r(t)] = \limsup\limits_{n\to\infty} \dfrac{1}{n} \sum\limits_{t=0}^{n-1} J_{rt} \end{cases} \tag{7}$$

with the following one-step-ahead payoff functions:

$$\begin{cases} J_{1t} = J_{1t}[u_1(t)\ \cdots\ u_r(t)] = \\ \qquad \mathrm{E}[(y_{t+1} - y^*_{1(t+1)})^2 + \lambda_1 u_1^2(t) | \mathscr{F}_t], \\ \quad \vdots \\ J_{rt} = J_{rt}[u_1(t)\ \cdots\ u_r(t)] = \\ \qquad \mathrm{E}[(y_{t+1} - y^*_{r(t+1)})^2 + \lambda_r u_r^2(t) | \mathscr{F}_t], \end{cases} \tag{8}$$

where $\lambda_1, \cdots, \lambda_r$ are positive weighting constants.

Remark 1 We will consider the case where the players are of 'bounded rationality'. Here the 'bound rationality' means that the players may not have the complete information about the system parameters because of the complexity and uncertainty of the environment, and that even though the players can get the complete parameter information, they may be able to only minimize the one-step-ahead payoff functions(8).

Furthermore, we assume that the players share the same estimator produced by the least-squares estimation. We will study whether or not there exists an adaptive strategy profile that can make the system stable and satisfy some other nice properties.

First, we give the definition of the global stabilization:

Definition 1 The stochastic system (1) is called globally stabilizable, if the strategy profile $(\{u_{1t}\}, \cdots, \{u_{rt}\})$ makes the system satisfy

$$\frac{1}{n} \sum_{t=0}^{n-1} (y_t^2 + u_{1t}^2 + \cdots + u_{rt}^2) = \mathrm{O}(1),\ \text{a.s.}, \tag{9}$$

for any initial value $y_0 \in \mathbb{R}$.

In addition to Condition A1), the following assumptions will be also needed:

A2) $D(z) \neq 0,\ \forall\, z : |z| \leqslant 1$, where $D(z) = \sum\limits_{i=1}^{r} \dfrac{b_{i1}}{\lambda_i} B_i(z)$.

A3) $E(z) \neq 0,\ \forall\ z : |z| \leqslant 1$, where $E(z) = A(z) + D(z)$.

A4) $\{y_{1t}^*\}, \cdots, \{y_{rt}^*\}$ are bounded random or deterministic reference sequences that are independent of $\{w_t\}$.

Remark 2 Conditions A2) and A3) are generalizations of the classic minimum-phase condition, which is necessary for the internal stability of the closed-loop system, even if the parameters are known[22].

To simplify the analysis, in this paper, we consider the case where the 'high-frequency gain' parameters $b_{11}, \cdots, b_{r1}$ are known. We collect the rest unknown parameters into a vector

$$\theta = [-a_1 \ \cdots \ -a_p \ \ b_{12} \cdots b_{1q_1} \cdots b_{r2} \cdots b_{rq_r}]^{\mathrm{T}},$$

and define the corresponding regressor

$$\begin{aligned}\varphi_t =&[y_t \ \cdots \ y_{t-p+1} \ \ u_{1(t-1)} \ \cdots \ u_{1(t-q_1+1)} \\ &\cdots \ u_{r(t-1)} \ \cdots \ u_{r(t-q_r+1)}]^{\mathrm{T}}.\end{aligned}$$

Thus, the system (1) can now be rewritten in the form:

$$y_{t+1} = \theta^{\mathrm{T}}\varphi_t + b_{11}u_{1t} + \cdots + b_{r1}u_{rt} + w_{t+1}. \quad (10)$$

The recursive least-squares (LS) method is used to give estimates for the unknown parameter θ in the model (10):

$$\begin{aligned}\theta_{t+1} =& \theta_t + a_t P_t\varphi_t(y_{t+1} - b_{11}u_{1t} - \cdots - \\ & b_{r1}u_{rt} - \theta_t^{\mathrm{T}}\varphi_t), \quad (11)\\ P_{t+1} =& P_t - a_t P_t\varphi_t\varphi_t^{\mathrm{T}}P_t, \quad (12)\\ a_t =& (1 + \varphi_t^{\mathrm{T}}P_t\varphi_t)^{-1}, \quad (13)\end{aligned}$$

where the initial value θ_0 and $P_0 > 0$ can be chosen arbitrarily.

Using Condition A1), we can get the expressions of payoff functions as follows:

$$\begin{cases} J_{1t}[u_{1t} \ \cdots \ u_{rt}] = \sigma^2 + (\theta^{\mathrm{T}}\varphi_t + b_{11}u_{1t} + \cdots + \\ \qquad\qquad b_{r1}u_{rt} - y_{1(t+1)}^*)^2 + \lambda_1 u_{1t}^2, \\ \qquad\qquad \vdots \\ J_{rt}[u_{1t} \ \cdots \ u_{rt}] = \sigma^2 + (\theta^{\mathrm{T}}\varphi_t + b_{11}u_{1t} + \cdots + \\ \qquad\qquad b_{r1}u_{rt} - y_{r(t+1)}^*)^2 + \lambda_r u_{rt}^2. \end{cases} \quad (14)$$

