Notes on "Playing Repeated Coopetitive Polymatrix Games with Small Manipulation Cost"

Zhijian Gao, Guanda Chen

June 2024

1 Introduction

The paper looks at the following problem: in a multi-player no-communication coopetitive game in which players are supposed to not just maximise their payoff, but also to win, traditional MAS based approaches may not work. This is because when communication isn't feasible, emerging cooperative behaviours based on observation is important, and that doesn't just involve finding the right partner, but also finding the right time to betray. To solve this problem, the paper considers 3-player game without communications in which a player can manipulate the underlying payoff matrix with a cost, with others satisfying stronger and stronger assumptions. The contents are

- show that there exist a set of dominance solvable policies that can guarantee the win for the manipulator, and that they can be found in polynomial time
- improve the results by batch coordination policies that can provably guarantee low manipulation cost, and prove that they can be found in polynomial time
- add more objectives, like winning with maximum margin, or socially good income, etc.
- show with example that the total manipulation cost in concrete game is small

This can be seen as a mechanism design problem, in which the matrix chosen by the manipulator is the mechanism, and the actions of the participants are information reported, similar to an online MD framework in which participants don't depart or arrive. However, the goal is to guide players into playing specific strategies rather than standard solution concepts such as incentive compatibility or social welfare for MD problems.

2 Preliminaries

Define $\Gamma = (\mathcal{N}, \mathcal{A}, u)$, where $\mathcal{N} = \{1, 2, 3\}$ denoting the players, $\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2 \times \mathcal{A}_3$ denoting the set of all combinations of actions that can be taken by the three players. Actions must be chosen simultaneously by the players from the corresponding sets, and randomising is allowed, i.e. player i may select any probability distribution s over $\mathcal{A}_i$ and sample actions randomly according to the distribution. If a strategy makes up of deterministic choices at each point, i.e. no randomness is involved what so ever, then we call it a **pure strategy**, otherwise we call it a **mixed strategy**.

Player 1 will be the manipulator, whose strategy chosen is denoted by probability vector x, and x_i and $x(a_i)$ representing the probability that action a_i is selected. Similarly, y, z represents the strategy chosen by players 2 and 3.

After strategies have been selected, each player i receives a reward given by the utility function $u_i : \mathcal{A}_i \to \mathbb{R}, u_i = \Sigma_{i \neq j} u_{ij}$. Here, $u_{ij} : \mathcal{A}_i \times \mathcal{A}_j \to \mathbb{R}$ describes the payoff gained by player i from interacting with player j, and corresponds to a payoff matrix $A^{(i,j)}$, i.e. $u_{ij}(a) = A^{(i,j)}(a_i, a_j)$, with a_i, a_j being the actions actually chosen by players i and j.

We consider the the combinations of actions as random variables in the probability space $(\mathcal{A}, \mathcal{F}, \mathbb{P})$, with $\mathcal{F} = \mathcal{P}(\mathcal{A})$ and $\mathbb{P}[(a_1, a_2, a_3)] = x(a_1)y(a_2)z(a_3)$. This makes the rewards of each player random variables too.

Player 1, as the manipulator, can alter the payoff matrices $A^{(2,1)}$ and $A^{(3,1)}$ with cost $\Sigma_{(i,j)\in P} ||A^{(i,j)} - A_0^{(i,j)}||_\infty$, where $P = \{(2, 1), (3, 1)\}$ Therefore, player 1's complete strategy will be the joint submission of a strategy and alternated payoff matrices, i.e. tuple $(x, (A^{(i,j)})_{(i,j)\in P})$. Here we denote A_0 the original payoff matrices, and $A_0^{(2,3)}, A_0^{(3,2)}$ can't be altered by player 1.

Hence, the expected utility for the manipulator is

$$\mathbb{E}[u_1(a)] = y^T A^{(1,2)} x + z^T A^{(1,3)} x - \Sigma_{(i,j)\in P} ||A^{(i,j)} - A_0^{(i,j)}||_\infty$$

which can be proved using definition of expectation of function of discrete random variable. Similarly, the expected utility of players 2 and 3 are

$$\mathbb{E}[u_2(a)] = x^T A^{(2,1)} y + z^T A^{(2,3)} y$$

and

$$\mathbb{E}[u_3(a)] = x^T A^{(3,1)} z + y^T A^{(3,2)} z$$

respectively.

