
Scalable Neural Incentive Design with Parameterized Mean-Field Approximation

Nathan Corecco*

Department of Computer Science
ETH Zurich
nathan.corecco@inf.ethz.ch

Batuhan Yardim*

Department of Computer Science
ETH Zurich
yardima@inf.ethz.ch

Vinzenz Thoma

Department of Computer Science
ETH Zurich & ETH AI Center
vinzenz.thoma@ai.ethz.ch

Zebang Shen

Department of Computer Science
ETH Zurich
zebang.shen@inf.ethz.ch

Niao He

Department of Computer Science
ETH Zurich
niao.he@inf.ethz.ch

Abstract

Designing incentives for a multi-agent system to induce a desirable Nash equilibrium is both a crucial and challenging problem appearing in many decision-making domains, especially for a large number of agents N . Under the exchangeability assumption, we formalize this incentive design (ID) problem as a parameterized mean-field game (PMFG), aiming to reduce complexity via an infinite-population limit. We first show that when dynamics and rewards are Lipschitz, the finite- N ID objective is approximated by the PMFG at rate $\mathcal{O}(1/\sqrt{N})$. Moreover, beyond the Lipschitz-continuous setting, we prove the same $\mathcal{O}(1/\sqrt{N})$ decay for the important special case of sequential auctions, despite discontinuities in dynamics, through a tailored auction-specific analysis. Built on our novel approximation results, we further introduce our Adjoint Mean-Field Incentive Design (AMID) algorithm, which uses explicit differentiation of iterated equilibrium operators to compute gradients efficiently. By uniting approximation bounds with optimization guarantees, AMID delivers a powerful, scalable algorithmic tool for many-agent (large N) ID. Across diverse auction settings, the proposed AMID method substantially increases revenue over first-price formats and outperforms existing benchmark methods.

1 Introduction

Setting the right incentives in a game with many participants is a challenging and high-stakes problem. Policymakers must frequently make choices that affect millions, for instance, planners must design rules for curtailing city traffic [46], set pricing to maximize effective bandwidth in telecommunications networks [3], design spectrum auctions between telecom operators [41] or manage supply and demand in energy grids [52].

We study the *incentive design* (ID) problem. Given an objective G , a parameterized N -player game \mathcal{G} , and player strategies π_1, \dots, π_N , ID solves the equilibrium constrained optimization problem:

*Equal contribution.

$$\text{Maximize } G(\theta, \pi_1, \dots, \pi_N) \text{ such that } (\pi_1, \dots, \pi_N) \in \text{Nash}_{\mathcal{G}}(\theta) \quad (\text{ID})$$

where $\text{Nash}_{\mathcal{G}}(\theta)$ denotes the set of Nash equilibria (NE) of a game for the incentive parameter θ (which is to be learned). (ID) is also referred to in the literature as mathematical programming with equilibrium constraints (MPEC) [51]. Despite their relevance, MPECs are notoriously computationally challenging and in the general case NP-hard [39, 51]. Simply computing a Nash equilibrium for a fixed incentive parameter θ is also PPA-complete [22], even for 2 players [17]. For games with many players, as in many real-world problems, the so-called *curse of many agents* [71] becomes an added challenge. In such cases, computing $\text{Nash}_{\mathcal{G}}(\theta)$ becomes prohibitively expensive, let alone tackling (ID).

In this work, we take the approach of mean-field approximation for games with agent exchangeability (i.e., symmetry) to tackle this problem. Instead of solving (ID) directly, we construct an appropriate mean-field game (MFG) approximation \mathcal{M} to \mathcal{G} and solve the mean-field ID (MID) problem:

$$\text{Maximize } G(\theta, \pi_{\text{MFG}}) \text{ such that } \pi_{\text{MFG}} \in \text{Nash}_{\mathcal{M}}(\theta). \quad (\text{MID})$$

For the mean-field approximation (MID) to be meaningful, its solution should closely match that of the original N -player incentive design problem (ID), with the approximation error vanishing as $N \rightarrow \infty$. We formalize this requirement in the following desideratum.

