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

Young Wu [1] **Jeremy McMahan** [1] **Yiding Chen** [2] **Yudong Chen** [1] **Xiaojin Zhu** [1] **Qiaomin Xie** [3]

Abstract

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.

1. Introduction

Consider a two-player zero-sum Markov game $G^\circ = (R^\circ, P^\circ)$ with payoff matrices R° and transition probability matrices P°. Let $\mathcal{S}$ be the finite state space, $\mathcal{A}_i$ the finite set of actions for player $i \in \{1, 2\}$, and H is the horizon. It is well known that such a game has at least one Markov Perfect (Nash) Equilibrium (MPE) $(\mathbf{p}^\circ, \mathbf{q}^\circ)$, where $\mathbf{p}^\circ$ is the Markov policy for player 1 and $\mathbf{q}^\circ$ for player 2 (Maskin & Tirole, 2001). Furthermore, all the MPEs of G° have the same game value v°, corresponding to the expected payoff for player 1 and loss for player 2 at equilibrium. In the special case with $H = 1$ stage, the Markov game reduces to a matrix normal form game and the MPE reduces to a Nash Equilibrium (NE).

There may be reasons for a third party to prefer an outcome with a different MPE and/or game value. For instance, a **benevolent** third party may want to achieve fairness. Many games are unfair in that $v^\circ \neq 0$ (an example, two-finger Morra, is given in the experiment section). The third party can modify the payoffs R° into R such that the new game given to the players is fair with value $v = 0$. Similarly, many games have non-intuitive MPEs, and players with bounded rationality (e.g., average people) may fail to find or implement them. For the benefit of such players, the third party may seek a new game whose MPE $(\mathbf{p}, \mathbf{q})$ is an intuitive strategy profile, such as uniform randomization among a set of actions.

In addition, one often desires an MPE consisting of stochastic policies (i.e., a *mixed* strategy equilibrium). If actions represent resources (roads, advertisement slots, etc.), the game designer might want all resources to be utilized; if actions represent customers, requests, or demands, the designer might want all of them to be served; if a board/video game is concerned, the designer might want the agents to take diverse actions so that the game is more entertaining. Conversely, a **malicious** third party may want to trick the players into playing an MPE $(\mathbf{p}, \mathbf{q})$ of its choice. As most games have mixed equilibria, the players may get suspicious if the modified game turns out to have a pure strategy MPE, whereas a mixed equilibrium is harder to detect. Furthermore, the adversary may want to control the game value v to favor one player over the other—this is the analog of adversarial attacks in supervised learning.

Regardless of intention, game modification typically incurs a cost to the third party, who seeks to minimize it. We assume that the cost is measured by some loss function $\ell(R, R^\circ)$ depending on the new and original games R and R°. For example, one may consider $\ell(R, R^\circ) = \|R - R^\circ\|$ for some norm $\|\cdot\|$.

The Game Modification Problem It is important to understand when efficient modification is possible and to understand malicious attacks so as to develop an effective defense. This motivates us to study the following *Game Modification* problem, specified by the tuple

$$(R^\circ, P^\circ, b, (\mathbf{p}, \mathbf{q}), [\underline{v}, \overline{v}], \ell).$$

[1]Department of Computer Sciences, University of Wisconsin–Madison, Madison, Wisconsin, United States [2]Department of Computer Science, Cornell University, Ithaca, New York, United States [3]Department of Industrial and Systems Engineering, University of Wisconsin–Madison, Madison, Wisconsin, United States. Correspondence to: Young Wu <yw@cs.wisc.edu>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024.

Here R° and P° are the payoff and transition matrices, respectively, of the original Markov game. A valid payoff value must be in $[-b, b]$. The third party has in mind an arbitrary (and potentially stochastic) target MPE $(\mathbf{p}, \mathbf{q})$, which is typically not the unique MPE of R°. The third party also has in mind a target game value range $[\underline{v}, \overline{v}]$. It is possible that $b = \infty$, $\underline{v} = -\infty$ or $\overline{v} = \infty$.

Definition 1 (Game Modification). Game modification is the following optimization problem to find R given $(R^\circ, P^\circ, b, (\mathbf{p}, \mathbf{q}), [\underline{v}, \overline{v}], \ell)$:

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

It is important to require that the modified game (R, P°) has a **unique** MPE. In this case, no matter what solver the players use, they will inevitably find $(\mathbf{p}, \mathbf{q})$ and not some other MPEs of R. Henceforth, we refer to a Markov game simply by its payoff matrices R and suppress reference to the transition matrices P°, which the third party cannot change.

To the best of our knowledge, the Game Modification problem in the generality of Definition 1 has not been studied in the literature. With a potentially mixed target MPE $(\mathbf{p}, \mathbf{q})$ and the constraints on uniqueness, game value, and payoffs, it is a priori unclear when the optimization problem (1) has a feasible solution. Moreover, in addition to just finding one feasible game or checking the feasibility of a specific game, we need to solve the harder problem of *optimizing* over all feasible games with a target strategy as the unique NE. The multi-step structure of Markov games further complicates the problem.

Our Contributions In this paper, we answer the above questions: we provide a sufficient and necessary condition for the feasibility of the game modification problem and develop an efficient algorithm that provably finds a near-optimal solution under convex losses ℓ. In particular, using an operational characterization of MPE uniqueness, we formulate the game modification problem as an optimization problem with linear and spectral constraints and completely characterize its feasibility. We further propose an efficient Relax and Perturb algorithm circumventing the spectral constraint's nonconvexity and establish the algorithm's correctness and near-optimality.

We first study the special case of normal form games in Section 3, followed by a generalization to Markov games in Section 4.

2. Related Work

Reward modification in single-agent reinforcement learning has been studied in Banihashem et al. (2022); Huang & Zhu (2019); Rakhsha et al. (2021a;b; 2020); Zhang et al. (2020). In this setting, a deterministic optimal policy always exists. Generalizing to the multi-agent setting, even in the zero-sum case, involves the complication of multiple equilibria and the non-existence of deterministic equilibrium policies.

Adversarial attacks on multi-agent reinforcement learners are studied in Wu et al. (2023b); Ma et al. (2021), who consider the setting where an attacker installs a target *dominant strategy equilibrium* by modifying the underlying bandit or Markov game. In general, mixed strategies that assign positive probabilities to multiple actions cannot be dominant (they are not dominated by at least one of the actions in the support). Therefore, the approach in Wu et al. (2023b); Ma et al. (2021) cannot be directly applied in our setting, which targets at a mixed strategy Nash equilibrium.

Our model is similar to Wu et al. (2023a), where an attacker installs a target Nash equilibrium by poisoning the training data set. Their work requires the target equilibrium to be a *deterministic* action profile (i.e., not mixed), and they assume the victims estimate confidence regions of the game payoff matrices based on a noisy data set. Since it is generally impossible for all games in the confidence region to have the same mixed strategy Nash equilibrium, the modification goal in our setting is infeasible under their setting. Similarly, data poisoning techniques in Ma et al. (2019); Rangi et al. (2022); Zhang & Parkes (2008); Zhang et al. (2009) do not apply to our setting. Instead, we consider the problem in which the players are provided with the exact payoff matrix by the game designer, so it is possible to install a mixed strategy as the unique equilibrium of the modified game. Monderer & Tennenholtz (2003); Anderson et al. (2010) explore the problem of installing a pure strategy equilibrium while minimizing the modification cost, but their method does not directly extend to mixed-strategy equilibria.

As our work concerns *optimizing* over the set of games with a target strategy as the unique NE, we need a *sufficient and necessary* condition for uniqueness. Related and partial results can be found in a line of prior work on matrix games with a unique Nash equilibrium (Kreps, 1974; Millham, 1972; Heuer, 1979; Quintas, 1988; Bohnenblust et al., 1950) and the related problem of unique optimal solutions to linear programs (Mangasarian, 1978; Appa, 2002; Szilágyi, 2006). Our work gives a form of sufficient and necessary condition that is amenable to being used as constraints in a cost minimization formulation, and we provide a short proof. We also go beyond prior work by studying when this condition is satisfiable under the additional value and payoff constraints in (1), with generalization to Markov games.

3. Modifying Normal Form Games

We begin with the game modification problem for matrix normal form games, which is a special case of Markov Games with horizon $H = 1$.

3.1. Preliminaries

Consider a finite two-player zero-sum game with action space $\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2$ and a b-bounded payoff matrix $R \in [-b, b]^{|\mathcal{A}_1| \times |\mathcal{A}_2|}$. When a joint action $(i, j) \in \mathcal{A}_1 \times \mathcal{A}_2$ is played, player 1 receives reward $[R]_{ij}$ and player 2 receives reward $-[R]_{ij}$. Let $(\mathbf{p}, \mathbf{q})$ denote a (possibly mixed) strategy profile, where $\mathbf{p} \in \Delta_{\mathcal{A}_1}$ and $\mathbf{q} \in \Delta_{\mathcal{A}_2}$, with $\Delta_{\mathcal{D}}$ denoting the probability simplex on $\mathcal{D}$. The expected reward for player 1 is given by $\mathbf{p}^\top R \mathbf{q}$.

NE can be defined in several equivalent ways. Most convenient for us is the following definition in terms of lack of incentive for unilateral deviation. A finite two-player game has at least one NE and possibly more (Nash Jr, 1950).

Definition 2 (Nash Equilibrium). $(\mathbf{p}, \mathbf{q})$ is a Nash Equilibrium of a game R if and only if $\mathbf{p}^\top R \mathbf{q} \geqslant \mathbf{p}'^\top R \mathbf{q}$ for all $\mathbf{p}' \in \Delta_{\mathcal{A}_1}$ and $\mathbf{p}^\top R \mathbf{q} \leqslant \mathbf{p}^\top R \mathbf{q}'$ for all $\mathbf{q}' \in \Delta_{\mathcal{A}_2}$.

3.2. Equivalent Formulation of Game Modification

As stated in Definition 1, the game designer seeks a least-cost game with a given $(\mathbf{p}, \mathbf{q})$ as the unique NE and satisfying the value and payoff constraints in (1). To understand when such a game exists and how to find the optimal game algorithmically, our first step is to provide an equivalent formulation where the uniqueness requirement is expressed explicitly as linear and spectral constraints.

This is done in the theorem below, for which some notations are needed. Let $\mathcal{I} = supp(\mathbf{p})$ and $\mathcal{J} = supp(\mathbf{q})$ denote the supports. We use $[R]_{\mathcal{I}\mathcal{J}}$ or $R_{\mathcal{I}\mathcal{J}}$ to denote 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 $\mathcal{J}$. Denotes by $\mathbf{1}_{|\mathcal{I}|}$ the $|\mathcal{I}|$-dimensional all-one vector.

Proposition 1 (Reformulation of Normal-Form Game Modification). For normal form games and a target policy $(\mathbf{p}, \mathbf{q})$ with supports $\mathcal{I}, \mathcal{J}$, the game modification problem (1) is equivalent to the following optimization problem:

$$\inf_{R, v} \quad \ell(R, R^\circ) \tag{2a}$$

$$\text{s.t. } R_{\mathcal{I}\bullet}\mathbf{q} = v\mathbf{1}_{|\mathcal{I}|} \quad \text{[row SII]} \tag{2b}$$

$$\mathbf{p}^\top R_{\bullet\mathcal{J}} = v\mathbf{1}_{|\mathcal{J}|}^\top \quad \text{[column SII]} \tag{2c}$$

$$R_{\mathcal{A}_1\setminus\mathcal{I}\bullet}\mathbf{q} < v\mathbf{1}_{|\mathcal{A}_1\setminus\mathcal{I}|} \quad \text{[row SOW]} \tag{2d}$$

$$\mathbf{p}^\top R_{\bullet\mathcal{A}_2\setminus\mathcal{J}} > v\mathbf{1}_{|\mathcal{A}_2\setminus\mathcal{J}|}^\top \quad \text{[column SOW]} \tag{2e}$$

$$\sigma_{\min}\left(\begin{bmatrix} R_{\mathcal{I}\mathcal{J}} & -\mathbf{1}_{|\mathcal{I}|} \\ \mathbf{1}_{|\mathcal{J}|}^\top & 0 \end{bmatrix}\right) > 0 \quad \text{[INV]} \tag{2f}$$

$$\underline{v} \leqslant v \leqslant \overline{v} \quad \text{[value range]} \tag{2g}$$

$$-b \leqslant R_{ij} \leqslant b, \forall (i, j) \in \mathcal{A} \quad \text{[payoff bound]} \tag{2h}$$

where $\sigma_{\min}(\cdot)$ denotes the smallest singular value.

Proposition 1 follows immediately from the lemma below, which shows that SIISOW and INV constitute a sufficient and necessary condition for a game R to admit a given $(\mathbf{p}, \mathbf{q})$ as the unique NE.

Lemma 2 (Uniqueness of NE). R has a unique Nash equilibrium $(\mathbf{p}, \mathbf{q})$ if and only if R satisfies both SIISOW (Condition 1) and INV (Condition 2) with respect to $(\mathbf{p}, \mathbf{q})$:

Condition 1 (SIISOW: Switch-In Indifferent, Switch-Out Worse). A game R satisfies SIISOW with respect to $(\mathbf{p}, \mathbf{q})$ if equations (2b), (2c), (2d) and (2e) hold.

Condition 2 (INV: Invertability). A game R satisfies INV with respect to $(\mathbf{p}, \mathbf{q})$ if equation (2f) holds, that is, the matrix $\begin{bmatrix} R_{\mathcal{I}\mathcal{J}} & -\mathbf{1}_{|\mathcal{I}|} \\ \mathbf{1}_{|\mathcal{J}|}^\top & 0 \end{bmatrix}$ is invertible.

If the strict inequalities in SIISOW were changed to weak inequalities, the four equations would be equivalent to Definition 2 of NE (Osborne, 2004). Therefore, SIISOW implies that $(\mathbf{p}, \mathbf{q})$ is an NE of R. Moreover, under this NE, if the other player switches to any pure strategy outside its NE support, its reward will be *strictly* worse by equations (2e) and (2d) ("switch-out worse"); if the other player uses any pure strategy within its support, it will achieve the same game value by equations (2c) and (2b) (known as the "switch-in indifference" principle).

We are not aware of an NE uniqueness result stated in this form in the literature, though several partial and related results exist. Lemma C.3 in Mertikopoulos et al. (2018) implies that the SIISOW condition is necessary for NE uniqueness. Several papers study the existence of (or explicitly construct) a game with a unique NE (Kreps, 1974; Millham, 1972; Bohnenblust et al., 1950; Nagarajan et al., 2020); our Lemma 2 characterizes *all* such games and thereby allow one to optimize over them, as done in the formulation (2). In Appendix A.1, we provide a short, self-contained proof for Lemma 2, noting that with some additional work one may also derive the lemma from the results in Szilágyi (2006) on unique solutions to linear program (LP). We remark that Appa (2002); Mangasarian (1978) also provide uniqueness results for LP, but they are in terms of perturbation stability of the solution and hence not in an operational form that can be used as constraints in a cost-minimization problem like (2).

3.3. Feasibility of Game Modification

We now study when Game Modification in normal-form games, as formulated in Proposition 1, is feasible. The following theorem provides a *sufficient and necessary* condition.

Theorem 3 (Feasibility of Game Modification). The Game Modification problem in Proposition 1 for normal-form games is feasible if and only if $(\mathbf{p}, \mathbf{q})$ satisfies $|\mathcal{I}| = |\mathcal{J}|$ and it holds that $(-b, b) \cap [\underline{v}, \overline{v}] \neq \varnothing$.

The equal-support condition $|\mathcal{I}| = |\mathcal{J}|$ arises due to the INV condition, which requires $R_{\mathcal{I}\mathcal{J}}$ to be a square matrix. The necessity of the equal-support condition is known in Kreps (1974); Millham (1972); Heuer (1979). Our lemma further establishes the necessity of the condition $(-b, b) \cap [\underline{v}, \overline{v}] \neq \varnothing$. Note that the game value cannot equal b or $-b$, because the SIISOW condition stipulates a strictly positive gap between the game value and the value of the off-support actions. The complete proof is provided in the Appendix A.4.

The other direction of our proof is constructive. 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.