3 Problem Setting

We repeat the polymatrix game for T time steps, and for each time step t, manipulator can manipulate $A_t^{(i,j)}$ for $(i, j) \in P$, and player 2, 3 gets feedback $u_{i,t}(x_t, y_t, z_t)$. Player 2 and 3 don't have access to A_0, but have access of history of feedback. Player 1, however, has full knowledge of A_0.

The histories observed by player 1 up to time step t are denoted by $H_t = (u_{1,t'}, x_{t'})_{t'=1}^{t}$, and $\mathcal{H}_t$ is the set of all histories of length t.

Also, the utility of each player now is $U_i = \frac{1}{T}\Sigma_{t=1}^{T} u_i(x_t, y_t, z_t)$. The manipulator wins if U_1 is the largest. Player 1 aims to win, and players 2 and 3 are assumed to be "consistent agents".

Definition 1 (consistent agent): Suppose a^* is the best action for all rounds, and the number of rounds in which the agent chooses a^* in T rounds is T^*. If $\mathbb{P}[lim_{T\to\infty} \frac{T^*}{T} = 1] = 1$, then the agent is a consistent agent. If such an action doesn't exist, then we don't make any assumption on the behaviour of this agent.

4 Winning Policies

First consider the situation in which player 1 chooses i^* every time, and players 2 and 3 have strictly dominant actions j^* and k^* respectively. The goal for player 1 is to alter the payoff matrices such that

$$U_1(e_{i^*}, e_{j^*}, e_{k^*}) \geq max\{U_2(e_{i^*}, e_{j^*}, e_{k^*}), U_3(e_{i^*}, e_{j^*}, e_{k^*})\}.$$

Here, e_m denotes the case in which the player definitely chooses its mth action, and therefore when represented by the probability vector, all entries except for the mth entry must be zero, with the mth entry being 1, and therefore gives e_m.

In this case, player 1 will win with sufficiently large T and small manipulation cost.

Then consider a new case, in which player 3 has a strict dominant action against player 1 and 2, i.e. player 3 is willing to wait for some action to eventually become the single optimal choice. Here we call player 3 the persistent player.

4.1 Designing Dominance Solvable Games

Here we show how to alter payoff matrices such that under the second setting mentioned above, j^* and k^* are indeed the actions that P2 and P3 will take.

Set $|\mathcal{A}_1| = n, |\mathcal{A}_2| = m, |\mathcal{A}_3| = l$, and let the fixed strategy played by P1 be x.

P1 must alter $A^{(2,1)}$ s.t. for some $v_2 \in \mathbb{R}^l$ and $\forall k \in [l]$

$$x^T A^{(2,1)} e_j + e_j^T A^{(2,3)} e_k = v_{2k} \text{ if } j = j^*$$

and

$$x^T A^{(2,1)} e_j + e_j^T A^{(2,3)} e_k < v_{2k} \text{ if } j \neq j^* \qquad (1)$$

and similarly, alter alter $A^{(3,1)}$ s.t. for some $v_3 \in \mathbb{R}^m$ and $\forall j \in [m]$

$$x^T A^{(3,1)} e_k + e_j^T A^{(3,2)} e_k = v_{3k} \text{ if } k = k^*$$

and

$$x^T A^{(3,1)} e_k + e_j^T A^{(3,2)} e_k < v_{3k} \text{ if } k \neq k^* \qquad (2)$$

Furthermore, assume now that the fixed strategy played by P2 is y. To ensure that k^* is strictly dominant against both x and y, P1 must alter $A^{(2,1)}$ (**shouldn't it be $A^{(3,1)}$?**) s.t. for some $v_0 \in \mathbb{R}$

$$x^T A^{(3,1)} e_k + y^T A^{(3,2)} e_k = v_0 \text{ if } k = k^*$$

and

$$x^T A^{(3,1)} e_k + y^T A^{(3,2)} e_k < v_0 \text{ if } k \neq k^* \qquad (3)$$

Proposition 1: Payoff matrices $A^{(2,1)}, A^{(3,1)}$ that satisfies both (1) and (2) exists.

