Minimally Modifying a Markov Game to Achieve Any Nash Equilibrium and Value

Young Wu, Jeremy McMahan, Yiding Chen, Yudong Chen, Xiaojin Zhu, Qiaomin Xie, with Joy Cheng

University of Wisconsin - Madison

Markov Game

- A finite-horizon two-player zero-sum Markov game $G^\circ = (R^\circ, P^\circ)$ has:

1. $\mathcal{S}$ is the finite state space,
2. $\mathcal{A}_i$ the finite set of actions for player $i \in \{1, 2\}$,
3. P° is the transition probability matrices,
4. R° is the payoff matrices,
5. H is the horizon,

The Game Modification Problem

- Game modification is the following optimization problem to find R given $(R^\circ, P^\circ, b, (\mathbf{p}, \mathbf{q}), [\underline{v}, \overline{v}], \ell)$:

$$\begin{aligned} \inf_{R} \quad & \ell(R, R^\circ) \\ \text{s.t.} \quad & (\mathbf{p}, \mathbf{q}) \text{ is the unique MPE of } (R, P^\circ) \\ & \text{value}(R, P^\circ) \in [\underline{v}, \overline{v}], \quad R \text{ has entries in } [-b, b]. \end{aligned} \tag{1}$$

- It is important to require that the modified game (R, P°) has a **unique** Markov Perfect (Nash) Equilibrium (MPE).
- The Game Modification problem (1) for Markov games is feasible if and only if $|\mathcal{I}_h(s)| = |\mathcal{J}_h(s)|$ for every $h \in [H], s \in \mathcal{S}$, and $(-Hb, Hb) \cap [\underline{v}, \overline{v}] \neq \emptyset$.
  - Here, $\mathcal{I} = supp(\mathbf{p})$ and $\mathcal{J} = supp(\mathbf{q})$ denote the supports (the set of actions used with non-zero probabilities) of the MPE.

Equivalent Formulation

- We consider the following optimization problem:

$$\begin{aligned} \min_{R, v, \mathbb{Q}} \quad & \ell(R, R^\circ) \\ \text{s.t.} \quad & [\mathbb{Q}_h(s)]_{\mathcal{I}_h(s)\bullet} \mathbf{q}_h(s) = v_h(s) \mathbf{1}_{|\mathcal{I}_h(s)|} \\ & \quad \forall\, h \in [H], s \in \mathcal{S} \qquad \text{[row SII]} \\ & \mathbf{p}_h^\top(s) [\mathbb{Q}_h(s)]_{\bullet\mathcal{J}_h(s)} = v_h(s) \mathbf{1}^\top_{|\mathcal{J}_h(s)|} \\ & \quad \forall\, h \in [H], s \in \mathcal{S} \qquad \text{[column SII]} \\ & [\mathbb{Q}_h(s)]_{\mathcal{A}_1 \setminus \mathcal{I}_h(s)\bullet} \mathbf{q}_h(s) \leqslant (v_h(s) - \iota) \mathbf{1}_{|\mathcal{A}_1 \setminus \mathcal{I}_h(s)|} \\ & \quad \forall\, h \in [H], s \in \mathcal{S} \qquad \text{[row SOW]} \\ & \mathbf{p}_h^\top(s) [\mathbb{Q}_h(s)]_{\bullet\mathcal{A}_2 \setminus \mathcal{J}_h(s)} \geqslant (v_h(s) + \iota) \mathbf{1}^\top_{|\mathcal{A}_2 \setminus \mathcal{J}_h(s)|} \\ & \quad \forall\, h \in [H], s \in \mathcal{S} \qquad \text{[column SOW]} \\ & \mathbb{Q}_h(s) = R_h(s) + \sum_{s' \in \mathcal{S}} P_h(s'|s) v_{h+1}(s') \\ & \quad \forall\, h \in [H-1], s \in \mathcal{S} \qquad \text{[Bellman]} \\ & \mathbb{Q}_H(s) = R_H(s), \forall\, s \in \mathcal{S} \\ & \underline{v} \leqslant \sum_{s \in \mathcal{S}} P_0(s) v_1(s) \leqslant \overline{v} \qquad \text{[value range]} \\ & -b + \lambda \leqslant [R_h(s)]_{ij} \leqslant b - \lambda \\ & \quad \forall\, (i, j) \in \mathcal{A}, h \in [H], s \in \mathcal{S} \quad \text{[reward bound]} \end{aligned} \tag{2}$$