Desideratum D1 (Approximation) *The solution of (MID) should be a good (approximate) solution for (ID), in particular when N is large, with explicit guarantees.*

MFGs are known to approximate finite-player NEs with explicit bounds for large N [63, 14, 19, 75]. Under Lipschitz continuity, a non-asymptotic bound of $\mathcal{O}(1/\sqrt{N})$ in exploitability is obtained with a propagation-of-chaos type argument. By contrast, our problem (MID) not only contains the classical MFG as a subproblem but also an incentive objective, accordingly, we must show that the approximation error still vanishes as $N \rightarrow \infty$. Establishing this result is our first contribution.

Contribution 1: Lipschitz PMFGs and Approximation. We formalize (MID) as a parameterized MFG (PMFG), and show that for Lipschitz PMFGs the (MID) problem approximates (ID) with a rate of $\mathcal{O}(1/\sqrt{N})$, both in exploitability and optimality of the design objective.

While Lipschitz continuous MFGs cover a broad class of real-world games and are well-studied, they exclude many important applications, notably large-scale auction design, where the transition dynamics are inherently correlated and non-Lipschitz. Motivated by the ubiquity and impact of auctions, we analyze (D1) beyond Lipschitz PMFGs, which is our second contribution.

Contribution 2: Approximation beyond Lipschitz. We identify a PMFG for sequential batched auctions (BA-MFG) with many bidders, and establish (D1) with an $\mathcal{O}(1/\sqrt{N})$ approximation rate, by identifying a set of “well-behaved” policies that is dense in the set of Nash equilibria.

Although (D1) guarantees that the mean-field approximation captures the fundamental solution structure of the N -player incentive design problem, realizing practical benefits from this framework requires the following desideratum.

Desideratum D2 (Optimization) *We would like (MID) to be computationally tractable or easier to solve than (ID), specifically, admit an efficient first-order oracle that can be used for optimization.*

Our algorithmic contribution is to satisfy this desideratum in both settings: (i) PMFGs under Lipschitz continuity assumptions, and (ii) mean-field auction design.

Contribution 3: Algorithmic Contribution. We formulate our adjoint mean-field incentive design (AMID) algorithm for solving (MID) efficiently. AMID is a modification of backpropagation based on the adjoint method for computing (approximate) derivatives with Nash equilibrium constraints. While naive autodiff-based approaches incur a large $\mathcal{O}(T)$ memory footprint, our reformulation of the gradient computations reduces the memory footprint to $\mathcal{O}(\sqrt{T})$, with up to 80% savings in practice.

Contribution 4: Experimental Contribution. Finally, we use AMID to (i) solve congestion pricing in the classical beach bar MFG, and (ii) design revenue-optimal mechanisms across a variety of sequential auctions with a neural network parameterization. Our method consistently outperforms standard first-price mechanisms used in practice and other optimization approaches to solve (MID).

Table 1: Comparison of selected works on ID in literature. **Large N** : results scale to many-agent (symmetric) \mathcal{G} , **Dynamic**: solves ID on games with dynamics, **Explicit diff.**: explicit differentiation for first-order information, **Approx.**: finite-agent approximation in N , **Auctions**: applies to auctions.

Work	Model	Large N	Dynamic	Approx.	Explicit diff.	Auctions
[48]	VI	✗	✗	-	✓	✗
[50]	VI	✓	✗	-	✗	✗
[23]	Cont. time	✓	✓	✓	✗	✗
[54]	LQ	✓	✓	✓	✗	✗
[64]	Maj.-min.	✓	✓	✓	✗	✗
Ours	PMFG	✓	✓	✓-Theorem 1	✓-Section 2.2	✓- Section 3

1.1 Related Works

We present the works most relevant to this paper, complemented by the discussion in Appendix B.

Mathematical Programming with Equilibrium Constraints (MPEC). Several works have studied gradient-based approaches for MPEC [48, 50, 70, 47, 26]. Some of these assume that the equilibrium problem satisfies strong monotonicity to compute the gradients. Others use explicit differentiation, an approach we also follow in this work. Motivated by the success of reinforcement learning (RL), many works have focused on the optimal design of (multi-agent) RL environments [67, 16, 80, 28, 21, 69]. An important instance of designing games with desirable outcomes is *automated mechanism design*, capturing many real-world economic problems [18, 20, 49].