Let the derivatives with respect to $u_{1t}, \cdots, u_{rt}$ be respectively set to zero in the above payoff functions. Then we get the following linear equations:

$$\begin{cases} (\lambda_1 + b_{11}^2)u_{1t} + \cdots + b_{11}b_{r1}u_{rt} = \\ -\, b_{11}(\theta^{\mathrm{T}}\varphi_t - y_{1(t+1)}^*), \\ \ \vdots \\ b_{r1}b_{11}u_{1t} + \cdots + (\lambda_r + b_{r1}^2)u_{rt} = \\ -\, b_{r1}(\theta^{\mathrm{T}}\varphi_t - y_{r(t+1)}^*). \end{cases} \quad (15)$$

It is easy to prove that the determinant of the coefficient matrix is $\lambda_1\lambda_2\cdots\lambda_r(1 + \sum\limits_{i=1}^{r}\dfrac{b_{i1}^2}{\lambda_i}) > 0$. Therefore, the solutions of these linear equations are unique, and the solutions can minimize the one-step-ahead payoff functions (8). In the case where the parameter θ is unknown, we will replace the θ in (15) by its LS estimate θ_t, and obtain the adaptive actions $u_{1t}, \cdots, u_{rt}$ from the following linear equations:

$$\begin{cases} (\lambda_1 + b_{11}^2)u_{1t} + \cdots + b_{11}b_{r1}u_{rt} = \\ -\, b_{11}(\theta_t^{\mathrm{T}}\varphi_t - y_{1(t+1)}^*), \\ \ \vdots \\ b_{r1}b_{11}u_{1t} + \cdots + (\lambda_r + b_{r1}^2)u_{rt} = \\ -\, b_{r1}(\theta_t^{\mathrm{T}}\varphi_t - y_{r(t+1)}^*). \end{cases} \quad (16)$$

For the multiple output situation, y_t and u_{it} are vectors, and A_i, B_{jk} are corresponding dimension coefficient matrices. The corresponding one-step-ahead payoff functions are $J_{it} = \mathrm{E}[(y_{t+1} - y_{i(t+1)}^*)^{\mathrm{T}}Q_i(y_{t+1} - y_{i(t+1)}^*) + u_{it}^{\mathrm{T}}R_i u_{it}|\mathscr{F}_t]$. The linear equations (15), where the coefficient matrix depends on A, B, Q, R, may have no solutions we need, then more conditions are needed. Moreover, the analysis of the stability and optimality where the coefficients are matrices is more complicated. Therefore, we just consider the single output situation in this paper.

For the sake of convenience, let $(\mathscr{U}_1, \cdots, \mathscr{U}_r)$ (where $\mathscr{U}_i = \{u_{it}, t \geqslant 0\},\ i = 1, \cdots, r$) be the adaptive strategy profile defined by the adaptive actions (16). We proceed to prove that this profile $(\mathscr{U}_1, \cdots, \mathscr{U}_r)$ can make the closed-loop system globally stable.

Now, we give a theorem on global stability of the closed-loop system.

Theorem 1 Consider the linear stochastic system (1) with unknown parameter θ. If Conditions A1)–A4) are fulfilled, and if the players adopt the strategy profile $(\mathscr{U}_1, \cdots, \mathscr{U}_r)$ defined by the adaptive actions (16), then the corresponding closed-loop system is globally stable, in the sense that for any initial value $y_0 \in \mathbb{R}$,

$$\frac{1}{n}\sum_{t=0}^{n-1}(y_t^2 + u_{1t}^2 + \cdots + u_{rt}^2) = \mathrm{O}(1),\ \text{a.s.}.$$

Next, we consider the optimality of the adaptive strategy profile $(\mathscr{U}_1, \cdots, \mathscr{U}_r)$. First, we give some notations. From the definition of the $(J_{1t}[u_{1t} \ \cdots \ u_{rt}], \cdots, J_{rt}[u_{1t} \ \cdots \ u_{rt}])$, it is easy to see that the payoff functions depend on not only the current actions $u_{it}(i = 1, \cdots, r)$ but also the previous actions $u_{i0}, u_{i1}, \cdots, u_{i(t-1)}(i = 1, \cdots, r)$. Then, we can rewrite the payoff functions as $(J_{1t}[U_{1t} \ \cdots \ U_{rt}], \cdots, J_{rt}[U_{1t} \ \cdots \ U_{rt}])$, where $U_{it} = (u_{i0}, u_{i1}, \cdots, u_{it})$, $i = 1, \cdots, r$. Moreover, let $(\Omega_{1t}, \cdots, \Omega_{rt})$ be corresponding admissible strategy set of $(u_{1t}, \cdots, u_{rt})$.