Definition 3 (Extended Rock-Paper-Scissors Game). Given strategy spaces $\mathcal{A}_1, \mathcal{A}_2$, and target strategy profile $(\mathbf{p}, \mathbf{q}) \in \Delta_{\mathcal{A}_1} \times \Delta_{\mathcal{A}_2}$ with equal supports $\mathcal{I} = \mathcal{J} = \{0, \dots, k-1\}$, where $1 \leqslant k \leqslant \min(|\mathcal{A}_1|, |\mathcal{A}_2|)$, the Extended Rock Paper Scissors Game $R^{\text{eRPS}\,(\mathbf{p},\mathbf{q})}$ is:

$$R_{ij}^{\text{eRPS}\,(\mathbf{p},\mathbf{q})} = \begin{cases} -\dfrac{c}{\mathbf{p}_i \mathbf{q}_j} & \text{if } \underset{j=(i+1) \mod k}{k>1, i, j<k} \\ \dfrac{c}{\mathbf{p}_i \mathbf{q}_j} & \text{if } \underset{j=(i+2) \mod k}{k>1, i, j<k} \\ 1 & \text{if } i < k, j \geqslant k \\ -1 & \text{if } i \geqslant k, j < k \\ 0 & \text{otherwise}\,, \end{cases} \tag{3}$$

where $c = \min_{i \in \mathcal{I}} \left(\mathbf{p}_i \mathbf{q}_{(i+1 \mod k)}, \mathbf{p}_i \mathbf{q}_{(i+2 \mod k)} \right)$ is a normalizing constant ensuring that all the entries of R^{eRPS} are between -1 and 1.

For support size $k = 1$, namely $(\mathbf{p}, \mathbf{q})$ is a pure strategy profile, the R^{eRPS} game is visualized in Table 1 (left). It is easy to check that the upper left corner $(0, 0)$ is indeed the unique pure Nash equilibrium.

For support size $k \geqslant 2$, namely $(\mathbf{p}, \mathbf{q})$ is a mixed strategy profile, the R^{eRPS} game is visualized in Table 1 (right) and Table 2. As a special case, for $\mathbf{p} = \mathbf{q} = (1/3, 1/3, 1/3)$, R^{eRPS} is the standard Rock-Paper-Scissors game.

Lemma 4. Given any $(\mathbf{p}, \mathbf{q})$ with equal support sizes, the Extended Rock-Paper-Scissors Game $R^{\text{eRPS}\,(\mathbf{p},\mathbf{q})}$ has $(\mathbf{p}, \mathbf{q})$ as the unique Nash equilibrium, and its game value is 0.

Note that applying any positive affine transformation to the reward matrix preserves the set of Nash equilibria of the game (Tewolde, 2023). Therefore, if we want the game R to be bounded between $[-b, b]$ for $b > 0$, we can simply scale

0	1	...	1
-1	0	...	0
...	...	...	
-1	0	...	0

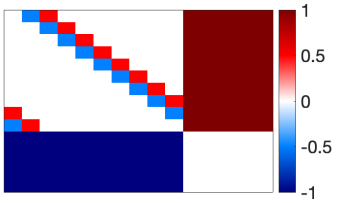

Table 1. R^{eRPS} when $k = 1$ (left) and $k \geqslant 2$ (right).

R^{eRPS} by b. More generally, for each $\iota > 0$ and $v \in \mathbb{R}$, the game $\iota R^{\text{eRPS}} + v$ has entries in $[v - \iota, v + \iota]$ and $(\mathbf{p}, \mathbf{q})$ as the unique Nash equilibrium with value v.

There exist other constructions of games that have some $(\mathbf{p}, \mathbf{q})$ as the unique NE; see Bohnenblust et al. (1950); Nagarajan et al. (2020). Nevertheless, our eRPS construction is simple and intuitive, generalizing the well-known rock paper scissors game. The eRPS game matrix also possesses certain cyclic symmetry and naturally has game value 0. As we will soon see, the eRPS game is also used in our game modification algorithm. The proof of Lemma 4 in Appendix A.2 for eRPS showcases an application of the sufficiency of the SIISOW and INV conditions for NE uniqueness.

3.4. An Efficient Algorithm for Game Modification in Normal Form Games

We now turn to the main result of this section: We describe an efficient algorithm to approximately solve Game Modification in normal-form games and provide guarantees on its correctness. In particular, we relax the invertibility constraints so that the remaining constraints are linear and perturb the solution in a way that maintains the feasibility of the linear constraints while making the perturbed solution satisfy invertibility with probability 1.

Thanks to Lemma 2, the requirement of R having $(\mathbf{p}, \mathbf{q})$ as the unique NE can be fulfilled by the equivalent SIISOW and the INV conditions, as done in reformulation in Proposition 1. If we ignore the INV condition therein for a moment and tighten the strict inequalities, we obtain an optimization problem with linear constraints:

$$\min_{R, v} \ell\left(R, R^{\circ}\right) \tag{4a}$$

$$\text{s.t. } R_{\mathcal{I}\bullet}\mathbf{q} = v\mathbf{1}_{|\mathcal{I}|} \tag{4b}$$

$$\mathbf{p}^{\top} R_{\bullet\mathcal{J}} = v\mathbf{1}_{|\mathcal{J}|}^{\top} \tag{4c}$$

$$R_{\mathcal{A}_1\setminus\mathcal{I}\bullet}\mathbf{q} \leqslant (v - \iota)\,\mathbf{1}_{|\mathcal{A}_1\setminus\mathcal{I}|} \tag{4d}$$

$$\mathbf{p}^{\top} R_{\bullet\mathcal{A}_2\setminus\mathcal{J}} \geqslant (v + \iota)\,\mathbf{1}_{|\mathcal{A}_2\setminus\mathcal{J}|}^{\top} \tag{4e}$$

$$\underline{v} \leqslant v \leqslant \overline{v} \tag{4f}$$

$$-b + \lambda \leqslant R_{ij} \leqslant b - \lambda, \forall\,(i, j) \in \mathcal{A}. \tag{4g}$$

In (4), the first four constraints (4b)–(4e) encode the SI-

$\mathcal{A}_1 \backslash \mathcal{A}_2$	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

Table 2. The R^{eRPS} game when $k \geqslant 2$, i.e. $(\mathbf{p}, \mathbf{q})$ is a mixed strategy

ISOW condition. Notice we introduced a small SIISOW margin parameter $\iota > 0$ in (4d) and (4e), tightening the strict inequalities in Proposition 1. Doing so ensures that the feasible set of the problem (4) is closed. A margin λ is also added to the reward bound (4g) for reasons that would become clear momentarily.

One can solve the linearly constrained program (4) for a solution R. To ensure R has a unique NE, it remains to satisfy the INV condition that the matrix $\begin{bmatrix} R_{\mathcal{I}\mathcal{J}} & -\mathbf{1}_{|\mathcal{I}|} \\ \mathbf{1}_{|\mathcal{J}|}^\top & 0 \end{bmatrix}$ must be invertible. However, enforcing INV directly by constraining the smallest singular value of the matrix leads to a nonlinear, nonconvex optimization problem that is difficult to solve.

We adopt an alternative approach: we take the solution R' to the program (4)—which may not satisfy the INV condition—and add a small special random matrix to R' in such a way that: (i) the resulting matrix R is invertible with probability 1; (ii) R still has $(\mathbf{p}, \mathbf{q})$ as its unique NE and satisfies the value constraint $v \in [\underline{v}, \overline{v}]$ in (4f). Moreover, by introducing a small margin λ in the reward bound (4g) and using a sufficiently small perturbation, we further ensure that the perturbed rewards remain in the original designated range $[-b, b]$. Specifically, the matrix we add is $\varepsilon R^{\text{eRPS}}$, where ε is a random number in $[-\lambda, \lambda]$ and R^{eRPS} the Extended Rock-Paper-Scissors game matrix, which has entries in $[-1, 1]$ and game value 0.

Combining the above ingredients, we have the complete procedure, Relax And Perturb (RAP), which is presented in Algorithm 1. RAP approximately solves the Game Modification problem, provably satisfying the constraints with probability 1 and achieving a near minimal cost $\ell(R, R^\circ)$ as long as the random perturbation is small (Theorem 5).

Remark 1. If the solution of problem (4) R' already satisfies the INV condition, that is $\begin{bmatrix} R'_{\mathcal{I}\mathcal{J}} & -\mathbf{1}_{|\mathcal{I}|} \\ \mathbf{1}_{|\mathcal{J}|}^\top & 0 \end{bmatrix}$ is invertible, then no perturbation is needed. In addition, the perturbation can also be put in a loop while the INV condition is not satisfied, although the perturbed solution satisfies the INV condition with probability one in theory.

Algorithm 1 Relax And Perturb (RAP)

Input: original game R°, cost function ℓ, target policy $(\mathbf{p}, \mathbf{q})$, target 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.

1: Solve the problem (4). Call the solution R'.
2: Sample $\varepsilon \sim \text{uniform}[-\lambda, \lambda]$
3: Return $R = R' + \varepsilon R^{\text{eRPS}(\mathbf{p},\mathbf{q})}$.

When the cost function ℓ is convex, the problem (4) is a convex program with linear constraints, for which efficient solvers exist (Wright, 2006). The program (4) is further reduced to a linear program when ℓ is piecewise linear, as shown in the following examples.

Example 1 (L^1 Cost). One may measure the cost of modifying the game from R° to R by the L^1 distance

$$\ell(R, R^\circ) = \|R - R^\circ\|_1 = \sum_{i \in \mathcal{A}_1, j \in \mathcal{A}_2} \left| R_{ij} - R^\circ_{ij} \right|.$$

Example 2 (Occupancy Weighted Cost). In some applications, the cost of modifying an entry is proportional to how often the entry is visited by the players at the equilibrium $(\mathbf{p}, \mathbf{q})$. We can use the following weighted cost function:

$$\ell(R, R^\circ) = \sum_{i \in \mathcal{A}_1, j \in \mathcal{A}_2} \mathbf{p}_i \mathbf{q}_j \left| R_{ij} - R^\circ_{ij} \right|. \tag{5}$$

Note that it is costless to modify the entries outside the product of the supports of $\mathbf{p}, \mathbf{q}$. Applications of this weighted cost include online reward poisoning in multi-agent reinforcement learning, where an attacker pays for the modified reward entry only when the online learners use the corresponding action profile.

3.5. Performance Guarantees for RAP

Below, we show that the RAP Algorithm has the desired feasibility and near-optimality properties with respect to the original Game Modification problem (2) in Proposition 1. Let $C^\star$ denote the optimal value of (2). We say that the cost function ℓ is L-Lipschitz if $|\ell(X, R^\circ) - \ell(Y, R^\circ)| \leqslant L\|X - Y\|_1, \forall X, Y$.

Theorem 5 (Feasibility and Optimality of RAP Algorithm). Suppose that the parameters ι, λ of Algorithm 1 satisfy

$$(-b + \lambda + \iota, b - \lambda - \iota) \cap [-\underline{v}, \overline{v}] \neq \varnothing$$

and let $R(\iota, \lambda) = R' + \varepsilon R^{\text{eRPS}}$ be the output of the Algorithm 1 with margin parameters ι, λ. The following hold.

1. (**Existence**) The solution R' to program (4) exists.
2. (**Feasibility**) With probability 1, $R(\iota, \lambda)$ is feasible for the original Game Modification problem (2).
3. (**Optimality**) If in addition the cost ℓ is L-Lipschitz with $L < \infty$, 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 (e.g., L^1 cost), then the optimality gap is at most linear in (ι, λ):
$$\ell(R(\iota, \lambda), R^\circ) = C^\star + O(\max\{\iota, \lambda\}).$$

In the result above, existence follows from Theorem 3. Feasibility holds because the matrix sum

$$\begin{bmatrix} R'_{\mathcal{I}\mathcal{J}} & -\mathbf{1}_{|\mathcal{I}|} \\ \mathbf{1}^\top_{|\mathcal{J}|} & 0 \end{bmatrix} + \varepsilon \begin{bmatrix} R^{eRPS}_{\mathcal{I}\mathcal{J}} & -\mathbf{1}_{|\mathcal{I}|} \\ \mathbf{1}^\top_{|\mathcal{J}|} & 0 \end{bmatrix}$$

is invertible with probability 1, as ε is a continuous random variable and the second matrix above is invertible.

To prove optimality, we take a feasible solution $R^{(\varepsilon)}$ to the original game modification problem (2) with a cost at most $C^\star + \varepsilon$, and then slightly and carefully modify its entries to get a new solution $R'^{(\varepsilon)}$ for which (i) the reward bound (4g) with λ margin is satisfied, (ii) the SIISOW properties (4b)–(4e) are preserved, and (iii) the game value is the same. The costs of $R'^{(\varepsilon)}$ and $R^{(\varepsilon)}$ are close thanks to the Lipschitz property of the cost. In particular, the difference $\ell(R'^{(\varepsilon)}, R^\circ) - \ell(R^{(\varepsilon)}, R^\circ)$, and in turn the optimality gap, vanish when the margin parameters ι, λ go to zero.

Part 4 of the theorem further shows that the optimality gap vanishes at a linear rate in (ι, λ) under piecewise linear cost ℓ. In this case (4) is a linear program with a full rank constraint matrix, and we can control the optimality gap using techniques from sensitivity analysis of linear programs (Bertsimas & Tsitsiklis, 1997; Jansen et al., 1997).

4. Markov Games Modification

In this section, we generalize to Markov games. We install a possibly stochastic policy as the unique Markov perfect equilibrium by installing a unique Nash equilibrium in every stage game defined by the Q functions.

4.1. Preliminaries

A finite-horizon two-player zero-sum Markov game can be described by a pair (P, R), given the finite state space $\mathcal{S}$, the finite joint action space $\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2$, and horizon H. Here $P = \left\{P_h : \mathcal{S} \times \mathcal{S} \to [0,1]^{|\mathcal{A}_1|\times|\mathcal{A}_2|}\right\}_{h=1}^H$ is the transition probabilities, $P_0 : \mathcal{S} \to [0,1]$ the initial state distribution, and $R = \left\{R_h : \mathcal{S} \to [-b, b]^{|\mathcal{A}_1|\times|\mathcal{A}_2|}\right\}_{h=1}^H$ the mean reward function. For each $h \in [H], s \in \mathcal{S}$, we treat $R_h(s)$ as an $|\mathcal{A}_1| \times |\mathcal{A}_2|$ matrix, where $[R_h(s)]_{ij}$ is the reward for the joint action profile $(i, j) \in \mathcal{A}_1 \times \mathcal{A}_2$. Similarly, the transition probabilities are given by an $|\mathcal{A}_1| \times |\mathcal{A}_2|$ matrix $P_h(s'|s)$, where $[P_h(s'|s)]_{ij}$ is the probability of transitioning from state $s \in \mathcal{S}$ in period $h \in [H]$ to state $s' \in \mathcal{S}$ when the joint action (i, j) is used. The above matrix representations are chosen to follow the convention used in the last section for normal-form matrix games.

A Markovian policy $(\mathbf{p}, \mathbf{q})$ is a pair of policies for the two players: $\mathbf{p} = \{\mathbf{p}_h : \mathcal{S} \to \Delta_{\mathcal{A}_1}\}_{h=1}^H$ and $\mathbf{q} = \{\mathbf{q}_h : \mathcal{S} \to \Delta_{\mathcal{A}_2}\}_{h=1}^H$. Here $\mathbf{p}_h(s)$ and $\mathbf{q}_h(s)$ are probability vectors; in period $h \in [H]$, state $s \in \mathcal{S}$, $[\mathbf{p}_h(s)]_i$ specifies the probability that player 1 takes action $i \in \mathcal{A}_1$; similarly for $[\mathbf{q}_h(s)]_j$.

A zero-sum Markov game has at least one Markov perfect equilibrium and a unique Nash value. The action-value or Q function of the MPE, denoted by $Q^\star$, satisfies the following Bellman equations: for each $h \in [H], s \in \mathcal{S}, (i, j) \in \mathcal{A}$,

$$Q^\star_h(s, (i, j)) := R_h(s, (i, j)) + \sum_{s' \in \mathcal{S}} P_h\left(s'|s, (i, j)\right) \max_{p' \in \Delta_{\mathcal{A}_1}} \min_{q' \in \Delta_{\mathcal{A}_2}} Q^\star_{h+1}\left(s', \left(p', q'\right)\right), \tag{6}$$

where for a possibly stochastic strategy profile $(p', q') \in \Delta_{\mathcal{A}_1} \times \Delta_{\mathcal{A}_2}$, we define

$$Q^\star_h\left(s, \left(p', q'\right)\right) := \sum_{i \in \mathcal{A}_1, j \in \mathcal{A}_2} p'_i q'_j Q^\star_h(s, (i, j)). \tag{7}$$

We use the convention that $Q^\star_{H+1}(s, (i, j)) = 0, \forall s, i, j$.

Under an MPE policy, the stage game of the Markov game in each period $h \in [H]$ and state $s \in \mathcal{S}$ is a normal form game with payoff matrix $\mathbb{Q}_h(s)$, whose (i, j) entry is

$$[\mathbb{Q}_h(s)]_{ij} := Q^\star_h(s, (i, j)) \tag{8}$$

and corresponds to the payoff under the action profile $(i, j) \in \mathcal{A}$. Consequently, an MPE can be defined recursively as the Nash equilibrium for every stage game.