Steering and Equilibrium Selection. Complementary to ID problems, a related strand of work has focused on steering and equilibrium selection. For mean-field games, [33] considers a problem of choosing equilibria with high social welfare. Steering learning dynamics towards desirable equilibria was studied by [35, 11, 78] for Markovian and no-regret learners and extended to MFGs by [72].

Mean-Field Games (MFG). MFGs, first formulated in the seminal works of Lasry & Lions [44] and Huang et al. [36], analyze symmetric competitive agents at the many-agent limit. Recently, many works have studied RL in mean-field settings, such as stationary MFGs [32, 77, 73, 19], monotone MFGs [58, 56, 57, 76], static MFGs [74], and mean-field control [15, 5]. While general MFGs remain a theoretical challenge [75], under structured settings, they have shown empirical and algorithmic efficiency [19, 15, 45]. MFGs have also been studied in Stackelberg equilibria, closer to our setting [13, 23, 1, 23]. While these works have similar objectives to ours, rather than letting a leader influence a population through interactions with a static environment, we aim to design parameterized MFGs directly by explicit differentiation. In this sense, (MID) can be seen to differ in objective from these works. Moreover, these results do not readily apply to auction design, a foundational problem for incentive design. In Table 1, we provide a comparison with selected works and our results.

2 Designing Games for Large Populations: Lipschitz Case

We first formalize parameterized N -player dynamic games and the corresponding ID problem.

Notation. We use \mathcal{S}, \mathcal{A} to denote (finite) state-actions spaces. For the horizon H , define policy space $\Pi_H := \{\pi : [H] \times \mathcal{S} \rightarrow \Delta_{\mathcal{A}}\}$, abbreviate $\pi_h(a|s) := \pi(h, s)(a)$ and also treat Π_H as a subspace of $\mathbb{R}^{[H] \times \mathcal{S} \times \mathcal{A}}$. For a finite set \mathcal{X} , define the ‘‘empirical distribution’’ $\sigma(\mathbf{x})(x') = 1/N \sum_{i=1}^N \mathbb{1}_{x_i=x'}$. We also provide a full reference table for our notation in Appendix A.

Definition 1 (Parameterized Dynamic Games) A *finite-horizon parameterized dynamic game (PDG)* is a tuple $\mathcal{G} := (N, \mathcal{S}, \mathcal{A}, H, \mu_0, \Theta, \{\bar{P}_{h,\theta}\}_{h=0}^{H-1}, \{\bar{R}_{h,\theta}\}_{h=0}^{H-1})$ of players $N \in \mathbb{N}_{\geq 1}$, discrete state actions sets \mathcal{S}, \mathcal{A} , parameter space Θ , parameterized transition dynamics $\bar{P}_{h,\theta} : \mathcal{S}^N \times \mathcal{A}^N \rightarrow \Delta_{\mathcal{S}^N}$, parameterized reward functions $\bar{R}_{h,\theta} : \mathcal{S}^N \times \mathcal{A}^N \rightarrow [0, 1]^N$, starting distribution $\mu_0 \in \Delta_{\mathcal{S}}$, and time horizon $H \in \mathbb{N}_{>0}$. For a strategy profile $\boldsymbol{\pi} \in \Pi_H^N$, $\tau \geq 0$ and some $\theta \in \Theta$, the expected

(entropy-regularized) sum of rewards of player $i \in [N]$ is defined as

$$J_{\mathcal{G}}^{\tau,i}(\boldsymbol{\pi}|\theta) := \mathbb{E} \left[\sum_{h=0}^{H-1} \bar{R}_{h,\theta}^i(\mathbf{s}_h, \mathbf{a}_h) + \tau \mathcal{H}(\pi_h^i(s_h^i)) \middle| \substack{\forall j: s_0^j \sim \mu_0, a_h^j \sim \pi_h^j(s_h^j), \mathbf{s}_h := \{s_h^j\}_j, \mathbf{a}_h := \{a_h^j\}_j, \\ \mathbf{s}_{h+1} \sim \bar{P}_{h,\theta}(\mathbf{s}_h, \mathbf{a}_h)} \right].$$

