Learning Theory in Games

Series II, Lesson 5

Lecturer: Changxi Li

(School of Mathematics, Shandong University)

Center of STP Theory and Its Applications
July 18-23, 2021

LiaoCheng University, LiaoCheng, Shandong, P.R. China

Outline

1. What is Learning Theory in Games?

"The theory of learning in games studies how, which and what kind of equilibria might arise as a consequence of a long-run non-equilibrium process of learning, adaptation and/or imitation."

Drew Fudenberg, David K. Levine, Learning and equilibrium, *Annual Review of Economics*, vol. 1, no. 1, 385-420, 2009.

Drew Fudenberg, David K. Levine, *The Theory of Learning in Games*, MIT Press, 1998.

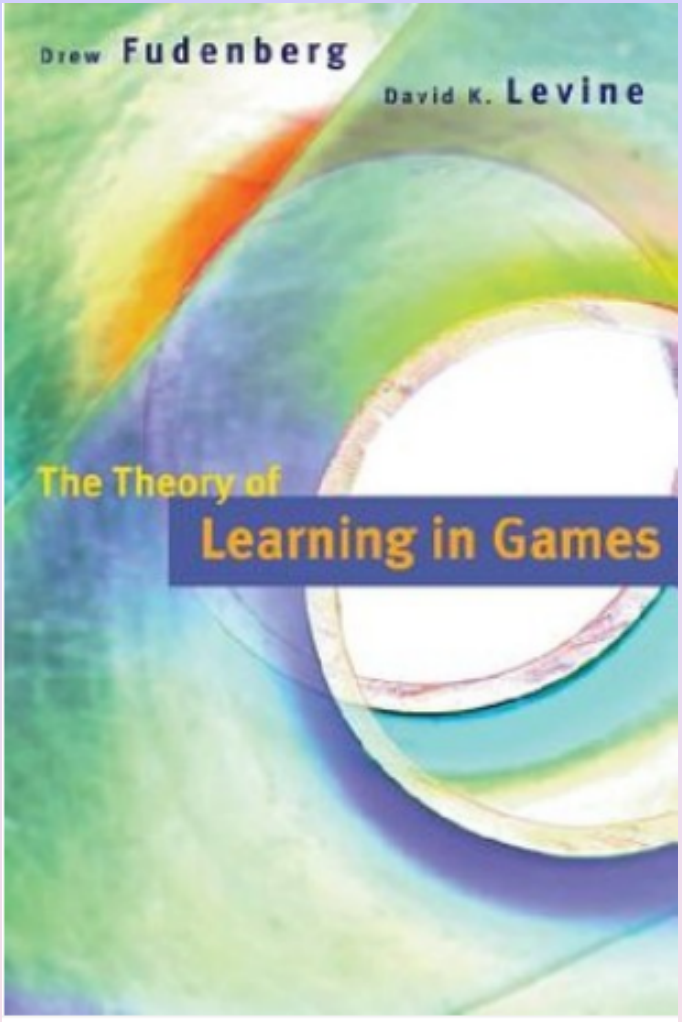
☞ History

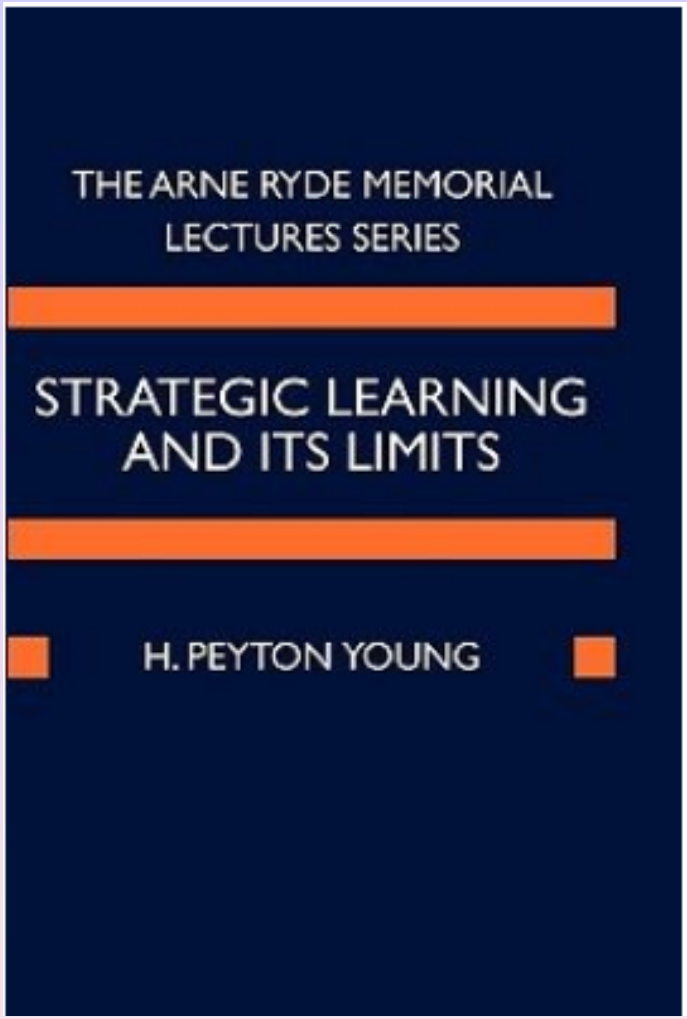
- Richards, PI, On Game-learning Machines, *Scientific Monthly*, 1952.
- Herbert A. Simon, A comparison of game theory and learning theory, *Journal Of The American Statistical Association*, vol. 21, 267 – 272, 1956.
- Feichtin. G, A Markov learning model for 2-person zero-sum games, vol. 11, no. 7, 322-331, 1962.

- Drew Fudenberg, David K. Levine, *The Theory of Learning in Games*, MIT Press, 1998.
- Drew Fudenberg, David K. Levine, Learning and equilibrium, *Annual Review of Economics*, vol. 1, no. 1, 385-420, 2009.

- R. Gopalakrishnan, J. R. Marden, and A. Wierman, "An architectural view of game theoretic control," *ACM SIGMETRICS Performance Evaluation Review*, vol. 38, no. 3, pp. 31-36, 2011.
- Ting Liu, Jinhuan Wang, Xiao Zhang, Daizhan Cheng, Game Theoretic Control of Multiagent Systems. *SIAM J. Control. Optim.* 57(3): 1691-1709, 2019.

☞ References





Drew Fudenberg, David K. Levine, *The Theory of Learning in Games*, MIT Press, 1998.

H. P. Young, *Strategic Learning and Its Limits*, Oxford, U.K.: Oxford Univ. Press, 2004.

2. Basic Concepts

☞ 2.1 Structure of Game Learning

Consider a repeated game $G = \{N, A, C\}$. A learning rule of games consists of two parts: *prediction* and *response*.

- Prediction: Consider a repeated one-shot game $G = \{N, A, C\}$. Player i predicts opponents' action according to available information $O_i(t)$.

$$f_i : O_i(t) \to A_{-i}, \tag{1}$$

- Response: Player i makes decisions according to prediction and available information $O_i(t)$

$$g_i : O_i(t) \times f_i(O_i(t)) \to A_i. \tag{2}$$


Nicolo Cesa-Bianchi and Gabor Lugosi, *Prediction, Learning and Games*, Cambridge University Press, 2006.


J.S. Jordan, Three problems in learning mixed-strategy Nash equilibria, *Games and Economic Behavior*, Vol. 5, No. 3, 368-386, 1993.

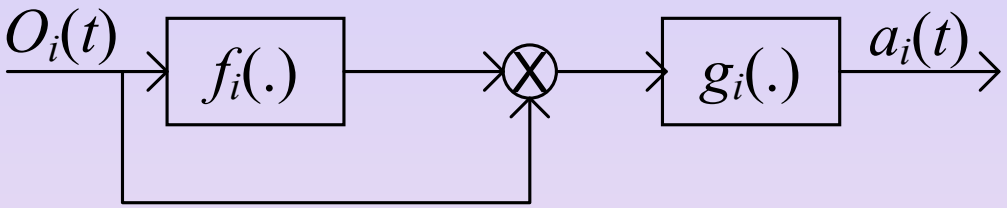


Figure 1: Structure of Game Learning

Figure 2: Factors in Learning Rule

☞ Information Structure

A learning rule is called

- *coupled*, if the available information of agent i is the payoff structure of all players and history sequence of the play, i.e.,

$$O_i(t) = \Big\{ \{a(\tau)\}_{\tau=0,1,\ldots,t-1};\ \{c_i(a)\}_{i\in N} \Big\}.$$

- *uncoupled*, if the available information of agent i is the payoff structure of himself and history sequence of the play, i.e.,

$$O_i(t) = \Big\{ \{a(\tau)\}_{\tau=0,1,\ldots,t-1};\ c_i(a) \Big\}.$$

- *completely uncoupled*, if the available information of agent i is his own past realized payoffs and actions, i.e.,

$$O_i(t) = \Big\{ \{a_i(\tau), c_i(a(\tau))\}_{\tau=0,1,\ldots,t-1,} \Big\}.$$


M.S. Talebi, "Uncoupled learning rules for seeking equilibria in repeated plays: An overview," *Computer Science*, 1-9, 2013.


Y. Babichenko, "Completely Uncoupled Dynamics and Nash Equilibria." *Games & Economic Behavior*, 76(1), 1-14, 2012.

☞ Updating Orders

- *Synchronous updating rule (SUR)*: all players update actions synchronously.

$$a_i(t+1) = f_i(a_1(t), a_2(t), \cdots, a_n(t)), \quad \forall i \in N.$$

- *Asynchronous updating rule (AUR)*: only one player is allowed to update his action at each time t.

$$\begin{cases} a_i(t+1) &= f_i(a_1(t), a_2(t), \cdots, a_n(t)). \\ a_j(t+1) &= a_j(t),\ j \neq i. \end{cases} \tag{3}$$

  ◇ *Deterministic AUR*: player updates his action according to given orders.

  ◇ *Stochastic AUR*: player i updates his action with probability $p_i > 0$, where $\sum_{i \in N} p_i = 1$.

- *Cascading updating rule (CUR)*: when player i updates his action, he knows j's action, $j < i$.

$$\begin{cases} a_1(t+1) &= f_1(a_1(t), a_2(t), \cdots, a_n(t)) \\ a_2(t+1) &= f_2(a_1(t+1), a_2(t), \cdots, a_n(t)) \\ & \vdots \\ a_n(t+1) &= f_n(a_1(t+1), \cdots, a_{n-1}(t-1), a_n(t)) \end{cases} \tag{4}$$

☞ Convergence Types

Consider the sequence of action profile $a(t, a_0), t = 1, 2, 3...$ generated by a learning rule.

- *Convergence*: $a(t)$ converges to NE a^*, if $\exists T > 0$

$$a(t, a_0) = a^*, \quad \forall t \geq T, \quad \forall a_0.$$

- *With probability one*: $a(t)$ converges to NE a^* with probability one

$$\lim_{t \to \infty} P(a(t, a_0) = a^*) = 1, \quad \forall a_0.$$

- *Almost surely*: $a(t)$ converges to a^* almost surely, if

$$P(\lim_{t \to \infty} a(t, a_0) = a^*) = 1, \quad \forall a_0.$$

- *With frequency*: $a(t)$ converges to a^* with frequency $1 - \epsilon$, if

$$\lim_{t \to \infty} \inf \frac{|\{1 \leq \tau \leq t : a(t, a_0) = a^*\}|}{t} \geq 1 - \epsilon, \quad \forall a_0.$$

☞ Types of equilibria:

NE, mixed NE, correlated equilibrium, coarse correlated equilibrium...