Definition 4 (Markov Perfect Equilibrium). A Markov perfect equilibrium policy $(\mathbf{p}, \mathbf{q})$ is a policy that satisfies, for every $h \in [H], s \in \mathcal{S}$,

$$(\mathbf{p}_h(s), \mathbf{q}_h(s)) \text{ is a Nash equilibrium of } \mathbb{Q}_h(s),$$

where $\mathbb{Q}_h(s)$ is defined by equations (6)–(8).

We remark that an alternative approach to studying the equilibria of Markov games is by converting it to a single, big normal-form game and considering the NEs of the latter. An NE defined in this way is, in general, not Markov perfect—it requires coordination and commitment to policies in stage games that are not visited along equilibrium paths. Such policies are often not realistic. Moreover, it is computationally intractable to manipulate such a big normal-form game. Therefore, we focus on MPEs and make use of their recursive characterization through the Bellman equations.

4.2. Reformulation and Feasibility of Markov Game Modification

A two-player zero-sum Markov game has a unique MPE if and only if every stage game $\mathbb{Q}_h(s)$ has a unique NE. Our results on the uniqueness of NE for normal form games (Lemma 2) apply to each stage game of the Markov game. Combining these two observations and the Bellman equations for $\mathbb{Q}_h(s)$'s, we can write the Game Modification problem in Definition 1 for Markov games equivalently as an optimization problem similar to (2), where SIISOW (Condition 1), INV (Condition 2) and the Bellman equations are imposed as constraints for every stage game. Due to space limit, this optimization problem is provided in the appendix.

We provide a sufficient and necessary condition for the feasibility of the above Game Modification problem for Markov games. Let $\mathcal{I}_h(s) = supp(\mathbf{p}_h(s))$ and $\mathcal{J}_h(s) = supp(\mathbf{q}_h(s))$.

Theorem 6 (Feasibility of Markov Game Modification). The Game Modification problem in Definition 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 $(-Hb, Hb) \cap [\underline{v}, \overline{v}] \neq \varnothing$.

The above theorem subsumes Theorem 3 for normal-form games. The sufficient condition above is proved by explicitly constructing a feasible Markov game, recursively using the Extended Rock-Paper-Scissors game.

4.3. Efficient Algorithm for Modifying Markov Games

To develop an efficient algorithm, we follow a similar strategy as in normal form games: we ignore the INV (invertibility) condition and retain only the linear constraints for the Markov game modification problem, and add small margins ι, λ to the SIISOW and reward bound constraints so that random perturbation can be added later. Doing so leads to a linearly constrained optimization problem, given in (9), which generalizes the program (4) for normal-form games.

$$\begin{aligned}
\min_{R,v,\mathbb{Q}} \ & \ell(R, R^\circ) && (9)\\
\text{s.t. } & [\mathbb{Q}_h(s)]_{\mathcal{I}_h(s)\bullet}\, \mathbf{q}_h(s) = v_h(s)\, \mathbf{1}_{|\mathcal{I}_h(s)|} \\
& \quad \forall\, h \in [H], s \in \mathcal{S} && \text{[row SII]}\\
& \mathbf{p}_h^\top(s)\, [\mathbb{Q}_h(s)]_{\bullet \mathcal{J}_h(s)} = v_h(s)\, \mathbf{1}^\top_{|\mathcal{J}_h(s)|} \\
& \quad \forall\, h \in [H], s \in \mathcal{S} && \text{[column SII]}\\
& [\mathbb{Q}_h(s)]_{\mathcal{A}_1 \setminus \mathcal{I}_h(s)\bullet}\, \mathbf{q}_h(s) \leqslant (v_h(s) - \iota)\, \mathbf{1}_{|\mathcal{A}_1 \setminus \mathcal{I}_h(s)|} \\
& \quad \forall\, h \in [H], s \in \mathcal{S} && \text{[row SOW]}\\
& \mathbf{p}_h^\top(s)\, [\mathbb{Q}_h(s)]_{\bullet \mathcal{A}_2 \setminus \mathcal{J}_h(s)} \geqslant (v_h(s) + \iota)\, \mathbf{1}^\top_{|\mathcal{A}_2 \setminus \mathcal{J}_h(s)|} \\
& \quad \forall\, h \in [H], s \in \mathcal{S} && \text{[column SOW]}\\
& \mathbb{Q}_h(s) = R_h(s) + \sum_{s' \in \mathcal{S}} P_h\left(s' | s\right) v_{h+1}\left(s'\right) \\
& \quad \forall\, h \in [H-1], s \in \mathcal{S} && \text{[Bellman]}\\
& \mathbb{Q}_H(s) = R_H(s), \forall\, s \in \mathcal{S} \\
& \underline{v} \leqslant \sum_{s \in \mathcal{S}} P_0(s)\, v_1(s) \leqslant \overline{v} && \text{[value range]}\\
& -b + \lambda \leqslant [R_h(s)]_{ij} \leqslant b - \lambda \\
& \quad \forall\ (i,j) \in \mathcal{A}, h \in [H], s \in \mathcal{S} && \text{[reward bound]}
\end{aligned}$$

Remark 2. If there is no value range constraint and the cost ℓ is decomposable across the states and periods (e.g., L^1 cost), then the program (9) can be broken into $H|\mathcal{S}|$ smaller optimization problems, one for each stage game, that can be solved sequentially by backward induction.

We present our algorithm, Relax And Perturb for Markov Games (RAP-MG), in Algorithm 2, which adds random perturbation to the reward matrix of every stage game.

Algorithm 2 Relax And Perturb for Markov Games (RAP-MG)

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 (9). Call the solution R'.
2: **for** $h \in [H], s \in \mathcal{S}$ **do**
3: Sample $\varepsilon \sim \text{uniform}[-\lambda, \lambda]$
4: Perturb the reward matrix in stage (h, s):
5: $R_h(s) = R'_h(s) + \varepsilon R^{\text{eRPS}(\mathbf{p}_h(s), \mathbf{q}_h(s))}$.
6: **end for**
7: Return (R, P).

In the theorem below we provide feasibility and optimality guarantees for Algorithm 2. These results are similar to

those normal form games in Theorem 5, but the proofs are more complicated due to the dependency across the stage games. Let $C^\star$ be the optimal objective value for the original game modification problem in Definition 1.

Theorem 7 (Feasibility and Optimality of the RAP-MG Algorithm). Let $R(\iota,\lambda) = R' + \varepsilon R^{\text{eRPS}}$ denote the output of Algorithm 2 with margin parameters ι, λ. If

$$(-b+\lambda+\iota, b-\lambda-\iota) \cap [-\underline{v}/H, \overline{v}/H] \neq \varnothing, \quad (10)$$

then the following hold.

1. (**Existence**) The solution R' to the program (9) exists.
2. (**Feasibility**) $R(\iota,\lambda)$ is feasible for the game modification problem in Definition 1 with probability 1.
3. (**Optimality**) If in addition the cost function ℓ is L-Lipschitz, then $R(\iota,\lambda)$ is asymptotically optimal:
$$\lim_{\max\{\iota,\lambda\}\to 0} \ell\left(R(\iota,\lambda), R^\circ\right) = C^\star,$$
4. (**Optimality Gap**) If ℓ is piecewise linear, then
$$\ell\left(R(\iota,\lambda), R^\circ\right) = C^\star + O(\max\{\iota,\lambda\}),$$

5. Experiments

5.1. Toy Experiments

We run Algorithm 1 on several small normal-form games such as two-finger Morra and five-action rock-paper-scissors games.

1. Given left below is the payoff matrix for the **simplified Two-finger Morra** game (Good, 1965), 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.

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

We provide another example of game modification for the classic Two-finger Morra game in the Appendix A.6.

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 (Tagiew, 2009). 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.

Note that an alternative 5-action game, Rock-Paper-Scissors-Spock-Lizard, also has the desired NE (more details are provided in the Appendix A.6). However, our modification has a lower modification cost 4, compared to the cost 8 for using the alternative game.

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

5.2. Approximation

Theorem 5 shows that Algorithm 1 approaches the optimal cost C^* as a linear function in $\max\{\iota,\lambda\}$ in the worst case. To see how fast the convergence happens in practice, we tested RAP on a fixed Game Modification instance with varying choices of ι and λ. In particular, we considered $(p,q) = ((.47,.53,0,0)^\top, (.42,.58,0,0)^\top)$,

$$R^\circ = \begin{pmatrix} -0.33 & -0.03 & 0.68 & -0.04 \\ 0.16 & -0.43 & 0.94 & -0.45 \\ 0.02 & 0.85 & -0.28 & -0.98 \\ -0.57 & 0.3 & -0.12 & -0.17 \end{pmatrix}, \quad (11)$$

and no reward bound or value constraints. We considered ι and λ of the form 10^{-i} for $i \in \{0,\dots,15\}$. However, convergence happened by 10^{-4} for both parameters and so we only report for parameters down to 10^{-4}.

To further explore which parameter had the largest effect on the cost, we ran RAP under three different configurations. The result is Figure 1. First, we fixed $\iota = 10^{-5}$ and varied λ to construct the λ curve. Second, we fixed $\lambda = 10^{-5}$ and varied ι to construct the ι curve. Lastly, we varied both equally, i.e. considered $(\iota,\lambda) = (10^{-i}, 10^{-i})$, to construct the $\lambda = \iota$ curve.

We observe that in all three cases, convergence happened even faster than the linear rate promised by Theorem 5. In addition, we see that λ was generally the bottleneck for convergence with the λ curve being very close to the $\lambda = \iota$ curve. In contrast, ι had less of an impact on convergence. We ran the same experiment on other, uniform-randomly generated instances and noticed a general trend of λ being the dominant factor.

5.3. Scale Benchmarks

We run Algorithm 1 and Algorithm 2 on several games to illustrate the efficacy of our techniques. We know our algorithm succeeds by checking that $(\mathbf{p},\mathbf{q})$ satisfies the SIISOW and INV conditions for $R^\dagger$.

We first show how our methods scale with the number of actions. For each $m \in \{2,4,8,\dots,512\}$ we generate $N = 5$ random matrices $R^\circ \sim \text{uniform}[-1,1]^{m\times m}$.

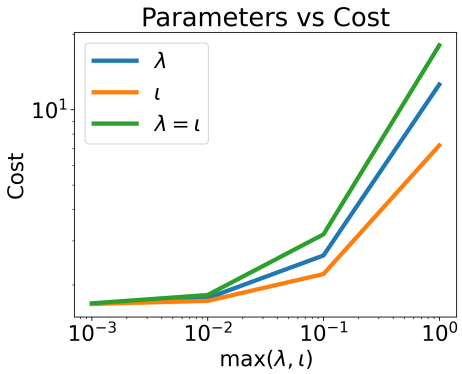

Figure 1. Convergence to Optimal Cost

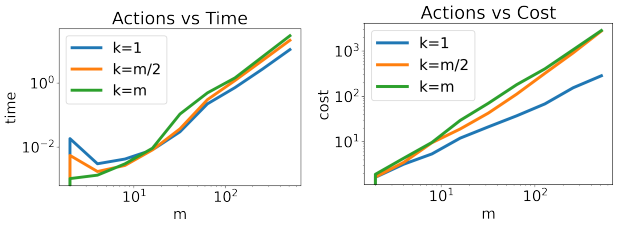

Figure 2. Scale Benchmark for Number of Actions

For each matrix, we also generate 3 random $(\mathbf{p}, \mathbf{q}) \sim \text{Dirichlet}(1, \ldots, 1)$ with support size (i) $k = 1$, (ii) $k = m/2$, and (iii) $k = m$ (full support). We run Algorithm 1 on each instance and report the worst running time (in seconds) and the worst cost encountered for each m in Figures 2. We see that the solving time grows linearly in the log, so the runtime is polynomial in the actions. Using the Gurobi LP solver, even on a laptop computer, the algorithm handles millions of variables (512^2) in roughly 10 seconds. The L^1 costs also appear to grow linearly, though with different slopes.

Next, we show how our methods scale with the horizon. We consider Markov games with $S = 10$, $A = 2$, random transitions and random reward matrices. Formally, for each $H \in \{1, 2, 4, \ldots, 512\}$, we generate $N = 5$ random Markov games and corresponding target NE pairs with full support. For any fixed H, we generate $R_h(s) \in \text{uniform}[-1, 1]^{2\times 2}$ for each h and s, and choose $P_h(s, a) \sim \text{Dirichlet}(1, \ldots 1)$ for each (h, s, a). We run Algorithm 2 on each instance and report the worst running time and cost encountered for each H in Figures 3. We observe the solutions are correct, and again, the algorithm is efficient.

6. Concluding Remarks

Our work points to several future directions: (i) It is interesting to study Markov game modification problems where the transition probabilities can also be changed and generalize to general-sum, multi-agent games with other equilibrium concepts. (ii) In many games, the rewards are constrained to take discrete values (e.g., $-1, 0, 1$). The feasibility and tractability of such constrained game modification problems require further investigation. (iii) It is non-trivial to extend the problem when the attacker's target is an infinite set of policies, for example, when the attacker only cares about the support of the target policies and not specific mixing probabilities or when the attacker only cares about the target policies of one of the players. (iv) Extending our results to data poisoning problems, where the players learn the true game from observational data, leads to interesting theoretical and algorithmic questions.

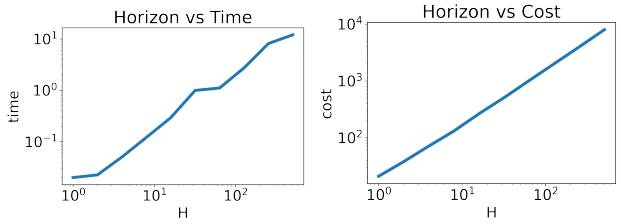

Figure 3. Scale Benchmark for Number of Periods

Acknowledgments

Xie was supported in part by National Science Foundation Awards CNS-1955997 and EPCN-2339794. Zhu was supported in part by NSF grants 1836978, 2023239, 2202457, 2331669, ARO MURI W911NF2110317, and AF CoE FA9550-18-1-0166. Chen is partially supported in part by NSF grant CCF-2233152. We would like to thank Lijun Ding for inspiring discussion. We would like to thank Joy Cheng for implementing the code for the RAP algorithm.

Impact Statement

This paper presents work whose goal is to advance the field of MARL. Our work is largely theoretical, so we do not see any immediate negative societal impacts.

References

Anderson, A., Shoham, Y., and Altman, A. Internal implementation. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1*, pp. 191–198. Citeseer, 2010.

Appa, G. On the uniqueness of solutions to linear programs. *Journal of the Operational Research Society*, 53:1127–1132, 2002.

Banihashem, K., Singla, A., Gan, J., and Radanovic, G.

Admissible policy teaching through reward design. *arXiv preprint arXiv:2201.02185*, 2022.

Bertsimas, D. and Tsitsiklis, J. N. *Introduction to linear optimization*, volume 6. Athena scientific Belmont, MA, 1997.

Bohnenblust, H., Karlin, S., and Shapley, L. Discrete solutions of two-person games. *Contributions to the Theory of Games*, (24):51, 1950.

Dantzig, G. *Linear programming and extensions*. Princeton university press, 1963.

Good, R. f-finger morra. *SIAM Review*, 7(1):81–87, 1965.

Heuer, G. Uniqueness of equilibrium points in bimatrix games. *International Journal of Game Theory*, 8:13–25, 1979.

Huang, Y. and Zhu, Q. Deceptive reinforcement learning under adversarial manipulations on cost signals. In *International Conference on Decision and Game Theory for Security*, pp. 217–237. Springer, 2019.

Jansen, B., De Jong, J., Roos, C., and Terlaky, T. Sensitivity analysis in linear programming: just be careful! *European Journal of Operational Research*, 101(1):15–28, 1997.

Kreps, V. Bimatrix games with unique equilibrium points. 1974.

Ma, Y., Zhang, X., Sun, W., and Zhu, J. Policy poisoning in batch reinforcement learning and control. *Advances in Neural Information Processing Systems*, 32:14570–14580, 2019.

Ma, Y., Wu, Y., and Zhu, X. Game redesign in no-regret game playing. *arXiv preprint arXiv:2110.11763*, 2021.

Mangasarian, O. Uniqueness of solution in linear programming. Technical report, University of Wisconsin-Madison Department of Computer Sciences, 1978.

Maskin, E. and Tirole, J. Markov perfect equilibrium: I. observable actions. *Journal of Economic Theory*, 100(2): 191–219, 2001.

Mertikopoulos, P., Papadimitriou, C., and Piliouras, G. Cycles in adversarial regularized learning. In *Proceedings of the twenty-ninth annual ACM-SIAM symposium on discrete algorithms*, pp. 2703–2717. SIAM, 2018.

