

2 Preliminary

2.1 Two-Player Normal-Form Game

Let us define two-player (of X and Y) $m(\in \mathbb{N})$ -action games (see illustration of Fig. 1-A). Player X and Y choose their actions from $\mathcal{A} = \{a_1, \dots, a_m\}$ and $\mathcal{B} = \{b_1, \dots, b_m\}$ in a single round. After they finish choosing their actions $a \in \mathcal{A}$ and $b \in \mathcal{B}$, each of them gains a payoff $U(a, b) \in \mathbb{R}$ and $V(a, b) \in \mathbb{R}$, respectively.

2.2 Two-Player Multi-Memory Repeated Game

We further consider two-player $n(\in \mathbb{N})$ -memory repeated games as an iteration of the two-player normal-form game (see illustration Fig. 1-A). The players are assumed to memorize their actions in the last n rounds. Since each player can take m actions, there are m^{2n} cases for possible memorized states, described as $\mathcal{S} = \prod_{k=1}^n (\mathcal{A} \times \mathcal{B})$. Under any memorized state, player X can choose any action stochastically. Such a stochastic choice of an action is described by a parameter $x^{a|s}$, which means the probability of choosing an action $a \in \mathcal{A}$ under memorized state $s \in \mathcal{S}$. Thus, X's strategy is represented by $|\mathcal{S}| (= m^{2n})$ -numbers of $(m-1)$ -dimension simplexes, $\mathbf{x} \in \prod_{s \in \mathcal{S}} \Delta^{m-1}$, while Y's is $\mathbf{y} \in \prod_{s \in \mathcal{S}} \Delta^{m-1}$.

2.3 Formulation as Markov Games

In order to handle this multi-memory repeated game as a Markov game [28, 29], we define a vector notation of memorized states;

$$\mathbf{s} = (\underbrace{a_1 b_1 \cdots a_1 b_1}_{\times n}, \underbrace{a_1 b_1 \cdots a_1 b_1}_{\times (n-1)} a_1 b_2, \dots, \underbrace{a_m b_m \cdots a_m b_m}_{\times n}), \quad (1)$$

which orders all the elements of \mathcal{S} as a vector. We also define a vector notation of utility function as

$$\mathbf{u} = (\underbrace{U(a_1, b_1), \dots, U(a_1, b_1)}_{\times m^{2n-2}}, \underbrace{U(a_1, b_2), \dots, U(a_1, b_2)}_{\times m^{2n-2}}, \dots, \underbrace{U(a_m, b_m), \dots, U(a_m, b_m)}_{\times m^{2n-2}}), \quad (2)$$

which orders all the last-round payoffs for \mathcal{S} as a vector. The utility function for Y, i.e., \mathbf{v} , is defined similarly. In addition, we denote an index for these vectors as $i \in \{1, \dots, m^{2n}\}$. For example, when player X is in state s_i , the last-round payoff for player X was u_i .

Let $\mathbf{p} \in \Delta^{|\mathcal{S}|-1}$ be a probability distribution on \mathbf{s} in a round. As the name Markov matrix implies, a distribution in the next round \mathbf{p}' is given by $\mathbf{p}' = \mathbf{M}\mathbf{p}$, where \mathbf{M} is a Markov transition matrix;

$$M_{i'i} = \begin{cases} x^{a|s_i} y^{b|s_i} & (s_{i'} = a b s_i^-) \\ 0 & (\text{otherwise}) \end{cases}, \quad (3)$$

which shows the transition probability from i -th state to i' -th one for $i, i' \in \{1, \dots, m^{2n}\}$. Here, note that s_i^- shows the state s_i except for the oldest two actions. See Fig. 1-B illustrating an example of Markov transition.

2.4 Nash Equilibrium

We now analyze the Nash equilibrium in multi-memory repeated games based on the formulation of Markov games. Let us assume that every agent uses a fixed strategy \mathbf{x} and \mathbf{y} or learns slowly enough for the timescale of the Markov transitions. If we further assume that the strategies are within the interiors of simplexes (i.e., the Markov matrix is ergodic), this stationary distribution is unique and described as $\mathbf{p}^{\text{st}}(\mathbf{x}, \mathbf{y})$. This stationary distribution satisfies $\mathbf{p}^{\text{st}} = \mathbf{M}\mathbf{p}^{\text{st}}$. We also denote each player's expected payoff in the stationary distribution as $u^{\text{st}}(\mathbf{x}, \mathbf{y}) = \mathbf{p}^{\text{st}} \cdot \mathbf{u}$ and $v^{\text{st}}(\mathbf{x}, \mathbf{y}) = \mathbf{p}^{\text{st}} \cdot \mathbf{v}$. The goal of learning in the multi-memory game is to search for the Nash equilibrium, denoted by $(\mathbf{x}^*, \mathbf{y}^*)$, where their payoffs are maximized as

$$\begin{cases} \mathbf{x}^* \in \operatorname{argmax}_{\mathbf{x}} u^{\text{st}}(\mathbf{x}, \mathbf{y}^*) \\ \mathbf{y}^* \in \operatorname{argmax}_{\mathbf{y}} v^{\text{st}}(\mathbf{x}^*, \mathbf{y}) \end{cases}. \quad (4)$$

Here, u^{st} and v^{st} are complex non-linear functions for high-dimensional variables of (\mathbf{x}, \mathbf{y}) . This Nash equilibrium is difficult to find in general.