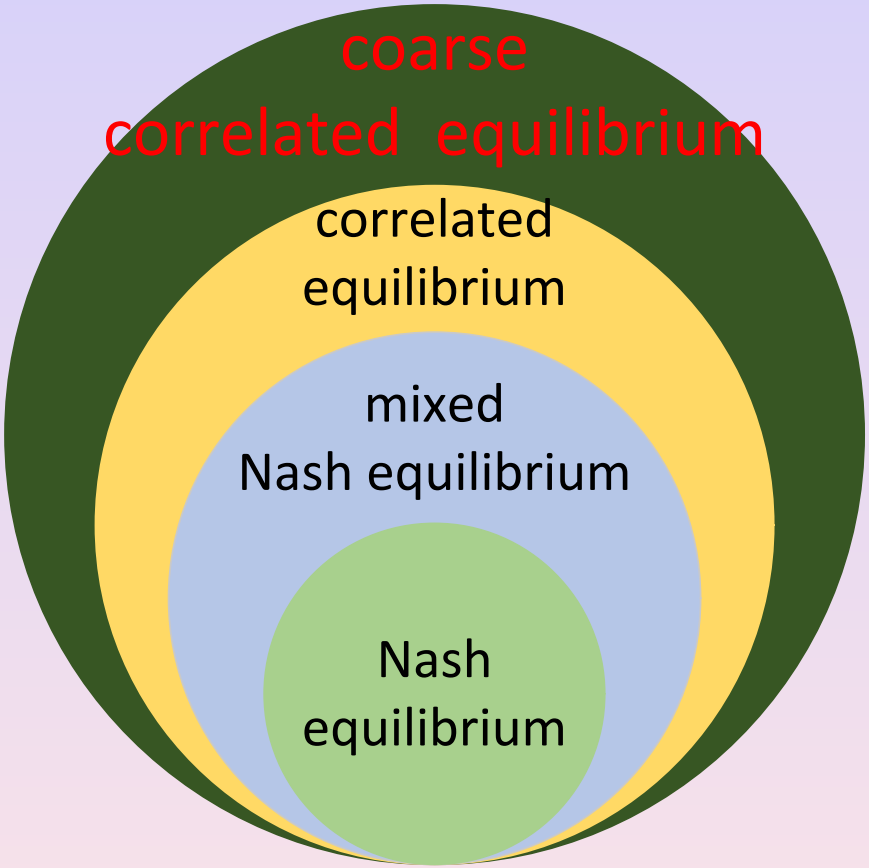


Figure 3: Different Equilibria

☞ Properties of games:

Potential games, zero-sum games, symmetric game...

$$\mathcal{G} = \underbrace{\mathcal{G}^{P_0} \quad \oplus \quad \overbrace{\mathcal{G}^{N}}^{\mathcal{G}^{H}}}_{\mathcal{G}^{P}} \quad \oplus \quad \mathcal{G}^{H_0}, \tag{5}$$

$$\mathcal{G} = \mathcal{S} \quad \oplus \quad \mathcal{K} \quad \oplus \quad \mathcal{A}. \tag{6}$$

☞ Important results:

Question: Are there simple dynamics that converge to NE for any game?

Theorem

- *There exist uncoupled dynamics converging to correlated equilibria.*
- *There are no "natural" dynamics that lead to NE in any game.*

- Natural: adaptive, simple, efficient, one memory (eg. fictitious play, best response...)
- Not natural: exhaustive search, mediator(中介者，调停者)...

S. Hart, A. Mas-Colell, Uncoupled dynamics do not lead to Nash equilibrium. *Amer. Econ. Rev.*, vol. 93, pp. 1830 – 1836, 2003.

3. Three Typical Learning Rules

3.1 Myopic Best Response Adjustment

Myopic Best Response Adjustment (MBRA)

At time $t+1$, player i is make decisions according to observed information $a_{-i}(t)$. Let

$$BR_i(t) := \text{argmax}_{s_i \in S_i} c_i(s_i, a_{-i}(t)) = \{j_1, \cdots, j_l \mid j_1 < \cdots < j_l\}.$$

BR_i is called best response set of player i.

- If $a_i(t) \in BR_i(t)$, then $a_i(t+1) = a_i(t)$.
- If $a_i(t) \notin BR_i(t)$, then
  - MBRA-D: $a_i(t+1) = j_1$;
  - MBRA-P: $\mathbf{Pr}(a_i(t+1) = j_d) = \frac{1}{|BR_i(t)|},\ d = 1, \cdots, l.$

Fiat A, Koutsoupias E, Ligett K, et al. Beyond myopic best response in Cournot competition. *Games and Economic Behavior*, 2012: 993-1005.

Question 1: Will MBRA converge to NE in games?

Question 2: Will MBRA stay at an NE if it reaches one?

Example 1

Consider a common interest game as follows

Table 1: Common Interest Game

$P_1 \backslash P_2$	a_2	b_2
a_1	$(\underline{10},\ \underline{10})$	$(0,\ 0)$
b_1	$(0,\ 0)$	$(\underline{5},\ \underline{5})$

Observe cycles

$$(b_1, a_2) \to (a_1, b_2) \to (b_1, a_2) \to (a_1, b_2) \to \cdots$$

Question 3: What are some potential remedies?

- Remedy 1: Only one agent updates at time. *Asynchronous MBRA* - Will converge to NE but requires a mechanism to coordinate who updates.
- Remedy 2: Introduce a probabilistic reluctance to change actions. *MBRA with inertia*:

$$a_i(t+1) := \begin{cases} a_i(t) & \text{with probability } \epsilon; \\ BR_i(t) & \text{with probability } 1-\epsilon. \end{cases} \tag{7}$$

Theorem

Consider a potential game $G = \{N, A, C\}$, if all players play the game according to asynchronous MBRA or MBRA with inertia, then it will converge to an NE.

Proof

- According to the definition of potential game, G has at least one pure NE (potential maximizer).
- Using asynchronous MBRA, each updating leads to a higher potential.
- As there are finite action profiles, after finite steps, MBRA will reach and stabilize at NE.

D. Monderer, L.S. Shapley, Potential games, *Games and Economic Behavior*, Vol. 14, 124-143, 1996.

Homework 1:

- What will happen for the cycle in Example 1 using asynchronous MBRA or MBRA with inertia?
- Prove the convergence of MBRA with inertia in potential games.
- Is the potential game condition necessary?
  (Hint: acyclic games)

Question 4: Will MBRA converge in any games with NE?

Example 2

Table 2: S. Hart Game

$V_i \backslash a$	111	112	113	121	122	123	131	132	133	211	212	213	221	222	223	231	232	233	311	312	313	321	322	323	331	332	333
V_1	0	4	2	0	4	3	2	3	0	4	0	3	4	0	2	3	2	0	1	1	0	1	1	0	0	0	6
V_2	0	0	2	4	4	3	1	1	0	4	4	3	0	0	2	1	1	0	2	3	0	3	2	0	0	0	6
V_3	0	4	1	4	0	1	2	3	0	0	4	1	4	0	1	3	2	0	2	3	0	3	2	0	0	0	6

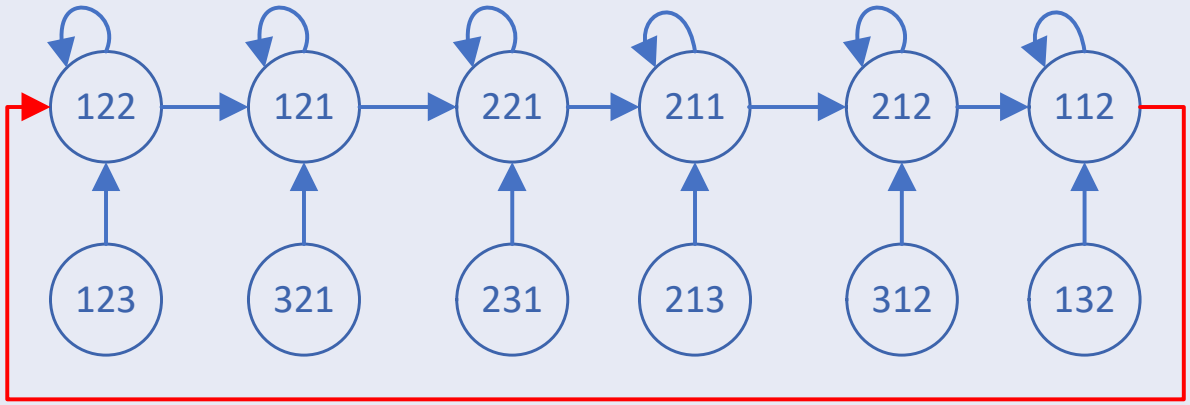


Figure 4: An Adjustment Cycle

S. Hart, Y. Mansour, Stochastic uncoupled dynamics and Nash equilibrium, *Games and Economic Behavior*, Vol. 57, No. 2, 286-303, 2006.

3.2 Log-Linear Learning

- Recall the common interest game in Example 1

$P_1 \backslash P_2$	a_2	b_2
a_1	$(\underline{10},\ \underline{10})$	$(0,\ 0)$
b_1	$(0,\ 0)$	$(\underline{5},\ \underline{5})$

- Asynchronous MBRA will converge to pure NE (a_1, a_2) or (b_1, b_2). However (a_1, a_2) is clearly better choice.
- Question:
  - Are there any learning rule that can converge to (a_1, a_2)?
  - Are there any learning rule that converge to NE that optimizes the potential function of potential games?
- Yes, but different notion of "convergence".

Log Linear Learning

- One player, say i, is drawn randomly from N and allowed to alter his action. All others must repeat previous action, i.e.,

$$a_{-i}(t+1) = a_{-i}(t).$$

- Player i selects action according to the probability

$$Pr^{\tau}(a_i(t+1)) = s_i | a(t)) = \frac{e^{\frac{1}{\tau}c_i(s_i, a_{-i}(t))}}{\sum_{a_i \in S_i} e^{\frac{1}{\tau}c_i(a_i, a_{-i}(t))}},$$

  where τ is referred to as the temperature coefficient.
- Repeat.

- Update rule also referred to as a noisy best response... Why?
- Questions:
  - As $\tau \to \infty$ what happens to the log linear learning?
  - As $\tau \to 0$ what happens to the log linear learning?
  - Will log-linear stay at a NE if it reaches one?
- The log linear learning process induces an aperiodic and irreducible Markov chain with state space X. State transition probability

$$P^{\tau}_{ss'} = Pr^{\tau}(a(t+1) = s' | a(t) = s).$$

- Let the unique stationary distribution μ^{τ}. A state s is called stochastically stable if

$$\lim_{\tau \to 0} \mu^{\tau}(s) > 0.$$

Homework 2: Recall the definition of aperiodic irreducible Markov chain.

Theorem

Consider a potential game $G = \{N, A, C\}$ with potential function ρ, if all players play the game according to log-linear learning, then the unique stationary distribution is

$$\mu^{\tau}(s) = \frac{e^{\frac{1}{\tau}\rho(s)}}{\sum_{s' \in S} e^{\frac{1}{\tau}\rho(s')}}. \tag{8}$$

Proof

- It is sufficient to prove that μ^{τ} satisfies the detailed balance condition.
- According to state transition probability, $P^{\tau}_{ss'} > 0$ if and only if there only one element different for s and s'.
- Without loss of generality, suppose $s_i \neq s'_i,\ s_{-i} = s'_{-i}$. Then

$$\mu^{\tau}(s)P^{\tau}_{ss'} = \frac{e^{\frac{1}{\tau}\rho(s)}}{\sum_{s'' \in S} e^{\frac{1}{\tau}\rho(s'')}} \cdot \frac{e^{\frac{1}{\tau}c_i(s'_i, s_{-i})}}{\sum_{s''_i \in S_i} e^{\frac{1}{\tau}c_i(a_i, a_{-i}(t))}} = \mu^{\tau}(s')P^{\tau}_{s's}.$$

C. Alós-Ferrer, N. Netzer, The logit-response dynamics, *Games & Economic Behavior*, Vol. 68, No. 2, 413-427, 2010.

- Stationary distribution = Probability that process will be in each action profile at large enough time.
- Illustration: $\mu^\tau(a_1, a_2)$ is the likelihood we will be at the action profile (a_1, a_2) if we use log-linear learning on the above game and wait long enough.
- Implications:
  - As $\tau \to \infty$ what happens to the stationary distribution?
  - As $\tau \to 0$ what happens to the stationary distribution?

Example 3

Recall the common interest game in Example 1

$P_1 \backslash P_2$	a_2	b_2
a_1	$(\underline{10},\ \underline{10})$	$(0,\ 0)$
b_1	$(0,\ 0)$	$(\underline{5},\ \underline{5})$

What is the asymptotic behavior under log-linear learning?

- If $\tau = 1000$ the unique stationary distribution is

$$\mu^{1000} = [0.25, 0.25, 0.25, 0.25].$$

- If $\tau = 10$ the unique stationary distribution is

$$\mu^{10} = [0.42, 0.16, 0.16, 0.26].$$

- If $\tau = 1$ the unique stationary distribution is

$$\mu^{1} = [0.993, 0.0, 0.0, 0.07].$$

- If $\tau = 0.1$ the unique stationary distribution is

$$\mu^{0.1} = [1, 0.0, 0.0, 0.0].$$

Theorem

Consider a potential game G with potential function ρ, if all players play the game according to log-linear learning, then the set of stochastically stable states is the potential maximizer.

