中国科学: 数学 2016 年 第 46 卷 第 10 期: 1367 ∼ 1382
SCIENTIA SINICA Mathematica

《中国科学》杂志社
SCIENCE CHINA PRESS

随机自适应动态博弈

献给陈翰馥教授 80 华诞

袁硕, 郭雷*

中国科学院数学与系统科学研究院系统控制重点实验室, 北京 100190
E-mail: yuanshuo15@mails.ucas.ac.cn, lguo@amss.ac.cn

收稿日期: 2015-11-15; 接受日期: 2016-01-07; 网络出版日期: 2016-09-30; * 通信作者
国家数学与交叉科学中心资助项目

摘要 在不确定性结构和环境下的动态博弈是重要的复杂系统问题, 然而现有文献少有理论研究. 本文将博弈论思想与自适应控制方法相结合, 考虑系数未知的线性随机系统在输出跟踪二次型指标下, 两人零和自适应博弈问题. 具体来讲, 先用最小二乘算法对未知参数进行在线估计, 再根据 "必然等价原则" 设计博弈策略组合. 本文在随机适应控制理论基础上, 证明了相应闭环博弈系统是全局稳定的, 并且行动组合在一定意义下是博弈问题的渐近 Nash 均衡解.

关键词 随机系统 动态博弈 适应控制 稳定性 渐近 Nash 均衡 最小二乘

MSC (2010) 主题分类 91A25, 93C40

1 引言

过去半个世纪, 博弈论和控制论无论在理论上还是应用上都取得了巨大发展, 显著提高了人类认识世界和改造世界的水平. 博弈论, 主要是微分博弈, 虽然与控制论有密切联系, 但是两者在很大程度上是平行发展的 (参见文献 [1–3]). 现有控制理论主要是研究被控对象没有博弈行为的系统控制问题, 而对于系统内存在博弈行为的控制问题, 迄今探讨很少 (参见文献 [4–11]), 尽管鲁棒控制问题也可以看作博弈问题[12]. 在现实生活中, 动态博弈问题比比皆是, 如追踪 - 逃避问题、市场经济中企业之间的竞争、军事上或网络上的攻防、病毒侵害与防毒杀毒等. 现有微分博弈论为了研究问题方便, 往往假设博弈模型中参数已知[1–3], 但这样的假设过于理想化. 现实中, 博弈双方不但会面临较大不确定性, 并且还有自己的目标追求或利益诉求, 因此, 我们既不能单纯套用传统控制论的观点来处理, 也无法只用经典博弈论的方法来解决. 这类问题的研究探索至少可以追溯到文献 [13–15]. 之后, 文献 [16,17] 研究了完全状态信息下由连续时间状态空间模型描述的系统的自适应博弈问题. 但由于该问题在数学上的困难性, 为了能得到保证系统稳定性的策略, 文献 [16,17] 需要的条件比较苛刻, 例如, 要求系统开环稳定和/或系统矩阵 A 信息已知等. 如何解决这些困难, 仍需要进一步探索.

本文将从新的角度, 采用由输入输出模型描述的线性随机博弈系统, 并在有两位参与者进行 "有限理性" 决策的情形下, 考虑关于输出跟踪二次型指标的两人零和博弈问题. 我们将在不假定系统开

英文引用格式: Yuan S, Guo L. Stochastic adaptive dynamical games (in Chinese). Sci Sin Math, 2016, 46: 1367–1382, doi: 10.1360/N012015-00355

© 2016《中国科学》杂志社 www.scichina.com mathcn.scichina.com

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

环稳定, 也不需要结构的过多先验信息情形下, 借鉴自适应控制方法, 先用最小二乘在线估计系统的未知参数, 再根据 "必然等价原则 (certainty equivalence principle)" 设计博弈策略组合, 最后证明相应的闭环博弈系统具有全局稳定性, 且该策略组合具有一定的优化性, 即行动组合在一定意义下是博弈问题的渐近 Nash 均衡解.

2 主要结果

考虑下列具有两个参与者输入的线性随机系统:

$$A(z)y_t = B(z)u_{t-1} + C(z)v_{t-1} + \omega_t, \quad t \geqslant 0, \tag{2.1}$$

其中 $\{y_t\}$ 和 $\{\omega_t\}$ 分别是系统的输出和噪声过程, $\{u_t\}$ 是参与者 1 的决策输入, $\{v_t\}$ 是参与者 2 的决策输入 (不失一般性, 假设 $y_t = \omega_t = u_t = v_t = 0, \forall\, t < 0$), $A(z)$、$B(z)$ 和 $C(z)$ 是后移算子 z 的多项式,

$$A(z) = 1 + a_1 z + \cdots + a_p z^p, \quad p \geqslant 0, \tag{2.2}$$

$$B(z) = b_1 + b_2 z + \cdots + b_q z^{q-1}, \quad q \geqslant 1, \tag{2.3}$$

$$C(z) = c_1 + c_2 z + \cdots + c_r z^{r-1}, \quad r \geqslant 1, \tag{2.4}$$

这里 a_i、b_j 和 c_k 是系数, 且 $b_1 \neq 0$, $c_1 \neq 0$, p、q 和 r 是系统阶数的已知上界. 假设 $\{\omega_t\}$ 满足条件:

(A1) 噪声 $\{\omega_t, \mathcal{F}_t\}$ 是鞅差序列 ($\mathcal{F}_t$ 是非降的子 σ- 代数序列), 且有

$$E[\omega_{t+1}^2 \mid \mathcal{F}_t] = \sigma^2 > 0, \quad \text{a.s.}, \tag{2.5}$$

并存在 $\kappa > 2$ 使得

$$\sup_t E[|\omega_{t+1}|^\kappa \mid \mathcal{F}_t] < \infty, \quad \text{a.s.} \tag{2.6}$$

引入参数向量

$$\theta = [-a_1 \cdots -a_p \ b_2 \cdots b_q \ c_2 \cdots c_r]^{\mathrm{T}}, \tag{2.7}$$

及相应的回归向量

$$\varphi_t = [y_t \cdots y_{t-p+1} \ u_{t-1} \cdots u_{t-q+1} \ v_{t-1} \cdots v_{t-r+1}]^{\mathrm{T}}, \tag{2.8}$$

则 (2.1) 可简写为

$$y_{t+1} = \theta^{\mathrm{T}} \varphi_t + b_1 u_t + c_1 v_t + \omega_{t+1}, \quad t \geqslant 0. \tag{2.9}$$

进一步, 考虑下列指标 (收益函数):

$$J(u, v) = \limsup_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n} J_t(u_t, v_t), \tag{2.10}$$

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

其中一步博弈指标为

$$J_t(u_t, v_t) = E[(y_{t+1} - y_{t+1}^*)^2 + \lambda_1 u_t^2 - \lambda_2 v_t^2 \mid \mathcal{F}_t], \tag{2.11}$$

这里 λ_1 和 λ_2 是大于 0 的加权常数, $\{y_t^*\}$ 是已知参考信号. 在上述指标中, 两个参与者在考虑各自能量耗费情形下分别极小化和极大化系统输出跟踪误差. 典型的追踪 - 逃避问题等就属于此类问题.