Define $\mathcal{E}_{\mathcal{G}}^{\tau}(\boldsymbol{\pi}|\theta) := \max_{i \in [N]} \mathcal{E}_{\mathcal{G}}^{\tau,i}(\boldsymbol{\pi}|\theta)$ where $\mathcal{E}_{\mathcal{G}}^{\tau,i}(\boldsymbol{\pi}|\theta) := \max_{\boldsymbol{\pi}' \in \Pi} J_{\mathcal{G}}^{\tau,i}((\boldsymbol{\pi}', \boldsymbol{\pi}^{-i})|\theta) - J_{\mathcal{G}}^{\tau,i}(\boldsymbol{\pi}|\theta)$, the exploitability. If $\mathcal{E}_{\mathcal{G}}^{\tau}(\boldsymbol{\pi}^*|\theta) = 0$ for $\boldsymbol{\pi}^* \in \Pi_H^N$, we call $\boldsymbol{\pi}^*$ a Nash equilibrium (DG-NE) with respect to parameter θ . The set of all Nash equilibria for $\theta \in \Theta$ is denoted $\text{Nash}_{\mathcal{G}}^{\tau}(\theta)$. One is typically interested in maximizing a function of the aggregated population behavior (e.g., revenue, negative congestion):

$$G(\theta, \boldsymbol{\pi}) := \mathbb{E}[g(\theta, \{\hat{L}_h\}_{h=0}^{H-1})|\boldsymbol{\pi}, \theta], \text{ where } \hat{L}_h := \frac{1}{N} \sum_{j=1}^N \mathbf{e}_{s_h^j, a_h^j}, \quad (\text{ID Objective})$$

given by some $g : \Theta \times \Delta_{\mathcal{S} \times \mathcal{A}}^H \rightarrow \mathbb{R}$, with the constraint that $\boldsymbol{\pi} \in \Pi_H^N$ is an (approximate) Nash equilibrium under θ (ignoring multiplicities). The parameter space Θ and the parameterizations of $\bar{P}_{h,\theta}, \bar{R}_{h,\theta}$ will dictate the implicit constraints on the design, such as the available information at time h . For such parameterizations, optimizing G will be nontrivial and incorporate an intractable many-agent NE computation as a subproblem. In the following, we reduce this problem to a lower-dimensional MFG (i.e. of size independent of N) and propose tractable alternatives.

2.1 Parameterized Mean Field Game Design

Below, we formalize PMFGs. Definition 2 generalizes the standard definition of MFGs to a parametric family of MFGs, and approximates Definition 1 on a continuum of infinitely many players.

Definition 2 (Parameterized Mean-Field Games) A finite-horizon parameterized mean-field game (PMFG) is a tuple $\mathcal{M} := (\mathcal{S}, \mathcal{A}, H, \mu_0, \Theta, \{P_{h,\theta}\}_{h=0}^{H-1}, \{R_{h,\theta}\}_{h=0}^{H-1})$ of discrete state actions sets \mathcal{S}, \mathcal{A} , parameterized transition dynamics $P_{h,\theta} : \mathcal{S} \times \mathcal{A} \times \Delta_{\mathcal{S} \times \mathcal{A}} \rightarrow \Delta_{\mathcal{S}}$, parameterized reward functions $R_{h,\theta} : \mathcal{S} \times \mathcal{A} \times \Delta_{\mathcal{S} \times \mathcal{A}} \rightarrow [0, 1]$, initial distribution $\mu_0 \in \Delta_{\mathcal{S}}$, and horizon $H \in \mathbb{N}_{>0}$. Define operators Γ_h, Λ as $\Gamma_h(L, \pi_h|\theta)(s', a') := \sum_{s,a} L(s, a) P_{h,\theta}(s'|s, a, L) \pi_h(a'|s')$ and $\Lambda(\pi|\theta) := \{\Gamma_{h-1}(\dots \Gamma_1(\Gamma_0(\mu_0 \cdot \pi_0, \pi_1|\theta)|\theta) \dots, \pi_{h-1}|\theta)\}_{h=0}^{H-1}$, called population operators². For $\pi \in \Pi_H, \tau \geq 0$ and $\mathbf{L} = \{L_h\}_{h=0}^{H-1} \in \Delta_{\mathcal{S} \times \mathcal{A}}^H$, the total expected (entropy regularized) reward is

$$V_{\mathcal{M}}^{\tau}(\mathbf{L}, \pi|\theta) := \mathbb{E} \left[\sum_{h=0}^{H-1} R_{h,\theta}(s_h, a_h, L_h) + \tau \mathcal{H}(\pi_h(s_h)) \middle| \substack{s_0 \sim \mu_0, a_h \sim \pi_h(s_h) \\ \mathbf{s}_{h+1} \sim P_{h,\theta}(\mathbf{s}_h, \mathbf{a}_h, L_h)} \right].$$