Proof

For any given profile $a \in A$

$$\begin{aligned} \lim_{\tau \to 0} \mu^{\tau}(s) &= \lim_{\tau \to 0} \frac{e^{\frac{1}{\tau}\rho(s)}}{\sum_{s' \in S} e^{\frac{1}{\tau}\rho(s')}} \\ &= \lim_{\tau \to 0} \frac{1}{\sum_{s' \in S} e^{\frac{1}{\tau}[\rho(s') - \rho(s)]}}. \end{aligned}$$

According to the analysis, only when $s^* \in \arg\max_{s \in S} \rho(s)$, we have $\lim_{\tau \to 0} \mu^{\tau}(s) > 0$.

3.3 Fictitious Play

- Consider a repeated finite game. Each player has subjective beliefs about other players' future behavior. In each period each player chooses actions according to his belief and updates his belief according to the past observations. We call such a process a belief-based learning process.
- A nature idea is: can we forecast the opponent's strategies using statistical approach (history information)?
- How to make a decision according to the prediction?
- Above learning rule is called fictitious play, a basic belief-based learning process.

Fictitious Play

- Empirical frequency of player i selecting a_i, denoted by $q_i^{a_i}(t)$, is

$$q_i^{a_i}(t) = \frac{1}{t}\sum_{\tau=0}^{t-1} I\{a_i(\tau) = a_i\}.$$

- Empirical frequency vector of player i

$$q_i(t) := [q_i^{a_1}(t), q_i^{a_2}(t), \cdots, q_i^{a_{k_i}}(t)]^{\mathrm{T}} \in \Upsilon_{k_i \times 1}.$$

- Expected utility for the action a_i is

$$u_i(a_i, q_{-i}(t)) := \sum_{a_{-i} \in A_{-i}} c_i(a_i, a_{-i}) \prod_{a_j \in a_{-i}} q_j^{a_j}(t).$$

- Select an optimal action $a_i(t+1) \in BR_i(q_{-i}(t))$, where

$$BR_i(q_{-i}(t)) := \arg\max_{a_i \in A_i} u_i(a_i, q_{-i}(t)).$$

Remark: The action of player i at each time is based on the (maybe incorrect) presumption that other players are playing randomly and independently according to their empirical frequencies.

Theorem

Consider a potential game G, if all players play the game according to fictitious play, then the empirical frequency $q(t)$ converges to mixed NE of G.


D. Monderer and L. S. Shapley, Potential games, *Games & Economic Behavior*, Vol. 14, 124 – 143, 1996.


D. Monderer and L. S. Shapley, Fictitious Play Property for Games with Identical Interests, *Journal of Economic Theory*, Vol. 68, 258-265, 1996.


X. Zhang, D. Cheng, Profile dynamic based fictitious play, *Science China Information Science*, Vol. 64, 169202:1-3, 2021.

4. Learning in State-Based Games

How to design a learning rule for games

in dynamic environment?

4.1 State-Based Games

Definition

A finite state-based game is a tuple $\mathcal{G} = \{N, \{A_i\}_{i\in N}, \{c_i\}_{i\in N}, X, P\}$, *where*

1. $N = \{1, 2, \cdots, n\}$: *the set of agents;*
2. $A_i = \{1, 2, \cdots, k_i\}$: *the set of actions of agent* i;
3. $c_i : A \times X \to \mathbb{R}$ *is the payoff function of agent* $i \in N$, *where* $A = \prod_{i=1}^{n} A_i$ *is the action profile set;*
4. $X = \{1, 2, \cdots, m\}$: *the set of underlying finite state;*
5. $P : X \times A \to \Upsilon(X)$: *the Markovian state transition function.* $\Upsilon(X)$ *denotes the set of probability distributions over the finite state space* X. $P(a; x, y)$ *is the probability of state* x *transiting to state* y *under the action of* a.

H.P. Young, *Strategic Learning and Its Limits*. Oxford, U.K.: Oxford Univ. Press, 2004.

J.R. Marden, "State-based potential games," *Automatica*, vol. 48, no. 12, pp. 3075-3088, 2012.

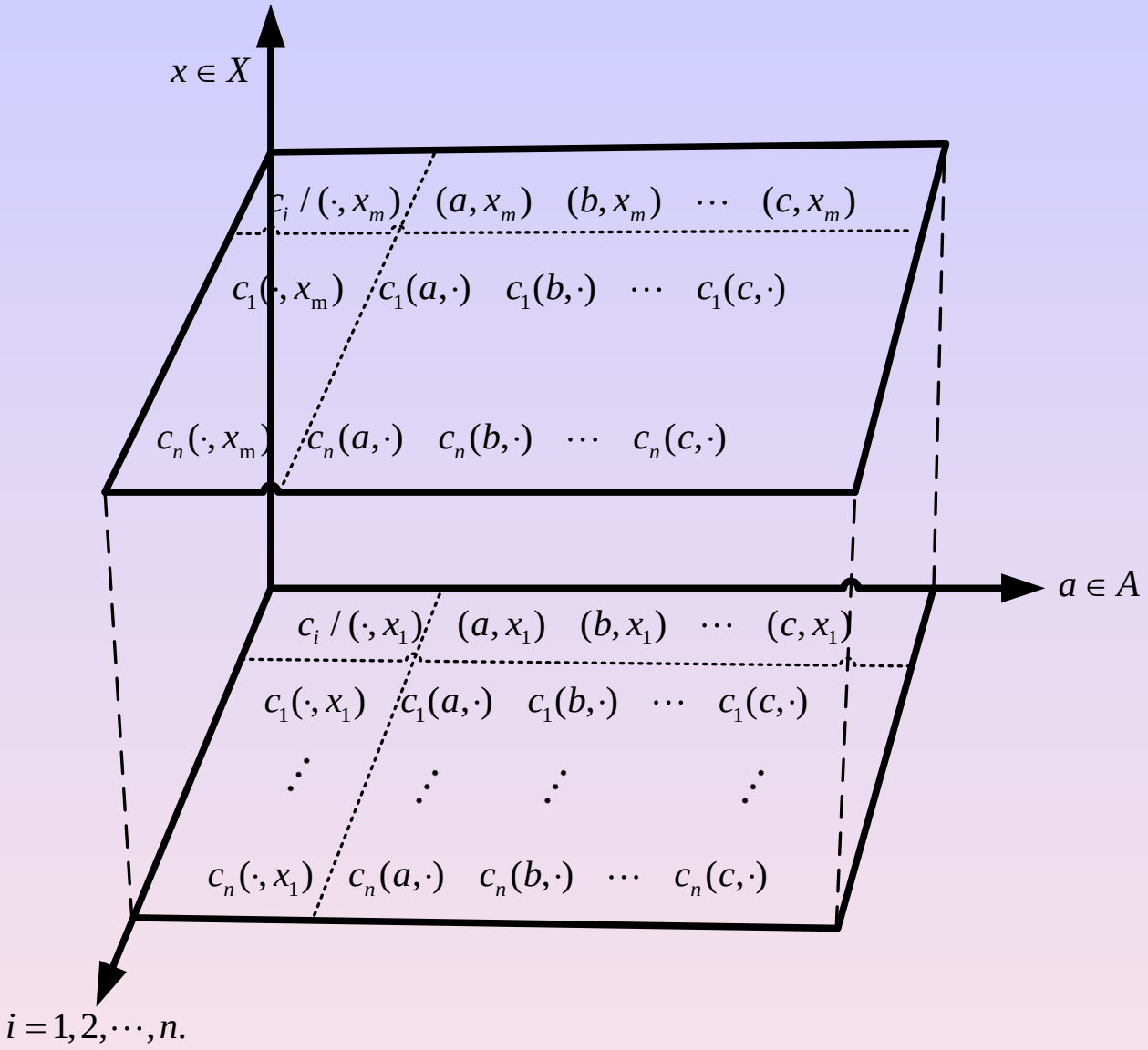


Figure 5: State-based games

Reachable set

- A state y is reachable from initial state x driven by an invariant action a, if and only if, there exists a time $t_y > 0$ such that

$$\mathbf{Pr}[x(t_y) = y] > 0,$$

  conditioned on the events $x(0) = x$ and $x(k+1) \in P(a, x(k), \cdot)$ for all $k \in \{0, 1, \cdots, t_y - 1\}$.

- The transition process can be illustrated as

$$x \xrightarrow{a} x(1) \xrightarrow{a} \cdots \xrightarrow{a} x(t_y - 1) \xrightarrow{a} x(t_y) = y.$$

- Denote by $X(a|x) \subseteq X$ the set of reachable states starting from initial state x driven by an invariant action a.

Definition

Consider a state-based game $G = \{N, \{A_i\}_{i\in N}, \{c_i\}_{i\in N}, X, P\}$. *The action state pair* $[a^*, x^*]$ *is a recurrent state equilibrium (RSE) with respect to the state transition process* $P(\cdot)$ *if the following two conditions are satisfied:*

1. $X(a^*|x^*)$ *is a recurrent state set of* $P(a^*; \cdot, \cdot)$*;*
2. *For each agent* $i \in N$ *and every state* $x \in X(a^*|x^*)$*,*

$$c_i(a_i^*, a_{-i}^*, x) \geqslant c_i(a_i, a_{-i}^*, x), \ \forall a_i \in A_i.$$

Example

Consider a state-based game with $N = \{1, 2\}$, $A_1 = A_2 = \{1, 2\}$, $X = \{1, 2, 3\}$. The followings are the payoff matrices.

Table 3: Coordination game of $x = 1$

$P_1 \backslash P_2$	1	2
1	$(4,\ 4)$	$(1,\ 3)$
2	$(3,\ 1)$	$(2,\ 2)$

Table 4: Prisoner's dilemma game of $x = 2$

$P_1 \backslash P_2$	1	2
1	$(2,\ 2)$	$(0,\ 3)$
2	$(3,\ 0)$	$(1,\ 1)$

Table 5: Matching pennies game of $x = 3$

$P_1 \backslash P_2$	1	2
1	$(-1,\ 1)$	$(1,\ -1)$
2	$(1,\ -1)$	$(-1,\ 1)$

Example (Contd)

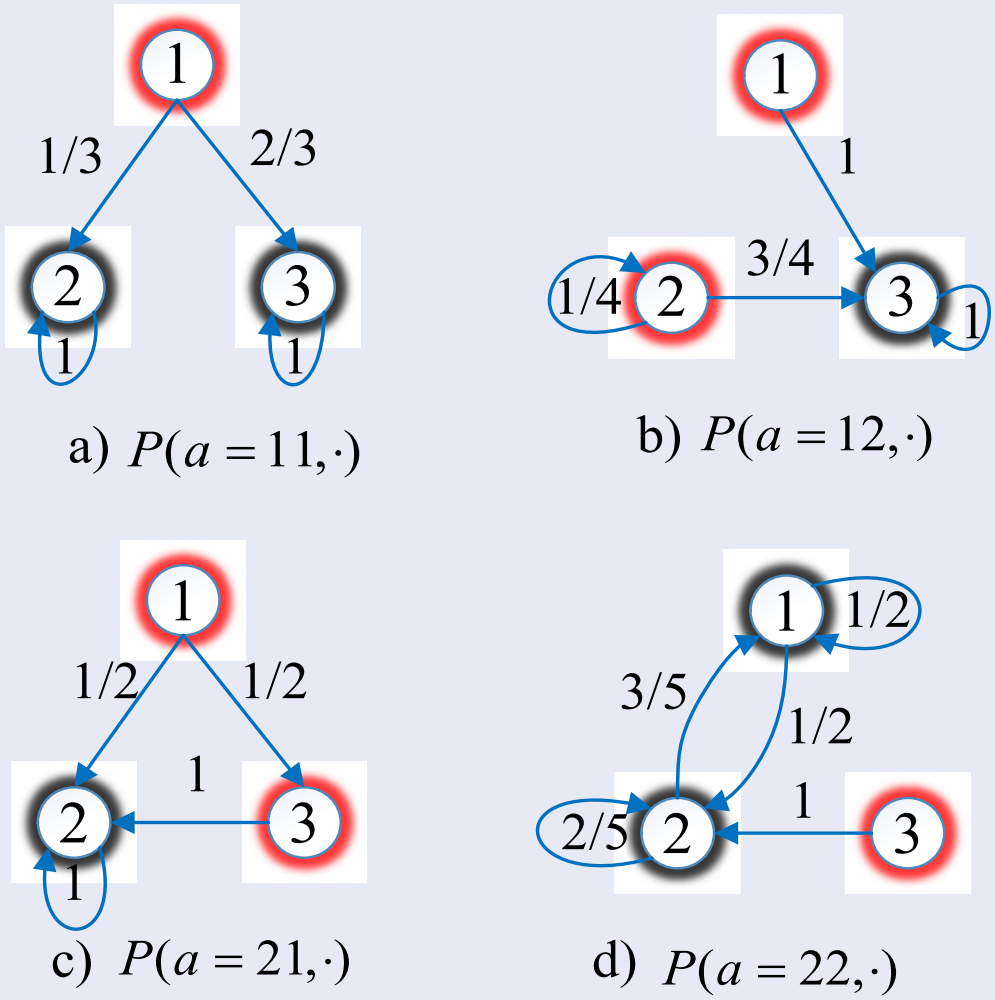


Figure 6: State transition diagram of the example

☞ State-Based Potential Games

Definition

A finite state-based game is called state-based potential game if there is a function $\phi : A \times X \to \mathbb{R}$*, such that for each* $[a, x] \in A \times X$*, the following conditions satisfied*

1. *for each* $i \in N$ *and* $a_i, b_i \in A_i$*,*

$$c_i(a_i, a_{-i}, x) - c_i(b_i, a_{-i}, x) = \phi(a_i, a_{-i}, x) - \phi(b_i, a_{-i}, x);$$

2. *for the state* y *which belongs to the support of* $P(a; x, \cdot)$

$$\phi(a, y) \geq \phi(a, x).$$

Theorem

Each state-based potential game has at least one RSE.