Millham, C. Constructing bimatrix games with special properties. *Naval Research Logistics Quarterly*, 19(4): 709–714, 1972.

Monderer, D. and Tennenholtz, M. k-implementation. In *Proceedings of the 4th ACM conference on Electronic Commerce*, pp. 19–28, 2003.

Nagarajan, S. G., Balduzzi, D., and Piliouras, G. From chaos to order: Symmetry and conservation laws in game dynamics. In *International Conference on Machine Learning*, pp. 7186–7196. PMLR, 2020.

Nash Jr, J. F. Equilibrium points in n-person games. *Proceedings of the national academy of sciences*, 36(1):48–49, 1950.

Osborne, M. J. *An introduction to game theory*, volume 3. Oxford university press New York, 2004.

Quintas, L. G. *Uniqueness of Nash equilibrium points in bimatrix games*. Center for Mathematical Studies in Economics and Management Science., 1988.

Rakhsha, A., Radanovic, G., Devidze, R., Zhu, X., and Singla, A. Policy teaching via environment poisoning: Training-time adversarial attacks against reinforcement learning. In *International Conference on Machine Learning*, pp. 7974–7984. PMLR, 2020.

Rakhsha, A., Radanovic, G., Devidze, R., Zhu, X., and Singla, A. Policy teaching in reinforcement learning via environment poisoning attacks. *Journal of Machine Learning Research*, 22(210):1–45, 2021a.

Rakhsha, A., Zhang, X., Zhu, X., and Singla, A. Reward poisoning in reinforcement learning: Attacks against unknown learners in unknown environments. *arXiv preprint arXiv:2102.08492*, 2021b.

Rangi, A., Xu, H., Tran-Thanh, L., and Franceschetti, M. Understanding the limits of poisoning attacks in episodic reinforcement learning. In Raedt, L. D. (ed.), *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pp. 3394–3400. International Joint Conferences on Artificial Intelligence Organization, 7 2022. doi: 10.24963/ijcai.2022/471. URL https://doi.org/10.24963/ijcai.2022/471. Main Track.

Szilágyi, P. On the uniqueness of the optimal solution in linear programming. *Revue d'analyse numérique et de théorie de l'approximation*, 35(2):225–244, 2006.

Tagiew, R. Hypotheses about typical general human strategic behavior in a concrete case. In *Congress of the Italian Association for Artificial Intelligence*, pp. 476–485. Springer, 2009.

Tewolde, E. Game transformations that preserve Nash equilibria or best response sets. *arxiv preprint arXiv:2111.00076*, 2023.

Wright, S. J. *Numerical optimization*. New York, NY: Wiley, 2006.

Wu, Y., McMahan, J., Zhu, X., and Xie, Q. On faking a Nash equilibrium. *arXiv preprint arXiv:2306.08041*, 2023a.

Wu, Y., McMahan, J., Zhu, X., and Xie, Q. Reward poisoning attacks on offline multi-agent reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 10426–10434, 2023b.

Zhang, H. and Parkes, D. C. Value-based policy teaching with active indirect elicitation. In *AAAI*, volume 8, pp. 208–214, 2008.

Zhang, H., Parkes, D. C., and Chen, Y. Policy teaching through reward function learning. In *Proceedings of the 10th ACM conference on Electronic commerce*, pp. 295–304, 2009.

Zhang, X., Ma, Y., Singla, A., and Zhu, X. Adaptive reward-poisoning attacks against reinforcement learning. In *International Conference on Machine Learning*, pp. 11225–11234. PMLR, 2020.

A. Appendix

In this appendix we provide omitted proofs and additional experiments.

A.1. Proof of Lemma 2

Proof. Lemma 2 states that the SIISOW and INV conditions are sufficient and necessary for $(\mathbf{p}, \mathbf{q})$ to be the unique NE of the game R. We prove sufficiency and necessity separately.

We exploit the well-established connection between Nash equilibrium and linear program duality. In particular, any $(\mathbf{p}, \mathbf{q})$ that is a Nash equilibrium of R is an optimal solution pair to the following pair of primal-dual linear programs (LPs), and vice versa (Dantzig, 1963).

Definition 5 (Linear Programs for NE).

$$\text{(Primal)} \quad \max_{\mathbf{p}' \in \Delta \mathcal{A}_1, v} v \quad \text{s.t. } \mathbf{p}'^\top R \geqslant v \mathbf{1}_{|\mathcal{J}|}^\top \tag{12}$$

$$\text{(Dual)} \quad \min_{\mathbf{q}' \in \Delta \mathcal{A}_2, v} v \quad \text{s.t. } R\mathbf{q}' \leqslant v \mathbf{1}_{|\mathcal{I}|} \tag{13}$$

The inequalities are elementwise.

The optimal values of the two linear programs both equal v^*, the value of the game.

We emphasize that these LPs are used only for characterizing the properties of the set of Nash equilibria of R and its uniqueness. We do *not* assume that the players must use LP to find an NE: they can use any other solvers and may find any one of the NEs if there are multiple ones. This reflects how NE solvers typically work in practice.

Conditions $\Rightarrow$ unique NE: We have already argued that $(\mathbf{p}, \mathbf{q})$ is an NE; see the discussion after the definition of SIISOW. Suppose $(\mathbf{r}, \mathbf{s})$ is another NE. We show that it must be the case $\mathbf{r} = \mathbf{p}, \mathbf{s} = \mathbf{q}$.

First of all, it is easy to see that $supp(\mathbf{r}) \subseteq \mathcal{I}, supp(\mathbf{s}) \subseteq \mathcal{J}$. Suppose there is a violation $\exists i \in supp(\mathbf{r}), i \notin \mathcal{I}$. By (2e), $\mathbf{e}_i^\top R\mathbf{q} < \mathbf{p}^\top R\mathbf{q} = v^*$ which leads to $\mathbf{r}^\top R\mathbf{q} < v^*$. But since $(\mathbf{r}, \mathbf{s})$ is another NE in a two-player zero-sum game, $(\mathbf{r}, \mathbf{q})$ is a third NE with $\mathbf{r}^\top R\mathbf{q} = v^*$, a contradiction. The case for $\mathbf{s}$ is similar.

Because $(\mathbf{r}, \mathbf{s})$ is an NE, it satisfies the primal-dual LP in Definition 5. Now with the support constraints, they satisfy the reduced LPs where the vectors and matrices are restricted to the appropriate support:

$$\max_{\mathbf{r}'_{\mathcal{I}} \in \Delta_{\mathcal{I}}, v} v \quad \text{s.t. } \mathbf{r}'^\top_{\mathcal{I}} R_{\mathcal{I}\cdot} \geqslant v \mathbf{1}_{|\mathcal{J}|}^\top \tag{14}$$

$$\min_{\mathbf{s}'_{\mathcal{J}} \in \Delta_{\mathcal{J}}, v} v \quad \text{s.t. } R_{\cdot\mathcal{J}} \mathbf{s}'_{\mathcal{J}} \leqslant v \mathbf{1}_{|\mathcal{I}|}. \tag{15}$$

We now show this must mean $\mathbf{s} = \mathbf{q}$. Consider two cases on the dual restricted LP:

(Case 1) At the solution $(\mathbf{s}, v^*)$, all constraints in $R_{\mathcal{I}\mathcal{J}}\mathbf{s}_{\mathcal{J}} \leqslant v^*$ are active, i.e. they are equalities $R_{\mathcal{I}\mathcal{J}}\mathbf{s}_{\mathcal{J}} = v^*$. Also $\mathbf{s}_{\mathcal{J}}$ sums to 1. We may write the two as a linear system:

$$\begin{bmatrix} R_{\mathcal{I}\mathcal{J}} & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{s}_{\mathcal{J}} \\ v^* \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{16}$$

By the invertability condition, $\mathbf{s}_{\mathcal{J}}$ has a unique solution and it must equal $\mathbf{q}_{\mathcal{J}}$ because $\mathbf{q}_{\mathcal{J}}$ is also a solution to this linear system. The rest of $\mathbf{s}$ and $\mathbf{q}$ are both zeros. Thus $\mathbf{s} = \mathbf{q}$.

(Case 2) At least one constraint in $R_{\mathcal{I}\mathcal{J}}\mathbf{s}_{\mathcal{J}} \leqslant v^*$ is inactive. Then there exists slack variables $\xi \in \mathbb{R}^{|\mathcal{J}|}, \xi \geqslant 0$ with at least one positive entry, such that

$$R_{\mathcal{I}\mathcal{J}}\mathbf{s}_{\mathcal{J}} = v^*\mathbf{1} - \xi.$$

Recall $(\mathbf{p}, \mathbf{q})$ is an NE. By the assumption that $(\mathbf{r}, \mathbf{s})$ is an NE, and the property of two-player zero-sum games, $(\mathbf{p}, \mathbf{s})$ is also an NE with the same value v^*. But $\mathbf{p}^\top R\mathbf{s} = \mathbf{p}_{\mathcal{I}}^\top R_{\mathcal{I}\mathcal{J}}\mathbf{s}_{\mathcal{J}} = v^* - \mathbf{p}_{\mathcal{I}}^\top \xi < v^*$, because all terms in $\mathbf{p}_{\mathcal{I}}$ are positive and at least one term in ξ is positive. This is a contradiction. So case 2 will not happen.

Taken together, $\mathbf{s} = \mathbf{q}$. Similarly, one can show $\mathbf{r} = \mathbf{p}$.

Unique NE $\Rightarrow$ conditions: Let $(\mathbf{p}, \mathbf{q})$ be the unique NE of R with value v^*, and let $\mathcal{I}, \mathcal{J}$ be their support.

We first show SIISOW. Equations (2c) and (2b) are immediate from NE definition. Since $(\mathbf{p}, \mathbf{q})$ is the only NE of the game, it satisfies Goldman and Tucker Corollary 3A. The corollary states that

$$\forall i \in \mathcal{A}_1, (\mathbf{e}_i^\top R\mathbf{q} = v^*) \Rightarrow (i \in \mathcal{I}) \tag{17}$$
$$\forall j \in \mathcal{A}_2, (\mathbf{p}^\top R\mathbf{e}_j = v^*) \Rightarrow (j \in \mathcal{J}). \tag{18}$$

Their contraposition is

$$\forall i \in \mathcal{A}_1, (i \notin \mathcal{I}) \Rightarrow (\mathbf{e}_i^\top R\mathbf{q} \neq v^*) \tag{19}$$
$$\forall j \in \mathcal{A}_2, (j \notin \mathcal{J}) \Rightarrow (\mathbf{p}^\top R\mathbf{e}_j \neq v^*). \tag{20}$$

But since v^* is the NE game value, these imply

$$\forall i \in \mathcal{A}_1, (i \notin \mathcal{I}) \Rightarrow (\mathbf{e}_i^\top R\mathbf{q} < v^*) \tag{21}$$
$$\forall j \in \mathcal{A}_2, (j \notin \mathcal{J}) \Rightarrow (\mathbf{p}^\top R\mathbf{e}_j < v^*). \tag{22}$$

Therefore, $(\mathbf{p}, \mathbf{q})$ satisfies the SIISOW condition.

We next show invertability by contradition. Suppose the matrix in Definition 2 is not invertable. Then either (i) $|\mathcal{I}| < |\mathcal{J}|$, (ii) $|\mathcal{I}| > |\mathcal{J}|$, or (iii) $|\mathcal{I}| = |\mathcal{J}| \geqslant 2$. Case (iii) is due to the fact that should $|\mathcal{I}| = |\mathcal{J}| = 1$, $R_{\mathcal{I}\mathcal{J}}$ is a scalar and the matrix $\begin{bmatrix} R_{\mathcal{I}\mathcal{J}} & -1 \\ 1 & 0 \end{bmatrix}$ with determinant 1 is always invertible. We show that any one of the three cases leads to a second NE, contradicting the uniqueness of $(\mathbf{p}, \mathbf{q})$. In what follows we give the proof for (i) or (iii); case (ii) is similar to (i) but with respect to $R_{\mathcal{I}\mathcal{J}}^\top$ and $\mathbf{p}$, and is omitted.

In cases (i) or (iii) the following homogeneous linear system has a nonzero solution:

$$\begin{bmatrix} R_{\mathcal{I}\mathcal{J}} & -\mathbf{1}_{|\mathcal{I}|} \\ \mathbf{1}_{|\mathcal{J}|}^\top & 0 \end{bmatrix} \begin{bmatrix} \delta \\ x \end{bmatrix} = 0, \tag{23}$$

where $\delta \in \mathbb{R}^{|\mathcal{J}|}, x \in \mathbb{R}$. This nonzero solution (δ, x) has some useful properties:

- δ sums to zero:
$$\mathbf{1}^\top \delta = 0. \tag{24}$$
This follows directly from the second equality of (23).

- $\delta \neq 0$. This follows from the first equality of (23)
$$R_{\mathcal{I}\mathcal{J}}\delta = x\mathbf{1}, \tag{25}$$
otherwise both δ and x would be zero, contradicting a nonzero solution.

- $x = 0$ and
$$R_{\mathcal{I}\mathcal{J}}\delta = \mathbf{0}. \tag{26}$$
We first show $x = 0$. Consider
$$\mathbf{p}^\top R \begin{bmatrix} \delta \\ \mathbf{0}_{|\mathcal{A}_2|-|\mathcal{J}|} \end{bmatrix} \tag{27}$$
$$= \sum_{j \in \mathcal{J}} \mathbf{p}^\top R\mathbf{e}_j \delta_j \tag{28}$$
$$= \sum_{j \in \mathcal{J}} v^* \delta_j \tag{29}$$
$$= 0, \tag{30}$$

where the second equality follows from the SIISOW condition $\mathbf{p}^\top R\mathbf{e}_j = \mathbf{p}^\top R\mathbf{q} = v^*,\ \forall j \in \mathcal{J}$. But at the same time, by the support of $\mathbf{p}$

$$\begin{aligned} & \mathbf{p}^\top R \begin{bmatrix} \delta \\ \mathbf{0}_{|\mathcal{A}_2|-|\mathcal{J}|} \end{bmatrix} && (31) \\ = \ & \mathbf{p}_{\mathcal{I}}^\top R_{\mathcal{I}\mathcal{J}}\delta && (32) \\ = \ & \mathbf{p}_{\mathcal{I}}^\top x\mathbf{1} = x. && (33) \end{aligned}$$

Therefore $x = 0$. Then use (25) to obtain (26).

We use this δ to construct another NE with the following steps:

1. We scale δ so its magnitute is sufficiently small. The desired scale is determined by two constants:
   (a) Since we are under cases (i) or (iii), $|\mathcal{J}| \geqslant 2$. Thus the entries of $\mathbf{q}_{\mathcal{J}}$ cannot be 0 or 1: $\exists c_1 > 0 : c_1 \leqslant q_j \leqslant 1 - c_1, \forall j \in \mathcal{J}$.
   (b) By the SIISOW condition, $\mathbf{e}_i^\top R\mathbf{q} < v^*$ for $i \notin \mathcal{I}$. Let $c_2 = v^* - \max_{i\notin\mathcal{I}} \mathbf{e}_i^\top R\mathbf{q}$.

   We choose the scale
   $$c = \min\left(\frac{c_1}{\|\delta\|_\infty}, \min_{i\notin I} \frac{c_2}{|R_{i\mathcal{J}}\delta|}\right). \tag{34}$$
2. Set $\mathbf{r} = \mathbf{q} + \begin{bmatrix} c\delta \\ 0 \end{bmatrix}$.