We define mean-field exploitability as $\mathcal{E}_{\mathcal{M}}^{\tau}(\pi|\theta) := \max_{\boldsymbol{\pi}' \in \Pi} V_{\mathcal{M}}^{\tau}(\Lambda(\boldsymbol{\pi}'), \boldsymbol{\pi}'|\theta) - V_{\mathcal{M}}^{\tau}(\Lambda(\boldsymbol{\pi}'), \boldsymbol{\pi}|\theta)$. If $\mathcal{E}_{\mathcal{M}}^{\tau}(\boldsymbol{\pi}^*|\theta) = 0$ for $\boldsymbol{\pi}^* \in \Pi_H$, we call $\boldsymbol{\pi}^*$ a MFG Nash equilibrium (MFG-NE) with respect to parameter θ . The set of all Nash equilibria for $\theta \in \Theta$ is denoted $\text{Nash}_{\mathcal{M}}^{\tau}(\theta)$.

In this section, we will make the following (standard) assumption on $P_{\theta}(s'|s, a, L), R_{\theta}(s, a, L)$, which holds in many relevant applications.

Assumption 1 (Lipschitz continuity) For all $s, s' \in \mathcal{S}, a \in \mathcal{A}$, the functions $P_{h,\theta}(s'|s, a, L), R_{h,\theta}(s, a, L)$, and $g(\theta, \mathbf{L})$ are Lipschitz continuous in θ, L .

Theorem 1 below demonstrates that by optimizing the objective $g(\theta, \Lambda(\boldsymbol{\pi}^*))$, one can obtain approximation guarantees (up to a bias of $\mathcal{O}(1/\sqrt{N})$) on the performance of the PDG that has independent and symmetric state transitions. We have therefore established (D1) for PMFGs.

Theorem 1 Let \mathcal{M} be a PMFG, Assumption 1 hold, and \mathcal{G} be the PDG such that $\bar{P}_{h,\theta}(\mathbf{s}, \mathbf{a}) := \bigotimes_{i \in [N]} P_{h,\theta}(s^i, a^i, \sigma(\mathbf{s}, \mathbf{a}))$ and $\bar{R}_{h,\theta}^i(\mathbf{s}, \mathbf{a}) = R_{h,\theta}(s^i, a^i, \sigma(\mathbf{s}, \mathbf{a}))$ for all i . Let $\boldsymbol{\pi}^* \in \text{Nash}_{\mathcal{M}}^{\tau}(\theta)$ and $\boldsymbol{\pi}^* := (\pi^*, \dots, \pi^*) \in \Pi_H^N$. Then:

1. $\mathcal{E}_{\mathcal{G}}^{\tau}(\boldsymbol{\pi}^*|\theta) \leq \mathcal{O}(1/\sqrt{N})$, that is, $\boldsymbol{\pi}^*$ is a $\mathcal{O}(1/\sqrt{N})$ -NE of \mathcal{G} under θ .

²Note that we define $\Lambda(\boldsymbol{\pi}|\theta)_0 := \mu_0 \cdot \pi_0$

$$2. G(\theta, \pi^*) \geq g(\theta, \Lambda(\pi^*|\theta)) - \mathcal{O}(1/\sqrt{N}).$$

Theorem 1 mirrors bounds in MFGs without ID [75] and Stackelberg MFGs in other settings [54]. For clarity, the theorem is stated as an approximation result for a PDG \mathcal{G} that exactly satisfies agent exchangeability, which might not always be the case. In some applications, finding the mean-field formulation \mathcal{M} given a PDF \mathcal{G} might be nontrivial. The *converse* problem of constructing an appropriate PMFG \mathcal{M} that approximates a given \mathcal{G} was studied in [76], and this work can be trivially generalized to this case with an additional approximation bias due to asymmetries in \mathcal{G} . The case of auction design, which also does not satisfy Assumption 1 and the assumption of $\bar{P}_{h,\theta}$ being a product measure, will require specific treatment in Section 3.