假设系统内部参数 θ 对两个参与者来讲均为未知信息, "有限理性" 的参与者针对指标 $J(u,v)$ 进行零和博弈, 在任何时刻 t, 基于测量数据 $\{y_0 \cdots y_t, u_0 \cdots u_{t-1}, v_0 \cdots v_{t-1}\}$ 构造出决策输入 u_t 和 v_t, 其中参与者 1 的目的是让指标极小, 参与者 2 的目的是让指标极大, 且参与者只能做一步最优决策. 假设两人共享同一参数估计器给出的信息, 我们感兴趣是, 是否存在自适应博弈策略组合, 使相应的闭环系统是全局稳定的, 且策略组合具有一定优化性.

这里所说的 "有限理性" 一则是指参与者面对的环境是复杂的、不确定的, 因而获得的信息不完全; 二则是指参与者对环境的计算能力和理解能力有限, 即使有了全部的信息也不一定能做出全局最优的决策. 本文假设参与者是 "有限理性" 的, 即参与者会在博弈中选择使自己当前 "局部" 指标 $J_t(\cdot,\cdot)$ 达到最优的策略. 由于各参与者之间相互制衡, 一般不能使每个参与者都能达到理想的最优值, 这跟最优控制是有所不同的, 但是, 往往可以达到某种 "均衡" 状态, 这种 "均衡" 称作博弈的解或结果.

众所周知, 稳定性是动态系统最重要的性质之一, 是保证系统正常运行并达到控制目标的首要条件. 对于随机系统 (2.1), 下面给出其全局稳定定义.

定义 2.1 考虑随机系统 (2.1), 如果策略组合 (u,v) (其中 $u=\{u_t\}$, $v=\{v_t\}$) 使得闭环系统对任何初始值 $y_0 \in \mathbb{R}$, 满足

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{t=0}^{n}(y_t^2 + u_t^2 + v_t^2) = O(1), \quad \text{a.s.},$$

则称闭环系统是全局稳定的.

为了建立相应理论, 除条件 (A1) 外, 再引入以下条件:

(A2) $D(z) \neq 0, \forall z: |z| \leqslant 1$, 其中

$$D(z) = B(z) - \frac{\lambda_1 c_1}{\lambda_2 b_1} C(z);$$

(A3) $E(z) \neq 0, \forall z: |z| \leqslant 1$, 其中

$$E(z) = \frac{\lambda_2 b_1}{\lambda_1\lambda_2 + \lambda_2 b_1^2 - \lambda_1 c_1^2}\left[D(z) + \frac{\lambda_1}{b_1} A(z)\right];$$

(A4) $\{y_t^*\}$ 是确定的或随机的有界参考信号序列, 并且与 $\{\omega_t\}$ 独立;

(A5) $\lambda_1\lambda_2 + \lambda_2 b_1^2 - \lambda_1 c_1^2 \neq 0$, 且 $c_1^2 - \lambda_2 < 0$.

注 2.1 条件 (A2) 和 (A3) 可看作经典最小相位条件的推广, 是保证闭环系统内部稳定性的必要条件, 即使在参数已知情形下, 这两个条件也是需要的 (参见文献 [18]). 此外, 从下面的分析容易看出, 条件 (A5) 第一个不等式是参数已知情形下 Nash 均衡存在的必要条件, 第二个不等式是保证两个参与者进行零和博弈的参数条件.

我们进一步假定 $\{d_t\}$ 是一列非降的确定性正数序列, 满足

$$\omega_t^2 = O(d_t), \quad \text{a.s.}, \quad d_{t+1} = O(d_t). \tag{2.12}$$

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

易证当 $\{\omega_t\}$ 满足条件 (A1) 时, $\{d_t\}$ 可取为

$$d_t = t^{\delta}, \quad \forall\, \delta \in \left(\frac{2}{\kappa}, 1\right), \tag{2.13}$$

其中 κ 由 (2.6) 给出 (参见文献 [19]).

下面介绍关于定常参数系统估计方法中最常见的一类自适应估计—最小二乘估计. 对于线性随机系统 (2.9), 用递推最小二乘方法[4,7] 可得到未知参数 θ 的递推估计值 θ_t,

$$\theta_{t+1} = \theta_t + a_t P_t \varphi_t (y_{t+1} - b_1 u_t - c_1 v_t - \theta_t^{\mathrm{T}} \varphi_t), \tag{2.14}$$

$$P_{t+1} = P_t - a_t P_t \varphi_t \varphi_t^{\mathrm{T}} P_t, \tag{2.15}$$

$$a_t = (1 + \varphi_t^{\mathrm{T}} P_t \varphi_t)^{-1}, \tag{2.16}$$

其中初值 θ_0 和 $P_0 > 0$ 任取. 定义参数向量

$$\theta_t = [-a_{1t} \cdots - a_{pt}\ b_{2t} \cdots b_{qt}\ c_{2t} \cdots c_{rt}]^{\mathrm{T}}. \tag{2.17}$$

在参与者只能做一步最优的假设下, 考虑到参与者 1 想让指标 $J(u,v)$ 极小且参与者 2 想让指标 $J(u,v)$ 极大的要求, 我们将通过下述方法设计策略组合.

由 (2.9)、(2.11)、条件 (A1) 和 (A4) 可得

$$J_t(u_t, v_t) = \sigma^2 + (\theta^{\mathrm{T}} \varphi_t + b_1 u_t + c_1 v_t - y_{t+1}^*)^2 + \lambda_1 u_t^2 - \lambda_2 v_t^2, \tag{2.18}$$

上式分别对 u_t 和 v_t 求偏导并令导数为零, 可得

$$\begin{cases} u_t = -\dfrac{\lambda_2 b_1 (\theta^{\mathrm{T}} \varphi_t - y_{t+1}^*)}{\lambda_1 \lambda_2 + \lambda_2 b_1^2 - \lambda_1 c_1^2}, \\ v_t = \dfrac{\lambda_1 c_1 (\theta^{\mathrm{T}} \varphi_t - y_{t+1}^*)}{\lambda_1 \lambda_2 + \lambda_2 b_1^2 - \lambda_1 c_1^2}. \end{cases} \tag{2.19}$$

再考虑到参数 θ 未知, 根据 "必然等价原则", 用最小二乘估计算法得到的 θ_t 取代未知参数 θ, 得到自适应行动组合

$$\begin{cases} u_t = -\dfrac{\lambda_2 b_1 (\theta_t^{\mathrm{T}} \varphi_t - y_{t+1}^*)}{\lambda_1 \lambda_2 + \lambda_2 b_1^2 - \lambda_1 c_1^2}, \\ v_t = \dfrac{\lambda_1 c_1 (\theta_t^{\mathrm{T}} \varphi_t - y_{t+1}^*)}{\lambda_1 \lambda_2 + \lambda_2 b_1^2 - \lambda_1 c_1^2}. \end{cases} \tag{2.20}$$