Homework: Prove above theorem.

☞ Learning in state-based potential games

State-Based Better Reply with Inertia (SBR with inertia)

- For each $[a, x]$, denote the better reply set of player i as

$$B_i(a, x) := \{b_i \in A_i \mid c_i(x, b_i, a_{-i}) > c_i(x, a_i, a_{-i})\}.$$

- If $B_i(a(t-1), x(t)) = \emptyset$, then the action of i at t is

$$a_i(t) = a_i(t-1).$$

- If $B_i(a(t-1), x(t)) \neq \emptyset$, then

$$\begin{cases} p_i^{a_i} = \epsilon, & a_i = a_i(t-1) \\ p_i^{a_i'} = \frac{1-\epsilon}{B_i(a(t-1), x(t))}, & a_i' \in B_i(a(t-1), x(t)) \\ p_i^{a_i} = 0, & \text{Otherwise} \end{cases}$$

Theorem

Consider a state-based potential game, if all players update his action according to state-based better reply with inertia, then $[a(t), x(t)]$ *will converge to the set of RSE almost surely.*

Homework: Prove above theorem.

J.R. Marden, “State-based potential games,” *Automatica*, vol. 48, no. 12, pp. 3075-3088, 2012.

4.2 Learning Design For State-Based Games

Question: Can we design a learning rule which can converge to RSE for general state-based games (at least two memory)?

☞ Information Structure

- Available information at time t

$$O_i(t) := \{a(t-2), a(t-1), x(t); c_i(a,x)\}.$$

- Recall the better reply set of player i

$$B_i(a,x) := \{b_i \in A_i : \ c_i(b_i, a_{-i}, x) > c_i(a,x)\};$$

- For simplicity, let

$$B_i(t) := B_i(a(t-1), x(t)), \ \forall t > 1.$$

The flow of the proposed learning rule

$$a(t-2) \stackrel{?}{=} a(t-1) \begin{cases} \text{YES}, \begin{cases} B_i(t) = \emptyset, \text{then } a_i(t) = a_i(t-1). \\ B_i(t) \neq \emptyset, \text{then } \begin{cases} p_i^{a_i(t-1)}(t) = \epsilon_i, 0 < \epsilon_i < 1 \\ p_i^{a_i}(t) = \frac{1-\epsilon_i}{|B_i(t)|}, \forall a_i \in B_i(t). \end{cases} \end{cases} \\ \text{NO}, \begin{cases} p_i^{a_i(t-1)}(t) = \epsilon_i, \\ p_i^{a_i}(t) = \frac{1-\epsilon_i}{|A_i|-1}, \forall a_i \neq a_i(t-1). \end{cases} \end{cases}$$

$\epsilon_i \in (0,1)$ is the inertia of agent i.

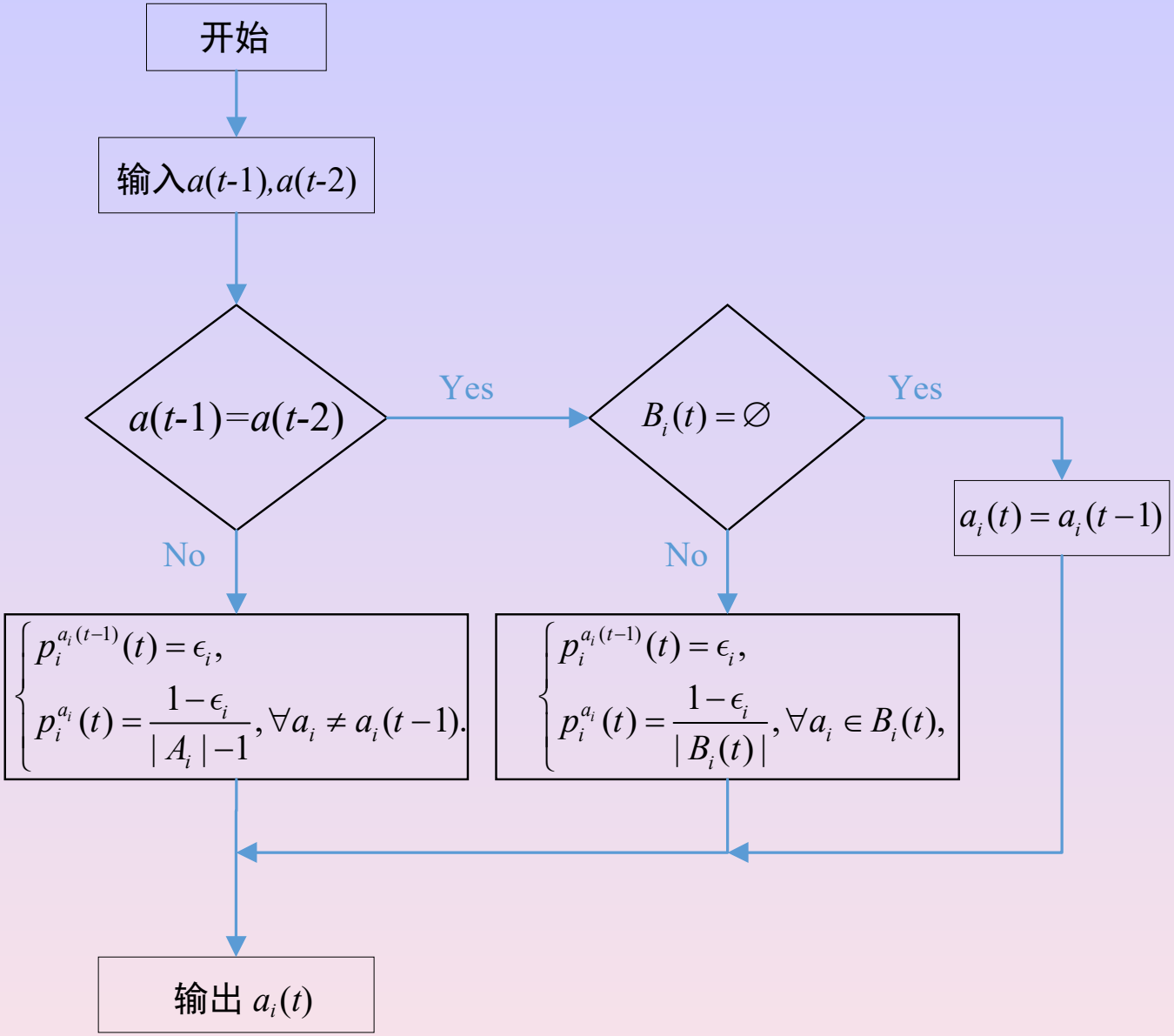


Figure 7: A two memory strategy learning rule

☞ Combined dynamics

- Combine state dynamics and profile dynamics

$$\begin{cases} x(t+1) = M_P x(t) a(t), \\ a(t+1) = M_F x(t+1) a(t) a(t-1). \end{cases}$$

- Let $\omega(t) := x(t) \ltimes a(t) \ltimes a(t-1), t \geq 1$, then

$$\omega(t+1) = M\omega(t).$$

  where $M = \{m_{j,i}\}$ and $m_{j,i}$ is the probability from i to j.
- A Markov Chain on $\Omega := A \times X \times A$. The initial distribution is

$$\begin{aligned} & \mathbf{Pr}\{\omega(1) = [a(1), x(1), a(0)] \mid x(0) = x_0\} \\ = & \left(\prod_{1 \leq i \leq n} \frac{1}{|A_i|}\right)^2 \mathsf{P}(x^0) P(a(0); x(0), x(1)) \\ = & \frac{1}{k^2} \mathsf{P}(x^0) P(a(0); x(0), x(1)). \end{aligned}$$

☞ State-transition probability

Consider any $\omega_1, \omega_2 \in \Omega$, where $\omega_1 = [a^1, x^1, a^2]$,$\omega_2 = [b^1, x^2, b^2]$.

(1) If $a^2 \neq b^1$, then

$$\mathbf{Pr}\{\omega(t+1) = \omega_2 | \omega(t) = \omega_1\} = 0.$$

(2) If $a^2 = b^1$, and $x^2 \notin P(a^2; x^1, \cdot)$, then

$$\mathbf{Pr}\{\omega(t+1) = \omega_2 | \omega(t) = \omega_1\} = 0.$$

(3) If $a^2 = b^1 \neq a^1$,and $x^2 \in P(a^2; x^1, \cdot)$, then

$$\begin{aligned}&\mathbf{Pr}\{\omega(t+1) = \omega_2 | \omega(t) = \omega_1\} \\ &= \epsilon^{n-|H(b^1,b^2)|} \cdot \prod_{i \in H(b^1,b^2)} \frac{1-\epsilon}{|A_i| - 1},\end{aligned}$$

where $H(a,b) := \{i \in N : a_i \neq b_i\}, a, b \in A$.

(4) If $a^2 = b^1 = a^1$, and $x^2 \in P(a^2; x^1, \cdot)$, then

$$\begin{aligned}&\mathbf{Pr}\{\omega(t+1) = \omega_2 | \omega(t) = \omega_1\} \\ &= \epsilon^{n-|H(b^1,b^2)|-|N(b^1,x^2)|} \cdot \prod_{i \in H(b^1,b^2)} \frac{1-\epsilon}{|B_i(b^1, x^2)|},\end{aligned}$$

where $N(a,x) := \{i \in N : B_i(a,x) = \emptyset\}$.

Theorem

(常返状态平衡的吸引性) Consider a state-based games with $[a^*, x^*]$ *being the RSE. For initial state* $x(0) = x^0$*, if there exists an integer* $K \geq 2$*, such that the sequence* $(a^0, x^1),\ (a^1, x^2),\ \cdots, (a^K, x^{K+1})$ *generated by the designed learning rule satisfy that*

(1) $P(a^2; x^2, x^3)P(a^3; x^3, x^4)\cdots P(a^K; x^K, x^{K+1}) > 0$*;*

(2) *if* $a^{k-1} = a^k$ *holds for some integer* $k \in [1, K)$*, then* $a^{k+1} \in BT(a^k, x^{k+1})$*;*

(3) $[a^K, x^{K+1}] = [a^*, x^*]$ *is the RSE,*

then the designed learning rule will converge to RSE almost surely.

Theorem

Consider a state-based games with $[a^*, x^*]$ *being the RSE. If the following two conditions are satisfied*

(1) $\bar{P} := \frac{1}{|A|}\sum_{a\in A} P(a;\cdot,\cdot)$ *is irreducible;*

(2) $P(a;x,x) > 0, \forall [a,x] \in A \times X.$

then the designed learning rule will converge to RSE almost surely for any initial state $x(0)$.