  - $[R]_{\mathcal{I}\mathcal{J}}$ or $R_{\mathcal{I}\mathcal{J}}$ denotes the $|\mathcal{I}| \times |\mathcal{J}|$ submatrix of R with rows in $\mathcal{I}$ and columns in $\mathcal{J}$. We write $R_{\mathcal{I}\bullet}$ for the $|\mathcal{I}| \times |\mathcal{A}_2|$ submatrix with rows in $\mathcal{I}$, and $R_{\bullet\mathcal{J}}$ for the $|\mathcal{A}_1| \times |\mathcal{J}|$ submatrix with columns in $\mathcal{J}$; and $\mathbf{1}_{|\mathcal{I}|}$ denotes the $|\mathcal{I}|$-dimensional all-one vector.

Relax And Perturb Algorithm

- **Input**: original game (R°, P), cost function ℓ, target policy $(\mathbf{p}, \mathbf{q})$ and value range $[\underline{v}, \overline{v}]$, reward bound $b \in \mathbb{R}^+ \cup \{\infty\}$.
- **Parameters**: margins $\iota \in \mathbb{R}^+$ and $\lambda \in \mathbb{R}^+$.
- **Output**: modified game (R, P).

1. Solve the problem (2). Call the solution R'.
2. For $h \in [H], s \in \mathcal{S}$ Sample $\varepsilon \sim \text{uniform}[-\lambda, \lambda]$
3. Perturb the reward matrix in stage (h, s): $R_h(s) = R'_h(s) + \varepsilon R^{\text{eRPS}\,(\mathbf{p}_h(s), \mathbf{q}_h(s))}$, where R^{eRPS} is the reward matrix for the extended Rock-Paper-Scissor game, which has $((\mathbf{p}_h(s), \mathbf{q}_h(s)))$ as its unique NE.
4. Return (R, P).

Existence, Feasibility, and Optimality

Let $R(\iota, \lambda) = R' + \varepsilon R^{\text{eRPS}}$ denote the output of the RAP Algorithm with margin parameters ι, λ. If

$$(-b + \lambda + \iota, b - \lambda - \iota) \cap [-\underline{v}/H, \overline{v}/H] \neq \emptyset, \tag{3}$$

then the following hold.

1. (**Existence**) The solution R' to the program (2) exists.
2. (**Feasibility**) $R(\iota, \lambda)$ is feasible for the game modification problem in (1) with probability 1.
3. (**Optimality**) If in addition the cost function ℓ is L-Lipschitz, then $R(\iota, \lambda)$ is asymptotically optimal:
$$\lim_{\max\{\iota, \lambda\} \to 0} \ell(R(\iota, \lambda), R^\circ) = C^\star,$$
4. (**Optimality Gap**) If ℓ is piecewise linear, then
$$\ell(R(\iota, \lambda), R^\circ) = C^\star + O(\max\{\iota, \lambda\}),$$

Extended Rock-Paper-Scissors Game

- We present a special matrix game called Extended Rock-Paper-Scissors (eRPS), which has the desired $(\mathbf{p}, \mathbf{q})$ as the unique NE. This game can be defined for arbitrary strategy space sizes $|\mathcal{A}_1|$ and $|\mathcal{A}_2|$. The standard rock paper scissors game is a special case when the sizes are 3, hence the name.