In addition, if a positive sequence $\{x_t \geqslant 0;\ t \geqslant 0\}$

satisfies $\lim\limits_{n\to\infty}\dfrac{1}{n}\sum\limits_{t=0}^{n-1}x_t=0$, then we simply write $x_t=\overline{\mathrm{o}(1)}$[19–20]. Now, we give the definition of asymptotic Nash equilibrium:

Definition 2 The strategy profile $(\mathscr{U}_1,\cdots,\mathscr{U}_r)$ is called an asymptotic Nash equilibrium, if it makes the one-step-ahead payoff functions (8) satisfy

$$\begin{cases} J_{1t}[U_{1t}\cdots U_{rt}]=\overline{\mathrm{o}(1)}+\inf\limits_{u'_{1t}\in\Omega_{1t}}J_{1t}[U'_{1t}\cdots U_{rt}],\\ \qquad\vdots\\ J_{rt}[U_{1t}\cdots U_{rt}]=\overline{\mathrm{o}(1)}+\inf\limits_{u'_{rt}\in\Omega_{rt}}J_{rt}[U_{1t}\cdots U'_{rt}], \end{cases}\tag{17}$$

where the $U'_{it}=(U_{i(t-1)},u'_{it}),i=1,\cdots,r$.

From the following theorem, we can see that the adaptive strategy profile $(\mathscr{U}_1,\cdots,\mathscr{U}_r)$ defined by (16) is an asymptotic Nash equilibrium.

Theorem 2 Consider the linear stochastic system (1). If Conditions A1)–A4) are fulfilled, then the strategy profile $(\mathscr{U}_1,\cdots,\mathscr{U}_r)$ defined by the adaptive actions (16) is an asymptotic Nash equilibrium of the non-cooperative games with the payoff functions $(J_{1t}[u_{1t}\ \cdots\ u_{rt}],\ \cdots,\ J_{rt}[u_{1t}\ \cdots\ u_{rt}])$, where the admissible strategy set of $(\mathscr{U}_1,\cdots,\mathscr{U}_r)$ is $(\mathbb{R},\cdots,\mathbb{R})$.

3 Analysis of stabilization and optimality

To prove the stability and optimality, we need some lemmas. First of all, we introduce some notations that will be used throughout the sequel:

$$\alpha_t\triangleq\frac{(\tilde{\theta}_t^{\mathrm{T}}\varphi_t)^2}{1+\varphi_t^{\mathrm{T}}P_t\varphi_t},\ \delta_t\triangleq\mathrm{tr}(P_t-P_{t+1}),\tag{18}$$

$$r_t\triangleq 1+\sum_{i=0}^{t}\|\varphi_i\|^2,\ \tilde{\theta}_t\triangleq\theta-\theta_t,\tag{19}$$

where $\|\cdot\|$ is the Euclidean norm.

Lemma 1 (see [23]) Consider the linear stochastic system (1). The prediction error $\{\tilde{\theta}_t^{\mathrm{T}}\varphi_t\}$ of the LS algorithm (11) − (13) has the following asymptotic property:

$$\sum_{i=0}^{t}\alpha_i=\mathrm{O}(\log r_t),\ \text{a.s.}\ t\to\infty\tag{20}$$

with α_i and r_t defined by (18) and (19), respectively.

Lemma 2 Consider the linear stochastic system (1). If Conditions A1)–A4) are fulfilled, then there exists a positive random process $\{L_t\}$ such that

$$y_t^2\leqslant L_t,\ \forall\, t\geqslant 0,\ \text{a.s.},\tag{21}$$

and $\{L_t\}$ satisfies the following 'linear time-varying relationship'

$$L_{t+1}\leqslant\beta L_t+M_0\sum_{i=0}^{t}\alpha^{t-i}\alpha_i\delta_iL_i+\xi_t,\tag{22}$$

where the constants $\alpha,\beta\in(0,1)$, $M_0>0$, α_t and δ_t are defined by (18), and $\{\xi_t\}$ is a positive random process satisfying

$$\xi_t=\mathrm{O}(d_t\log r_t)\tag{23}$$

with d_t and r_t defined, respectively, by (5) and (19).

Proof 1 Combining (10) and (16), we can get

$$\begin{cases} u_{1t}=-\dfrac{b_{11}}{\lambda_1}(y_{t+1}-y^*_{1(t+1)}-\tilde{\theta}_t^{\mathrm{T}}\varphi_t-w_{t+1}),\\ \qquad\vdots\\ u_{rt}=-\dfrac{b_{r1}}{\lambda_r}(y_{t+1}-y^*_{r(t+1)}-\tilde{\theta}_t^{\mathrm{T}}\varphi_t-w_{t+1}). \end{cases}$$

Substituting $u_{it}(i=1,\cdots,r)$ into (1), we get

$$E(z)y_{t+1}=D(z)\tilde{\theta}_t^{\mathrm{T}}\varphi_t+\sum_{i=1}^{r}\frac{b_{i1}}{\lambda_i}B_i(z)y^*_{i(t+1)}+(D(z)+1)w_{t+1},\tag{24}$$

where $D(z)$,$E(z)$ are defined by Conditions(A2)(A3).