Proof: Consider

$$A^{(2,1)}_{i^*j} = \begin{cases} 2||A_0^{(2,3)}||_\infty + 1 & \text{if } j = j^* \\ 0 & \text{otherwise} \end{cases}$$

and

$$A^{(3,1)}_{i^*k} = \begin{cases} 2||A_0^{(3,2)}||_\infty + 1 & \text{if } k = k^* \\ 0 & \text{otherwise} \end{cases}$$

Note that $||A||_\infty$ equals the maximum absolute row sum of the matrix. Then, $\forall k \in [l]$ and $j \neq j^*$,

$$(x^T A^{(2,1)} e_{j^*} + e_{j^*}^T A^{(2,3)} e_k) - (x^T A^{(2,1)} e_j + e_j^T A^{(2,3)} e_k)$$

$$= (\Sigma_{p \in [n], p \neq i^*} x_p (A^{(2,1)}_{pj^*} - A^{(2,1)}_{pj})) + (2||A^{(2,3)}||_\infty + 1) + (A^{(2,3)}_{j^*k} - A^{(2,3)}_{jk})$$

$$> 0$$

(**Why?**)

If (2) is satisfied, then (3) will also be satisfied.

4.2 Dominance Solvable Policies

Definition 2 (dominance solvable type-1 policies): Let $A^{(2,1)}, A^{(3,1)}$ satisfy (1) and (2). Then the policy $\rho_t(H_t) = (e_{i^*}, A^{(2,1)}, A^{(3,1)})$ is a dominance solvable type-1 policy at time step t.

Definition 3 (dominance solvable type-2 policies): Let $A^{(2,1)}, A^{(3,1)}$ satisfy (1) and (3). Then the policy $\rho_t(H_t) = (e_{i^*}, A^{(2,1)}, A^{(3,1)})$ is a dominance solvable type-2 policy at time step t.

Definition (dominance solvable game): If one uses iterated elimination of strictly dominated strategies and there is only one strategy left for each player, the game is called dominance solvable

We say that a dominance solvable policy is winning if player 1 wins the corresponding single-shot game when $(e_{i^*}, e_{j^*}, e_{k^*})$ is played, i.e.

$$u_1(e_{i^*}, e_{j^*}, e_{k^*}) \geq u_2(e_{i^*}, e_{j^*}, e_{k^*}) \text{ and } u_1(e_{i^*}, e_{j^*}, e_{k^*}) \geq u_3(e_{i^*}, e_{j^*}, e_{k^*}) \qquad (4)$$

Theorem 1: If P1 uses dominance solvable type-1 policy against consistent agents in infinitely repeating game, then

$$\mathbb{P}[U_1(x_t, y_t, z_t)_{t=1}^{\infty} \geq U_2(x_t, y_t, z_t)_{t=1}^{\infty} \text{ and } U_1(x_t, y_t, z_t)_{t=1}^{\infty} \geq U_3(x_t, y_t, z_t)_{t=1}^{\infty}] = 1$$

This can be easily proved by substituting x_t, y_t, z_t with $e_{i^*}, e_{j^*}, e_{k^*}$ for all time steps, which is allowed since P2 and P3 are consistent agent.

Definition 4 (persistent agent): The agent who is consistent on the eventually best action once it is revealed.

It's important to note that all persistent agent are also consistent agent, and therefore all dominance solvable type-2 policies are also dominance solvable type-1 policies. In both cases, the opposite implication doesn't work.

Theorem 2: If P1 uses dominance solvable type-2 policy against consistent P2 and persistent P3 in infinitely repeating game, then

$$\mathbb{P}[U_1(x_t, y_t, z_t)_{t=1}^{\infty} \geq U_2(x_t, y_t, z_t)_{t=1}^{\infty} \text{ and } U_1(x_t, y_t, z_t)_{t=1}^{\infty} \geq U_3(x_t, y_t, z_t)_{t=1}^{\infty}] = 1$$

Theorem 3: If winning dominance solvable policies exist, then there exists an algorithm that can find such policies with running time that is polynomial in the number of actions of the players.