2.2 Approximating the First Order Derivatives

Having satisfied (D1) for ID with Lipschitz dynamics in the previous section, we turn to (D2)–formulating algorithmic methods to solve (MID). We state the standard definitions of value and q-functions for PMFGs, which will be important for learning NEs:

$$V_h^\tau(s|\mathbf{L}, \pi, \theta) := \mathbb{E} \left[\sum_{h'=h}^{H-1} R_\theta(s_{h'}, a_{h'}, L_{h'}) + \tau \mathcal{H}(\pi_{h'}(s_{h'})) \middle| \begin{array}{l} s_h=s, a_{h'} \sim \pi_{h'}(s_{h'}) \\ s_{h'+1} \sim P_\theta(s_{h'}, a_{h'}, L_{h'}) \end{array} \right]$$

$$q_h^\tau(s, a|\mathbf{L}, \pi, \theta) := R_\theta(s, a, L_h) + \mathbb{E} \left[V_{h+1}^\tau(s_{h+1}|\mathbf{L}, \pi, \theta) \middle| \begin{array}{l} s_h=s, a_h=a, a_{h'} \sim \pi_{h'}(s_{h'}) \\ s_{h'+1} \sim P_\theta(s_{h'}, a_{h'}, L_{h'}) \end{array} \right]$$

We define the commonly used online mirror descent update rule $\bar{F}_{\text{omd}} : \Theta \times \Pi_H \rightarrow \Pi_H$ with

$$\bar{F}_{\text{omd}}(\theta, \pi)(h, s) := \arg \max_{u \in \Delta_{\mathcal{A}}} \langle q_h^\tau(s, \cdot | \Lambda(\pi), \pi, \theta), u \rangle + \tau \mathcal{H}(u) - \eta^{-1}(1 - \tau\eta) D_{\text{kl}}(u | \pi(s)),$$

for some given learning rate $\eta > 0$ and entropy regularization $\tau \geq 0$. \bar{F}_{omd} has received particular attention in MFG literature due to its theoretical and empirical properties. Abbreviating $\bar{F}_{\text{omd}}^{(T)}(\theta, \zeta) := \bar{F}_{\text{omd}}(\theta, \bar{F}_{\text{omd}}(\theta, \dots \bar{F}_{\text{omd}}(\theta, \zeta) \dots))$, i.e., $\bar{F}_{\text{omd}}(\theta, \cdot)$ applied T times, the repeated iterations $\bar{F}_{\text{omd}}^{(T)}$ for $T > 0$ are known to convert to NE for monotone MFGs theoretically [56, 79, 37], and empirically find good approximations to NE [19] for general MFGs. Furthermore, any $\pi^* \in \text{Nash}_{\mathcal{M}}(\theta)$ is guaranteed to be a fixed point of the map $\bar{F}_{\text{omd}}(\theta, \cdot)$ for some learning rate η . We formulate an explicit differentiation scheme for the PMFG using these properties of \bar{F}_{omd} . Defining the softmax transform $\text{softmax} : \mathbb{R}^{[H] \times \mathcal{S} \times \mathcal{A}} \rightarrow \Pi$ as $\text{softmax}(\zeta)(h, s, a) := \frac{\exp\{\zeta_{h,s,a}\}}{\sum_{a'} \exp\{\zeta_{h,s,a'}\}}$, the above OMD update rule can be reformulated in terms of log probabilities:

$$F_{\text{omd}}(\theta, \zeta)(h, s, a) := (1 - \eta\tau)\zeta_{h,s,a} + \eta q_h^\tau(s, a | \Lambda(\text{softmax}(\zeta)), \text{softmax}(\zeta), \theta).$$

For fixed θ , the repeated iterations F_{omd} will converge (under technical conditions) to $\log \pi^*$ where π^* is an NE of the MFG induced by θ . With this motivation, we reformulate the PMFG design problem (MID) as a maximization of the T -step approximate objective

$$G_{\text{approx}}^T : \theta \rightarrow g \left(\theta, \Lambda(\text{softmax}(F_{\text{omd}}^{(T)}(\theta, \zeta_0)) | \theta) \right). \quad (T\text{-approx.})$$