We claim $(\mathbf{p}, \mathbf{r})$ is another NE:

- Since δ sums to zero, $\mathbf{q}_{\mathcal{J}} + c\delta$ remains normalized; since $c \leqslant \frac{c_1}{\|\delta\|_\infty}$, all entries of $\mathbf{q}_{\mathcal{J}} + c\delta$ remains in $[0, 1]$. Therefore $\mathbf{r} \in \Delta_{\mathcal{A}_2}$ is a proper strategy.
- $\mathbf{r}$ is a best response to $\mathbf{p}$:
  $$\mathbf{p}^\top R\mathbf{r} = \mathbf{p}^\top R\mathbf{q} + \mathbf{p}^\top R \begin{bmatrix} c\delta \\ 0 \end{bmatrix} = v^*, \tag{35}$$
  where we used (27). Therefore, $\mathbf{p}^\top R\mathbf{r} = v^* \leqslant \mathbf{p}^\top R\mathbf{q}', \forall \mathbf{q}' \in \Delta_{\mathcal{A}_2}$ because $\mathbf{p}$ is part of an NE.
- $\mathbf{p}$ is a best response to $\mathbf{r}$: $\forall \mathbf{p}' \in \Delta_{\mathcal{A}_1}$,
  $$\begin{aligned} & \mathbf{p}'^\top R\mathbf{r} && (36) \\ = \ & \sum_{i\in\mathcal{I}} p'_i \mathbf{e}_i^\top R\mathbf{q} + \mathbf{p}'^\top_{\mathcal{I}} R_{\mathcal{I}\mathcal{J}} c\delta + \sum_{i\notin\mathcal{I}} p'_i \left(\mathbf{e}_i^\top R\mathbf{q} + \mathbf{e}_i^\top R \begin{bmatrix} c\delta \\ 0 \end{bmatrix}\right) \\ = \ & \sum_{i\in\mathcal{I}} p'_i \mathbf{e}_i^\top R\mathbf{q} + \sum_{i\notin\mathcal{I}} p'_i \left(\mathbf{e}_i^\top R\mathbf{q} + \mathbf{e}_i^\top R \begin{bmatrix} c\delta \\ 0 \end{bmatrix}\right) \\ = \ & \sum_{i\in\mathcal{I}} p'_i v^* + \sum_{i\notin\mathcal{I}} p'_i \left(\mathbf{e}_i^\top R\mathbf{q} + \mathbf{e}_i^\top R \begin{bmatrix} c\delta \\ 0 \end{bmatrix}\right) \\ \leqslant \ & \sum_{i\in\mathcal{I}} p'_i v^* + \sum_{i\notin\mathcal{I}} p'_i \left(v^* - c_2 + \mathbf{e}_i^\top R \begin{bmatrix} c\delta \\ 0 \end{bmatrix}\right) \\ = \ & \sum_{i\in\mathcal{I}} p'_i v^* + \sum_{i\notin\mathcal{I}} p'_i \left(v^* - c_2 + cR_{i\mathcal{J}}\delta\right). && (37) \end{aligned}$$

  where the second equality follows from (26), the next two lines from SIISOW. Because $c \leqslant \min_{i\notin I} \frac{c_2}{|R_{i\mathcal{J}}\delta|}$,
  $$\begin{aligned} & \mathbf{p}'^\top R\mathbf{r} && (38) \\ \leqslant \ & \sum_{i\in\mathcal{I}} p'_i v^* + \sum_{i\notin\mathcal{I}} p'_i \left(v^* - c_2 + c_2\right) = v^* = \mathbf{p}^\top R\mathbf{r}. \end{aligned}$$

Because $\delta \neq 0$, $\mathbf{r} \neq \mathbf{q}$. Thus $(\mathbf{p}, \mathbf{r}) \neq (\mathbf{p}, \mathbf{q})$ is indeed a second NE, contradicting uniqueness.

□

A.2. Proof of Lemma 4

Proof. To show uniqueness, we check that the conditions in Lemma 2 is satisfied,

$$\begin{aligned}
\mathbf{e}_i^\top R\mathbf{q} &= 0 = \mathbf{p}^\top R\mathbf{q}, \forall\ i \in \mathcal{I}, \\
\mathbf{e}_i^\top R\mathbf{q} &= -1 < 0 = \mathbf{p}^\top R\mathbf{q}, \forall\ i \notin \mathcal{I}, \\
\mathbf{p}^\top R\mathbf{e}_j &= 0 = \mathbf{p}^\top R\mathbf{q}, \forall\ j \in \mathcal{J}, \\
\mathbf{p}^\top R\mathbf{e}_j &= 1 > 0 = \mathbf{p}^\top R\mathbf{q}, \forall\ j \notin \mathcal{J},
\end{aligned} \tag{39}$$

and we have $\begin{bmatrix} R_{\mathcal{I}\mathcal{J}} & -\mathbf{1}_{|\mathcal{I}|} \\ \mathbf{1}_{|J|}^\top & 0 \end{bmatrix}$ is invertible.

To simplify the notations, we omit the modulo k operation for the indices of $\mathbf{p}$ and $\mathbf{q}$. Observe that

$$\begin{aligned}
\mathbf{e}_i^\top R\mathbf{q} &= -\frac{c}{\mathbf{p}_i\mathbf{q}_{i+1}}\mathbf{q}_{i+1} + \frac{c}{\mathbf{p}_i\mathbf{q}_{i+2}}\mathbf{q}_{i+2} = 0, \forall\ i \in \mathcal{I}, \\
\mathbf{e}_i^\top R\mathbf{q} &= \sum_{j\in\mathcal{J}} -1\mathbf{q}_j = -1, \forall\ i \notin \mathcal{I},
\end{aligned} \tag{40}$$

and similarly,

$$\begin{aligned}
\mathbf{p}^\top R\mathbf{e}_j &= \frac{c}{\mathbf{p}_{j-2}\mathbf{q}_j}\mathbf{p}_{j-2} - \frac{c}{\mathbf{p}_{j-1}q_j}\mathbf{p}_{j-1} = 0, \forall\ j \in \mathcal{J}, \\
\mathbf{p}^\top R\mathbf{e}_j &= \sum_{i\in\mathcal{I}} 1\mathbf{p}_i = 1, \forall\ j \notin \mathcal{J}.
\end{aligned} \tag{41}$$

In addition, we have,

$$\mathbf{p}^\top R\mathbf{q} = \sum_{i\in\mathcal{I}} \mathbf{p}_i \left(\mathbf{e}_i^\top R\mathbf{q}\right) = 0. \tag{42}$$

Therefore, the SIISOW conditions are satisfied.

We now turn to the invertibility condition. For $k = 1$, $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ is invertible. For fixed $\mathbf{p}, \mathbf{q}$, for $k = 2$, we have,

$$\begin{aligned}
&\det \begin{bmatrix} \frac{c}{\mathbf{p}_0\mathbf{q}_0} & -\frac{c}{\mathbf{p}_0\mathbf{q}_1} & -1 \\ -\frac{c}{\mathbf{p}_1\mathbf{q}_0} & \frac{c}{\mathbf{p}_1\mathbf{q}_1} & -1 \\ 1 & 1 & 0 \end{bmatrix} \\
&= \det \begin{bmatrix} \frac{1}{\mathbf{p}_0} & 0 & 0 \\ 0 & \frac{1}{\mathbf{p}_1} & 0 \\ 0 & 0 & \frac{1}{c} \end{bmatrix} \det \begin{bmatrix} 1 & -1 & \mathbf{p}_0 \\ -1 & 1 & \mathbf{p}_1 \\ \mathbf{q}_0 & \mathbf{q}_1 & 0 \end{bmatrix} \det \begin{bmatrix} \frac{1}{\mathbf{q}_0} & 0 & 0 \\ 0 & \frac{1}{\mathbf{q}_1} & 0 \\ 0 & 0 & \frac{1}{c} \end{bmatrix} \\
&= c\left(\mathbf{p}_0 + \mathbf{p}_1\right)\frac{\mathbf{q}_0 + \mathbf{q}_1}{\mathbf{p}_0\mathbf{p}_1\mathbf{q}_0\mathbf{q}_1} \\
&> 0,
\end{aligned} \tag{43}$$

therefore it is invertible, similarly for $k = 3$,

$$\begin{aligned}
&\det\begin{bmatrix} 0 & -\frac{c}{\mathbf{p}_0\mathbf{q}_1} & \frac{c}{\mathbf{p}_0\mathbf{q}_2} & -1 \\ \frac{c}{\mathbf{p}_1\mathbf{q}_0} & 0 & -\frac{c}{\mathbf{p}_1\mathbf{q}_2} & -1 \\ -\frac{c}{\mathbf{p}_2\mathbf{q}_0} & \frac{c}{\mathbf{p}_2\mathbf{q}_1} & 0 & -1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \\
&= \det\begin{bmatrix} \frac{1}{\mathbf{p}_0} & 0 & 0 & 0 \\ 0 & \frac{1}{\mathbf{p}_1} & 0 & 0 \\ 0 & 0 & \frac{1}{\mathbf{p}_2} & 0 \\ 0 & 0 & 0 & \frac{1}{c} \end{bmatrix} \det\begin{bmatrix} 0 & -1 & 1 & -\mathbf{p}_0 \\ 1 & 0 & -1 & -\mathbf{p}_1 \\ -1 & 1 & 0 & -\mathbf{p}_2 \\ \mathbf{q}_0 & \mathbf{q}_1 & \mathbf{q}_2 & 0 \end{bmatrix} \det\begin{bmatrix} \frac{1}{\mathbf{q}_0} & 0 & 0 & 0 \\ 0 & \frac{1}{\mathbf{q}_1} & 0 & 0 \\ 0 & 0 & \frac{1}{\mathbf{q}_2} & 0 \\ 0 & 0 & 0 & \frac{1}{c} \end{bmatrix} \\
&= c^2 \frac{(\mathbf{p}_0 + \mathbf{p}_1 + \mathbf{p}_2)(\mathbf{q}_0 + \mathbf{q}_1 + \mathbf{q}_2)}{\mathbf{p}_0\mathbf{p}_1\mathbf{p}_2\mathbf{q}_0\mathbf{q}_1\mathbf{q}_2} \\
&> 0,
\end{aligned} \tag{44}$$

and for $k = 4$,

$$\begin{aligned}
&\det\begin{bmatrix} 0 & -\frac{c}{\mathbf{p}_0\mathbf{q}_1} & \frac{c}{\mathbf{p}_0\mathbf{q}_2} & 0 & -1 \\ 0 & 0 & -\frac{c}{\mathbf{p}_1\mathbf{q}_2} & \frac{c}{\mathbf{p}_1\mathbf{q}_3} & -1 \\ \frac{c}{\mathbf{p}_2\mathbf{q}_0} & 0 & 0 & -\frac{c}{\mathbf{p}_2\mathbf{q}_3} & -1 \\ -\frac{c}{\mathbf{p}_3\mathbf{q}_0} & \frac{c}{\mathbf{p}_3\mathbf{q}_1} & 0 & 0 & -1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \\
&= \det\begin{bmatrix} \frac{1}{\mathbf{p}_0} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\mathbf{p}_1} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\mathbf{p}_2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\mathbf{p}_3} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{c} \end{bmatrix} \det\begin{bmatrix} 0 & -1 & 1 & 0 & -\mathbf{p}_0 \\ 0 & 0 & -1 & 1 & -\mathbf{p}_1 \\ 1 & 0 & 0 & -1 & -\mathbf{p}_2 \\ -1 & 1 & 0 & 0 & -\mathbf{p}_3 \\ \mathbf{q}_0 & \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 & 0 \end{bmatrix} \det\begin{bmatrix} \frac{1}{\mathbf{q}_0} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\mathbf{q}_1} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\mathbf{q}_2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\mathbf{q}_3} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{c} \end{bmatrix} \\
&= c^3 \frac{(\mathbf{p}_0 + \mathbf{p}_1 + \mathbf{p}_2 + \mathbf{p}_3)(\mathbf{q}_0 + \mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3)}{\mathbf{p}_0\mathbf{p}_1\mathbf{p}_2\mathbf{p}_3\mathbf{q}_0\mathbf{q}_1\mathbf{q}_2\mathbf{q}_3} \\
&> 0,
\end{aligned} \tag{45}$$

and in general, we can write $\begin{bmatrix} R_{\mathcal{I}\mathcal{J}} & -\mathbf{1}_{|\mathcal{I}|} \\ \mathbf{1}_{|\mathcal{J}|}^\top & 0 \end{bmatrix}$ as the product of $\text{diag}\left(\frac{1}{\mathbf{p}_1}, \frac{1}{\mathbf{p}_2}, ..., \frac{1}{\mathbf{p}_k}, \frac{1}{c}\right)$, $\begin{bmatrix} R' & \mathbf{p} \\ \mathbf{q}^\top & 0 \end{bmatrix}$, and $\text{diag}\left(\frac{1}{\mathbf{q}_1}, \frac{1}{\mathbf{q}_2}, ..., \frac{1}{\mathbf{q}_k}, \frac{1}{c}\right)$, where R' is a matrix with entries,

$$R'_{ij} = \begin{cases} -1 & \text{if } j = (i+1) \mod k \\ 1 & \text{if } j = (i+2) \mod k \\ 0 & \text{otherwise} \end{cases}, \tag{46}$$

with the above examples provided for $k = 2, 3, 4$,

and the determinant is given by,

$$\begin{aligned}
&\det \begin{bmatrix} R_{\mathcal{I}\mathcal{J}} & -\mathbf{1}_{|\mathcal{I}|} \\ \mathbf{1}_{|\mathcal{J}|}^{\top} & 0 \end{bmatrix} \\
&= \det \ \text{diag} \left(\frac{1}{\mathbf{p}_1}, \frac{1}{\mathbf{p}_2}, ..., \frac{1}{\mathbf{p}_k}, \frac{1}{c} \right) \det \begin{bmatrix} R' & \mathbf{p} \\ \mathbf{q}^{\top} & 0 \end{bmatrix} \det \ \text{diag} \left(\frac{1}{\mathbf{q}_1}, \frac{1}{\mathbf{q}_2}, ..., \frac{1}{\mathbf{q}_k}, \frac{1}{c} \right) \\
&= c^{k-1} \frac{\sum_{i=1}^{k} \mathbf{p}_i \sum_{j=1}^{k} \mathbf{q}_j}{\prod_{i=1}^{k} \mathbf{p}_i \prod_{j=1}^{k} \mathbf{q}_j} \\
&> 0.
\end{aligned} \tag{47}$$

This verifies the INV condition and completes the proof. □

A.3. The Markov Game Modification Problem as An Optimization Problem

Here we instantiate the general Game Modification problem (Definition 1) to Markov games as an optimization problem.

Definition 6 (Game Modification for Two-Player Zero-Sum Markov Game). Given the cost function ℓ, the target policy $(\mathbf{p}, \mathbf{q})$ with supports $\mathcal{I}, \mathcal{J}$, target value range $[\underline{v}, \overline{v}]$, the game modification for Markov games can be written as the following optimization problem,

$$\begin{aligned}
&\inf_{R, v, \mathbb{Q}} \ \ell\left(R, R^{\circ}\right) \\
&\text{s.t.} \left[\mathbb{Q}_h\left(s\right)\right]_{\mathcal{I}_h(s)\bullet} \mathbf{q}_h\left(s\right) = v_h\left(s\right) \mathbf{1}_{|\mathcal{I}_h(s)|} \\
&\qquad \forall\ h \in [H], s \in \mathcal{S} \\
&\quad \mathbf{p}_h^{\top}\left(s\right) \left[\mathbb{Q}_h\left(s\right)\right]_{\bullet \mathcal{J}_h(s)} = v_h\left(s\right) \mathbf{1}_{|\mathcal{J}_h(s)|}^{\top} \\
&\qquad \forall\ h \in [H], s \in \mathcal{S} \\
&\quad \left[\mathbb{Q}_h\left(s\right)\right]_{\mathcal{A}_1 \setminus \mathcal{I}_h(s)\bullet} \mathbf{q}_h\left(s\right) < v_h\left(s\right) \mathbf{1}_{|\mathcal{A}_1 \setminus \mathcal{I}_h(s)|} \\
&\qquad \forall\ h \in [H], s \in \mathcal{S} \\
&\quad \mathbf{p}_h^{\top}\left(s\right) \left[\mathbb{Q}_h\left(s\right)\right]_{\bullet \mathcal{A}_2 \setminus \mathcal{J}_h(s)} > v_h\left(s\right) \mathbf{1}_{|\mathcal{A}_2 \setminus \mathcal{J}_h(s)|}^{\top} \\
&\qquad \forall\ h \in [H], s \in \mathcal{S} \\
&\quad \sigma_{\min}\left(\begin{bmatrix} \left[\mathbb{Q}_h\left(s\right)\right]_{\mathcal{I}_h(s)\mathcal{J}_h(s)} & -\mathbf{1}_{|\mathcal{I}_h(s)|} \\ \mathbf{1}_{|\mathcal{J}_h(s)|}^{\top} & 0 \end{bmatrix} \right) > 0 \\
&\qquad \forall\ h \in [H], s \in \mathcal{S} \\
&\quad \mathbb{Q}_h\left(s\right) = R_h\left(s\right) + \sum_{s' \in \mathcal{S}} P_h\left(s' | s\right) v_{h+1}\left(s'\right) \\
&\qquad \forall\ h \in [H-1], s \in \mathcal{S} \\
&\quad \mathbb{Q}_H\left(s\right) = R_H\left(s\right), \forall\ s \in \mathcal{S} \\
&\quad \underline{v} \leqslant \sum_{s \in \mathcal{S}} P_0\left(s\right) v_1\left(s\right) \leqslant \overline{v} \\
&\quad -b \leqslant \left[R_h\left(s\right)\right]_{ij} \leqslant b \\
&\qquad \forall\ (i,j) \in \mathcal{A}, h \in [H], s \in \mathcal{S}.
\end{aligned} \tag{48}$$

A.4. Proof of Theorem 3 and Theorem 6

Theorem 3 concerns the feasibility of modifying normal-form games in Proposition 1, and Theorem 6 concerns the feasibility of modifying H-period Markov games in Definition 6. Below we prove Theorem 6, from which Theorem 3 follows as an special case with $H = 1$.