This is because the problem of finding such a policy can be converted to the problem of solving the linear systems (1) and (2) or (1) and (3) restricted to (4), and this problem is solvable in linear time

4.3 Batch Coordination Policies

Dominance winning policies may exist, but the it may be very costly to alter the payoff matrices. To maximise total utility, P1 may also choose to beat one player by cheap alteration to the payoff matrix and lose to the other, and at the right time, change the strategy to beat the other one and lose to the one it had been beating with another cheap alteration to the corresponding payoff matrix.

Definition 5 (winning batch coordination policy): Suppose matrices $\hat{A}^{(2,1)}$ and $\hat{A}^{(3,1)}$ satisfy systems (1) and (2) for some $i_2 \in [n], j_2 \in [m], k_2 \in [l]$ and that the matrices $\tilde{A}^{(2,1)}$ and $\tilde{A}^{(3,1)}$ satisfy systems (1) and (2) for some $i_3 \in [n], j_3 \in [m], k_3 \in [l]$ such that for $i \neq 1$,

$$\mathbb{E}[u_1(e_{i_2}, e_{j_2}, e_{k_2})] + \mathbb{E}[u_1(e_{i_3}, e_{j_3}, e_{k_3})] > \mathbb{E}[u_i(e_{i_3}, e_{j_3}, e_{k_3})] + \mathbb{E}[u_i(e_{i_3}, e_{j_3}, e_{k_3})]$$

(Shouldn't it be i_2, j_2, k_2 and i_3, j_3, k_3 on the RHS?)
Then the policy

$$\rho_t = \begin{cases} (e_{i_1}, \hat{A}^{(2,1)}, \hat{A}^{(3,1)}) & \text{if } 1 \leq t \leq \frac{T}{2} \\ (e_{i_2}, \hat{A}^{(2,1)}, \hat{A}^{(3,1)}) & \text{if } \frac{T}{2} \leq t \leq T \end{cases}$$

(Shouldn't it be i_2 and i_3?)
is called a winning batch coordination policy, i.e., choosing different "local" dominance solvable policies for each half of the game. They may not be "globally" winning by themselves.

Definition 6 (regret): The regret of any sequence of strategies $(y_1, \ldots, y_T)$ chosen by P2 wrt a fixed strategy y is

$$R_{T,y} = \Sigma_{t=1}^{T} x_t A_t^{(2,1)} y_t + y_t^T A_0^{(2,3)} z_t - \Sigma_{t=1}^{T} x_t^T A_t^{(2,1)} y + y^T A_0^{(2,3)} z_t$$

i.e., the difference between the payoff accumulated by $(y_1, \ldots, y_T)$ and the payoff accumulated by $(y, \ldots, y)$.

Definition (no-regret player): A player is called no-regret if $lim_{T \to \infty} max_y \frac{R_{T,y}}{T} = 0$.

Theorem 4: If the manipulator uses a winning batch coordination policy against no-regret players then

$$\mathbb{P}[U_1(x_t, y_t, z_t)_{t=1}^{T} \geq U_2(x_t, y_t, z_t)_{t=1}^{T} \text{ and } U_1(x_t, y_t, z_t)_{t=1}^{T} \geq U_3(x_t, y_t, z_t)_{t=1}^{T} \text{ eventually}] = 1$$

(Why do we need eventually here?)

Theorem 5: If winning batch coordination policies exist, then there exists an algorithm that can find such policies with running time that is polynomial in the number of actions of the players.

Proof of these two theorems are analogous to the cases in section 4.2.

Proposition 2: All persistent players are consistent. Furthermore, all no-regret players are persistent.

Proposition 3: If an agent uses the Follow the Leader algorithm, then it's persistent but not no-regret.

5 Additional Objectives

Adding additional objectives can be done by adding a corresponding linear objective function and turn the problem into a sequence of LP problems.