G_{approx}^T in general is a well-defined differentiable function (see Lemma 4, Appendix D). In particular, standard autograd tools can be used to compute $\nabla G_{\text{approx}}^T$. While the behavior of $\nabla G_{\text{approx}}^T$ when $T \rightarrow \infty$ is not immediate, Lemma 1 below shows that under technical conditions G_{approx}^T is a meaningful objective function to maximize, and produces low-bias estimates of the derivatives of the true NE with respect to θ for sufficiently large T .

Lemma 1 (Differentiability of F_{omd}^∞) *Let $\zeta \in \mathbb{R}^{[H] \times \mathcal{S} \times \mathcal{A}}$, $\theta \in \Theta$ be such that the following hold:*

1. $F_{\text{omd}}^\infty(\theta', \zeta) := \lim_{T \rightarrow \infty} F_{\text{omd}}^{(T)}(\theta', \zeta)$ for $\theta' \in U$ for a neighborhood U of θ ,
2. For ζ^* such that $\zeta^* = F_{\text{omd}}^\infty(\theta, \zeta)$, q_h^τ is C^1 in a neighborhood of (θ, ζ^*) , and $\rho(\partial_\zeta q^\tau(\cdot, \cdot | \Lambda(\text{softmax}(\zeta^*)), \text{softmax}(\zeta^*), \theta)) < \tau$ for all $h \in [H]$.

Then, $\text{softmax}(F_{\text{omd}}^\infty(\theta', \zeta)) \in \text{Nash}_{\mathcal{M}}^\tau(\theta')$ on $\theta' \in U$, F_{omd}^∞ is partially differentiable in θ at (θ, ζ) , and $\lim_{T \rightarrow \infty} \partial_\theta F_{\text{omd}}^{(T)}(\theta, \zeta) = \partial_\theta F_{\text{omd}}^\infty(\theta, \zeta)$.

Lemma 1 justifies the use of G_{approx}^T (and subsequently the explicit differentiation scheme) as an objective function under mild technical conditions. If G is also C^1 , $\nabla G_{\text{approx}}^T$ converges to the derivative of a map $\theta' \rightarrow G(\theta', \pi^{\theta'})$ where $\pi^{\theta'}$ is a function of θ' such that $\pi^{\theta'} \in \text{Nash}_{\mathcal{M}}^{\tau}(\theta')$ locally, that is, for all θ' in some neighborhood of θ . Moreover, Lemma 1 provides intuition on how to tune the parameters η, T, τ and characterizes their impact on the quality of explicit differentiation.

One major challenge from a computational point of view of backpropagating (T -approx.) will be the size of the computational graph, growing with $\mathcal{O}(T)$. In many MFGs, finding a good approximate MFG-NE will require a large T , which will incur a large computational overhead.

Algorithm 1, which we call **adjoint mean-field incentive design** (AMID), reduces the potentially memory-intensive backpropagation through a complex computational graph to a simple forward and backward pass in t . Crucially, the update operator F is typically quite complicated for PMFGs: for instance, F_{omd} itself involves solving forward (Λ) and backward equations (q_h^{τ}) in h . Therefore, for large T , naive autograd will be inefficient due to the storage of many intermediate values and a large graph.

Algorithm 1 AMID

Input: Update rule F , objective $G, T, \eta, \tau, \theta, \zeta_0$
1: **for** $t \in 0, \dots, T$ **do** ▷ Forward pass
2: $\zeta_{t+1} = (1 - \eta\tau)\zeta_t + \eta F(\theta, \zeta_t)$
3: **end for**
4: $s_{T+1} = \partial_{\theta} G(\theta, \zeta_{T+1}), \mathbf{a}_T = -\partial_{\zeta} G(\zeta_{T+1})$
5: **for** $t \in T, \dots, 0$ **do** ▷ Backward pass
6: $a_{t-1} = (1 - \eta\tau)a_t + \eta a_t \partial_{\zeta} F(\theta, \zeta_t)$
7: $s_{t-1} = s_t - \eta a_t \partial_{\theta} F(\theta, \zeta_t)$
8: **end for**
9: **return** s_0