由此可知策略组合 (u,v) 满足

$$\theta_t^{\mathrm{T}} \varphi_t + b_1 u_t + c_1 v_t - y_{t+1}^* = -\frac{\lambda_1}{b_1} u_t = \frac{\lambda_1 \lambda_2 (\theta_t^{\mathrm{T}} \varphi_t - y_{t+1}^*)}{\lambda_1 \lambda_2 + \lambda_2 b_1^2 - \lambda_1 c_1^2}, \tag{2.21}$$

$$v_t = -\frac{\lambda_1 c_1}{\lambda_2 b_1} u_t. \tag{2.22}$$

定理 2.1 考虑系统 (2.1), 并设条件 (A1)–(A5) 满足, 参数 θ 未知, 则两位参与者在采取自适应行动组合 (2.20) 所定义的策略组合时, 相应的闭环系统是全局稳定的, 即对任何初始值 $y_0 \in \mathbb{R}$, 都有

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n} (y_t^2 + u_t^2 + v_t^2) = O(1), \quad \text{a.s.} \tag{2.23}$$

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

上面我们利用自适应控制的方法, 处理含未知参数的两人线性随机系统零和博弈问题, 得到了使得相应闭环系统具有全局稳定性的策略组合. 那么, 一个自然的问题是, 该策略组合是否具有一定的优化性呢? 事实上, 从指标 $J_t(u_t,v_t)$ 的定义可以看出, $J_t(u_t,v_t)$ 不仅依赖于参与者当前的行动 u_t 和 v_t, 还与参与者之前的行动 $u_0,u_1,\dots,u_{t-1}$ 和 $v_0,v_1,\dots,v_{t-1}$ 有关. 接下来, 我们将考虑采取自适应行动组合 (2.20), 在时刻 t 时, 参与者之前的行动分别为 $u_0,u_1,\dots,u_{t-1}$ 和 $v_0,v_1,\dots,v_{t-1}$ 的情形下, 行动组合 (u_t,v_t) 是否是以 $J_t(u_t,v_t)$ 为零和博弈指标的渐近 Nash 均衡解, 答案是肯定的. 下面给出渐近 Nash 均衡的定义.

首先引进一个符号 $\overline{o(1)}$: 设 $\{x_t,t\geqslant 0\}$ 是一个实数序列, 如果当 $n\to\infty$ 时满足

$$\frac{1}{n}\sum_{t=0}^{n}x_t=o(1),$$

则记 $x_t=\overline{o(1)}$.

为叙述方便, 我们假设两个参与者进行有记忆的重复博弈, 其策略组合为 (u,v), $u=\{u_t\}$, $v=\{v_t\}$, (u_t,v_t) 的容许行动集为 (Ω_u,Ω_v), 指标为 $J_t(U_t,V_t)$, 其中

$$U_t=(u_1,u_2,\dots,u_t),\quad V_t=(v_1,v_2,\dots,v_t),$$

参与者 1 的目的是极小化指标, 参与者 2 的目的是极大化指标.

定义 2.2 如果下列渐近最优等式成立,

$$J_t(U_t,V_t)+\overline{o(1)}=\inf_{u_t'\in\Omega_u}J_t((U_{t-1},u_t'),V_t)=\sup_{v_t'\in\Omega_v}J_t(U_t,(V_{t-1},v_t')),$$

则称行动组合 (u_t,v_t) 是渐近 Nash 均衡.

下面的定理说明自适应行动组合 (2.20) 具有一定的优化性.

定理 2.2 考虑系统 (2.1) 和自适应博弈行动组合 (2.20), 假设条件 (A1)–(A5) 满足, 那么, 行动组合 (u_t,v_t) 是以 $J_t(u_t,v_t)$ 为零和博弈指标的渐近 Nash 均衡解, 其中 (u_t,v_t) 的容许行动集为 $(\mathbb{R},\mathbb{R})$.

3 稳定性分析

为证明定理 2.1, 我们需要几个引理.

首先, 引入全文通用的几个记号:

$$\alpha_t\triangleq\frac{(\widetilde{\theta}_t^{\mathrm{T}}\varphi_t)^2}{1+\varphi_t^{\mathrm{T}}P_t\varphi_t},\quad \delta_t\triangleq\mathrm{tr}(P_t-P_{t+1}),\tag{3.1}$$

$$r_t\triangleq 1+\sum_{i=0}^{t}\|\varphi_i\|^2,\quad \widetilde{\theta}_t^{\mathrm{T}}\triangleq\theta-\theta_t.\tag{3.2}$$

本文中范数 $\|\cdot\|$ 的定义, 除特别说明之处外, 均指 Euclid 范数.

下面的引理给出基于最小二乘估计的预报误差性质.

引理 3.1 [20] 对于线性回归随机系统 (2.1), 设 (A1)–(A4) 满足, 则由最小二乘算法 (2.14)–(2.16) 给出的预报误差 $\{\widetilde{\theta}_t^{\mathrm{T}}\varphi_t\}$ 有以下渐近性质:

$$\sum_{i=0}^{t}\alpha_i=O(\log r_t),\quad \text{a.s. } t\to\infty,\tag{3.3}$$

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

其中 α_t 和 r_t 分别由 (3.1) 和 (3.2) 所定义.

引理 3.2 考虑具有未知参数的系统 (2.1), 假设条件 (A1)–(A5) 满足, 则对自适应行动组合 (2.20) 所定义的策略组合, 一定存在正随机序列 $\{L_t\}$ 使得

$$y_t^2 \leqslant L_t, \quad \text{a.s.}\ \forall t, \tag{3.4}$$

且 $\{L_t\}$ 满足下述线性时变关系:

$$L_{t+1} \leqslant \beta L_t + k\sum_{i=0}^{t}\alpha^{t-i}\sum_{j=i-g+1}^{i}\alpha_j\delta_j L_j + \xi_t, \tag{3.5}$$

其中常数 $\alpha,\beta\in(0,1)$, 常数 $k>0$, α_t 和 δ_t 由 (3.1) 定义, $\{\xi_t\}$ 是正随机序列, 且

$$\xi_t = O(d_t\log r_t), \tag{3.6}$$

r_t 由 (3.2) 定义.

证明 采取行动组合 (2.20) 时, 由 (2.9)、(2.20) 和 (2.21) 得

$$\begin{aligned} y_{t+1} &= \widetilde{\theta}_t^{\mathrm{T}}\varphi_t + \theta_t^{\mathrm{T}}\varphi_t + b_1u_t + c_1v_t - y_{t+1}^* + \omega_{t+1} + y_{t+1}^* \\ &= \widetilde{\theta}_t^{\mathrm{T}}\varphi_t + \frac{\lambda_1\lambda_2(\theta^{\mathrm{T}}\varphi_t - y_{t+1}^*)}{\lambda_1\lambda_2+\lambda_2b_1^2-\lambda_1c_1^2} + \frac{\lambda_1\lambda_2(\theta_t-\theta)^{\mathrm{T}}\varphi_t}{\lambda_1\lambda_2+\lambda_2b_1^2-\lambda_1c_1^2} + \omega_{t+1} + y_{t+1}^*. \end{aligned} \tag{3.7}$$

进一步, 由 (2.2)–(2.4)、(2.7) 和 (2.8) 得

$$\theta^{\mathrm{T}}\varphi_t - y_{t+1}^* = (1-A(z))y_{t+1} + (B(z)-b_1)u_t + (C(z)-c_1)v_t - y_{t+1}^*. \tag{3.8}$$

