Open Problem: Optimal Instance-Dependent Sample Complexity for finding Nash Equilibrium in Two Player Zero-Sum Matrix games

Arnab Maiti ARNABM2@UW.EDU
University of Washington

Editors: Nika Haghtalab and Ankur Moitra

Abstract

Optimal instance-dependent sample complexity is a well-studied topic in the multi-armed bandit literature. However, the analogous question in the setting of two-player zero-sum matrix games, where the payoff matrix can only be accessed through noisy samples, remains largely unexplored despite being a natural generalization of the multi-armed bandit problem. In this write-up, we pose a simple open question: *What is the optimal instance-dependent sample complexity to find an approximate Nash equilibrium in two-player zero-sum matrix games?*

1. Introduction

In single-player stochastic settings, such as multi-armed bandits and reinforcement learning, instance-dependent sample complexity bounds for identifying a good policy are well studied (Jamieson et al., 2014; Wagenmaker et al., 2022). In contrast, in multi-player stochastic settings, such as two-player zero-sum matrix games, such bounds remain relatively unexplored. Given the wide range of applications of two-player zero-sum games in machine learning, artificial intelligence, optimization, and game theory, where noise naturally arises, there is strong motivation to understand their instance-dependent statistical complexity. In the following section, we formalize this direction by posing a concrete open problem.

2. Problem Setting and Open Problem

We begin by formalizing the problem setting. A two-player zero-sum matrix game is defined by an input payoff matrix $A \in \mathbb{R}^{n\times m}$. In each round of the game, the row player selects a row i and the column player selects a column j; the row player receives a reward of $A_{i,j}$, and the column player receives $-A_{i,j}$.

To analyze strategic behavior in this setting, we consider the concept of a Nash equilibrium. Let $\triangle_n$ denote the n-dimensional probability simplex. A pair $(x_\star, y_\star) \in \triangle_n \times \triangle_m$ is said to be a Nash equilibrium of the game defined by A if and only if

$$\langle x, Ay_\star\rangle \le \langle x_\star, Ay_\star\rangle \le \langle x_\star, Ay\rangle,$$

for all $x \in \triangle_n$ and $y \in \triangle_m$.

A natural and well-studied relaxation of the Nash equilibrium is the ε-approximate Nash equilibrium. A pair $(x_\star, y_\star) \in \triangle_n \times \triangle_m$ is said to be an ε-approximate Nash equilibrium of A if and only if

$$\langle x, Ay_\star\rangle \le \langle x_\star, Ay_\star\rangle + \varepsilon \quad \text{and} \quad \langle x_\star, Ay\rangle \ge \langle x_\star, Ay_\star\rangle - \varepsilon,$$

for all $x \in \triangle_n$ and $y \in \triangle_m$.

In this work, we focus on a noisy observation setting, where the payoff matrix A is initially unknown. At each round, one can sample an entry (i, j) and observe a noisy value of the form $A_{i,j} + \eta_t$, where η_t is zero-mean, 1-subgaussian noise. The goal is to identify an ε-approximate Nash equilibrium, with probability at least $1 - \delta$, using as few samples as possible.

A standard approach is to simulate a repeated game where the row and column players run no-regret learning algorithms like EXP3-IX, which leads to a sample complexity of $\mathcal{O}(n \log(n/\delta)/\varepsilon^2)$. While this is nearly minimax-optimal even for the multi-armed bandit setting, which is a special case, instance-dependent sample complexity bounds are known to significantly improve performance in the multi-armed bandit setting.

For example, in the best-arm identification problem with K arms and means $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_K$, the near-optimal instance-dependent sample complexity for identifying an ε-best arm is

$$\mathcal{O}\left(\sum_{i=2}^{K} \frac{\log(1/\delta)}{\Delta_i^2}\right) + \widetilde{\mathcal{O}}\left(\sum_{i=2}^{K} \frac{1}{\Delta_i^2}\right),$$

where $\Delta_i := \max\{\varepsilon, \mu_1 - \mu_i\}$ and $\widetilde{\mathcal{O}}(\cdot)$ hides logarithmic factors.

Motivated by this, we seek to understand whether analogous **instance-dependent guarantees** can be derived for two-player zero-sum games. In particular, we ask the following:

Open Problem. What is the optimal instance-dependent sample complexity for identifying an ε-approximate Nash equilibrium, with probability at least $1 - \delta$, in a two-player zero-sum game defined by a payoff matrix $A \in \mathbb{R}^{n \times m}$? More precisely, what is the tight, instance-dependent characterization of the sample complexity

$$\Theta\left(\mathbf{H_1} \log(1/\delta) + \mathbf{H_2}\right),$$

where $\mathbf{H_1}$ and $\mathbf{H_2}$ are functions of the payoff matrix A and the approximation parameter ε, capturing the instance's inherent difficulty.