5.1 Winning by the Largest Margin

Definition 7 (margin of policy): The margin of a policy $\rho_t(H_t) = (x_t, (A_t^{(i,j)})_{(i,j)\in P})$ for $t \in \mathbb{N}$ when playing against player 2 and player 3's no-regret sequence of strategies $(y_t)_{t=1}^{\infty}$ and $(y_t)_{t=1}^{\infty}$ is defined to be

$$min\{\mathbb{E}[U_1(x_t, y_t, z_t)_{t=1}^{\infty} - U_2(x_t, y_t, z_t)_{t=1}^{\infty}], \mathbb{E}[U_1(x_t, y_t, z_t)_{t=1}^{\infty} - U_3(x_t, y_t, z_t)_{t=1}^{\infty}]\}$$

Theorem 6: If winning dominance solvable policies exist, then there exists an algorithm that can find the largest margin dominance solvable policy, with running time that is polynomial in the number of actions of the players.

5.2 Winning with the Lowest Inefficiency Ratio

Definition 8 (efficiency ratio of a policy): The efficiency ratio of a policy $\rho_t(H_t) = (x_t, (A_t^{(i,j)})_{(i,j)\in P})$ for $t \in \mathbb{N}$ when playing against player 2 and player 3's no-regret sequence of strategies $(y_t)_{t=1}^{\infty}$ and $(y_t)_{t=1}^{\infty}$ is defined to be

$$\frac{lim_{T\to\infty}\frac{1}{T}\Sigma_{t=1}^{T}\Sigma_{(i,j)\in P}||A_t^{(i,j)} - A_0^{(i,j)}||_{\infty}}{\mathbb{E}[lim_{T\to\infty}\frac{1}{T}\Sigma_{t=1}^{T}(x_t^T A_t^{(1,2)} y_t + x_t^T A_t^{(1,3)} z_t)] - K}$$

where $K = min_{i,j,k}(A^{(1,2)(i,j)} + A^{(1,3)}(j,k))$ (Why not $A^{(1,3)}(i,k)$) is the minimum revenue for P1.

Theorem 7: If winning dominance solvable policies exist, then there exists an algorithm that can find the winning dominance solvable policy with the lowest inefficiency ratio, with running time that is polynomial in the number of actions of the players.

5.3 Maximising the Egalitarian Social Welfare

Definition 9 (egalitarian social welfare): The egalitarian social welfare of a strategy profile (x, y, z) is defined to be

$$S(x, y, z) = min\{U_1(x, y, z), U_2(x, y, z), U_3(x, y, z)\}$$

Theorem 8: There exists an algorithm that can find the dominance solvable policy that maximizes egalitarian social welfare with running time that is polynomial in the number of actions of the players.

6 Global Optimum Manipulation Cost

Objective function:

$$min_{s,(|t_u|)_{u=1}^s}\frac{1}{T}\Sigma_{u=1}^s(|t_u|\Sigma_{j=2}^n||A_{t_u}^{(1,j)} - A_0^{(1,j)}||_\infty)$$

Here $(|t_u|)_{u=1}^s$ is a partition of $[T]$ of size s.

7 Game Setting for n Players

7.1 Version 1

$\Gamma = (\mathcal{N}, \mathcal{A}, u)$, where $\mathcal{N} = (1, \dots, n), \mathcal{A} = \mathcal{A}_1 \times \dots \times \mathcal{A}_n$.

The strategy of player i is denoted by vector $x_i \in [0,1]^{|\mathcal{A}_i|}$, and $\mathbb{P}[a] = \Pi_{i=1}^n x_i(a_i)$

$\forall i \in [n], a = (a_1, \dots, a_n), \in \mathcal{A}, u_i(a) = \Sigma_{j \neq i} u_{ij}(a) = \Sigma_{j \neq i} A^{(i,j)}(a_i, a_j)$

$\mathbb{E}[u_1(a)] = \Sigma_{j=2}^n(x_j^T A^{(1,j)} x_1 - ||A^{(1,j)} - A_0(1,j)||_\infty)$