由 (2.1) 和 (2.22) 可得

$$u_t = D^{-1}(z)(A(z)y_{t+1}-\omega_{t+1}), \tag{3.9}$$

$$v_t = -\frac{\lambda_1c_1}{\lambda_2b_1}D^{-1}(z)(A(z)y_{t+1}-\omega_{t+1}), \tag{3.10}$$

其中 $D(z)=B(z)-\frac{\lambda_1c_1}{\lambda_2b_1}C(z)$. 把 (3.8)–(3.10) 代入 (3.7) 并移项化简得

$$\left[D(z)+\frac{\lambda_1}{b_1}A(z)\right]y_{t+1} = D(z)\widetilde{\theta}_t^{\mathrm{T}}\varphi_t + \left(D(z)+\frac{\lambda_1}{b_1}\right)\omega_{t+1} + D(z)y_{t+1}^*. \tag{3.11}$$

记

$$E(z) = \frac{\lambda_2b_1}{\lambda_1\lambda_2+\lambda_2b_1^2-\lambda_1c_1^2}\left[D(z)+\frac{\lambda_1}{b_1}A(z)\right], \tag{3.12}$$

$$\eta_t = \frac{\lambda_2b_1}{\lambda_1\lambda_2+\lambda_2b_1^2-\lambda_1c_1^2}\left[\left(D(z)+\frac{\lambda_1}{b_1}\right)\omega_{t+1} + D(z)y_{t+1}^*\right], \tag{3.13}$$

$$\overline{\eta_t} = \frac{\lambda_2b_1}{\lambda_1\lambda_2+\lambda_2b_1^2-\lambda_1c_1^2}D(z)\widetilde{\theta}_t^{\mathrm{T}}\varphi_t + \eta_t, \tag{3.14}$$

则 (3.11) 可简记为

$$E(z)y_{t+1} = \overline{\eta_t}. \tag{3.15}$$

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

再记 $Y_{t+1}=[y_{t+1}\ y_t\ \cdots\ y_{t-h+2}]^{\mathrm{T}}$, 其中 $h=\max\{p,q-1,r-1\}$. 利用 (3.15), 可知存在矩阵 A 和 B, 使得

$$Y_{t+1}=AY_t+B\overline{\eta_t}, \tag{3.16}$$

其中 A 是 $h\times h$ 维方阵, $B=[1\ 0\ \cdots\ 0]^{\mathrm{T}}$ 是 h 维的列向量. 易证 $|zI-A|=z^hE(z^{-1})$, 由条件 (A3) 得矩阵 A 的特征值均在单位圆内, 所以矩阵 A 是稳定阵, 因此存在一个诱导范数 $\|\cdot\|$ 使得 $\|A\|<1$, 不妨设 $\alpha=\|A\|<1$ (参见文献 [21]).

进一步, 利用矩阵 A 的稳定性, 从 (3.16) 可得

$$\begin{aligned}\|Y_{t+1}\|^2&=\bigg\|A^{t+1}Y_0+\sum_{i=0}^{t}A^{t-i}B\overline{\eta_i}\bigg\|^2=O\bigg(\sum_{i=0}^{t}\alpha^{t-i}\overline{\eta_i}^2\bigg)+O(\alpha^{t+1})\\&=O\bigg(\sum_{i=0}^{t}\alpha^{t-i}\eta_i^2\bigg)+O\bigg(\sum_{i=0}^{t}\alpha^{t-i}(D(z)\widetilde{\theta}_i^{\mathrm{T}}\varphi_i)^2\bigg).\end{aligned} \tag{3.17}$$

注意到 η_i 的定义 (3.13), 上式第一项有估计

$$\begin{aligned}\sum_{i=0}^{t}\alpha^{t-i}\eta_i^2&=O\bigg(\sum_{i=0}^{t}\alpha^{t-i}\sum_{j=i-g+1}^{i}(\omega_{j+1}^2+{y_{j+1}^*}^2)\bigg)\\&=O\bigg(\sum_{i=0}^{t}\alpha^{t-i}\max_{i-g+1\leqslant j\leqslant i}\{\omega_{j+1}^2+{y_{j+1}^*}^2\}\bigg)\\&=O\bigg(\sum_{i=0}^{t}\alpha^{t-i}d_{i+1}\bigg)=O(d_t),\end{aligned} \tag{3.18}$$

其中 $g=\max\{q,r\}$. 根据 (3.1), (3.17) 第二项有估计

$$\begin{aligned}\sum_{i=0}^{t}\alpha^{t-i}(D(z)\widetilde{\theta}_i^{\mathrm{T}}\varphi_i)^2&=O\bigg(\sum_{i=0}^{t}\alpha^{t-i}\sum_{j=i-g+1}^{i}(\widetilde{\theta}_j^{\mathrm{T}}\varphi_j)^2\bigg)\\&=O\bigg(\sum_{i=0}^{t}\alpha^{t-i}\sum_{j=i-g+1}^{i}\alpha_j[1+\varphi_j^{\mathrm{T}}P_{j+1}\varphi_j+\varphi_j^{\mathrm{T}}(P_j-P_{j+1})\varphi_j]\bigg)\\&=O\bigg(\sum_{i=0}^{t}\alpha^{t-i}\sum_{j=i-g+1}^{i}\alpha_j[2+\delta_j\|\varphi_j\|^2]\bigg)\\&=O\bigg(\sum_{i=0}^{t}\alpha^{t-i}\sum_{j=i-g+1}^{i}\alpha_j\delta_j\|\varphi_j\|^2\bigg)+O(\log r_t),\end{aligned} \tag{3.19}$$

其中我们用到了性质 $\varphi_j^{\mathrm{T}}P_{j+1}\varphi_j\leqslant 1$ 和 (3.3).

根据条件 (A2), 采取行动组合 (2.20) 时, 从 (2.1) 易知, 存在 $\beta\in(0,1)$, 使得

$$u_{t-j}^2=O\bigg(\sum_{i=0}^{t}\beta^{t-i}y_i^2\bigg)+O(d_t),\quad j=1,2,\dots,q-1, \tag{3.20}$$

$$v_{t-j}^2=O\bigg(\sum_{i=0}^{t}\beta^{t-i}y_i^2\bigg)+O(d_t),\quad j=1,2,\dots,r-1, \tag{3.21}$$

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

因此,

$$\|\varphi_t\|^2=\sum_{i=0}^{p-1}y_{t-i}^2+\sum_{i=1}^{q-1}u_{t-i}^2+\sum_{i=1}^{r-1}v_{t-i}^2=O\bigg(\sum_{i=0}^{t}\beta^{t-i}y_i^2\bigg)+O(d_t). \tag{3.22}$$

定义

$$L_t=\sum_{i=0}^{t}\beta^{t-i}\|Y_i\|^2, \tag{3.23}$$

显然, (3.4) 成立, 且由 (3.22) 得

$$\|\varphi_t\|^2=O(L_t)+O(d_t). \tag{3.24}$$

进一步, 由

$$\sum_{j=0}^{\infty}\delta_j=\sum_{j=0}^{\infty}(\mathrm{tr}P_j-\mathrm{tr}P_{j+1})\leqslant \mathrm{tr}P_0<\infty, \tag{3.25}$$