Remark

根据定理中的条件 *(1)* 和 *(2)* 可知 $\bar{P}$ 为非周期不可约的马尔科夫链. 但 $\bar{P}$ 的非周期不可约是否能保证学习规则的收敛性还是个未知的问题.

Changxi Li, Yu Xing, Fenghua He, Daizhan Cheng, "A strategic learning algorithm for state-based games," *Automatica*, vol. 113, 108615, 2020.

Example

Consider the following state-based game with $N = \{1, 2\}$, $A_1 = A_2 = \{1, 2\}$, $X = \{1, 2, 3, 4\}$. The payoff matrices are shown as follows.

Table 6: Payoff Bi-Matrix for $x = 1, 2, 3, 4$

Player 1\Player 2	1	2
1	$(5,\ 4)$	$(2,\ 3)$
2	$(4,\ 2)$	$(3,\ 1)$

Player 1\Player 2	1	2
1	$(2,\ 2)$	$(3,\ 1)$
2	$(0,\ 3)$	$(2,\ 1)$

Player 1\Player 2	1	2
1	$(-1, 1)$	$(1, -1)$
2	$(1, -1)$	$(-1, 1)$

Player 1\Player 2	1	2
1	$(2,\ 2)$	$(2,\ 3)$
2	$(0,\ 3)$	$(3,\ 1)$

Example (Contd)

The Markov transition matrices under different actions have the following form.

$$P(a,\cdot) = \begin{bmatrix} p_{11}(a), & p_{12}(a), & 0, & 0 \\ p_{21}(a), & p_{22}(a), & 0, & 0 \\ 0, & 0, & p_{33}(a), & p_{34}(a) \\ 0, & 0, & p_{43}(a), & p_{44}(a) \end{bmatrix}.$$

where $0 < p_{ij}(a) < 1$ is the probability that state i transfers to state j, $\forall a \in \{11, 12, 21, 22\}$. RSE: $[a = 11, x = 1$ or $x = 2]$.

Claim: *If for all Markov chain $P(a,\cdot), \forall a \in A$ there exists a common closed set, denoted by $X^c \subseteq X$, s.t., for all $x \in X^c$ there do not exist a joint action $a \in A$ such that $[a, x]$ is a RSE. Then there does not exist learning rule that converges to a recurrent state equilibrium for state-based games where such a equilibrium exists.*

Corollary

- 当状态演化博弈退化为有限非合作博弈时, 常返状态平衡退化为纳什均衡. 故本文提出的算法可以用于寻找有限博弈中的纯纳什均衡.
- 本文提出的学习规则是一个具有两个记忆时刻的算法, 不同于文献中已有的算法. 例如, 梯度算法适合于具有连续策略集合的博弈, 本文的算法适合于离散形式. 虚拟对策规则需要所有个体能够记住所有的历史信息. 至于最优响应规则, 则会陷入到调整循环中
- *J. R. Marden* 针对状态势博弈提出了一种有限记忆学习规则, 并且证明了只需一个记忆就能保证几乎必然收敛到状态势博弈的常返状态平衡. 我们的结果则显示了对于一般的状态演化博弈, 需要两个记忆才能保证几乎必然收敛到状态演化博弈的常返状态平衡.

Example

Recall S. Hart game.

Table 7: S. Hart Game

$V_i \backslash a$	111	112	113	121	122	123	131	132	133	211	212	213	221	222	223	231	232	233	311	312	313	321	322	323	331	332	333
V_1	0	4	2	0	4	3	2	3	0	4	0	3	4	0	2	3	2	0	1	1	0	1	1	0	0	0	6
V_2	0	0	2	4	4	3	1	1	0	4	4	3	0	0	2	1	1	0	2	3	0	3	2	0	0	0	6
V_3	0	4	1	4	0	1	2	3	0	0	4	1	4	0	1	3	2	0	2	3	0	3	2	0	0	0	6

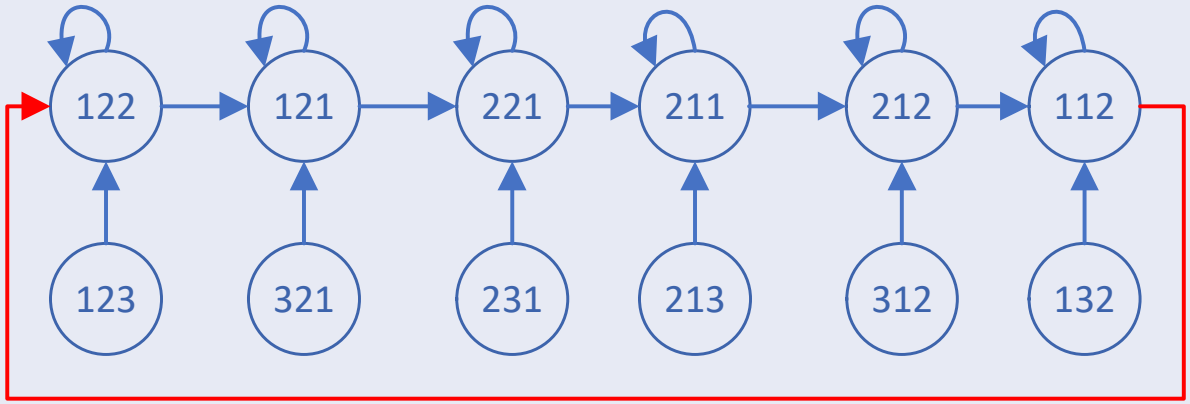
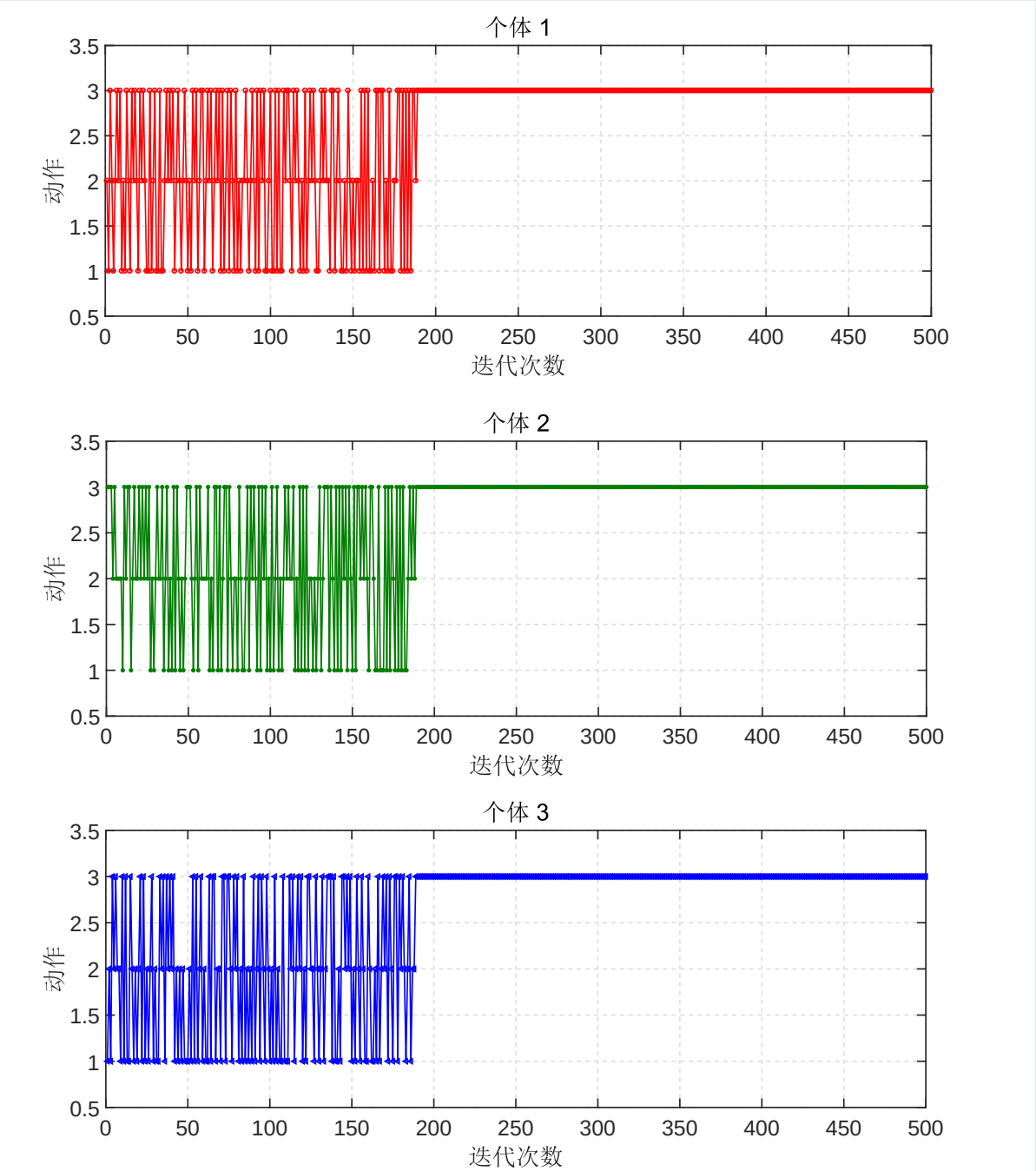


Figure 8: An Adjustment Cycle

S. Hart, Y. Mansour, Stochastic uncoupled dynamics and Nash equilibrium, *Games and Economic Behavior*, Vol. 57, No. 2, 286-303, 2006.

Example (Contd)



Example

Consider a networked game with $N = \{1, 2, 3\}$, and each agent has two actions. Three states $X = \{x_1, x_2, x_3\}$.

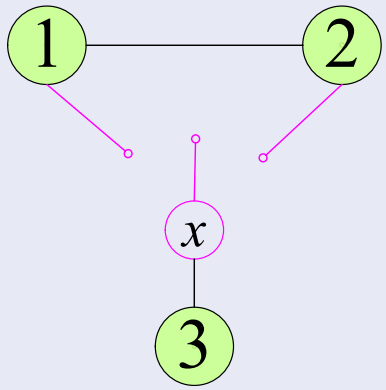


Figure 10: 时变多智能体的通信结构图

Example (Contd)

- *State evolution process*:

Table 8: State evolution process

$P(a,x_1)\backslash a$	111	112	121	122	211	212	221	222
x_1	$1/3$	$1/4$	$1/2$	1	$1/2$	0	$1/3$	$1/3$
x_2	$1/3$	$1/4$	0	0	$1/4$	1	0	$1/3$
x_3	$1/3$	$1/2$	$1/2$	0	$1/4$	0	$2/3$	$1/3$
$P(a,x_2)\backslash a$	111	112	121	122	211	212	221	222
x_1	1	0	$2/3$	0	0	$1/2$	0	$1/3$
x_2	0	1	$1/3$	$1/6$	$5/6$	$1/2$	0	$1/3$
x_3	0	0	0	$5/6$	$1/6$	0	1	$1/3$
$P(a,x_3)\backslash a$	111	112	121	122	211	212	221	222
x_1	$1/2$	$1/2$	1	0	$1/4$	0	1	$1/3$
x_2	0	$1/2$	0	$1/2$	0	1	0	$1/3$
x_3	$1/2$	0	0	$1/2$	$3/4$	0	0	$1/3$

$\Rightarrow P(a = 222; \cdot, \cdot)$: aperiodic, irreducible.