$\mathbb{E}[u_i(a)] = \Sigma_{j \neq i} x_j^T A^{(i,j)} x_i \forall i > 1$

$\forall j > 1$, we must choose $A^{(j,1)}$ such that for some $v_j \in \mathbb{R}^{|M_j|}$, where $M_j = \{(l_k)_{k \neq 1, k \neq j} : l_k \in \mathcal{A}_k\}$, we have

$$x_1^T A^{(j,1)} e_{l_j} + \Sigma_{k \neq 1, k \neq j} e_{l_k}^T A^{(j,k)} e_{l_j} = v_{j,m} \forall m \in [|M_j|] \text{ and } l_j = l_j^*$$

$$x_1^T A^{(j,1)} e_{l_j} + \Sigma_{k \neq 1, k \neq j} e_{l_k}^T A^{(j,k)} e_{l_j} < v_{j,m} \forall m \in [|M_j|] \text{ and } l_j \neq l_j^*$$

where l_j^* is the unique best action in all rounds for player j against the rest, and $m \in [|M_j|]$ is the index of the mth largest vector $(l_k) \in M_j$ when they are interpreted as $n-2$-digit integers.
If Player j is persistent, then the problem will be find matrix $A^{(j,1)}$ such that for some $v^{(j)} \in \mathbb{R}$ we have

$$\Sigma_{k \neq j} x_k^T A^{(j,k)} e_{l_j} = v_j \forall l_j = l_j^*$$

$$\Sigma_{k \neq j} x_k^T A^{(j,k)} e_{l_j} < v_j \forall l_j \neq l_j^*$$

7.2 Version 2 with objective function

$\Gamma = (\mathcal{N}, \mathcal{A}, u)$, where $\mathcal{N} = (1, \dots, n), \mathcal{A} = \mathcal{A}_1 \times \cdots \times \mathcal{A}_n$.

The strategy of player i is denoted by vector $x_i \in [0,1]^{|\mathcal{A}_i|}$, and $\mathbb{P}[a] = \Pi_{i=1}^n x_i(a_i)$

$\forall i \in [n], a = (a_1, \dots, a_n), \in \mathcal{A}, u_i(a) = \Sigma_{j \neq i} u_{ij}(a) = \Sigma_{j \neq i} A^{(i,j)}(a_i, a_j)$

$\mathbb{E}[u_1(a)] = \Sigma_{j=2}^n (x_j^T A^{(1,j)} x_1 - ||A^{(1,j)} - A_0(1,j)||_\infty)$

$\mathbb{E}[u_i(a)] = \Sigma_{j \neq i} x_j^T A^{(i,j)} x_i \forall i > 1$

$\forall j > 1$, we must choose $A^{(j,1)}$ such that for some $v_j \in \mathbb{R}^{|M_j|}$, where $M_j = \{(l_k)_{k \neq 1, k \neq j} : l_k \in \mathcal{A}_k\}$, we have

$$x_1^T A^{(j,1)} e_{l_j} + \Sigma_{k \neq 1, k \neq j} e_{l_k}^T A^{(j,k)} e_{l_j} = v_{j,m} \forall m \in [|M_j|] \text{ and } l_j = l_j^*$$

$$x_1^T A^{(j,1)} e_{l_j} + \Sigma_{k \neq 1, k \neq j} e_{l_k}^T A^{(j,k)} e_{l_j} < v_{j,m} \forall m \in [|M_j|] \text{ and } l_j \neq l_j^* \qquad (1)$$

where l_j^* is the unique best action in all rounds for player j against the rest, and $m \in [|M_j|]$ is the index of the mth largest vector $(l_k) \in M_j$ when they are interpreted as $n-2$-digit integers.

If Player j is persistent, then the problem will be find matrix $A^{(j,1)}$ such that for some $v^{(j)} \in \mathbb{R}$ we have

$$\Sigma_{k \neq j} x_k^T A^{(j,k)} e_{l_j} = v_j \forall l_j = l_j^*$$

$$\Sigma_{k \neq j} x_k^T A^{(j,k)} e_{l_j} < v_j \forall l_j \neq l_j^* \qquad (2)$$

Above are the settings for a single shot game. We will repeat it for T time steps.

Objective function:

$$min_{s,(|t_u|)_{u=1}^s} \frac{1}{T} \Sigma_{u=1}^s (|t_u| \Sigma_{j=2}^n ||A_{t_u}^{(1,j)} - A_0^{(1,j)}||_\infty)$$

subject to

$$\Sigma_{u=1}^s \mathbb{E}[u_1(l_1^{(u)}, \dots, l_n^{(u)})] > \Sigma_{u=1}^s \mathbb{E}[u_j(l_1^{(u)}, \dots, l_n^{(u)})]$$

and (1)(or (2) if the corresponding agent is persistent) $\forall j > 1$.