$\mathcal{A}_1 \backslash \mathcal{A}_2$	0	1	2	3	...	$k-2$	$k-1$	k	...	$\lvert\mathcal{A}_2\rvert - 1$
0	0	$-\frac{c}{\mathbf{p}_0\mathbf{q}_1}$	$\frac{c}{\mathbf{p}_0\mathbf{q}_2}$	0	...	0	0	1	...	1
1	0	0	$-\frac{c}{\mathbf{p}_1\mathbf{q}_2}$	$\frac{c}{\mathbf{p}_1\mathbf{q}_3}$	...	0	0	1	...	1
2	0	0	0	$-\frac{c}{\mathbf{p}_2\mathbf{q}_3}$	...	0	0	1	...	1
3	0	0	0	0	...	0	0	1	...	1
...	...	...	...	...	...	...	...	...	...	...
$k-2$	$\frac{c}{\mathbf{p}_{k-2}\mathbf{q}_0}$	0	0	0	...	0	$-\frac{c}{\mathbf{p}_{k-2}\mathbf{q}_{k-1}}$	1	...	1
$k-1$	$-\frac{c}{\mathbf{p}_{k-1}\mathbf{q}_0}$	$\frac{c}{\mathbf{p}_{k-1}\mathbf{q}_1}$	0	0	...	0	0	1	...	1
k	-1	-1	-1	-1	...	-1	-1	0	...	0
...	...	...	...	...	...	...	...	...	...	...
$\lvert\mathcal{A}_1\rvert - 1$	-1	-1	-1	-1	...	-1	-1	0	...	0

Experiments

1. Given left below is the payoff matrix for the **simplified Two-finger Morra** game, which has a unique NE $(\mathbf{p}, \mathbf{q}) = (\frac{7}{12}, \frac{5}{12})$ and value $-\frac{1}{12}$. On the right, we minimally modify the game to keep the same unique NE but make the game fair with a value of 0.

$$\text{Original: } \begin{pmatrix} 2 & -3 \\ -3 & 4 \end{pmatrix} \qquad \text{Modified: } \begin{pmatrix} 2.04 & -2.86 \\ -2.86 & 4 \end{pmatrix}$$

2. The **Rock-Paper-Scissors-Fire-Water** game, given on the left below, is a generalization of the Rock-Paper-Scissor game to five actions. The unique NE is $\mathbf{p} = \mathbf{q} = (\frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{3}, \frac{1}{3})$ and has value 0. We desire the NE to be simpler for humans, so we redesign the game to have a uniformly mixed NE $\mathbf{p} = \mathbf{q} = (\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5})$. The resultant game is given below.

$$\overset{\text{Original}}{\begin{pmatrix} 0 & -1 & 1 & -1 & 1 \\ 1 & 0 & -1 & -1 & 1 \\ -1 & 1 & 0 & -1 & 1 \\ 1 & 1 & 1 & 0 & -1 \\ -1 & -1 & -1 & 1 & 0 \end{pmatrix}} \overset{\text{Modified}}{\begin{pmatrix} 0 & -1 & 1 & -1 & 1 \\ 1 & 0 & -1 & -1 & 1 \\ -1 & 1 & 0 & -1 & 1 \\ 1 & 1 & 1 & 0 & -3 \\ -1 & -1 & -1 & 3 & 0 \end{pmatrix}}$$

Summary

- We study the game modification problem, where a benevolent game designer or a malevolent adversary modifies the reward function of a zero-sum Markov game so that a target deterministic or stochastic policy profile becomes the unique Markov perfect Nash equilibrium and has a value within a target range, in a way that minimizes the modification cost. We characterize the set of policy profiles that can be installed as the unique equilibrium of a game and establish sufficient and necessary conditions for successful installation. We propose an efficient algorithm that solves a convex optimization problem with linear constraints and then performs random perturbation to obtain a modification plan with a near-optimal cost.