Let us now introduce the following notations:

$$\eta_t\triangleq\sum_{i=1}^{r}\frac{b_{i1}}{\lambda_i}B_i(z)y^*_{i(t+1)}+(D(z)+1)w_{t+1},\tag{25}$$

$$\bar{\eta}_t\triangleq D(z)\tilde{\theta}_t^{\mathrm{T}}\varphi_t+\eta_t.\tag{26}$$

Then, (24) can be simply written

$$E(z)y_{t+1}=\bar{\eta}_t.\tag{27}$$

Let $Y_{t+1}=[y_{t+1}\ y_t\ \cdots\ y_{t-h+2}]^{\mathrm{T}}$, where $h=\max\{p,q_1-1,\cdots,q_r-1\}$. By (27) and Condition A3), there exists a stable matrix $\mathcal{A}\in\mathbb{R}^{h\times h}$ and a column vector $\mathcal{B}\in\mathbb{R}^h$ such that

$$Y_{t+1}=\mathcal{A}Y_t+\mathcal{B}\bar{\eta}_t.\tag{28}$$

By the stability of $\mathcal{A}$, there is a matrix norm $\|\cdot\|$ such that $\alpha=\|\mathcal{A}\|<1$. From (28) we have

$$\begin{aligned}&\|Y_{t+1}\|^2=\|\mathcal{A}^{t+1}Y_0+\sum_{i=0}^{t}\mathcal{A}^{t-i}\mathcal{B}\bar{\eta}_i\|^2=\\ &\mathrm{O}(\alpha^{t+1})+\mathrm{O}(\sum_{i=0}^{t}\alpha^{t-i}\|\bar{\eta}_i\|^2)=\\ &\mathrm{O}(\sum_{i=0}^{t}\alpha^{t-i}(D(z)\tilde{\theta}_i^{\mathrm{T}}\varphi_i)^2)+\\ &\mathrm{O}(\sum_{i=0}^{t}\alpha^{t-i}\|\eta_i\|^2).\end{aligned}\tag{29}$$

Note the definition of η_i. The second term of (29) has the following estimation:

$$\begin{aligned}&\mathrm{O}(\sum_{i=0}^{t}\alpha^{t-i}\|\eta_i\|^2)=\\ &\mathrm{O}(\sum_{i=0}^{t}\alpha^{t-i}\{w_{i+1}^2+\sum_{k=1}^{r}y^*_{k(i+1)}{}^2\})=\\ &\mathrm{O}(\sum_{i=0}^{t}\alpha^{t-i}d_{i+1})=\mathrm{O}(d_t).\end{aligned}\tag{30}$$

According to (20) and the property of $\varphi_i^{\mathrm{T}}P_{i+1}\varphi_i\leqslant 1$, we can estimate the first term of (29) as follows:

$$\mathrm{O}(\sum_{i=0}^{t}\alpha^{t-i}(D(z)\tilde{\theta}_i^{\mathrm{T}}\varphi_i)^2)=$$

$$\mathrm{O}(\sum_{i=0}^{t}\alpha^{t-i}(\tilde{\theta}_i^{\mathrm{T}}\varphi_i)^2)=$$

$$\mathrm{O}(\sum_{i=0}^{t}\alpha^{t-i}\alpha_i[1+\varphi_i^{\mathrm{T}}P_{i+1}\varphi_i+\varphi_i^{\mathrm{T}}(P_i-P_{i+1})\varphi_i])=$$

$$\mathrm{O}(\log r_t)+\mathrm{O}(\sum_{i=0}^{t}\alpha^{t-i}\alpha_i\delta_i\|\varphi_i\|^2). \tag{31}$$

By the adaptive actions (16), it is easy to see the relationship of u_{it} and u_{jt}:

$$\lambda_i b_{j1}u_{it}-\lambda_j b_{i1}u_{jt}=b_{i1}b_{j1}(y^*_{i(t+1)}-y^*_{j(t+1)}). \tag{32}$$

Substituting (32) into (1), we can get

$$u_{it}=\frac{b_{i1}}{\lambda_i}D(z)^{-1}(A(z)y_{t+1}+\sum_{j=1}^{r}\frac{b_{j1}}{\lambda_j}(y^*_{i(t+1)}-y^*_{j(t+1)})-w_{t+1}). \tag{33}$$

Under Condition A2), from (33) it is known that there exists $\beta\in(0,1)$ such that

$$\begin{cases} u^2_{1(t-j)}=\mathrm{O}(\sum_{i=0}^{t}\beta^{t-i}y_i^2)+\mathrm{O}(d_t),\ j=1,\cdots,q_1-1,\\ \quad\vdots\\ u^2_{r(t-j)}=\mathrm{O}(\sum_{i=0}^{t}\beta^{t-i}y_i^2)+\mathrm{O}(d_t),\ j=1,\cdots,q_r-1. \end{cases}$$

Therefore,

$$\|\varphi_t\|^2=\sum_{i=0}^{p-1}y^2_{t-i}+\sum_{i=0}^{q_1-1}u^2_{1(t-i)}+\cdots+\sum_{i=0}^{q_r-1}u^2_{r(t-i)}=\mathrm{O}(\sum_{i=0}^{t}\beta^{t-i}y_i^2)+\mathrm{O}(d_t).$$