Example (Contd)

- *Utility design*: $c_i(a,x) = c_i(a_{N_i}, a_i, x), \forall i \in N.$

Table 9: Designed utility function

$c_i(a,x_1)\backslash a$	111	112	121	122	211	212	221	222
c_1	1	0	0	-1	1	1	2	3
c_2	1	1	2	2	1	1	3	3
c_3	-1	0	-1	0	1	3	1	3
$c_i(a,x_2)\backslash a$	111	112	121	122	211	212	221	222
c_1	1	1	3	3	0	2	5	5
c_2	1	0	3	4	5	2	4	7
c_3	1	0	-1	2	1	0	-1	2
$c_i(a,x_3)\backslash a$	111	112	121	122	211	212	221	222
c_1	1	1	0	0	-1	-1	4	4
c_2	2	2	3	3	1	1	5	5
c_3	2	3	2	3	2	3	2	3

Example (Contd)

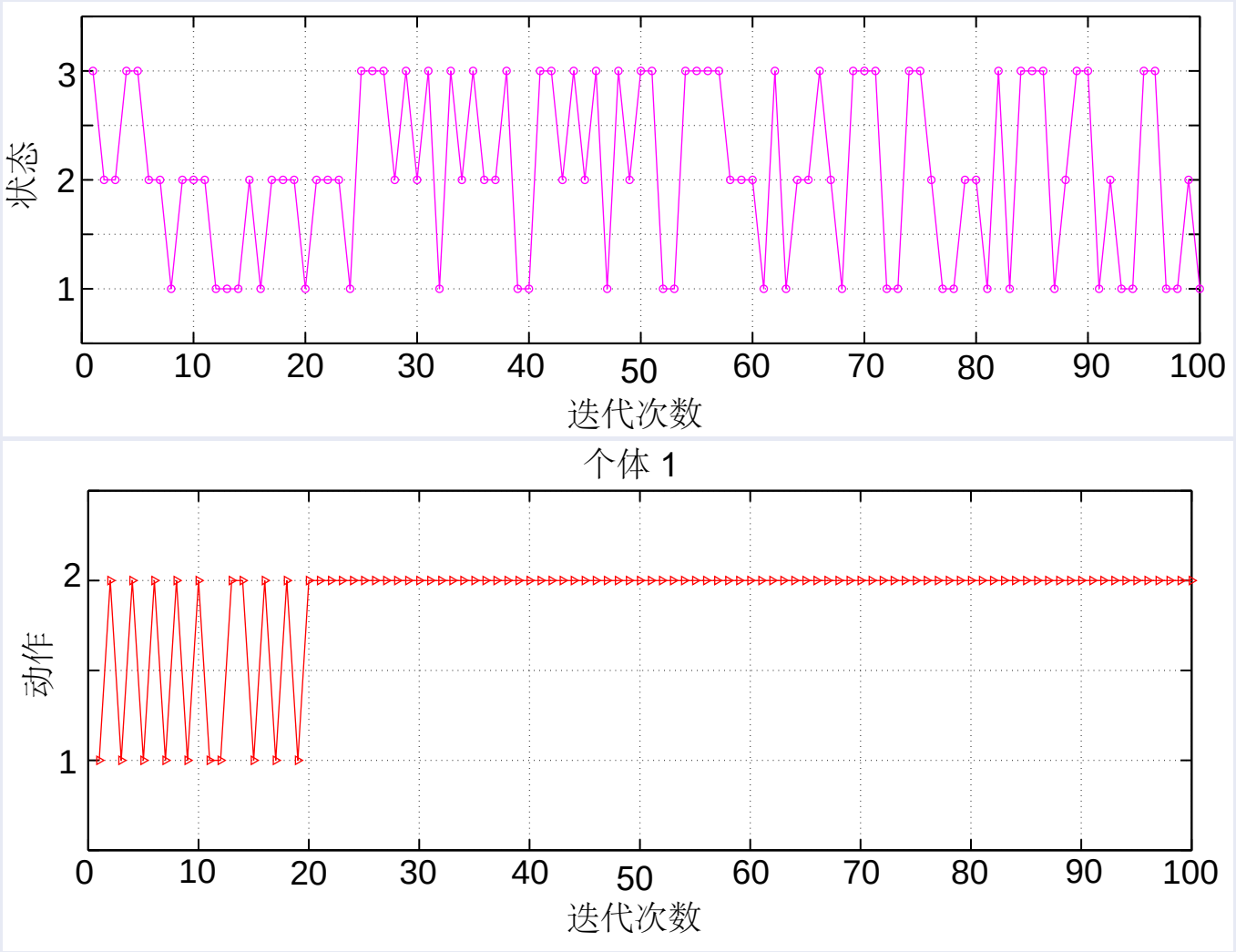


Figure 11: Dynamics of states and actions of agent 1

Example (Contd)

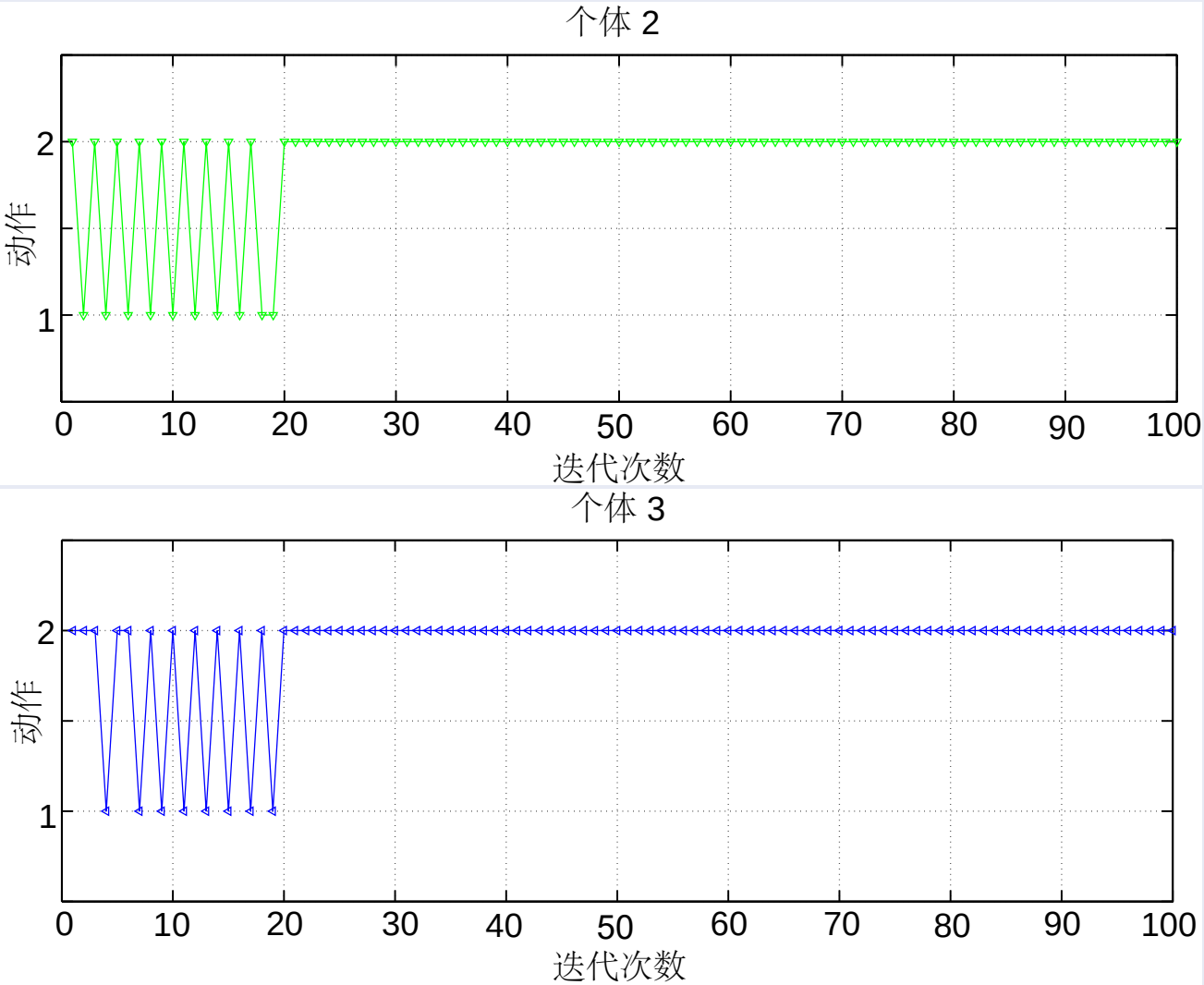


Figure 12: Dynamics of actions for agent 2 and 3

5. Application in Game Theoretic Control

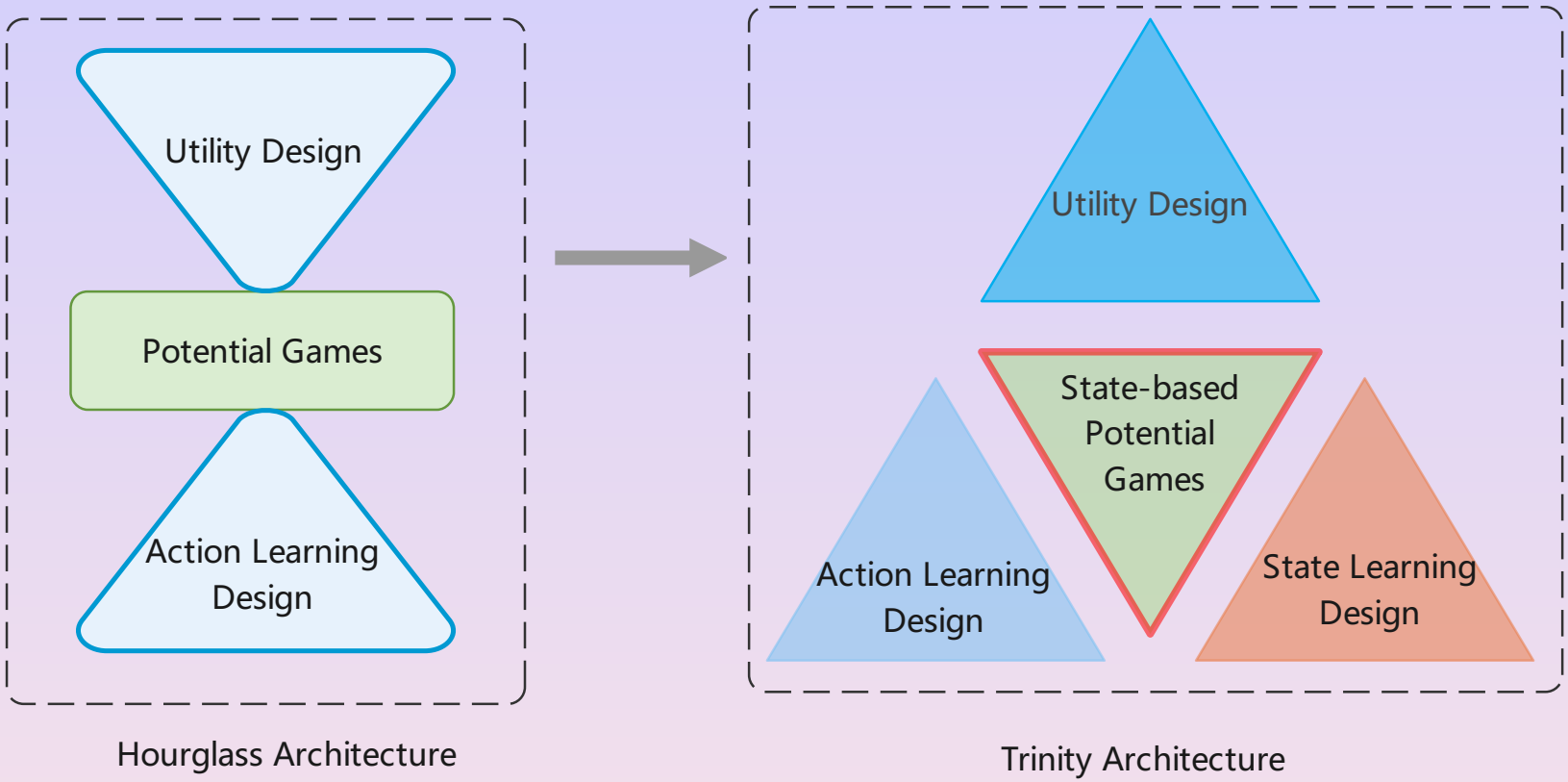


Figure 13: Architecture of Game Theoretic Control

R. Gopalakrishnan, J. R. Marden, and A. Wierman, "An architectural view of game theoretic control," *ACM SIGMETRICS Performance Evaluation Review*, vol. 38, no. 3, pp. 31-36, 2011.

5.1 Game Theoretic Control of Multi-Agent Systems

- Consider the following optimization problem of an MAS

$$\begin{array}{ll} \max & \phi(a_1, a_2, \cdots, a_n) \\ \text{s.t.} & a_i \in A_i, \forall i \in N, x \in X \\ & \mathcal{G} = (N, \mathcal{E}). \end{array} \tag{9}$$

  where $\phi : A \times X \to \mathbb{R}$ is a system level function that the system wants to optimize.

- How to update the strategy for each player to realize the optimization?


Ting Liu, Jinhuan Wang, Xiao Zhang, Daizhan Cheng, Game Theoretic Control of Multiagent Systems. *SIAM J. Control. Optim.* 57(3): 1691-1709, 2019.