3. Challenges and Partial progress

In the simple case of multi-armed bandits, it is straightforward to characterize the instances that are "easy": those where the best arm has a mean significantly larger than the means of the other arms. However, the analogous question, what makes an instance of a two-player zero-sum game easy, is itself challenging. Formally, which parameters of the payoff matrix A govern the instance-dependent sample complexity terms $\mathbf{H_1}$ and $\mathbf{H_2}$? Ideally, we want to identify parameters that are large in naturally arising games, analogous to large gap parameters in multi-armed bandits that make instances easier. Even if we momentarily set aside the goal of precisely characterizing $\mathbf{H_1}$ and $\mathbf{H_2}$, identifying families of instances where one can improve upon the generic $\mathcal{O}(n \log(n/\delta)/\varepsilon^2)$ upper bound remains a fundamental question.

As a first step, one may consider the important class of input matrices with a unique Nash equilibrium $(x_\star, y_\star)$. In such games, the support sizes of the row and column strategies at equilibrium are equal; that is, $|\text{supp}(x_\star)| = |\text{supp}(y_\star)|$, where $\text{supp}(v) := \{i : v_i \neq 0\}$. Let $k := |\text{supp}(x_\star)|$. If the row and column supports can be identified quickly, then one might hope to improve upon the generic upper bound. However, it is not obvious what instance-dependent parameters indicate whether such efficient identification is possible.

Maiti et al. (2023) initiated progress in this direction by analyzing $n \times 2$ games with a unique Nash equilibrium $(x_\star, y_\star)$. They proposed an instance-dependent parameter

$$\Delta_g = \min_{i \notin \text{supp}(x_\star)} \langle x_\star - e_i, Ay_\star \rangle,$$

(ignoring some rescaling terms here for simplicity), and showed that the difficulty of the instance depends crucially on Δ_g. When Δ_g is large, the instance is easier, and when it is small, the problem is harder. Specifically, they provided an upper bound that scales as $1/\varepsilon^2 + n/\Delta_g^2$, which improves over the standard n/ε^2 bound when Δ_g is large. They also proved a lower bound of $\Omega(1/\Delta_g^2)$ for 3×2 games using a change-of-measure argument, showing that the dependence on Δ_g is unavoidable.

These results build on a classic technical lemma by Bohnenblust et al. (1950), which shows that in any matrix game with a unique Nash equilibrium, the quantities

$$\Delta_{r,i} := \langle x_\star - e_i, Ay_\star \rangle > 0 \quad \text{and} \quad \Delta_{c,j} := \langle x_\star, A(e_j - y_\star) \rangle > 0$$

for all $i \notin \text{supp}(x_\star)$ and $j \notin \text{supp}(y_\star)$. This implies a strict margin of separation between support and non-support strategies, which can potentially be exploited to derive sharper sample complexity bounds for $n \times m$ games with unique equilibrium. Recently, Maiti et al. (2025) used this lemma to identify the equilibrium support using only $n \cdot \text{poly}(k)$ queries to the input matrix in the noiseless setting. This naturally raises the question: what is the sample complexity of identifying the row and column supports of the equilibrium when observations are noisy?

In this regard, progress has been made in the special case where $k = 1$, i.e., the input matrix $A \in \mathbb{R}^{n \times m}$ admits a strict pure strategy Nash equilibrium (PSNE). A strict PSNE is an entry $(i_\star, j_\star)$ such that $A_{i,j_\star} < A_{i_\star,j_\star} < A_{i_\star,j}$ for all $i \neq i_\star$ and $j \neq j_\star$. Zhou et al. (2017) proved a lower bound of $\Omega(\mathbf{H}_\star \log(1/\delta))$, where $\mathbf{H}_\star := \sum_{i \neq i_\star} \frac{1}{\Delta_i^2} + \sum_{j \neq j_\star} \frac{1}{\Delta_j^2}$, and proposed an LUCB-based algorithm achieving an upper bound of

$$\mathcal{O}\left(\mathbf{H}_\star \log(\mathbf{H}_\star/\delta) + \frac{nm}{\widetilde{\Delta}}\right),$$

where $\widetilde{\Delta}$ is a matrix-dependent parameter. However, the additive $nm/\widetilde{\Delta}$ term can dominate. For example, when δ and all relevant gaps are constant, the lower bound scales as $\Omega(n+m)$, whereas the upper bound scales as $\mathcal{O}(nm)$.

This gap raised the question of whether the upper bound can be improved. In the noiseless setting, it is known that a query complexity of $\widetilde{\mathcal{O}}(n+m)$ is achievable (Bienstock et al., 1991; Maiti et al., 2025; Dallant et al., 2024a,b). Recently, Maiti et al. (2024) proposed a non-trivial algorithm that avoids sampling the entire matrix and uses a median-of-means approach to achieve a sample complexity of $\widetilde{\mathcal{O}}(\mathbf{H}_\star \log(1/\delta))$, closing the gap up to logarithmic factors.