可知 $\delta_j\to 0$. 将 (3.24) 代入 (3.19), 并通过 (3.17) 和 (3.18) 得

$$\begin{aligned}\|Y_{t+1}\|^2&=O\bigg(\sum_{i=0}^{t}\alpha^{t-i}\sum_{j=i-g+1}^{i}\alpha_j\delta_j(L_j+d_j)\bigg)+O(\log r_t+d_t)\\&=O\bigg(\sum_{i=0}^{t}\alpha^{t-i}\sum_{j=i-g+1}^{i}\alpha_j\delta_jL_j\bigg)+O\Big(\max_{0\leqslant j\leqslant t}\{\delta_jd_j\}\log r_t\Big)+O(\log r_t+d_t)\\&\leqslant k\sum_{i=0}^{t}\alpha^{t-i}\sum_{j=i-g+1}^{i}\alpha_j\delta_jL_j+O(d_t\log r_t),\end{aligned} \tag{3.26}$$

其中 $k>0$ 是某个常数. 根据上式以及 L_t 的定义, 得

$$L_{t+1}=\beta L_t+\|Y_{t+1}\|^2\leqslant \beta L_t+k\sum_{i=0}^{t}\alpha^{t-i}\sum_{j=i-g+1}^{i}\alpha_j\delta_jL_j+O(d_t\log r_t). \tag{3.27}$$

引理得证. □

引理 3.3 在引理 3.2 的条件下, 下式成立:

$$\|\varphi_t\|^2=O(r_t^{\epsilon}d_t),\quad \text{a.s. }\forall\,\epsilon>0, \tag{3.28}$$

其中 r_t 和 d_t 分别由 (3.2) 和 (2.13) 定义.

证明 记 $S_i=k\sum_{j=i-g+1}^{i}\alpha_j\delta_jL_j$, 则由引理 3.2 有

$$L_{t+1}\leqslant \beta L_t+\sum_{i=0}^{t}\alpha^{t-i}S_i+\xi_t, \tag{3.29}$$

其中 ξ_t 由 (3.6) 定义. 记

$$K_{t+1}=\alpha K_t+S_t,\quad K_0=0, \tag{3.30}$$

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

易证

$$K_{t+1}=\sum_{i=0}^{t}\alpha^{t-i}S_i,$$

因此, 由 (3.29) 有

$$L_{t+1}\leqslant \beta L_t+K_{t+1}+\xi_t=\beta L_t+\alpha K_t+S_t+\xi_t. \tag{3.31}$$

现定义序列 $\{\widehat{L}_t\}$ 如下:

$$\widehat{L}_{t+1}=\beta\widehat{L}_t+\alpha K_t+S_t+\xi_t,\quad \widehat{L}_0=L_0, \tag{3.32}$$

显然,

$$L_{t+1}\leqslant \widehat{L}_{t+1}. \tag{3.33}$$

进一步, 由 (3.30) 和 (3.32) 得

$$\begin{pmatrix}\widehat{L}_{t+1}\\ K_{t+1}\end{pmatrix}=\begin{pmatrix}\beta & \alpha\\ 0 & \alpha\end{pmatrix}\begin{pmatrix}\widehat{L}_t\\ K_t\end{pmatrix}+\begin{pmatrix}1\\ 1\end{pmatrix}S_t+\begin{pmatrix}1\\ 0\end{pmatrix}\xi_t. \tag{3.34}$$

不妨设 $C=\left(\begin{smallmatrix}\beta & \alpha\\ 0 & \alpha\end{smallmatrix}\right)$, 则易知矩阵 C 的特征值均在单位圆内, 因此 C 是稳定阵, 故存在一个诱导范数 $\|\cdot\|$ 使得 $\|C\|<1$, 不妨记 $\gamma=\|C\|$. 再记 $Z_t=\left(\begin{smallmatrix}\widehat{L}_t\\ K_t\end{smallmatrix}\right)$, 定义范数 $\|Z_t\|_1=\widehat{L}_t+K_t$, 由于 $K_t\geqslant 0$, 故有 $\|Z_t\|_1\geqslant\widehat{L}_t$, 再根据范数的性质 [21], 一定存在常数 $k_1>0$, 使得 $\|Z_t\|_1\leqslant k_1\|Z_t\|$, 因此有

$$\widehat{L}_t\leqslant k_1\|Z_t\|. \tag{3.35}$$

由上式和 (3.34) 得

$$\begin{aligned}\|Z_{t+1}\|&\leqslant\|C\|\cdot\|Z_t\|+\left\|\begin{pmatrix}1\\ 1\end{pmatrix}\right\|\cdot S_t+\left\|\begin{pmatrix}1\\ 0\end{pmatrix}\right\|\cdot\xi_t\\ &=\gamma\|Z_t\|+k_2S_t+\xi_t\\ &\leqslant\gamma\|Z_t\|+k_2k\sum_{i=t-g+1}^{t}\alpha_i\delta_iL_i+\xi_t\\ &\leqslant\gamma\|Z_t\|+k_1k_2k\sum_{i=t-g+1}^{t}\alpha_i\delta_i\|Z_i\|+\xi_t\\ &=\gamma\|Z_t\|+k_3\sum_{i=t-g+1}^{t}\alpha_i\delta_i\|Z_i\|+\xi_t,\end{aligned} \tag{3.36}$$

其中 k_2 和 k_3 均是大于 0 的常数. 定义正序列 $\{\widehat{Z}_t\}$ 如下:

$$\widehat{Z}_{t+1}=\gamma\widehat{Z}_t+k_3\sum_{i=t-g+1}^{t}\alpha_i\delta_i\widehat{Z}_i+\xi_t,\quad \widehat{Z}_0=\|Z_0\|, \tag{3.37}$$

显然,

$$\|Z_{t+1}\|\leqslant\widehat{Z}_{t+1}. \tag{3.38}$$

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

下证

$$\widehat{Z}_{t+1}=O(r_t^{\epsilon}d_t),\quad \forall\,\epsilon>0. \tag{3.39}$$

记 $M_t=[\widehat{Z}_t\ \ \widehat{Z}_{t-1}\ \cdots\ \widehat{Z}_{t-g+1}]^{\mathrm{T}}$, 其中 $g=\max\{q,r\}$, 由 (3.37) 得

$$\begin{aligned}
M_{t+1}&=\begin{pmatrix}
\gamma+k_3\alpha_t\delta_t & k_3\alpha_{t-1}\delta_{t-1} & \cdots & \cdots & k_3\alpha_{t-g+1}\delta_{t-g+1}\\
1 & 0 & \cdots & \cdots & 0\\
0 & 1 & \ddots & \ddots & \vdots\\
\vdots & \vdots & \ddots & \ddots & \vdots\\
0 & 0 & \cdots & 1 & 0
\end{pmatrix}M_t+\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}\xi_t\\
&=\begin{pmatrix}
k_3\alpha_t\delta_t & k_3\alpha_{t-1}\delta_{t-1} & \cdots & k_3\alpha_{t-g+1}\delta_{t-g+1}\\
0 & 0 & \cdots & 0\\
\vdots & \vdots & \ddots & \vdots\\
0 & 0 & \cdots & 0
\end{pmatrix}M_t\\
&\quad+\begin{pmatrix}
\gamma & 0 & \cdots & \cdots & 0\\
1 & 0 & \cdots & \cdots & 0\\
0 & 1 & \ddots & \ddots & \vdots\\
\vdots & \vdots & \ddots & \ddots & \vdots\\
0 & 0 & \cdots & 1 & 0
\end{pmatrix}M_t+\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}\xi_t\\
&\triangleq(D_t+E)M_t+F\xi_t.
\end{aligned} \tag{3.40}$$