Let $L_t=\sum_{i=0}^{t}\beta^{t-i}\|Y_i\|^2$. It is obvious that (21) is satisfied. Then,

$$\|\varphi_t\|^2=\mathrm{O}(L_t)+\mathrm{O}(d_t). \tag{34}$$

From $\sum_{j=0}^{\infty}\delta_j=\sum_{j=0}^{\infty}(\mathrm{tr}P_j-\mathrm{tr}P_{j+1})=\mathrm{tr}P_0<\infty$, we know that $\delta_t\to 0$. According to (29)–(31), we have that

$$\|Y_{t+1}\|^2=\mathrm{O}(\log r_t+d_t)+\mathrm{O}(\sum_{i=0}^{t}\alpha^{t-i}\alpha_i\delta_i(L_i+d_i))=$$
$$\mathrm{O}(\log r_t+d_t)+\mathrm{O}(\sum_{i=0}^{t}\alpha^{t-i}\alpha_i\delta_i L_i)+$$
$$\mathrm{O}(\max_{0\leqslant j\leqslant t}\{\delta_j d_j\}\log r_t)\leqslant$$
$$M_0\sum_{i=0}^{t}\alpha^{t-i}\alpha_i\delta_i L_i+\mathrm{O}(d_t\log r_t).$$

Then, we arrive at (22), i.e.

$$L_{t+1}=\beta L_t+\|Y_{t+1}\|^2\leqslant$$
$$\beta L_t+M_0\sum_{i=0}^{t}\alpha^{t-i}\alpha_i\delta_i L_i+\mathrm{O}(d_t\log r_t).$$

The proof is complete. QED.

Lemma 3 Under the conditions of Lemma 2, we have

$$\|\varphi_t\|^2=\mathrm{O}(r_t^{\varepsilon}d_t),\ \text{a.s. }\forall\,\varepsilon>0, \tag{35}$$

where r_t and d_t are defined by (19) and (5), respectively.

Proof 2 Define

$$K_{t+1}=M_0\sum_{i=0}^{t}\alpha^{t-i}\alpha_i\delta_i L_i,\ K_0=0, \tag{36}$$

then

$$K_{t+1}=\alpha K_t+M_0\alpha_t\delta_t L_t, \tag{37}$$

and from (22), we have

$$L_{t+1}\leqslant\beta L_t+K_{t+1}+\xi_t=$$
$$\beta L_t+\alpha K_t+M_0\alpha_t\delta_t L_t+\xi_t. \tag{38}$$

Then, we define

$$\hat{L}_{t+1}=\beta\hat{L}_t+\alpha K_t+M_0\alpha_t\delta_t L_t+\xi_t,\ \hat{L}_0=L_0. \tag{39}$$

From (38), it is easy to see

$$L_{t+1}\leqslant\hat{L}_{t+1}.$$

Therefore, we have the following iterative equations:

$$\begin{bmatrix}\hat{L}_{t+1}\\ K_{t+1}\end{bmatrix}=\begin{bmatrix}\beta&\alpha\\ 0&\alpha\end{bmatrix}\begin{bmatrix}\hat{L}_t\\ K_t\end{bmatrix}+\begin{bmatrix}1\\1\end{bmatrix}M_0\alpha_t\delta_t L_t+\begin{bmatrix}1\\0\end{bmatrix}\xi_t. \tag{40}$$

Let $\mathcal{C}=\begin{bmatrix}\beta&\alpha\\0&\alpha\end{bmatrix}$ and $Z_t=\begin{bmatrix}\hat{L}_t\\K_t\end{bmatrix}$. Since $\mathcal{C}$ is a stable matrix, there exists some norm $\|\cdot\|$ such that $\lambda=\|\mathcal{C}\|<1$. Consider 1-norm $\|Z_t\|_1=\hat{L}_t+K_t$. According to the norm property, there exists a constant $M_2>0$ such that $\|Z_t\|_1\leqslant M_2\|Z_t\|$, then $\hat{L}_t\leqslant M_2\|Z_t\|$. Denote $M_1=\|\begin{bmatrix}1\\1\end{bmatrix}\|$, from (40), we have

$$\|Z_{t+1}\|\leqslant\|\mathcal{C}\|\cdot\|Z_t\|+\left\|\begin{bmatrix}1\\1\end{bmatrix}\right\|M_0\alpha_t\delta_t L_t+\left\|\begin{bmatrix}1\\0\end{bmatrix}\right\|\xi_t\leqslant$$
$$\lambda\|Z_t\|+M_1M_0\alpha_t\delta_t\hat{L}_t+\xi_t\leqslant$$
$$\lambda\|Z_t\|+M_2M_1M_0\alpha_t\delta_t\|Z_t\|+\xi_t=$$
$$(\lambda+M_2M_1M_0\alpha_t\delta_t)\|Z_t\|+\xi_t. \tag{41}$$

Let $M_3=M_2M_1M_0$, then

$$\|Z_{t+1}\|\leqslant\prod_{i=0}^{t}(\lambda+M_3\alpha_i\delta_i)\|Z_0\|+\sum_{i=0}^{t}\prod_{j=i+1}^{t}(\lambda+M_3\alpha_j\delta_j)\xi_j=$$
$$\lambda^{t+1}\prod_{i=0}^{t}(1+\lambda^{-1}M_3\alpha_i\delta_i)\|Z_0\|+\sum_{i=0}^{t}\lambda^{t-i}\prod_{j=i+1}^{t}(1+\lambda^{-1}M_3\alpha_j\delta_j)\xi_j. \tag{42}$$