While much of the literature has focused on such special cases, the general problem of identifying an ε-approximate Nash equilibrium in $n \times m$ matrix games, where the equilibrium may not be unique, remains largely open from a sample complexity perspective. Recently, Ito et al. (2025) initiated work in this direction by studying instance-dependent external regret when two no-regret learners (row and column players) are run independently. Specifically, they show that when both players follow Tsallis-INF, the regret after T rounds scales as

$$\widetilde{\mathcal{O}}\left(\sqrt{(|\mathcal{I}| + |\mathcal{J}| - 2)T} + \sum_{i \notin \mathcal{I}} \frac{\log(nT)}{\Delta(i)} + \sum_{j \notin \mathcal{J}} \frac{\log(mT)}{\Delta'(j)}\right),$$

where $(x_\star, y_\star)$ is a Nash equilibrium with maximum support, $\mathcal{I} = \text{supp}(x_\star)$, $\mathcal{J} = \text{supp}(y_\star)$, $\Delta = (\langle x_\star, Ay_\star \rangle)\mathbf{1} - Ay_\star$, and $\Delta' = A^\top x_\star - (\langle x_\star, Ay_\star \rangle)\mathbf{1}$. These results provide hope that similar instance-dependent guarantees might be achievable for sample complexity as well.

However, even in the seemingly simple case of matrices with a strict PSNE, the regret guarantees from Ito et al. (2025) are suboptimal by a factor of $\sqrt{n+m}$ compared to the sample complexity bounds of Maiti et al. (2024). Whether uncoupled no-regret learners can achieve optimal sample complexity in this special case remains an open question.

4. Potential Consequences

From a theoretical perspective, any progress on this problem could lead to new algorithmic frameworks for multi-agent learning. While two-player zero-sum games represent a simple and foundational setting, they capture many essential challenges of multi-agent interaction under uncertainty. Yet, their sample complexity characteristics remain poorly understood. Understanding their instance-dependent sample complexity may serve as a stepping stone toward more general multi-agent learning problems. Given the growing interest in multi-agent systems within the broader context of AI, such foundational insights could open up a new line of research at the intersection of game theory and learning theory, potentially leading to algorithms with real-world impact.

References

Daniel Bienstock, Fan Chung, Michael L Fredman, Alejandro A Schäffer, Peter W Shor, and Subhash Suri. A note on finding a strict saddlepoint. *The American mathematical monthly*, 98(5): 418–419, 1991.

HF Bohnenblust, S Karlin, and LS Shapley. Solutions of discrete, two-person games. *Contributions to the Theory of Games*, 1:51–72, 1950.

Justin Dallant, Frederik Haagensen, Riko Jacob, László Kozma, and Sebastian Wild. Finding the saddlepoint faster than sorting. In *2024 Symposium on Simplicity in Algorithms (SOSA)*, pages 168–178. SIAM, 2024a.

Justin Dallant, Frederik Haagensen, Riko Jacob, László Kozma, and Sebastian Wild. An optimal randomized algorithm for finding the saddlepoint. In *32nd Annual European Symposium on Algorithms (ESA 2024)*, pages 44–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2024b.

Shinji Ito, Haipeng Luo, Taira Tsuchiya, and Yue Wu. Instance-dependent regret bounds for learning two-player zero-sum games with bandit feedback. *arXiv preprint arXiv:2502.17625 (To appear at COLT 2025)*, 2025.

Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck. lil'ucb: An optimal exploration algorithm for multi-armed bandits. In *Conference on Learning Theory*, pages 423–439. PMLR, 2014.

Arnab Maiti, Kevin Jamieson, and Lillian Ratliff. Instance-dependent sample complexity bounds for zero-sum matrix games. In *International Conference on Artificial Intelligence and Statistics*, pages 9429–9469. PMLR, 2023.

Arnab Maiti, Ross Boczar, Kevin Jamieson, and Lillian Ratliff. Near-optimal pure exploration in matrix games: A generalization of stochastic bandits & dueling bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 2602–2610. PMLR, 2024.

Arnab Maiti, Ross Boczar, Kevin Jamieson, and Lillian Ratliff. Query-efficient algorithm to find all nash equilibria in a two-player zero-sum matrix game. *ACM Transactions on Economics and Computation*, 2025.

Andrew J Wagenmaker, Max Simchowitz, and Kevin Jamieson. Beyond no regret: Instance-dependent pac reinforcement learning. In *Conference on Learning Theory*, pages 358–418. PMLR, 2022.

Yichi Zhou, Jialian Li, and Jun Zhu. Identify the nash equilibrium in static games with random payoffs. In *International Conference on Machine Learning*, pages 4160–4169. PMLR, 2017.