由 $|zI-E|=(z-\gamma)z^{g-1}$ 知, 矩阵 E 的特征值均在单位圆内, 因此, E 是稳定阵, 故存在一个诱导范数 $\|\cdot\|$ 使得 $\|E\|<1$, 记 $\lambda=\|E\|$. 由 $\|\cdot\|_1$ 定义知, $\|D_t\|_1=k_3\sum_{i=t-g+1}^{t}\alpha_i\delta_i$, 根据范数的性质, 一定存在常数 $k_4>0$, 使得 $\|D_t\|\leqslant\frac{k_4}{k_3}\|D_t\|_1$, 因此, 由 (3.40) 可得

$$\|M_{t+1}\|\leqslant\left(\lambda+k_4\sum_{i=t-g+1}^{t}\alpha_i\delta_i\right)\|M_t\|+\xi_t. \tag{3.41}$$

记 $\varepsilon_t=\sum_{i=t-g+1}^{t}\alpha_i\delta_i$, 将上式递推可得

$$\begin{aligned}
\|M_{t+1}\|&\leqslant\prod_{j=0}^{t}(\lambda+k_4\varepsilon_j)\|M_0\|+\sum_{i=0}^{t}\prod_{j=i+1}^{t}(\lambda+k_4\varepsilon_j)\xi_i\\
&=\lambda^{t+1}\prod_{j=0}^{t}(1+\lambda^{-1}k_4\varepsilon_j)\|M_0\|+\sum_{i=0}^{t}\lambda^{t-i}\prod_{j=i+1}^{t}(1+\lambda^{-1}k_4\varepsilon_j)\xi_i.
\end{aligned} \tag{3.42}$$

下面分析 $\prod_{j=i}^{t}(1+\lambda^{-1}k_4\varepsilon_j)$. 利用不等式 $(1+x+y)\leqslant(1+x)(1+y)$, $x\geqslant0$, $y\geqslant0$, 得

$$\prod_{j=i}^{t}(1+\lambda^{-1}k_4\varepsilon_j)=\prod_{j=i}^{t}\left(1+\lambda^{-1}k_4\sum_{r=j-g+1}^{j}\alpha_r\delta_r\right)$$

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

$$
\begin{aligned}
&\leqslant \prod_{j=i}^{t}(1+\lambda^{-1}k_4\alpha_j\delta_j)\prod_{j=i}^{t}(1+\lambda^{-1}k_4\alpha_{j-1}\delta_{j-1})\cdots \\
&\quad\times\prod_{j=i}^{t}(1+\lambda^{-1}k_4\alpha_{j-g+1}\delta_{j-g+1}).
\end{aligned}\tag{3.43}
$$

由 (3.25) 知 $\delta_j\to 0$. 由引理 3.1 知, 对任给 $\epsilon>0$, 存在充分大的 i_0 使得

$$\lambda^{-1}k_4\sum_{j=i}^{t}\alpha_j\delta_j\leqslant\frac{\epsilon}{g}\log r_t,\quad \forall\, t\geqslant i\geqslant i_0.\tag{3.44}$$

根据上式及不等式 $1+x\leqslant \mathrm{e}^x$, $x\geqslant 0$, 可得

$$\prod_{j=i}^{t}(1+\lambda^{-1}k_4\alpha_j\delta_j)\leqslant\exp\left(\lambda^{-1}k_4\sum_{j=i}^{t}\alpha_j\delta_j\right)=O(r_t^{\frac{\epsilon}{g}}),\quad \forall\, t\geqslant i\geqslant i_0.\tag{3.45}$$

将 (3.45) 代入 (3.43), 可得

$$\prod_{j=i}^{t}(1+\lambda^{-1}k_4\varepsilon_j)=O(r_t^{\epsilon}),\quad \forall\, t\geqslant i\geqslant i_0+g-1.\tag{3.46}$$

进一步, 将此式代入 (3.42), 利用 (3.6), 并由 ϵ 的任意性得

$$\|M_{t+1}\|=O(r_t^{\epsilon}d_t),\quad \forall\,\epsilon>0,\tag{3.47}$$

从而自然有

$$\widehat{Z}_{t+1}=O(r_t^{\epsilon}d_t),\quad \forall\,\epsilon>0.\tag{3.48}$$

因此, 根据 (3.33)、(3.35)、(3.38) 和 (3.48), 可得

$$L_{t+1}=O(\widehat{L}_{t+1})=O(\|Z_{t+1}\|)=O(\widehat{Z}_{t+1})=O(r_t^{\epsilon}d_t),\quad \forall\,\epsilon>0.\tag{3.49}$$

最后, 再根据 (3.24), 可得

$$\|\varphi_t\|^2=O(r_t^{\epsilon}d_t),\quad \text{a.s. }\forall\,\epsilon>0.\tag{3.50}$$

引理得证. □

定理 2.1 的证明　由 (3.17) 得

$$\sum_{t=0}^{n}\|Y_{t+1}\|^2=O\left(\sum_{t=0}^{n}\sum_{i=0}^{t}\alpha^{t-i}\overline{\eta_i}^2\right)=O\left(\sum_{i=0}^{n}\overline{\eta_i}^2\right).\tag{3.51}$$

根据条件 (A1), 易知 (参见文献 [7])

$$\sum_{i=0}^{n}\omega_{i+1}^2=O(n).\tag{3.52}$$

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

进一步, 利用 (3.13)、(3.14)、(3.51)、(3.52) 和引理 3.3 得

$$
\begin{aligned}
\sum_{t=0}^{n}\|Y_{t+1}\|^2 &= O\bigg(\sum_{i=0}^{n}\overline{\eta_i}^2\bigg) = O\bigg(\sum_{i=0}^{n}(D(z)\widetilde{\theta}_i^{\mathrm{T}}\varphi_i)^2\bigg) + O\bigg(\sum_{i=0}^{n}\eta_i^2\bigg) \\
&= O\bigg(\sum_{i=0}^{n}\sum_{j=i-g+1}^{i}(\widetilde{\theta}_j^{\mathrm{T}}\varphi_j)^2\bigg) + O\bigg(\sum_{i=0}^{n}\sum_{j=i-g+1}^{i}(\omega_{j+1}^2 + {y_{j+1}^*}^2)\bigg) \\
&= O\bigg(\sum_{i=n-g+1}^{n}\sum_{j=0}^{i}\alpha_j\bigg) + O\bigg(\sum_{i=n-g+1}^{n}\sum_{j=0}^{i}\alpha_j\delta_j\|\varphi_j\|^2\bigg) \\
&\quad + O\bigg(\sum_{i=n-g+1}^{n}\sum_{j=0}^{i}(\omega_{j+1}^2 + {y_{j+1}^*}^2)\bigg) \\
&= O(\log r_n) + O\Big(\max_{0\leqslant j\leqslant n}\{\delta_j\|\varphi_j\|^2\}\log r_n\Big) + O(n) \\
&= O(\log r_n) + O(r_n^{\epsilon}d_n\log r_n) + O(n) \\
&= O(r_n^{\epsilon}d_n\log r_n) + O(n).
\end{aligned}
\tag{3.53}
$$