Direction $\Rightarrow$. If $\pi = (\mathbf{p}, \mathbf{q})$ is the unique Nash in stage game in period $h \in [H]$, state $s \in \mathcal{S}$, then by Lemma 2, $\begin{bmatrix} R_{\mathcal{I}_h(s)\mathcal{J}_h(s)} & -\mathbf{1}_{\mathcal{I}_h(s)} \\ \mathbf{1}^{\top}_{\mathcal{J}_h(s)} & 0 \end{bmatrix}$ is an invertible square matrix, therefore, $|\mathcal{I}_h(s)| = |\mathcal{J}_h(s)|$.
Now to show that $(-Hb, Hb) \cap [\underline{v}, \overline{v}] =$ empty leads to infeasibility, note that either,

$$\overline{v} \geqslant Hb, \tag{49}$$

or,

$$\underline{v} \leqslant -Hb, \tag{50}$$

meaning the value of at least one stage game at least b or at most $-b$, and the SIISOW conditions imply that there are some entries of $R_h(s)$ that are strictly larger than b or strictly smaller than $-b$, which contradicts the reward bound conditions.

Direction $\Leftarrow$. Fix a stage game in period $h \in [H]$, state $s \in \mathcal{S}$, if $|\mathcal{I}_h(s)| = |\mathcal{J}_h(s)| = k$ for some k, then without loss of generality, we can rename the actions so that $\mathcal{I}_h(s) = \mathcal{J}_h(s) = \{0, 1, 2, ..., k-1\}$ and Lemma 4 provides a game with the unique Nash equilibrium $(\mathbf{p}_h(s), \mathbf{q}_h(s))$. Note that since the value of R^{eRPS} is 0, all stage games have value 0, so we have, for every $h \in [H], s \in \mathcal{S}$,

$$\mathbb{Q}_h(s) = R_h(s). \tag{51}$$

The $(-Hb, Hb) \cap [\underline{v}, \overline{v}] \neq \varnothing$ condition guarantees the existence of some $v^\star \in [\underline{v}, \overline{v}]$ that satisfies,

$$-Hb < v^\star < Hb. \tag{52}$$

Now consider the Markov game (R, P) with rewards defined by,

$$R_h(s) = \delta R^{\text{eRPS}(\mathbf{p}_h(s), \mathbf{q}_h(s))} + \frac{1}{H} v^\star. \tag{53}$$

This implies that the Q matrices can be computed as recursively for $h = H-1, H-2, ..., 1$,

$$\begin{aligned} v_h(s) &= \frac{H-h+1}{H} v^\star, \\ \mathbb{Q}_h(s) &= R_h(s) + \sum_{s' \in \mathcal{S}} P_h(s'|s) v_{h+1}(s) \\ &= R_h(s) + \sum_{s' \in \mathcal{S}} P_h(s'|s) \frac{H-h}{H} v^\star \\ &= R_h(s) + \frac{H-h}{H} v^\star \\ &= \delta R^{\text{eRPS}(\mathbf{p}_h(s), \mathbf{q}_h(s))} + \frac{H-h+1}{H} v^\star, \end{aligned} \tag{54}$$

which is an affine transformation of R^{eRPS}, so it has unique Nash $(\mathbf{p}_h(s), \mathbf{q}_h(s))$ with value $\frac{H-h+1}{H} v^\star$. In particular, the value of this game is given by,

$$\begin{aligned} v_0 &:= \sum_{s \in \mathcal{S}} P_0(s) v_1(s) \\ &= \sum_{s \in \mathcal{S}} P_0(s) \frac{H-1+1}{H} v^\star \\ &= v^\star, \end{aligned} \tag{55}$$

which satisfies the value range constraint. In addition, for δ sufficiently small, the entry bound conditions are satisfied as well. In particular, if $\delta < \min\{Hb - v^\star, Hb + v^\star\}$, for which the righthand side is strictly positive due to the condition $(-Hb, Hb) \cap [\underline{v}, \overline{v}] \neq \varnothing$, we have $R_h(s) \in [-b, b]$.

A.5. Proof of Feasibility/Optimality for RAP and RAP-MG Algorithms (Theorem 5 and Theorem 7)

Theorem 5 concerns the feasibility and optimality of the RAP algorithm for normal form games. This result is a special case of Theorem 7 below for the RAP-MG algorithm for Markov games.

Proof. We show the general result for H-period Markov games, and Theorem 5 is the special case when $H = 1$.

Existence. Existence of a solution is implied by Theorem 6 with value bounds $[-Hb + H\lambda, Hb - H\lambda]$, and due to (10), we have,

$$(-Hb + H\lambda, Hb - H\lambda) \cap [\underline{v}, \overline{v}] \neq \varnothing, \tag{56}$$

and therefore, Theorem 6 implies the feasible of the problem thus existence of a solution.

Feasibility. We only have to check the INV constraints since $\iota, \lambda > 0$ implies that the other constraints in the original problem are satisfied. We check that for every stage game $\mathbb{Q}$ in period $h \in [H], s \in \mathcal{S}$, we have $\mathbb{Q}_h(s) = \mathbb{Q}'_h(s) + \varepsilon R^{\text{eRPS}}(\mathbf{p}_h(s), \mathbf{q}_h(s))$ satisfies INV, where $\mathbb{Q}'_h(s)$ is the solution to the optimization. To simplify the notations, we drop the (h, s) indices.

We use the following properties of R^{eRPS} from the proof of Lemma 4,

$$\begin{aligned} R^{\text{eRPS}}_{\mathcal{I}\bullet}\mathbf{q} &= \mathbf{0}_{|\mathcal{I}|}, \\ \mathbf{p}^\top R^{\text{eRPS}}_{\bullet\mathcal{J}} &= \mathbf{0}_{|\mathcal{J}|}, \\ R^{\text{eRPS}}_{\mathcal{A}_1\setminus\mathcal{I}\bullet}\mathbf{q} &= -\mathbf{1}_{|\mathcal{A}_1\setminus\mathcal{I}|}, \\ \mathbf{p}^\top R^{\text{eRPS}}_{\bullet\mathcal{A}_2\setminus\mathcal{J}} &= \mathbf{1}_{|\mathcal{A}_2\setminus\mathcal{J}|}. \end{aligned} \tag{57}$$

Now we check the three conditions of the attacker's problem are satisfied. We have

$$\begin{aligned} \mathbb{Q}_{\mathcal{I}\bullet}\mathbf{q} &= \mathbb{Q}'_{\mathcal{I}}\mathbf{q} + \varepsilon R^{\text{eRPS}}_{\mathcal{I}\bullet}\mathbf{q} \\ &= v\mathbf{1}_{|\mathcal{I}|} \\ &= v'\mathbf{1}_{|\mathcal{I}|}, \end{aligned} \tag{58}$$

and similarly,

$$\begin{aligned} \mathbf{p}^\top\mathbb{Q}_{\bullet\mathcal{J}} &= \mathbf{p}^\top\mathbb{Q}'_{\bullet\mathcal{J}} + \varepsilon\mathbf{p}R^{\text{eRPS}}_{\bullet\mathcal{J}} \\ &= v\mathbf{1}_{|\mathcal{J}|} \\ &= v'\mathbf{1}_{|\mathcal{J}|}. \end{aligned} \tag{59}$$

We also have

$$\begin{aligned} \mathbb{Q}_{\mathcal{A}_1\setminus\mathcal{I}\bullet}\mathbf{q} &= \mathbb{Q}'_{\mathcal{A}_1\setminus\mathcal{I}\bullet}\mathbf{q} + \varepsilon R^{\text{eRPS}}_{\mathcal{A}_1\setminus\mathcal{I}}\mathbf{q} \\ &< v\mathbf{1}_{|\mathcal{A}_1\setminus\mathcal{I}|} - \varepsilon\mathbf{1}_{|\mathcal{A}_1\setminus\mathcal{I}|} \\ &< v'\mathbf{1}_{|\mathcal{A}_1\setminus\mathcal{I}|}. \end{aligned} \tag{60}$$

and similarly,

$$\begin{aligned} \mathbf{p}^\top\mathbb{Q}_{\bullet\mathcal{A}_2\setminus\mathcal{J}} &= \mathbf{p}^\top\mathbb{Q}'_{\bullet\mathcal{A}_2\setminus\mathcal{J}} + \varepsilon R^{\text{eRPS}}_{\mathcal{A}_2\setminus\mathcal{J}}\mathbf{q} \\ &> v\mathbf{1}_{|\mathcal{A}_2\setminus\mathcal{J}|} + \varepsilon\mathbf{1}_{|\mathcal{A}_2\setminus\mathcal{J}|} \\ &> v'\mathbf{1}_{|\mathcal{A}_2\setminus\mathcal{J}|}. \end{aligned} \tag{61}$$

Next we show that $\begin{bmatrix} \mathbb{Q}_{\mathcal{I}\mathcal{J}} & -\mathbf{1}_{|\mathcal{I}|} \\ \mathbf{1}^\top_{|\mathcal{J}|} & 0 \end{bmatrix}$ is invertible with probability 1, in particular, since $\begin{bmatrix} R^{\text{eRPS}} & -\mathbf{1}_{|\mathcal{I}|} \\ \mathbf{1}^\top_{|\mathcal{J}|} & 0 \end{bmatrix}$ is invertible by Lemma 4, we can write its singular value decomposition,

$$\begin{bmatrix} R^{\text{eRPS}} & -\mathbf{1}_{|\mathcal{I}|} \\ \mathbf{1}^\top_{|\mathcal{J}|} & 0 \end{bmatrix} = U\Sigma V^\top, \tag{62}$$

for some orthonormal $U, V \in \mathbb{R}^{(|\mathcal{I}|+1)\times(|\mathcal{J}|+1)}$ and nonsingular diagonal matrix $\Sigma \in \mathbb{R}^{(|\mathcal{I}|+1)\times(|\mathcal{J}|+1)}$. Consider the event

$\begin{bmatrix} \mathbb{Q}_{\mathcal{I}\mathcal{J}} & -\mathbf{1}_{|\mathcal{I}|} \\ \mathbf{1}_{|\mathcal{J}|}^\top & 0 \end{bmatrix}$ is singular. Then $\begin{bmatrix} \mathbb{Q}_{\mathcal{I}\mathcal{J}} & -(1+\varepsilon)\mathbf{1}_{|\mathcal{I}|} \\ (1+\varepsilon)\mathbf{1}_{|\mathcal{J}|}^\top & 0 \end{bmatrix}$ is singular, and the following matrix is also singular:

$$
\begin{aligned}
&\Sigma^{-1/2}U^\top \begin{bmatrix} \mathbb{Q}_{\mathcal{I}\mathcal{J}} & -(1+\varepsilon)\mathbf{1}_{|\mathcal{I}|} \\ (1+\varepsilon)\mathbf{1}_{|\mathcal{J}|}^\top & 0 \end{bmatrix} V\Sigma^{-1/2} \\
&= \Sigma^{-1/2}U^\top \begin{bmatrix} \mathbb{Q}'_{\mathcal{I}\mathcal{J}} + \varepsilon R^{\text{eRPS}} & -(1+\varepsilon)\mathbf{1}_{|\mathcal{I}|} \\ (1+\varepsilon)\mathbf{1}_{|\mathcal{J}|}^\top & 0 \end{bmatrix} V\Sigma^{-1/2} \\
&= \Sigma^{-1/2}U^\top \begin{bmatrix} \mathbb{Q}'_{\mathcal{I}\mathcal{J}} + \dfrac{\varepsilon'}{1-\varepsilon'} R^{\text{eRPS}} & -\dfrac{1}{1-\varepsilon'}\mathbf{1}_{|\mathcal{I}|} \\ \dfrac{1}{1-\varepsilon'}\mathbf{1}_{|\mathcal{J}|}^\top & 0 \end{bmatrix} V\Sigma^{-1/2} \\
&\qquad \text{where } \varepsilon' := \frac{\varepsilon}{1+\varepsilon} = 1 - \frac{1}{1+\varepsilon}, \\
&\qquad \text{which implies } \varepsilon = \frac{1}{1-\varepsilon'} - 1 = \frac{\varepsilon'}{1-\varepsilon'}, \\
&= \frac{1}{1-\varepsilon'}\Sigma^{-1/2}U^\top \begin{bmatrix} (1-\varepsilon')\,\mathbb{Q}'_{\mathcal{I}\mathcal{J}} + \varepsilon' R^{\text{eRPS}} & -\mathbf{1}_{|\mathcal{I}|} \\ \mathbf{1}_{|\mathcal{J}|}^\top & 0 \end{bmatrix} V\Sigma^{-1/2} \\
&= \Sigma^{-1/2}U^\top \begin{bmatrix} \mathbb{Q}'_{\mathcal{I}\mathcal{J}} & -\mathbf{1}_{|\mathcal{I}|} \\ \mathbf{1}_{|\mathcal{J}|}^\top & 0 \end{bmatrix} V\Sigma^{-1/2} \\
&\quad + \frac{\varepsilon'}{1-\varepsilon'}\Sigma^{-1/2}U^\top \begin{bmatrix} R^{\text{eRPS}}_{\mathcal{I}\mathcal{J}} & -\mathbf{1}_{|\mathcal{I}|} \\ \mathbf{1}_{|\mathcal{J}|}^\top & 0 \end{bmatrix} V\Sigma^{-1/2} \\
&= \Sigma^{-1/2}U^\top \begin{bmatrix} \mathbb{Q}'_{\mathcal{I}\mathcal{J}} & -\mathbf{1}_{|\mathcal{I}|} \\ \mathbf{1}_{|\mathcal{J}|}^\top & 0 \end{bmatrix} V\Sigma^{-1/2} + \varepsilon I.
\end{aligned}
\tag{63}
$$

Consequently, there exists a nonzero vector $x \in \mathbb{R}^{|\mathcal{I}|+1} = \mathbb{R}^{|\mathcal{J}|+1}$ such that,

$$
\Sigma^{-1/2}U^\top \begin{bmatrix} \mathbb{Q}'_{\mathcal{I}\mathcal{J}} & -\mathbf{1}_{|\mathcal{I}|} \\ \mathbf{1}_{|\mathcal{J}|}^\top & 0 \end{bmatrix} V\Sigma^{-1/2} x = -\varepsilon x. \tag{64}
$$

This means that $-\varepsilon$ is an eigenvalue of the following deterministic matrix,

$$
\Sigma^{-1/2}U^\top \begin{bmatrix} \mathbb{Q}'_{\mathcal{I}\mathcal{J}} & -\mathbf{1}_{|\mathcal{I}|} \\ \mathbf{1}_{|\mathcal{J}|}^\top & 0 \end{bmatrix} V\Sigma^{-1/2}, \tag{65}
$$

which happens with probability 0 since $\varepsilon \sim$ Unif $[-\lambda, \lambda]$ is continuous.

Optimality. Fix $\varepsilon > 0$. Consider a feasible solution to (48), $\left(R^{(\varepsilon)}, v^{(\varepsilon)}\right)$, that satisfies

$$
\ell\left(R^{(\varepsilon)}, R^\circ\right) - C^\star < \frac{\varepsilon}{2}. \tag{66}
$$

In particular, feasibility of $R^{(\varepsilon)}$ implies, for every $h \in [H], s \in \mathcal{S}$,

$$
\begin{aligned}
\left[\mathbb{Q}_h^{(\varepsilon)}(s)\right]_{\mathcal{I}\bullet}\mathbf{q} &= v_h^{(\varepsilon)}(s)\,\mathbf{1}_{|\mathcal{I}|} \\
\mathbf{p}^\top\left[\mathbb{Q}_h^{(\varepsilon)}(s)\right]_{\bullet\mathcal{J}} &= v_h^{(\varepsilon)}(s)\,\mathbf{1}_{|\mathcal{J}|}^\top \\
\left[\mathbb{Q}_h^{(\varepsilon)}(s)\right]_{\mathcal{A}_1\setminus\mathcal{I}\bullet}\mathbf{q} &< v_h^{(\varepsilon)}(s)\,\mathbf{1}_{|\mathcal{A}_1\setminus\mathcal{I}|} \\
\mathbf{p}^\top\left[\mathbb{Q}_h^{(\varepsilon)}(s)\right]_{\bullet\mathcal{A}_2\setminus\mathcal{J}} &> v_h^{(\varepsilon)}(s)\,\mathbf{1}_{|\mathcal{A}_2\setminus\mathcal{J}|} \\
\sigma_{\min}\left(\begin{bmatrix} \mathbb{Q}_h^{(\varepsilon)}(s) & \mathcal{I}\mathcal{J} \\ -\mathbf{1}_{|\mathcal{I}|} & \mathbf{1}_{|\mathcal{J}|} \end{bmatrix} 0\right) &> 0 \\
\mathbb{Q}_h^{(\varepsilon)}(s) &= R_h^{(\varepsilon)}(s) + \sum_{s'\in\mathcal{S}} P_h\left(s'|s\right) v_{h+1}^{(\varepsilon)}\left(s'\right) \\
-b \leqslant \left[R_h^{(\varepsilon)}(s)\right]_{ij} &\leqslant b, \forall\ (i,j) \in \mathcal{A}.
\end{aligned}
\tag{67}
$$

Due to the strict SOW inequality in (48), we can find the $\iota^{(\varepsilon)} > 0$ such that the SOW conditions in (9) is also satisfied,

$$\iota^{(\varepsilon)} := \min_{h\in[H],s\in\mathcal{S}} \left\{ v_h^{(\varepsilon)}(s)\,\mathbf{1}_{|\mathcal{A}_1\setminus\mathcal{I}|} - \left[\mathbb{Q}_h^{(\varepsilon)}(s)\right]_{\mathcal{A}_1\setminus\mathcal{I}\bullet}\mathbf{q}, \mathbf{p}^\top \left[\mathbb{Q}_h^{(\varepsilon)}(s)\right]_{\bullet\mathcal{A}_2\setminus\mathcal{J}} - v_h^{(\varepsilon)}(s)\,\mathbf{1}_{|\mathcal{A}_2\setminus\mathcal{J}|} \right\}, \tag{68}$$

where the $\min$ is element-wise for the vectors.