C. Li, F. He, H. Qi, et. al., "Potential Games Design Using Local Information," *Proeedings of 57th IEEE Conference on Decision and Control*, pp. 1911-1916, 2018.

☞ Steps of game theoretic control

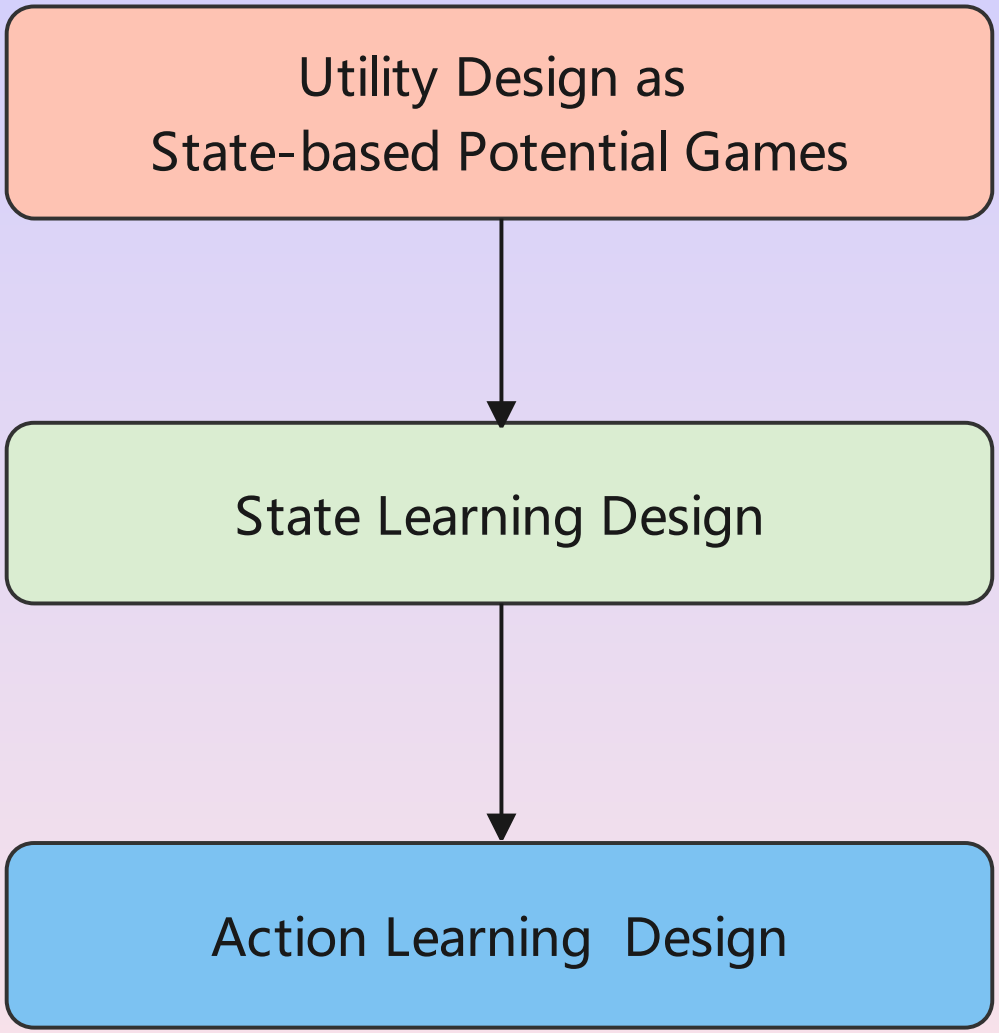


Figure 14: Steps of game theoretic control

☞ Step 1: Utility Design

- A finite game G is potential if and only if there exist functions $d_i : A^{-i} \times X \to \mathbb{R}, i \in N$ such that for every $a \in A$

$$c_i(a,x) = \phi(a,x) + d_i(a^{-i},x),\ \forall i \in N, \forall x \in X, \tag{10}$$

  where $P(a)$ is the potential function, and $a^{-i} \in A^{-i}$.
- If the utility function is local information-based, then

$$c_i(a,x) = V_i \ltimes_{j\in N_i} a_j \ltimes x = V_i \Gamma_{N_i} \ltimes_{j=1}^{n} a_j \ltimes x.$$

  Rewrite (10) into vector form yields

$$\begin{array}{rcl} V_\phi \ltimes_{i=1}^{n} a_i \ltimes x & = & V_i \ltimes_{j\in N_i} a_j \ltimes x - V_i^d \ltimes_{j\neq i} a_j \ltimes x, \\ & = & V_i \Gamma_{N_i} \ltimes_{j=1}^{n} a_j \ltimes x - V_i^d \Gamma_{-i} \ltimes_{j=1}^{n} a_j \ltimes x, \end{array}$$

  where V_ϕ and V_i^d are the structure vectors of ϕ and d_i.
- Since $a_i \in \Delta_{k_l}, i \in N,\ x$ are arbitrary, we have

$$V_i \Gamma_{N_i} = V_\phi + V_i^d \Gamma_{-i},\ \ \forall i \in N. \tag{11}$$

Theorem

Consider the optimization problem in (9) with the following global objective function ϕ

$$\phi(a) = V_\phi \ltimes_{j=1}^{n} a_j \ltimes x,$$

where V_ϕ *is the structure vector of* ϕ*. The optimization problem can be modeled as a local information based state-based potential game with the global objective function* ϕ *as its potential function, if and only if, all the following equations have a solution* $\xi_i, \forall i$

$$T_i \cdot \xi_i^r = V_\phi^{\mathrm{T}}(x = r), \quad \forall r \in X \tag{12}$$

where $T_i = \Gamma_{N_i}^{\mathrm{T}}, \Gamma_{-i}^{\mathrm{T}}$, $\xi_i^r = [(\xi_{i,1}^r)^{\mathrm{T}}, (\xi_{i,2}^r)^{\mathrm{T}}]^{\mathrm{T}}$, $\xi_{i,1}^r \in \mathbb{R}^{k_{N_i}}$, $\xi_{i,2}^r \in \mathbb{R}^{k_{-i}}$, $k_{N_i} = \prod_{j \in N_i} k_j$, $k_{-i} = \prod_{j \neq i} k_j$, *and* $N_i = U(i) \cup \{i\}, \forall i \in N$.
Moreover if the solution $\xi_i^r, \forall i \in N$ *exists, the local information based utility function of agent* i *is*

$$c_i(a, x = r) = (\xi_{i,1}^r)^{\mathrm{T}} \Gamma_{N_i} \ltimes_{j=1}^{n} a_j \ltimes x, \forall i \in N. \tag{13}$$

Subset drawing matrix

- Consider a subset player $U \subset N$. Set $\Gamma_U := \otimes_{i=1}^{n} \gamma_i$, where

$$\gamma_i = \begin{cases} I_{k_i}, i \in U \\ \mathbf{1}_{k_i}^T, \text{ Otherwise.} \end{cases}$$

- Γ_U can "draw" the strategies of players in U from the profile, that is

$$\ltimes_{j \in U} a_j = \Gamma_U \ltimes_{i=1}^{n} a_i.$$

Homework: Try to prove above equation.

☞ Step 2: State evolutionary process (SEP) design

$$x(t+1) = M_P x(t) a(t).$$

- SEP-1 (Remaining Priority): Construct

$$B_1(x(t)|a(t)) := \{x_j \mid \phi(x_j, a(t)) > \phi(x(t), a(t))\}.$$

Then

$$\begin{cases} x(t+1) = x(t), & \text{if } B_1(x(t)|a(t)) = \emptyset, \\ P(x(t+1) = x_j)) = \frac{1}{|B_1(x(t)|a(t))|}, & \text{if } x_j \in B_1(x(t)|a(t)). \end{cases}$$

- SEP-2 (Equal Probability): Construct

$$B_2(x(t)|a(t)) := \{x_j \mid \phi(x_j, a(t)) \geq \phi(x(t), a(t))\}.$$

Then

$$P(x(t+1) = x_j)) = \frac{1}{|B_2(x(t)|a(t))|}, \text{ if } x_j \in B(x(t)|a(t)).$$

☞ Step 3. Action Learning: SBR with inertia

$$a(t+1) = M_a x(t+1) a(t).$$

- For each $[a, x]$, denote the better reply set of player i as

$$B_i(a, x) := \{b_i \in A_i \mid c_i(x, b_i, a_{-i}) > c_i(x, a_i, a_{-i})\}.$$

- If $B_i(a(t-1), x(t)) = \emptyset$, then the action of i at t is

$$a_i(t) = a_i(t-1).$$

- If $B_i(a(t-1), x(t)) \neq \emptyset$, then

$$\begin{cases} p_i^{a_i} = \epsilon, & a_i = a_i(t-1) \\ p_i^{a_i'} = \frac{1-\epsilon}{B_i(a(t-1), x(t))}, & a_i' \in B_i(a(t-1), x(t)) \\ p_i^{a_i} = 0, & \text{Otherwise} \end{cases}$$

Theorem

Both the SEP-1 and the SEP-2 assure the conditions of state-based potential games' definition.

Why?

Example

Consider a consensus problem of a MAS with network graph Fig. 18. $N = \{1, 2, 3, 4\}$ with a common action set $S_i = \{1, 2\}, i \in N$. Assume all players can only communicate with their neighbors. Additionally, there is a switch, denoted by x, which can link agent 1 with 2, or agent 1 with 3, or neither of them. The system objective function is

$$\phi(a, x) = 2 \sum_{i \in N} \mathbf{1}_{\{a_i = 1\}} + \sum_{(i,j) \in E(x)} \frac{\mathbf{1}_{\{a_i = a_j\}}}{2}.$$

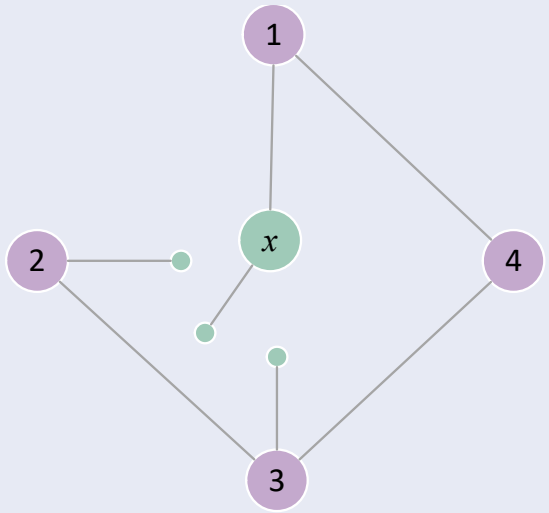


Figure 15: Network Graph

Example (Contd)

Define the state set $X = \{x_1, x_2, x_3\}$, where x_1 means the switch x is open; x_2 means the switch x is connected with node 3; x_3 means the switch x is connected with node 2. Then there are 3 states shown in Fig. 16.