下证 $r_n = O(n)$. 由 (3.53) 和 ϵ 的任意性知,

$$
\sum_{t=0}^{n} y_{t+1}^2 = O(r_n^{\epsilon}d_n) + O(n),
\tag{3.54}
$$

据此和条件 (A2), 易知当采取行动组合 (2.20) 时, 由 (2.1) 可得

$$
\sum_{t=0}^{n} u_t^2 = O(r_n^{\epsilon}d_n) + O(n),
\tag{3.55}
$$

$$
\sum_{t=0}^{n} v_t^2 = O(r_n^{\epsilon}d_n) + O(n),
\tag{3.56}
$$

从而,

$$
\sum_{t=0}^{n}\|\varphi_t\|^2 = \sum_{t=0}^{n}\bigg(\sum_{i=0}^{p-1}y_{t-i}^2 + \sum_{i=1}^{q-1}u_{t-i}^2 + \sum_{i=1}^{r-1}v_{t-i}^2\bigg) = O(r_n^{\epsilon}d_n) + O(n).
\tag{3.57}
$$

所以,

$$
\begin{aligned}
r_n &= 1 + \sum_{t=0}^{n}\|\varphi_t\|^2 = O(r_n^{\epsilon}d_n) + O(n) \\
&= O(r_n^{\epsilon}n^{\delta}) + O(n), \quad \forall\,\delta\in\Big(\frac{2}{\kappa},1\Big).
\end{aligned}
\tag{3.58}
$$

取 ϵ 充分小使得 $\epsilon+\delta<1$, 得

$$
\begin{aligned}
\frac{r_n}{n} &= O\bigg(\Big(\frac{r_n}{n}\Big)^{\epsilon}\frac{1}{n^{1-\epsilon-\delta}}\bigg) + O(1) \\
&= O(1) + o\bigg(\Big(\frac{r_n}{n}\Big)^{\epsilon}\bigg),
\end{aligned}
\tag{3.59}
$$

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

因此,

$$r_n = O(n). \tag{3.60}$$

最后, 由 (3.60) 和 (3.54)–(3.56) 得

$$\sum_{t=0}^{n}(y_t^2 + u_t^2 + v_t^2) = O(n). \tag{3.61}$$

从而 (2.23) 成立, 定理得证. □

至此我们得到, 对于有两个参与者的线性随机系统 (2.1), 当 "有限理性" 的参与者均不知道系统内部参数 θ, 在以 $J(u,v)$ 为指标的零和博弈问题中, 采取自适应行动组合 (2.20) 所定义的策略组合时, 相应的闭环系统具有全局稳定性.

4 优化性分析

定理 2.2 的证明 根据定理 2.2 的条件, 在时刻 t, 我们定义

$$J_t(u_t', v_t') = J_t(u_0, \dots, u_{t-1}, u_t', v_0, \dots, v_{t-1}, v_t'), \tag{4.1}$$

其中 (u_t', v_t') 是任意行动组合, (u_i, v_i) $(i = 0, 1, \dots, t-1)$ 由 (2.20) 定义.

因此, 类似于 (2.18), 我们有

$$J_t(u_t', v_t') = \sigma^2 + (\widetilde{\theta}_t^{\mathrm{T}}\varphi_t + \theta_t^{\mathrm{T}}\varphi_t + b_1 u_t' + c_1 v_t' - y_{t+1}^*)^2 + \lambda_1 u_t'^2 - \lambda_2 v_t'^2. \tag{4.2}$$

要证定理条件下, 在时刻 t, 对于任意行动组合 (u_t', v_t'), 有

$$J_t(u_t, v_t') \leqslant J_t(u_t, v_t) + \overline{o(1)} \leqslant J_t(u_t', v_t). \tag{4.3}$$

我们先来证

$$(\widetilde{\theta}_t^{\mathrm{T}}\varphi_t)^2 = \overline{o(1)}. \tag{4.4}$$

事实上, 利用引理 3.3 和 (3.60) 得

$$\begin{aligned}
\sum_{t=0}^{n}(\widetilde{\theta}_t^{\mathrm{T}}\varphi_t)^2 &= \sum_{t=0}^{n}\alpha_t[1 + \varphi_t^{\mathrm{T}}P_{t+1}\varphi_t + \varphi_t^{\mathrm{T}}(P_t - P_{t+1})\varphi_t] \\
&\leqslant \sum_{t=0}^{n}\alpha_t[2 + \delta_t\|\varphi_t\|^2] \\
&= O\left(\sum_{t=0}^{n}\alpha_t\right) + O\left(\sum_{t=0}^{n}\alpha_t\delta_t\|\varphi_t\|^2\right) \\
&= O(\log r_n) + O\left(\max_{0\leqslant t\leqslant n}\{\delta_t r_t^{\epsilon} d_t\}\log r_n\right) \\
&= O(r_n^{\epsilon} d_n \log r_n) \\
&= O(n^{\epsilon+\delta}\log r_n), \quad \forall \epsilon > 0.
\end{aligned} \tag{4.5}$$

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

取充分小的 ϵ 使得 $\epsilon+\delta<1$, 则

$$\frac{1}{n}\sum_{t=0}^{n}(\widetilde{\theta}_t^{\mathrm{T}}\varphi_t)^2=o(1), \tag{4.6}$$

根据定义可知 (4.4) 成立.

下面证明对于任意行动 u_t', 有

$$J_t(u_t,v_t)+\overline{o(1)}\leqslant J_t(u_t',v_t). \tag{4.7}$$

由 (4.2) 和 (2.21) 得

$$\begin{aligned}J_t(u_t',v_t)&=(b_1^2+\lambda_1)u_t'^2+2b_1(\widetilde{\theta}_t^{\mathrm{T}}\varphi_t+\theta_t^{\mathrm{T}}\varphi_t+c_1v_t-y_{t+1}^*)u_t'\\&\quad+(\theta^{\mathrm{T}}\varphi_t+c_1v_t-y_{t+1}^*)^2-\lambda_2v_t^2+\sigma^2\\&=(b_1^2+\lambda_1)\left(u_t'-u_t+\frac{b_1\widetilde{\theta}_t^{\mathrm{T}}\varphi_t}{b_1^2+\lambda_1}\right)^2+\frac{\lambda_1(\theta^{\mathrm{T}}\varphi_t+c_1v_t-y_{t+1}^*)^2}{b_1^2+\lambda_1}-\lambda_2v_t^2+\sigma^2.\end{aligned} \tag{4.8}$$