Since $v^{(\varepsilon)} \in (-Hb, Hb)$, we can find the value gap $\lambda^{(\varepsilon)} > 0$,

$$\lambda^{(\varepsilon)} := b - \min_{h\in[H],s\in\mathcal{S},(i,j)\in\mathcal{A}} \left| v_h(s) - P_{ij}\left(s'|s\right) v_{h+1}\left(s'\right)\right|, \tag{69}$$

by noting that if $\lambda^{(\varepsilon)} = 0$, then $\left|v^{(\varepsilon)}\right| \geqslant Hb$ which contradicts our assumption.

Now we define the following δ,

$$\delta := \min\left\{ \frac{\iota^{(\varepsilon)}}{2}, \frac{\varepsilon\lambda^{(\varepsilon)}}{2LbH\left(H+1\right)|\mathcal{S}|\,|\mathcal{A}|} \right\}. \tag{70}$$

Note that $R^{(\varepsilon)}$ does not satisfy (9) due the tighter bounds on the entries, meaning $-b+\lambda \leqslant \left[R_h^{(\varepsilon)}(s)\right]_{ij} \leqslant b-\lambda$ may not be satisfied for some $h\in[H], s\in\mathcal{S}, (i,j)\in\mathcal{A}$. We define $R'^{(\varepsilon)}$ as follows and show that $\left(R'^{(\varepsilon)}, v^{(\varepsilon)}\right)$ is feasible to (9), for every $h\in[H], s\in\mathcal{S}, (i,j)\in\mathcal{A}$,

$$\left[R_h'^{(\varepsilon)}(s)\right]_{ij} := \begin{cases} \left(1-\dfrac{\delta}{\lambda^{(\varepsilon)}}\right)\left[\mathbb{Q}_h^{(\varepsilon)}(s)\right]_{ij} + \dfrac{v_h^{(\varepsilon)}(s)\,\delta}{\lambda^{(\varepsilon)}} - \displaystyle\sum_{s'\in\mathcal{S}}\left[P_h\left(s'|s\right)\right]_{ij} v_{h+1}^{(\varepsilon)}\left(s'\right) & \text{if } i\in\mathcal{I}_h(s), j\in\mathcal{J}_h(s) \\ \min\left\{\max\left\{\left[R_h^{(\varepsilon)}(s)\right]_{ij}, -b+\delta\right\}, b-\delta\right\} & \text{otherwise} \end{cases}. \tag{71}$$

In particular, we have for $i\in\mathcal{I}_h(s), j\in\mathcal{J}_h(s)$,

$$\left[\mathbb{Q}_h'^{(\varepsilon)}(s)\right]_{ij} = \left(1-\frac{\delta}{\lambda^{(\varepsilon)}}\right)\left[\mathbb{Q}_h^{(\varepsilon)}(s)\right]_{ij} + \frac{v_h^{(\varepsilon)}(s)\,\delta}{\lambda^{(\varepsilon)}}. \tag{72}$$

Now, to check the feasibility of $\left(R'^{(\varepsilon)}, v^{(\varepsilon)}\right)$ to (9), fix $h\in[H], s\in\mathcal{S}$. To simplify the notations, we drop the (h,s) indices. Observe that

$$\begin{aligned} \mathbb{Q}_{\mathcal{I}\bullet}'^{(\varepsilon)}\mathbf{q} &= \left(\left(1-\frac{\delta}{\lambda^{(\varepsilon)}}\right)\mathbb{Q}_{\mathcal{I}\bullet}^{(\varepsilon)} + \frac{v^{(\varepsilon)}\delta}{\lambda^{(\varepsilon)}}\right)\mathbf{q}, \text{ since } \mathbf{q}_{\mathcal{A}_2\setminus\mathcal{J}} = \mathbf{0}_{|\mathcal{A}_2\setminus\mathcal{J}|} \\ &= \left(1-\frac{\delta}{\lambda^{(\varepsilon)}}\right)\mathbb{Q}_{\mathcal{I}\bullet}^{(\varepsilon)}\mathbf{q} + \frac{v^{(\varepsilon)}\delta}{\lambda^{(\varepsilon)}}\mathbf{1}_{\mathcal{I}\mathcal{J}}\mathbf{q} \\ &= \left(1-\frac{\delta}{\lambda^{(\varepsilon)}}\right)v^{(\varepsilon)}\mathbf{1}_{|\mathcal{I}|} + \frac{v^{(\varepsilon)}\delta}{\lambda^{(\varepsilon)}}\mathbf{1}_{\mathcal{I}\mathcal{J}}\mathbf{q}, \text{ since } \left(R^{(\varepsilon)}, v^{(\varepsilon)}\right) \text{ is feasible} \\ &= \left(1-\frac{\delta}{\lambda^{(\varepsilon)}}\right)v^{(\varepsilon)} + \frac{v^{(\varepsilon)}\delta}{\lambda^{(\varepsilon)}} \\ &= v^{(\varepsilon)}, \end{aligned} \tag{73}$$

and similarly,

$$\begin{aligned} \mathbf{p}^\top\mathbb{Q}_{\bullet\mathcal{J}}'^{(\varepsilon)} &= \mathbf{p}^\top\left(\left(1-\frac{\delta}{\lambda^{(\varepsilon)}}\right)\mathbb{Q}_{\bullet\mathcal{J}}^{(\varepsilon)} + \frac{v^{(\varepsilon)}\delta}{\lambda^{(\varepsilon)}}\right) \\ &= \left(1-\frac{\delta}{\lambda^{(\varepsilon)}}\right)v^{(\varepsilon)} + \frac{v^{(\varepsilon)}\delta}{\lambda^{(\varepsilon)}} \\ &= v^{(\varepsilon)}. \end{aligned} \tag{74}$$

Consider any $\iota < \delta$, we have,

$$\begin{aligned}
\mathbb{Q}'^{(\varepsilon)}_{\mathcal{A}_1\setminus\mathcal{I}\bullet}\mathbf{q} &\leqslant \left(\mathbb{Q}^{(\varepsilon)}_{\mathcal{A}_1\setminus\mathcal{I}\bullet} + \frac{\iota^{(\varepsilon)}}{2}\right)\mathbf{q} \\
&\leqslant \mathbb{Q}^{(\varepsilon)}_{\mathcal{A}_1\setminus\mathcal{I}\bullet}\mathbf{q} + \frac{\iota^{(\varepsilon)}}{2}\mathbf{1}_{|\mathcal{A}_1\setminus\mathcal{I}|} \\
&\leqslant \left(v^{(\varepsilon)} - \iota^{(\varepsilon)}\right)\mathbf{1}_{|\mathcal{A}_1\setminus\mathcal{I}|} + \frac{\iota^{(\varepsilon)}}{2}\mathbf{1}_{|\mathcal{A}_1\setminus\mathcal{I}|} \\
&\leqslant \left(v^{(\varepsilon)} - \frac{\iota^{(\varepsilon)}}{2}\right)\mathbf{1}_{|\mathcal{A}_1\setminus\mathcal{I}|} \\
&\leqslant \left(v^{(\varepsilon)} - \delta\right)\mathbf{1}_{|\mathcal{A}_1\setminus\mathcal{I}|} \\
&\leqslant \left(v^{(\varepsilon)} - \iota\right)\mathbf{1}_{|\mathcal{A}_1\setminus\mathcal{I}|},
\end{aligned} \tag{75}$$

and similarly,

$$\begin{aligned}
\mathbf{p}^\top\mathbb{Q}'^{(\varepsilon)}_{\bullet\mathcal{A}_2\setminus\mathcal{J}} &\geqslant \mathbf{p}^\top\left(\mathbb{Q}^{(\varepsilon)}_{\bullet\mathcal{A}_2\setminus\mathcal{J}} - \frac{\iota^{(\varepsilon)}}{2}\right) \\
&\geqslant \left(v^{(\varepsilon)} + \iota^{(\varepsilon)}\right)\mathbf{1}_{|\mathcal{A}_2\setminus\mathcal{J}|} - \frac{\iota^{(\varepsilon)}}{2}\mathbf{1}_{|\mathcal{A}_2\setminus\mathcal{J}|} \\
&\geqslant \left(v^{(\varepsilon)} + \frac{\iota^{(\varepsilon)}}{2}\right)\mathbf{1}_{|\mathcal{A}_2\setminus\mathcal{J}|} \\
&\geqslant \left(v^{(\varepsilon)} + \iota\right)\mathbf{1}_{|\mathcal{A}_2\setminus\mathcal{J}|}.
\end{aligned} \tag{76}$$

Now to show that $\begin{bmatrix}\mathbb{Q}'^{(\varepsilon)}_{\mathcal{I}\mathcal{J}} & -\mathbf{1}_{|\mathcal{J}|} \\ \mathbf{1}^\top_{|\mathcal{I}|} & 0\end{bmatrix}$ is invertible, since $\begin{bmatrix}\mathbb{Q}^{(\varepsilon)}_{\mathcal{I}\mathcal{J}} & -\mathbf{1}_{|\mathcal{J}|} \\ \mathbf{1}^\top_{|\mathcal{I}|} & 0\end{bmatrix}$ is invertible, there exists vector $\begin{bmatrix}x \\ t\end{bmatrix} \neq \mathbf{0}_{|\mathcal{J}|+1}$, such that

$$\begin{aligned}
&\begin{bmatrix}\mathbb{Q}^{(\varepsilon)}_{\mathcal{I}\mathcal{J}} & -\mathbf{1}_{|\mathcal{J}|} \\ \mathbf{1}^\top_{|\mathcal{I}|} & 0\end{bmatrix}\begin{bmatrix}x \\ t\end{bmatrix} = \mathbf{0}_{|\mathcal{J}|+1} \\
\Rightarrow &\begin{cases}\mathbb{Q}^{(\varepsilon)}_{\mathcal{I}\mathcal{J}}x - t\mathbf{1}_{|\mathcal{J}|} = \mathbf{0}_{|\mathcal{J}|} \\ \mathbf{1}^\top_{|\mathcal{I}|}x = 0\end{cases} \\
\Rightarrow &\begin{cases}\left(1 - \frac{\delta}{\lambda^{(\varepsilon)}}\right)\mathbb{Q}^{(\varepsilon)}_{\mathcal{I}\mathcal{J}}x - \left(1 - \frac{\delta}{\lambda^{(\varepsilon)}}\right)t\mathbf{1}_{|\mathcal{J}|} = \mathbf{0}_{|\mathcal{J}|} \\ \mathbf{1}^\top_{|\mathcal{I}|}x = 0\end{cases} \\
\Rightarrow &\begin{cases}\left(1 - \frac{\delta}{\lambda^{(\varepsilon)}}\right)\mathbb{Q}^{(\varepsilon)}_{\mathcal{I}\mathcal{J}}x + \frac{v^{(\varepsilon)}\delta}{\lambda^{(\varepsilon)}}\mathbf{1}_{\mathcal{I}\mathcal{J}}x - \left(1 - \frac{\delta}{\lambda^{(\varepsilon)}}\right)t\mathbf{1}_{|\mathcal{J}|} = \mathbf{0}_{|\mathcal{J}|} & \text{, since } \mathbf{1}^\top_{|\mathcal{I}|}x = 0 \\ \mathbf{1}^\top_{|\mathcal{I}|}x = 0\end{cases} \\
\Rightarrow &\begin{cases}\left(\left(1 - \frac{\delta}{\lambda^{(\varepsilon)}}\right)\mathbb{Q}^{(\varepsilon)}_{\mathcal{I}\mathcal{J}} + \frac{v^{(\varepsilon)}\delta}{\lambda^{(\varepsilon)}}\right)x - \left(1 - \frac{\delta}{\lambda^{(\varepsilon)}}\right)t\mathbf{1}_{|\mathcal{J}|} = \mathbf{0}_{|\mathcal{J}|} \\ \mathbf{1}^\top_{|\mathcal{I}|}x = 0\end{cases} \\
\Rightarrow &\begin{cases}\mathbb{Q}'^{(\varepsilon)}_{\mathcal{I}\mathcal{J}}x - \left(1 - \frac{\delta}{\lambda^{(\varepsilon)}}\right)t\mathbf{1}_{|\mathcal{J}|} = \mathbf{0}_{|\mathcal{J}|} \\ \mathbf{1}^\top_{|\mathcal{I}|}x = 0\end{cases} \\
\Rightarrow &\begin{bmatrix}\mathbb{Q}'^{(\varepsilon)}_{\mathcal{I}\mathcal{J}} & -\mathbf{1}_{|\mathcal{J}|} \\ \mathbf{1}^\top_{|\mathcal{I}|} & 0\end{bmatrix}\begin{bmatrix}x \\ \left(1 - \frac{\delta}{\lambda^{(\varepsilon)}}\right)t\end{bmatrix} = \mathbf{0}_{|\mathcal{J}|+1}.
\end{aligned} \tag{77}$$

Since $\begin{bmatrix}x \\ \left(1 - \frac{\delta}{\lambda^{(\varepsilon)}}\right)t\end{bmatrix} \neq \mathbf{0}_{|\mathcal{J}|+1}$, we have $\begin{bmatrix}\mathbb{Q}'^{(\varepsilon)}_{\mathcal{I}\mathcal{J}} & -\mathbf{1}_{|\mathcal{J}|} \\ \mathbf{1}^\top_{|\mathcal{I}|} & 0\end{bmatrix}$ is invertible.