By $\delta_t\to 0$, we know that for any $\varepsilon>0$, there exists i_0 large enough that

$$\lambda^{-1}M_3\sum_{j=i}^{t}\alpha_j\delta_j\leqslant\varepsilon\log r_t,\ \forall\, t\geqslant i\geqslant i_0.$$

According to this and inequality $1+x\leqslant\mathrm{e}^x(x\geqslant 0)$,

$\forall\, t \geqslant i \geqslant i_0$, we have

$$\prod_{j=i+1}^{t}(1+\lambda^{-1}M_3\alpha_j\delta_j) \leqslant$$

$$\exp(\lambda^{-1}M_3\sum_{j=i}^{t}\alpha_j\delta_j) \leqslant \exp(\varepsilon\log r_t) = r_t^{\varepsilon}.$$

Substituting this into (42) and using (23), we have

$$L_{t+1} \leqslant \hat{L}_{t+1} \leqslant M_2\|Z_{t+1}\| = \mathrm{O}(r_t^{\varepsilon}d_t\log r_t).$$

By the arbitrariness of ε, we have $L_{t+1} = \mathrm{O}(r_t^{\varepsilon}d_t)$. Then, from (34), we have

$$\|\varphi_t\|^2 = \mathrm{O}(r_t^{\varepsilon}d_t).$$

The proof is complete. QED.

Proof of Theorem 1

From (29), we have

$$\sum_{t=0}^{n-1}\|Y_{t+1}\|^2 = \mathrm{O}(\sum_{t=0}^{n-1}\sum_{i=0}^{t}\alpha^{t-i}\|\bar{\eta}_i\|^2) = \mathrm{O}(\sum_{t=0}^{n-1}\|\bar{\eta}_t\|^2). \tag{43}$$

Under Condition A1), it is easy to know[5] that

$$\sum_{t=0}^{n-1}w_{t+1}^2 = \mathrm{O}(n). \tag{44}$$

Moreover, from (25)–(26)(43)–(44), Lemmas 1 and 3, we have

$$\begin{aligned}&\sum_{t=0}^{n-1}\|Y_{t+1}\|^2 =\\ &\mathrm{O}(\sum_{t=0}^{n-1}(D(z)\tilde{\theta}_t^{\mathrm{T}}\varphi_t)^2) + \mathrm{O}(\sum_{t=0}^{n-1}\|\eta_t\|^2) =\\ &\mathrm{O}(\sum_{t=0}^{n-1}\alpha_t) + \mathrm{O}(\sum_{t=0}^{n-1}\alpha_t\delta_t\|\varphi_t\|^2) + \mathrm{O}(n) =\\ &\mathrm{O}(\max_{0\leqslant i\leqslant n}\{\delta_i r_i^{\varepsilon}d_i\}\log r_n) + \mathrm{O}(n) =\\ &\mathrm{O}(r_n^{\varepsilon}d_n) + \mathrm{O}(n).\end{aligned} \tag{45}$$

By the definition of Y_{t+1}, we have

$$\sum_{t=0}^{n-1}y_{t+1}^2 = \mathrm{O}(r_n^{\varepsilon}d_n) + \mathrm{O}(n).$$

From this and Condition A2), we have from (33)

$$\sum_{t=0}^{n-1}u_{it}^2 = \mathrm{O}(r_n^{\varepsilon}d_n) + \mathrm{O}(n),\ i = 1,\cdots,r.$$

Therefore,

$$\sum_{t=0}^{n-1}\|\varphi_t\|^2 = \sum_{t=0}^{n-1}(\sum_{i=0}^{p-1}y_{t-i}^2 + \sum_{k=1}^{r}\sum_{i=0}^{q_k-1}u_{k(t-i)}^2) = \mathrm{O}(r_n^{\varepsilon}d_n) + \mathrm{O}(n). \tag{46}$$

Then, by (46) and (6) we have for any $\varepsilon > 0$,

$$r_n = 1 + \sum_{t=0}^{n-1}\|\varphi_t\|^2 = \mathrm{O}(r_n^{\varepsilon}d_n) + \mathrm{O}(n) = \mathrm{O}(r_n^{\varepsilon}n^{\delta}) + \mathrm{O}(n),\ \forall\,\delta \in (\frac{2}{\kappa}, 1). \tag{47}$$

Take ε small enough that $\varepsilon + \delta < 1$. We have

$$\frac{r_n}{n} = \mathrm{O}((\frac{r_n}{n})^{\varepsilon}\frac{1}{n^{1-\varepsilon-\delta}}) + \mathrm{O}(1) = \mathrm{O}(1) + \mathrm{o}((\frac{r_n}{n})^{\varepsilon}).$$

Hence, $r_n = \mathrm{O}(n)$. That is

$$\frac{1}{n}\sum_{t=0}^{n-1}(y_t^2 + u_{1t}^2 + \cdots + u_{rt}^2) = \mathrm{O}(1),\ \text{a.s..}$$

The proof is complete. QED.