因此, 当且仅当 $u_t'=u_t-\frac{b_1\widetilde{\theta}_t^{\mathrm{T}}\varphi_t}{b_1^2+\lambda_1}$ 时, $J_t(u_t',v_t)$ 取到极小值

$$\min J_t(u_t',v_t)=\frac{\lambda_1(\theta^{\mathrm{T}}\varphi_t+c_1v_t-y_{t+1}^*)^2}{b_1^2+\lambda_1}-\lambda_2v_t^2+\sigma^2. \tag{4.9}$$

当 $u_t'=u_t$ 时, 由 (4.8) 和 (4.4) 得

$$\begin{aligned}J_t(u_t,v_t)&=\frac{b_1^2(\widetilde{\theta}_t^{\mathrm{T}}\varphi_t)^2}{b_1^2+\lambda_1}+\frac{\lambda_1(\theta^{\mathrm{T}}\varphi_t+c_1v_t-y_{t+1}^*)^2}{b_1^2+\lambda_1}-\lambda_2v_t^2+\sigma^2\\&=\overline{o(1)}+\min J_t(u_t',v_t),\end{aligned} \tag{4.10}$$

从而,

$$\min J_t(u_t',v_t)=J_t(u_t,v_t)+\overline{o(1)}, \tag{4.11}$$

因此, (4.7) 得证.

类似地, 我们可以证明在满足定理 2.2 的条件下, 对于任意行动 v_t', 有

$$J_t(u_t,v_t')\leqslant J_t(u_t,v_t)+\overline{o(1)}. \tag{4.12}$$

综上, 在满足定理 2.2 条件的情形下, 对于任意行动组合 (u_t',v_t'), (4.3) 成立, 即行动组合 (u_t,v_t) 是渐近 Nash 均衡解, 定理得证. □

5 小结

博弈论是一种针对有利益冲突的个体如何做决策的问题的有效建模方法, 传统博弈论主要研究参与者对系统的参数和结构已知的情形, 但在实际中, 这样的假设往往过于理想化, 因为总是存在不确定性情形, 这既给参与者正确决策增加了难度, 也给理论研究增加了复杂性. 另一方面, 在控制理论

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

中, 自适应控制正是处理系统内部参数或结构不确定性问题的有效方法. 如何将博弈论的思想与自适应控制的方法相结合, 开展随机自适应动态博弈理论问题研究, 是本文研究的出发点和落脚点.

本文首次提出并研究了离散时间由输入输出模型描述的线性系统的随机自适应动态博弈问题, 考虑了在博弈系统内部对参与者有未知参数的情形下, 利用自适应控制的方法处理参数不确定性的问题. 具体来讲, 就是先采用最小二乘方法得到未知参数的在线估计值, 再根据 "必然等价原则" 用之代替未知参数, 从而设计参与者的自适应策略组合, 最后证明了该策略组合相应的闭环系统不但具有全局稳定性而且具有一定的优化性.

值得指出的是, 作为一个新的研究方向, 自适应博弈系统中尚有大量科学问题有待解决. 仅就与本文密切相关的未解决问题也有一些, 例如, 参与者 "高频增益" 未知情形, 以及博弈双方采用不同估计信息情形等, 都值得进一步研究.

致谢　感谢审稿人对本文提出的修改意见.

参考文献

1 Basar T, Olsder G J. Dynamic Noncooperative Game Theory. 2nd ed. Philadelphia: SIAM, 1999
2 Yeung D W K, Petrosyan L A. Cooperative Stochastic Differential Games. New York: Springer, 2006
3 Filar J A, Gaitsgory V, Mizukami K. Advances in Dynamic Games and Applications. Boston: Birkhäuser, 2000
4 郭雷, 程代展, 冯德兴. 控制理论导论: 从基本概念到研究前沿. 北京: 科学出版社, 2005
5 Åström K J, Wittenmark B. Adaptive Control. Boston: Addison-Wesley Publishing Company, 1994
6 Kumar P R, Varaja P. Stochastic Systems: Estimation, Identification and Adaptive Control. New York: Prentice-Hall, 1985
7 Chen H F, Guo L. Identification and Stochastic Adaptive Control. Boston: Birkhäuser, 1991
8 Sragovich V G. Mathematical Theory of Adaptive Control. Singapore: World Scientific Publishing, 2006
9 Ioannou P A, Sun J. Robust Adaptive Control. New Jersey: Prentical-Hall, 1996
10 Krstić M, Kanellakopoulos I, Kokotović P. Nonlinear and Adaptive Control Design. New York: Wiley, 1995
11 Zhou K, Doyle J C, Glover K. Robust and Optimal Control. New Jersey: Prentice-Hall, 1996
12 Basar T, Bernhard P. H_∞ Optimal Control and Related Minimax Design Problems. Boston: Birkhäuser, 1995
13 Mu Y F, Guo L. Optimization and identification in a non-equilibrium dynamic game. In: Proceedings of the 48th IEEE Conference on Decision and Control. Shanghai: IEEE, 2009, 5750–5755
14 Mu Y F, Guo L. How cooperation arises from rational players? Sci China Inf Sci, 2013, 56: 112201
15 Mu Y F, Guo L. Towards a theory of game-based non-equilibrium control systems. J Syst Sci Complex, 2012, 25: 209–226
16 Li Y, Guo L. Towards a theory of stochastic adaptive differential games. In: Proceedings of the 50th IEEE Chinese Conference on Decision and Control and European Control Conference. Orlando: IEEE, 2011, 5041–5046
17 Li Y, Guo L. Convergence of adaptive linear stochastic differential games: Nonzero-sum case. In: Proceedings of the 10th World Congress on Intelligent Control and Automation. Beijing: IEEE, 2012, 3543–3548
18 Goodwin G C, Sin K S. Adaptive Filtering, Prediction and Control. New Jersey: Prentice-Hall, 1984
19 Guo L, Chen H F. The Åström-Wittenmark self-tuning regulator revisited and ELS based adaptive trackers. IEEE Trans Automat Control, 1991, 36: 802–812
20 Guo L. Convergence and logarithm laws of self-tuning regulators. Automatica, 1995, 31: 435–450
21 郭雷. 时变随机系统 — 稳定性、估计与控制. 长春: 吉林科技出版社, 1993

Stochastic adaptive dynamical games

YUAN Shuo & GUO Lei

Abstract　In this paper, we consider two players zero-sum dynamical games in linear stochastic systems de-

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

scribed by input-output model with unknown parameters, and with an ergodic output-tracking quadratic index. We combine the ideas of game theory and control theory as a modeling method, and investigate the corresponding stochastic adaptive game problems. There is few existing results related to such stochastic adaptive game problems, since uncertainty raises the difficulty of making decisions for both players. We attempt to cope with this problem by the ideas and methods of adaptive control. We will firstly use the standard least squares to estimate the unknown parameters, and then construct the adaptive strategy profile for both players by using the so-called "certainty equivalence principle". We will prove that, the adaptive strategic profile makes the system globally stable, and that the action profile is an asymptotic Nash equilibrium solution to our game problem in a certain sense.

Keywords stochastic systems, dynamic game, adaptive control, stability, asymptotic Nash equilibrium, least-squares

MSC(2010) 91A25, 93C40

doi: 10.1360/N012015-00355

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net