Since we did not change the value $v^{(\varepsilon)}$, the value range constraint is still satisfied,

$$\underline{v} \leqslant v^{(\varepsilon)} \leqslant \overline{v}. \tag{78}$$

For the range condition, we use the short-hand notation,

$$\Delta_{ij} v_h^{(\varepsilon)}(s) := v_h^{(\varepsilon)}(s) - \sum_{s' \in \mathcal{S}} \left[P_h\left(s'|s\right)\right]_{ij} v_{h+1}^{(\varepsilon)}\left(s'\right). \tag{79}$$

Note that we have,

$$\begin{aligned}
R_h'^{(\varepsilon)}(s) &= \mathbb{Q}_h'^{(\varepsilon)}(s) - \sum_{s' \in \mathcal{S}} P_h\left(s'|s\right) v_{h+1}^{(\varepsilon)}\left(s'\right) \\
&= \left(1 - \frac{\delta}{\lambda^{(\varepsilon)}}\right) \mathbb{Q}_h^{(\varepsilon)}(s) + v_h^{(\varepsilon)}(s) \frac{\delta}{\lambda^{(\varepsilon)}} - \sum_{s' \in \mathcal{S}} P_h\left(s'|s\right) v_{h+1}^{(\varepsilon)}\left(s'\right) \\
&= \left(1 - \frac{\delta}{\lambda^{(\varepsilon)}}\right) \left(R_h^{(\varepsilon)}(s) + \sum_{s' \in \mathcal{S}} P v_{h+1}^{(\varepsilon)}\left(s'\right)\right) + v_h^{(\varepsilon)}(s) \frac{\delta}{\lambda^{(\varepsilon)}} - \sum_{s' \in \mathcal{S}} P_h\left(s'|s\right) v_{h+1}^{(\varepsilon)}\left(s'\right) \\
&= \left(1 - \frac{\delta}{\lambda^{(\varepsilon)}}\right) R^{(\varepsilon)} + \left(v_h^{(\varepsilon)}(s) - \sum_{s' \in \mathcal{S}} P_h\left(s'|s\right) v_{h+1}^{(\varepsilon)}\left(s'\right)\right) \frac{\delta}{\lambda^{(\varepsilon)}} \\
&= \left(1 - \frac{\delta}{\lambda^{(\varepsilon)}}\right) R^{(\varepsilon)} + \Delta v_h^{(\varepsilon)}(s) \frac{\delta}{\lambda^{(\varepsilon)}},
\end{aligned} \tag{80}$$

where we drop the indices (h, s) as before. Now for any $\lambda < \delta$, we have, for every $i \in \mathcal{I}, j \in \mathcal{J}$,

$$\begin{aligned}
& -b \leqslant R_{ij}^{(\varepsilon)} \leqslant b \\
& \Rightarrow \left(1 - \frac{\delta}{\lambda^{(\varepsilon)}}\right)(-b) + \frac{\delta}{\lambda^{(\varepsilon)}} \Delta_{ij} v^{(\varepsilon)} \leqslant \left(1 - \frac{\delta}{\lambda^{(\varepsilon)}}\right) R_{ij}^{(\varepsilon)} + \frac{\delta}{\lambda^{(\varepsilon)}} \Delta_{ij} v^{(\varepsilon)} \leqslant \left(1 - \frac{\delta}{\lambda^{(\varepsilon)}}\right) b + \frac{\delta}{\lambda^{(\varepsilon)}} \Delta_{ij} v^{(\varepsilon)} \\
& \Rightarrow -b + \delta \frac{b + \Delta_{ij} v^{(\varepsilon)}}{\lambda^{(\varepsilon)}} \leqslant R_{ij}'^{(\varepsilon)} \leqslant b - \delta \frac{b - \Delta_{ij} v^{(\varepsilon)}}{\lambda^{(\varepsilon)}} \\
& \Rightarrow -b + \delta \frac{b + \Delta_{ij} v^{(\varepsilon)}}{b - \min\limits_{i'j'} \left|\Delta_{i'j'} v^{(\varepsilon)}\right|} \leqslant R_{ij}'^{(\varepsilon)} \leqslant b - \delta \frac{b - \Delta_{ij} v^{(\varepsilon)}}{b - \min\limits_{i'j'} \left|\Delta_{i'j'} v^{(\varepsilon)}\right|} \\
& \Rightarrow -b + \delta \leqslant R_{ij}'^{(\varepsilon)} \leqslant b - \delta, \text{ since } b + \Delta_{ij} v^{(\varepsilon)} \geqslant b - \min_{i'j'} \left|\Delta_{i'j'} v^{(\varepsilon)}\right| \geqslant b - \Delta_{ij} v^{(\varepsilon)}, \\
& \Rightarrow -b + \lambda \leqslant R_{ij}'^{(\varepsilon)} \leqslant b - \lambda,
\end{aligned} \tag{81}$$

and for any other $(i, j) \in \mathcal{A}$,

$$\begin{aligned}
& -b + \delta \leqslant \min\left\{\max\left\{R_{ij}^{(\varepsilon)}, -b + \delta\right\}, b - \delta\right\} \leqslant b - \delta \\
& \Rightarrow -b + \delta \leqslant R_{ij}'^{(\varepsilon)} \leqslant b - \delta \\
& \Rightarrow -b + \lambda \leqslant R_{ij}'^{(\varepsilon)} \leqslant b - \lambda.
\end{aligned} \tag{82}$$

In addition, we show that each entry changes by less than $\frac{\varepsilon}{2LH|\mathcal{S}||\mathcal{A}|}$, for $i \in \mathcal{I}, j \in \mathcal{J}$. In particular, we have

$$\begin{aligned}
&\left|R_{ij}'^{(\varepsilon)} - R_{ij}^{(\varepsilon)}\right| \\
&\leqslant \left|\mathbb{Q}_{ij}'^{(\varepsilon)} - \mathbb{Q}_{ij}^{(\varepsilon)}\right| \\
&\leqslant \left|\left(1 - \frac{\delta}{\lambda^{(\varepsilon)}}\right)\mathbb{Q}_{ij}^{(\varepsilon)} + \frac{v^{(\varepsilon)}\delta}{\lambda^{(\varepsilon)}} - \mathbb{Q}_{ij}^{(\varepsilon)}\right| \\
&= \left|-\frac{\delta}{\lambda^{(\varepsilon)}}\mathbb{Q}_{ij}^{(\varepsilon)} + \frac{v^{(\varepsilon)}\delta}{\lambda^{(\varepsilon)}}\right| \\
&\leqslant \left|\frac{\delta}{\lambda^{(\varepsilon)}}\mathbb{Q}_{ij}^{(\varepsilon)}\right| + \left|\frac{v^{(\varepsilon)}\delta}{\lambda^{(\varepsilon)}}\right| \\
&\leqslant \left|\frac{bH\delta}{\lambda^{(\varepsilon)}}\right| + \left|\frac{b\delta}{\lambda^{(\varepsilon)}}\right| \\
&\leqslant \frac{(H+1)\,b}{\lambda^{(\varepsilon)}}\frac{\varepsilon}{LH\,(H+1)\,|\mathcal{S}|\,|\mathcal{A}|}\frac{1}{2}\frac{\lambda^{(\varepsilon)}}{b}, \text{ due to the definition of } \delta, \\
&= \frac{\varepsilon}{2LH\,|\mathcal{S}|\,|\mathcal{A}|},
\end{aligned} \tag{83}$$

and for other $(i,j) \in \mathcal{A}$,

$$\begin{aligned}
&\left|R_{ij}'^{(\varepsilon)} - R_{ij}^{(\varepsilon)}\right| \\
&\leqslant \left|\min\left\{\max\left\{R_{ij}^{(\varepsilon)}, -b\right\}, b\right\} - R_{ij}^{(\varepsilon)}\right| \\
&\leqslant \delta \\
&\leqslant \frac{\varepsilon\lambda^{(\varepsilon)}}{2LbH\,(H+1)\,|\mathcal{S}|\,|\mathcal{A}|} \\
&\leqslant \frac{\varepsilon}{2LH\,|\mathcal{S}|\,|\mathcal{A}|}, \text{ since } \lambda^{(\varepsilon)} \leqslant b.
\end{aligned} \tag{84}$$

Therefore we have,

$$\begin{aligned}
\ell\left(R^\star\right) - C^\star &\leqslant \ell\left(R'^{(\varepsilon)}\right) - C^\star \\
&\leqslant \ell\left(R'^{(\varepsilon)} - R^{(\varepsilon)} + R^{(\varepsilon)}\right) - C^\star \\
&\leqslant \ell\left(R^{(\varepsilon)}\right) - C^\star + L\left\|R'^{(\varepsilon)} - R^{(\varepsilon)}\right\|_1 \\
&\leqslant \frac{\varepsilon}{2} - C^\star + LH\,|\mathcal{S}|\,|\mathcal{A}|\,\frac{\varepsilon}{2LH\,|\mathcal{S}|\,|\mathcal{A}|} \\
&\leqslant \frac{\varepsilon}{2} + L\frac{\varepsilon}{2L} \\
&= \varepsilon,
\end{aligned} \tag{85}$$

which concludes the proof.

Optimality Gap. To obtain the result in the linear case, we note that if the cost function is linear, (9) (restated below) is a linear program (since it does not have the invertibility constraint),

$$\begin{aligned}
&\min_{R,v,\mathbb{Q}} \ \ell\left(R, R^\circ\right) \\
&\quad \text{s.t. } \left[\mathbb{Q}_h\left(s\right)\right]_{\mathcal{I}_h(s)\bullet}\mathbf{q}_h\left(s\right) = v_h\left(s\right)\mathbf{1}_{|\mathcal{I}_h(s)|}, \forall\, h \in [H], s \in \mathcal{S} \\
&\quad\ \ \mathbf{p}_h^\top\left(s\right)\left[\mathbb{Q}_h\left(s\right)\right]_{\bullet\mathcal{J}_h(s)} = v_h\left(s\right)\mathbf{1}_{|\mathcal{J}_h(s)|}^\top, \forall\, h \in [H], s \in \mathcal{S} \\
&\quad\ \ \left[\mathbb{Q}_h\left(s\right)\right]_{\mathcal{A}_1\setminus\mathcal{I}_h(s)\bullet}\mathbf{q}_h\left(s\right) \leqslant \left(v_h\left(s\right) - \iota\right)\mathbf{1}_{|\mathcal{A}_1\setminus\mathcal{I}_h(s)|}, \forall\, h \in [H], s \in \mathcal{S} \\
&\quad\ \ \mathbf{p}_h^\top\left(s\right)\left[\mathbb{Q}_h\left(s\right)\right]_{\bullet\mathcal{A}_2\setminus\mathcal{J}_h(s)} \geqslant \left(v_h\left(s\right) + \iota\right)\mathbf{1}_{|\mathcal{A}_2\setminus\mathcal{J}_h(s)|}^\top, \forall\, h \in [H], s \in \mathcal{S}
\end{aligned} \tag{86}$$

$$
\begin{aligned}
&\mathbb{Q}_h(s) = R_h(s) + \sum_{s' \in \mathcal{S}} P_h\left(s'|s\right) v_{h+1}\left(s'\right), \forall\ h \in [H-1], s \in \mathcal{S} \\
&\mathbb{Q}_H(s) = R_H(s), \forall\ s \in \mathcal{S} \\
&\underline{v} \leqslant \sum_{s \in \mathcal{S}} P_0(s)\, v_1(s) \leqslant \overline{v} \\
&-b + \lambda \leqslant [R_h(s)]_{ij} \leqslant b - \lambda, \forall\ (i,j) \in \mathcal{A}, h \in [H], s \in \mathcal{S}
\end{aligned}
$$

which we can rewrite it in the standard form for the case when $\iota = \lambda = 0$,

$$
\begin{aligned}
\min_{x(R,v,\mathbb{Q})}\ & \ell(R, R^\circ) \\
& Ax = \boldsymbol{b}, \\
& x \geqslant 0,
\end{aligned}
\tag{87}
$$

and since ι and λ enters the constraint through $\boldsymbol{b}$ linearly, we can write the problem for $\theta = \max\{\iota, \lambda\}$,

$$
\begin{aligned}
\min_{x(R,v,\mathbb{Q})}\ & \ell(R, R^\circ) \\
& Ax = \boldsymbol{b}' := \boldsymbol{b} + \theta \boldsymbol{d}, \\
& x \geqslant 0,
\end{aligned}
\tag{88}
$$

for some fixed vector $\boldsymbol{d}$. By (Bertsimas & Tsitsiklis, 1997), in particular, equation (5.2) in section 5.2, assuming the optimal solution to (9) is always finite for every ι, λ satisfying (10), which is true due to our previous feasibility proof and the fact that the costs are bounded by $bH|\mathcal{S}||\mathcal{A}|$, we have that the optimal solution can be written as a finite collection of linear functions in the form,

$$
\ell(R, R^\circ; \boldsymbol{b}) = \max_{i \in [N]} y_i^\top \boldsymbol{b},
\tag{89}
$$

where y_i is the dual optimal solution in a region where $\ell(R, R^\circ; \boldsymbol{b})$ is linear, and we have,

$$
\begin{aligned}
\ell\left(R, R^\circ; \boldsymbol{b}'\right) &= \ell(R, R^\circ; \boldsymbol{b} + \theta \boldsymbol{d}) \\
&= \max_{i \in [N]} y_i^\top (\boldsymbol{b} + \theta \boldsymbol{d}) \\
&= \ell(R, R^\circ; \boldsymbol{b}) + \theta \max_{i \in [N]} y_i^\top \boldsymbol{d} \\
&= \ell(R, R^\circ; \boldsymbol{b}) + O(\theta).
\end{aligned}
\tag{90}
$$

When $\iota = \lambda = 0$, the problem is a relaxation of (1), thus we have that the optimal solution to (9), denoted by R', satisfies,

$$
\ell\left(R', R^\circ; \boldsymbol{b}\right) \leqslant C^\star,
\tag{91}
$$

and due to our previous feasibility proof, for ι, λ satisfying (10),

$$
\ell\left(R', R^\circ; \boldsymbol{b}'\right) \geqslant C^\star,
\tag{92}
$$

and combined with the previous result,

$$
\begin{aligned}
\ell\left(R', R^\circ; \boldsymbol{b}'\right) &= \ell\left(R', R^\circ; \boldsymbol{b}\right) + O(\theta) \\
&\leqslant C^\star + O(\theta),
\end{aligned}
\tag{93}
$$

we have,

$$
\begin{aligned}
\ell(R(\iota, \lambda), R^\circ) &= C^\star + O(\theta). \\
&= C^\star + O(\max\{\iota, \lambda\}).
\end{aligned}
\tag{94}
$$

□

A.6. Additional Experiments

Code Details. We conducted our experiments using standard python3 libraries and the gurobi optimization package. We provide our code in a jupyter notebook with an associated database folder so that our experiments can be easily reproduced. The notebook already reads in the database by default so no file management is needed. Simply ensure the notebook is in the same directory as the database folder (like we have arranged in our uploaded zip). We note that for our benchmark tests, the database was too large to upload directly. Instead we will upload that database on github. However, the scale experiments can be reproduced by using the generation code we included in the notebook. Our code is available at: https://github.com/YoungWu559/game-modification.

Classic Two-finger Morra. Consider the classic Two-finger Morra game. The game's payoff matrix is described in (95). Note that this game is different from the simplified two-finger morra game considered in the main text.

$$TFM := \begin{pmatrix} 0 & 2 & -3 & 0 \\ -2 & 0 & 0 & 3 \\ 3 & 0 & 0 & -4 \\ 0 & -3 & 4 & 0 \end{pmatrix} \tag{95}$$

TFW has infinitely many NEs: each player's strategy can be any convex combination of $(0, 4/7, 3/7, 0)^\top$ and $(0, 3/5, 2/5, 0)^\top$. Since people often naively use uniform mixing, it may be desirable to derive a similar game where uniform mixing is NE. Applying Algorithm 1 with $p = q = (1/4, 1/4, 1/4, 1/4)^\top$ produces the new payoff matrix (96).

$$TFM^\dagger := \begin{pmatrix} 0 & 2 & -3 & 0 \\ -2 & 0 & -2 & 3 \\ 3 & 0 & 0 & -4 \\ -2 & -3 & 4 & 0 \end{pmatrix} \tag{96}$$

Observe that $TFW^\dagger$ is an unfair game with value $-.25$, unlike the original game whose value was 0. The total cost for the change was 4.

5-action RPSSL. Consider the generalization of the rock-paper-scissors (RPS) game where each player now has 5 strategies rock, paper, scissors, spock, and lizard (RPSSL) that we mentioned in the main text. The game's payoff matrix is described in (97). Note that this game is different from the 5-action Rock-Paper-Scissor-Fire-Water (RPSFW) game considered in the main text.

$$RPSSL := \begin{pmatrix} 0 & -1 & 1 & -1 & 1 \\ 1 & 0 & -1 & 1 & -1 \\ -1 & 1 & 0 & -1 & 1 \\ 1 & -1 & 1 & 0 & -1 \\ -1 & 1 & -1 & 1 & 0 \end{pmatrix} \tag{97}$$

Similar to RPS, the unique NE for RPSSL is the uniformly mixed strategy pair $\mathbf{p} = \mathbf{q} = (1/5, 1/5, 1/5, 1/5, 1/5)^\top$. Suppose that instead, we wish to skew the distribution to favor the new actions, spock and lizard. Specifically, if $\mathbf{p} = \mathbf{q} = (1/9, 1/9, 1/9, 1/3, 1/3)^\top$, running Algorithm 1 produces the new payoff matrix (98).

$$RPSSL^\dagger := \begin{pmatrix} 0 & -1 & 1 & -1 & 1 \\ 1 & 0 & -1 & 1 & -1 \\ -1 & 1 & 0 & -1 & 1 \\ 1 & -1 & 1 & 0 & -1/3 \\ -1 & 1 & -1 & 1/3 & 0 \end{pmatrix} \tag{98}$$

We observe the resultant NE is fair with value 0. The total cost for the change is 1.33.

Note on Other Cost Functions For general cost functions, one may use Frank-Wolfe-type algorithms, which call a linear programming (LP) oracle in each iteration. Faster specialized solvers can be used if the cost has additional structures. For example, if the cost function is representable by an SDP or other conic programs, one can use interior methods as

implemented in MOSEK or other libraries. If the cost function is quadratic (e.g., squared Euclidean norm), one can use a quadratic program (QP) solver based on the QP simplex methods, such as those implemented in Gurobi. If the cost function is piecewise linear (e.g., L1 norm), one can use an LP solver such as Gurobi or GLPK.