Proof of Theorem 2

First, we prove

$$(\tilde{\theta}_t^{\mathrm{T}}\varphi_t)^2 = \overline{\mathrm{o}(1)}.$$

In fact, using Lemma 3 we have

$$\begin{aligned}&\sum_{t=0}^{n-1}(\tilde{\theta}_t^{\mathrm{T}}\varphi_t)^2 = \mathrm{O}(\sum_{t=0}^{n-1}\alpha_t) + \mathrm{O}(\sum_{t=0}^{n-1}\alpha_t\delta_t\|\varphi_t\|^2) =\\ &\mathrm{O}(\log r_n) + \mathrm{O}(\max_{0\leqslant t\leqslant n-1}\{\delta_t r_t^{\varepsilon}d_t\}\log r_t) =\\ &\mathrm{O}(r_n^{\varepsilon}d_n\log r_n) = \mathrm{O}(n^{\varepsilon+\delta}\log r_n).\end{aligned} \tag{48}$$

Take ε small enough that $\epsilon + \delta < 1$, it is easy to know that

$$\frac{1}{n}\sum_{t=0}^{n-1}(\tilde{\theta}_t^{\mathrm{T}}\varphi_t)^2 = \mathrm{o}(1).$$

Next, we prove that for any action $u'_{1t} \in \mathbb{R}$,

$$J_{1t}[U_{1t}\ \cdots\ U_{rt}] = \overline{\mathrm{o}(1)} + \inf_{u'_{1t}\in\mathbb{R}}J_{1t}[U'_{1t}\ \cdots\ U_{rt}].$$

From (14) and (16), we have

$$\begin{aligned}&J_{1t}[U'_{1t}\ \cdots\ U_{rt}] =\\ &\sigma^2 + (\theta^{\mathrm{T}}\varphi_t + b_{11}u'_{1t} + b_{21}u_{2t} + \cdots + b_{r1}u_{rt} -\\ &y^*_{1(t+1)})^2 + \lambda_1 u'^2_{1t} =\\ &(\lambda_1 + b_{11}^2)(u'_{1t} + \frac{b_{11}}{\lambda_1 + b_{11}^2}(\theta_t^{\mathrm{T}}\varphi_t + b_{21}u_{2t} + \cdots +\\ &b_{r1}u_{rt} - y^*_{1(t+1)} + \tilde{\theta}_t^{\mathrm{T}}\varphi_t))^2 + \min =\\ &(\lambda_1 + b_{11}^2)(u'_{1t} - u_{1t} + \frac{b_{11}\tilde{\theta}_t^{\mathrm{T}}\varphi_t}{\lambda_1 + b_{11}^2})^2 + \min,\end{aligned} \tag{49}$$

where $\min = \dfrac{\lambda_1}{\lambda_1 + b_{11}^2}(\theta_t^{\mathrm{T}}\varphi_t + b_{21}u_{2t} + \cdots + b_{r1}u_{rt} - y^*_{1(t+1)} + \tilde{\theta}_t^{\mathrm{T}}\varphi_t)^2 + \sigma^2$.

From this, it is obvious that

$$\min_{u'_{1t}\in\mathbb{R}}J_{1t}[U'_{1t}\ \cdots\ U_{rt}] = \min.$$

Then, the adaptive strategy profile can make the first player's one-step-ahead payoff function satisfy

$$\begin{aligned}&J_{1t}[U_{1t}\ \cdots\ U_{rt}] =\\ &\frac{b_{11}^2(\tilde{\theta}_t^{\mathrm{T}}\varphi_t)^2}{\lambda_1 + b_{11}^2} + \min_{u'_{1t}}J_{1t}[U'_{1t}\ \cdots\ U_{rt}] =\\ &\overline{\mathrm{o}(1)} + \min_{u'_{1t}}J_{1t}[U'_{1t}\ \cdots\ U_{rt}].\end{aligned} \tag{50}$$

Similarly, we can also prove that for any $i = 1,\cdots,r$, for its any action u'_{it},

$$J_{it}[U_{1t}\ \cdots\ U_{rt}] =$$

$$\overline{\mathrm{o}(1)} + \min_{u'_{it}\in\mathbb{R}} J_{it}[U_{1t}\ \cdots\ U'_{it}\ \cdots\ U_{rt}].$$

The proof is complete. QED.

4 Concluding remarks

Dynamic game theory has been investigated extensively over the past several decades from various aspects, and also has been widely used in the study of many practical systems. However, less research attention has been paid in theory to the case where the mathematical model contains parametric or nonparametric uncertainties. This paper is a continuation and extension of the authors' research on adaptive game theory. Due to the theoretical difficulties for analysing complicated nonlinear stochastic dynamical systems resulted from general adaptive game problems, there are still a number of problems remain to be investigated in the future, for examples, it would be interesting to extend the results of the paper to the case of general multi-input and multi-output linear stochastic systems, the case of time-varying unknown parameters and nonlinear uncertain systems, and to the situations of other type of payoff functions, etc.

References:

[1] FUDENBERG D, TIROLE J. *Game Theory* [M]. England: The MIT Press, 1991.

[2] BASAR T, OLSDER G J. *Dynamic Noncooperative Game Theory* [M]. New York: Academic Press, 1982.

[3] YEUNG D W K, PETROSYAN L A. *Cooperative Stochastic Differential Games* [M]. New York: Spinger, 2006.