Here $(t_u)_{u=1}^s$ is a partition of $[T]$ of size s, and $(|t_u|)_{u=1}^s$ is the sequence of the lengths of the partitioned parts.

Explanation for the objective function: The objective is to win in long term with the smallest total cost for manipulation. From the fact that batch coordination policies can guarantee victory with lower manipulation cost, it's probably possible to win with even smaller manipulation cost if we break it into more "periods". However, the exact number s and the exact lengths $(|t_u|)_{u=1}^{s}$ of the periods can't be pre-determined. Maybe $s = n - 1$, which will be the analogous version of the batch coordination case for n players, or maybe it's larger. Therefore we treat them as variables that we minimise the cost with respect to. i,e they both changes, and the goal is to find the best s and $(|t_u|)_{u=1}^{s}$ that minimise the total manipulation cost.

The first restraint makes sure that player 1 wins in long term. It's analogous to the restraint on winning batch coordination policy. The second restraint makes sure that for each partitioned period, the player we want to beat has a single best action.

This is, I believe, a likely-over-complex version of objective function, because it's very likely that batch coordination policy on n players will give us the optimal cost. However, since there's a chance it won't and that even if it does, this objective function can get that optimal cost, I believe this over-complexity is meaningful.

8 add on

Theorem 1: If player 1 (P1) uses a dominance solvable type-1 policy against $n - 1$ consistent agents in an infinitely repeating game, then

$$\mathbb{P}\left[U_1(x_t, y_t, \ldots, z_t)_{t=1}^{\infty} \geq \max_{i=2}^{n} U_i(x_t, y_t, \ldots, z_t)_{t=1}^{\infty}\right] = 1$$

Proof: This can be easily proved by substituting $x_t, y_t, \ldots, z_t$ with $e_{i^*}, e_{j^*}, \ldots, e_{k^*}$ for all time steps, which is allowed since players P_2 to P_n are consistent agents.

Theorem 2: If player 1 (P1) uses a dominance solvable type-2 policy against $n - 2$ consistent agents and one persistent agent in an infinitely repeating game, then

$$\mathbb{P}\left[U_1(x_t, y_t, \ldots, z_t)_{t=1}^{\infty} \geq \max_{i=2}^{n} U_i(x_t, y_t, \ldots, z_t)_{t=1}^{\infty}\right] = 1$$

Theorem 3: If winning dominance solvable policies exist for n players, then there exists an algorithm that can find such policies with running time that is polynomial in the number of actions of the players.
Since manipulating the matrix could be operated using polynomial time algorithm such as gaussian elimination, and the size of the linear system is proportional to the number of actions and interactions ensuring tractability.

Question: Is it possible to find winning dominance solvable policies in logn time?
We need to make sure that there exists an l_j^* such that for all combination of

actions, as long as the action of player j is not that, it will have lower utility. To satisfy (1) we need to solve a sequence of $(|\mathcal{A}_j| - 1)|M_j|$ inequalities, which can be solved in polynomial time.
Since we definitely need to find $n-1$ matrices, the best we can hope for would be linear time, which still seems very unlikely. Therefore, we should maybe try to find a solution that is polynomial time but faster then the original one.

9 Random Ideas

1. Can we maybe add a special agent called the "supervisor"? The supervisor does not directly join the game, but at each time step, each player can pay the supervisor to gain the ability to manipulate the payoff matrices in the next round (or next few rounds), but only the highest bidder will get that ability, and the amount each player pays is unknown to other players (seal-bidding game). The supervisor's utility, i.e. the total amount paid to it by the highest bidders of each time step, will be taken into account at the end, and therefore it may win eventually without considering any strategies.

2. We may add behavioral modelling to our system, such as bounded rationality, risk-aversion, evolutionary strategies. Analyze how these would influence manipulator's ability.

3. We could discuss applications in artificial intelligence and robotics, where autonomous agents must cooperate or compete to achieve specific goals. We might also explore how the manipulator's strategies can guide agent behavior and optimize collective outcomes.