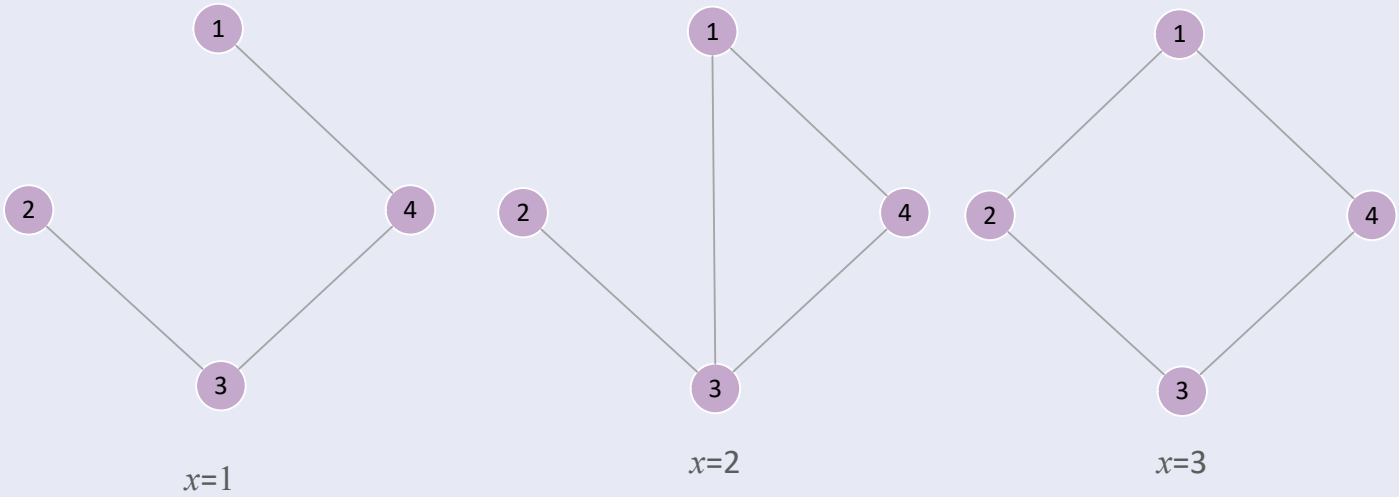


Figure 16: network graph

Then

$$\phi(x_i, a) = V^{\phi(x_i,\cdot)}ax.$$

where

$$\begin{aligned}
V^{\phi(x_1,\cdot)} &= \delta_{16}[11,7,7,5,8,4,6,4,8,6,4,4,5,3,3,3],\\
V^{\phi(x_2,\cdot)} &= \delta_{16}[12,8,7,5,9,5,6,4,8,6,5,5,5,3,4,4],\\
V^{\phi(x_3,\cdot)} &= \delta_{16}[12,8,8,6,8,4,6,4,8,6,4,4,6,4,4,4].
\end{aligned}$$

Example (Contd)

Design state evolutionary process using SEP-2:

$$V^{P(x_1)} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix},$$

$$V^{P(x_2)} = \begin{bmatrix} 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{3} & \frac{1}{3} & 1 & 1 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 1 & 1 & \frac{1}{3} & \frac{1}{3} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{2} & \frac{1}{2} \end{bmatrix},$$

$$V^{P(x_3)} = \begin{bmatrix} 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 1 & 1 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix},$$

Example (Contd)

State-depending utility design:

Construct

$$\begin{bmatrix} \Gamma_{U_r(i)} \\ E_i^T \end{bmatrix}, \ r = x_1, x_2, x_3.$$

where

$$\begin{aligned}
&\Gamma_{U_1(1)} = \Gamma_{U_2(2)} = I_2 \otimes \mathbf{1}_4^T \otimes I_2, \\
&\Gamma_{U_1(2)} = \mathbf{1}_2^T \otimes I_4 \otimes \mathbf{1}_4^T, \ \Gamma_{U_1(3)} = \Gamma_{U_3(3)} = \mathbf{1}_2^T \otimes I_8, \\
&\Gamma_{U_1(4)} = \Gamma_{U_2(1)} = \Gamma_{U_2(4)} = \Gamma_{U_3(4)} = I_2 \otimes \mathbf{1}_2^T \otimes I_4, \\
&\Gamma_{U_2(3)} = I_{16}, \ \Gamma_{U_3(1)} = I_4 \otimes \mathbf{1}_2^T \otimes I_2, \ \Gamma_{U_3(2)} = I_8 \otimes \mathbf{1}_2^T.
\end{aligned}$$

It is easy to verify that

$$V^{\phi(x_i=r,\cdot)} \in \bigcap_{i=1}^{4} \text{Span} \begin{bmatrix} \Gamma_{U_r(i)} \\ E_i^T \end{bmatrix}, \ i = 1, 2, 3.$$

Local-information based utilities:

$$c_i(x, a) = 2 \cdot \mathbf{1}_{a_i=1} + \sum_{j \in U_r(i)} \mathbf{1}_{a_i=a_j}, \quad i = 1, 2, 3, 4.$$

Example (Contd)

- State evolutionary dynamics:

$$x(t+1) = M_P x(t) a(t),$$

where

$$M_P = [V^{P(x_1)},\ V^{P(x_2)},\ V^{P(x_3)}].$$

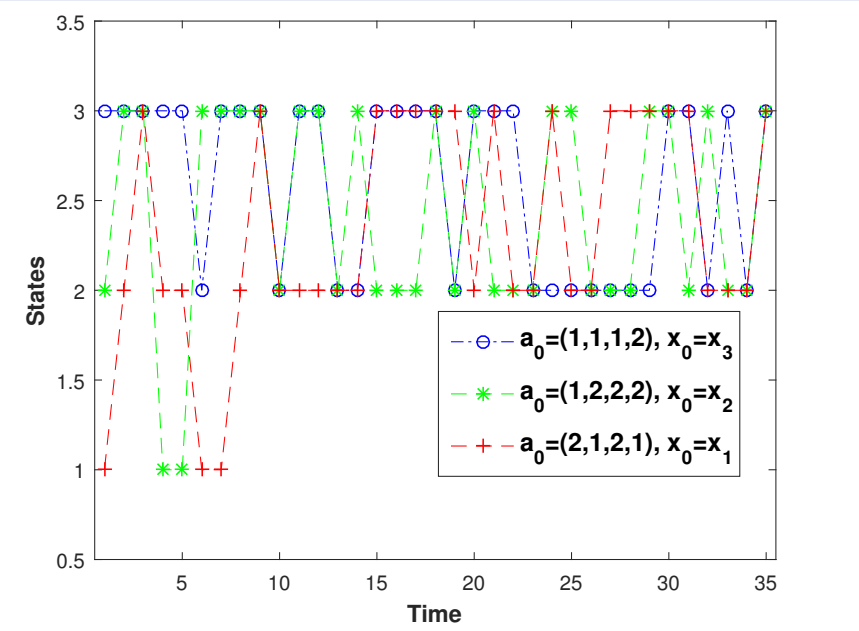
- Profile dynamics under SBR with inertia $\epsilon = 0.1$

$$a(t+1) = M_F x(t+1) a(t),$$

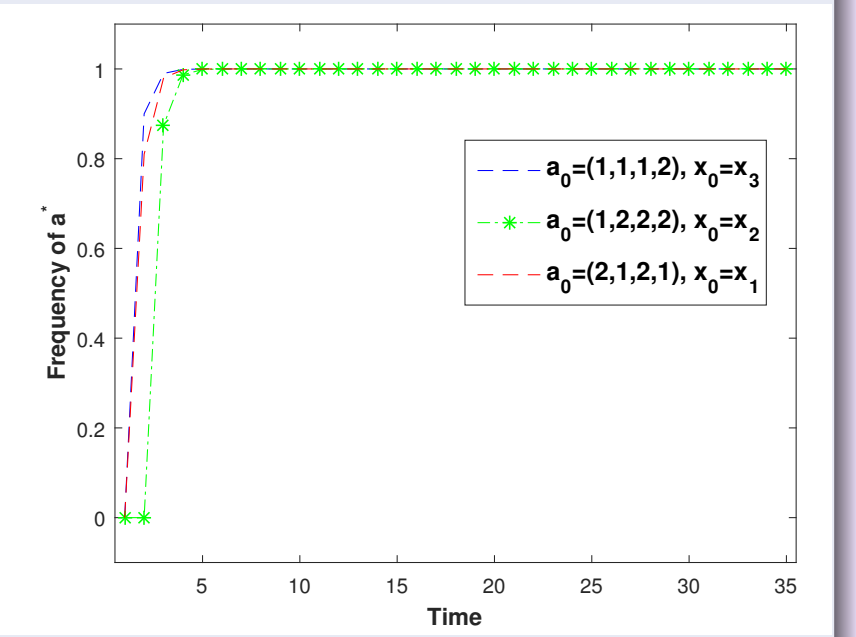
where

$$M_F = \begin{bmatrix} 1 & 0.9 & 0.9 & 0.81 & \cdots & 0 & 0 \\ 0 & 0.1 & 0 & 0.09 & \cdots & 0 & 0 \\ 0 & 0 & 0.1 & 0.09 & \cdots & 0 & 0 \\ \vdots & & & & & & \\ 0 & 0 & 0 & 0 & \cdots & 0.09 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0.01 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \in \mathcal{M}_{16\times 48},$$

Example (Contd)



(a) 状态的动态



(b) 策略 $a^* = (1, 1, 1, 1)$ 的频率曲线

Figure 17: Dynamics of state and profile

6. Future Directions

- Bring the connection between learning and STP.
- Design learning rule for mixed NE, correlated equilibrium...
- Combine utility design (using STP) with learning design...
- Distributed learning rule design (for networked evolutionary game)...
- Design learning rule for incomplete games...


Daizhan Cheng, Changxi Li, "Matrix expression of Bayesian games," submitted to *Applied Mathematics and Computation*, under review, arxiv:2106.12161, 2021.


Balcan, MF, Procaccia, AD and Zick, Y, "Learning Cooperative Games," *Proceedings Of The Twenty-fourth International Joint Conference On Artificial Intelligence*, 2015.

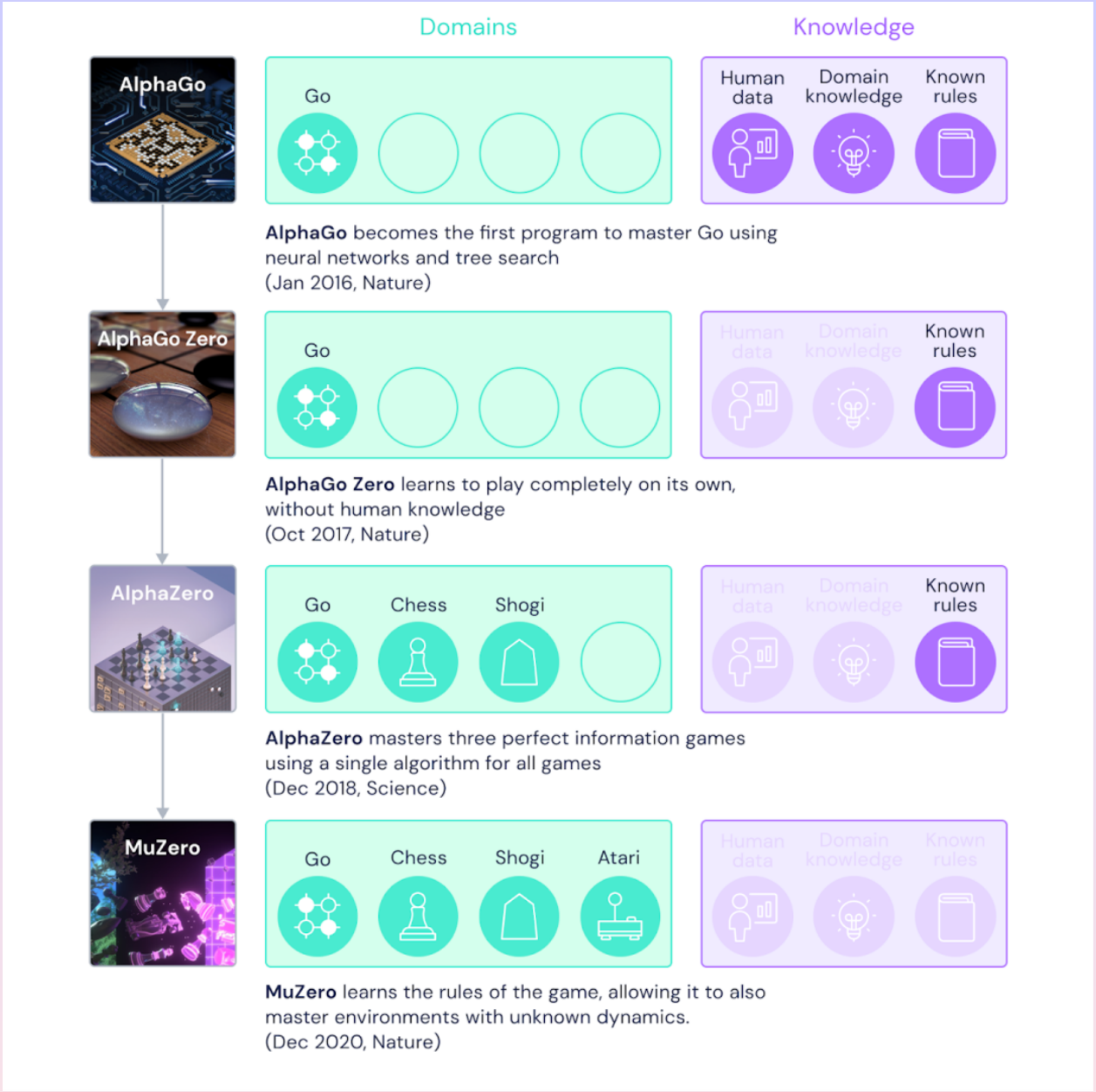


Figure 18: From complete information game to incomplete information game in AI

Thanks for your attention!

Q & A