[4] ÅSTRÖM K J, WITTENMARK B. *Adaptive Control* [M]. New Jersey: Addison-Wesley Publishing Company, 1994.

[5] KUMAR P R, VARAJA P. *Stochastic Systems: Estimation, Identification and Adaptive Control* [M]. Englewood Cliffs, NJ: Prentice Hall, 1986.

[6] CHEN H F, GUO L. *Identification and Stochastic Adaptive Control* [M]. Boston: Birkhäuser, 1991.

[7] KRSTIĆ M, KANELLAKOPOULOS I, KOKOTOVIĆ P. *Nonlinear and Adaptive Control Design* [M]. New York: John Wiley & Sons, 1995.

[8] BASAR T, BERNHARD P. H_∞ *Optimal Control and Related Minimax Design Problem* [M]. Boston: Birkhäuser, 2008.

[9] MU Y, GUO L. Optimization and identification in a non-equilibrium dynamic game [C] //*Proceedings of the 48th IEEE Conference on Decision and Control*. Shanghai: IEEE, 2009: 5750 – 5755.

[10] MU Y, GUO L. Towards a theory of game-based non-equilibrium control systems [J]. *Journal of Systems Science and Complexity*, 2012, 25(2): 209 – 226.

[11] LI Y, GUO L. Towards a theory of stochastic adaptive differential games [C] //*Proceedings of the 50th IEEE Chinese Conference on Decision and Control and European Control Conference*. Orlando, FL, USA: IEEE, 2011: 5041 – 5046.

[12] LI Y, GUO L. Convergence of adaptive linear stochastic differential games: nonzero-sum case [C] //*Proceedings of the 10th World Congress on Intelligent Control and Automation*. Beijing: IEEE, 2012: 3543 – 3548.

[13] HO Y C. Differential games, dynamic optimization, and generalized control theory [J]. *Journal of Optimization Theory and Applications*, 1979, 6(3): 179 – 209.

[14] BAGCHI A. *Stackelberg Differential Games in Economic Models* [M]. Berlin; New York: Springer-Verlag, 1984.

[15] SIMTH J M. *Evolution and the Theory of Games* [M]. Cambridge; New York: Cambridge University Press, 1982.

[16] LU Qiang, CHEN Laijun, MEI Shengwei. Typical applications and prospects of game theory in power system [J]. *Proceedings of the Chinese Society for Electrical Engineering*, 2014, 34(29): 5009 – 5017.
(卢强, 陈来军, 梅生伟. 博弈论在电力系统中典型应用及若干展望 [J]. 中国电机工程学报, 2014, 34(29): 5009 – 5017.)

[17] SAAD W, HAN Z, POOR H V, et al. Game-theoretic methods for the smart grid [J]. *IEEE Signal Proceeding Magazine*, 2012, 29(5): 86 – 105.

[18] MEI Shengwei, GUO Wentao, WANG Yingying, et al. A game model for robust optimization of power systems and its application [J]. *Proceedings of the Chinese Society for Electrical Engineering*, 2013, 33(19): 47 – 56.
(梅生伟, 郭文涛, 王莹莹, 等. 一类电力系统鲁棒优化问题的博弈模型及应用实例 [J]. 中国电机工程学报, 2013, 33(19): 47 – 56.)

[19] YUAN Shuo, GUO Lei. Stochastic adaptive dynamical games[J]. *Scientia Sinica Mathematica*, 2016, 46(10): 1367 – 1382.
(袁硕, 郭雷. 随机自适应动态博弈 [J]. 中国科学: 数学, 2016, 46(10): 1367 – 1382.)

[20] HU H, GUO L. Stability and convergence of non-cooperative stochastic adaptive games [C] //*Proceedings of the 35th Chinese Control Conference*. Chengdu: IEEE, 2016: 10396 – 10401.

[21] GUO L, CHEN H F. The Åström-Wittenmark self-tuning regulator revisited and ELS based adaptive trackers [J]. *IEEE Transaction on Automatic Control*, 1991, 36(7): 802 – 812.

[22] GOODWIN G C, SIN K S. *Adaptive Filtering, Prediction and Control* [M]. Englewood Cliffs, NJ: Prentice-Hall, 1984.

[23] GUO L. Convergence and logarithm laws of self-tuning regulators [J]. *Automatica*, 1995, 31(3): 435 – 450.

[24] GUO Lei. *Time-varing Stochastic System: Stability, Estimation and Control* [M]. Changchun: Jilin Science and Technology Press, 1994.
(郭雷. 时变随机系统: 稳定性, 估计与控制 [M]. 长春: 吉林科技出版社, 1994.)

作者简介:

胡浩洋 (1989–), 男, 博士研究生, 目前主要研究方向为随机自适应博弈, E-mail: hhylff@amss.ac.cn;

郭 雷 (1961–), 男, 中国科学院数学与系统科学研究院研究员, 中国科学院院士, 发展中国家科学院院士, 瑞士皇家工程科学院外籍院士, IEEE Fellow, IFAC Fellow, 目前主要研究兴趣包括博弈控制系统、分布式自适应滤波、反馈机制最大能力、PID控制理论、量子控制系统等, E-mail: lguo@amss.ac.cn.