Lemma 2 (Adjoint method) *Let $\Theta \subset \mathbb{R}^{d'}$ be an open set and $F : \Theta \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $G : \Theta \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ be differentiable functions. Assume Algorithm 1 is run with inputs F, G , and $T \in \mathbb{N}_{>0}, \eta, \tau > 0, \zeta_0 \in \mathbb{R}^d$. Then its return value s_0 is equal to $\nabla_{\theta} G(\theta, F^{(T)}(\theta, \zeta_0))$.*

Remark 1 *In Algorithm 1, $\{\zeta_t\}_t$ will need to be stored for the backward pass in memory, however, the memory footprint can be reduced to $\mathcal{O}(\sqrt{T})$ by caching every $\mathcal{O}(\sqrt{T})$ timesteps of the forward process and recomputing ζ_t as required, maintaining a time complexity of $\mathcal{O}(T)$. Furthermore, Algorithm 1 can be generalized to arbitrary Bregman divergences, which permits a variety of inner loop operators to be used (Appendix D.6).*

3 Beyond Lipschitz: Mechanism Design for Large-Scale Auctions

Auction design is a ubiquitous and well-studied problem of extraordinary economic interest [43]. To analyze auctions at the mean-field regime, we move beyond the Lipschitz PMFG assumptions. Designing auctions with maximum revenue in the resulting equilibrium can be framed as an instance of (ID) (see our discussion in Appendix C), and with an appropriate PMFG, tackled using AMID. Specifically, we consider the following sequential batched auction setting motivated by real-world formats used for selling government debt, broadcast licenses, mineral rights, art, fish, timber [27], and including transactions in mined Ethereum blocks [62].

(Parameterized) batched auctions. An H -round N -player batched auction with incentive parameter θ is a PDG \mathcal{G}_{auc} with state space $\mathcal{S} = \mathcal{V} \cup \{\perp\}$ (where the value space $\mathcal{V} := \{0, \dots, (|\mathcal{V}| - 1)/|\mathcal{V}|\}$ represents possible valuations and \perp denotes non-participation in the current round) and action space $\mathcal{A} = \{0, \dots, (|\mathcal{A}| - 1)/|\mathcal{A}|\}$ (possible bids). Each bidder $i \in [N]$ at round $h \in [H]$ has a private state $s_h^i \in \mathcal{S}$ not revealed to the auctioneer or other bidders. Overall, the auctioneer sells at most $\lceil \alpha_{\max} N \rceil$ goods (for some $\alpha_{\max} \in (0, 1)$) and chooses θ , parameterizing the allocations and payments as outlined below. The auction evolves for $h \in [H]$ as follows:

1. Initial states at $h = 0$ are independently sampled from distribution $\mu_0 \in \Delta_{\mathcal{V}}$.
2. At every round $h \in [H]$, bidders for which $s_h^i \neq \perp$ submit their bids $a_h^i \in \mathcal{A}$.
3. Observing the bid distributions $\widehat{\nu}_h^{-\perp} := \frac{1}{N} \sum_{i \in [N]} \mathbf{e}_{a_h^i} \mathbb{1}_{s_h^i \neq \perp}$, the mechanism decides on a ratio of goods to be sold this round, $\alpha_h^{\theta}(\widehat{\nu}_h^{-\perp})$. Items are allocated to the highest $\lfloor \alpha_h^{\theta}(\widehat{\nu}_h^{-\perp}) N \rfloor$ bidders, with ties broken uniformly at random.
4. Each bidder i who receives an item, makes a payment $p_h^i = p_h^{\theta}(a_h^i, \widehat{\nu}_h^{-\perp}) \in \mathbb{R}_{\geq 0}$. A winning bidder receives utility $u_h(s_h^i, p_h^i) \in \mathbb{R}$ and transitions to state \perp , while non-winning and non-participating bidders receive zero utility.

- Before proceeding to round $h+1$, each bidder transitions independently to a new state according to a dynamics function $w_h : \mathcal{S} \times \Delta_{\mathcal{S}} \rightarrow \Delta_{\mathcal{S}}$, which maps the agent's current state and the empirical population distribution (after the allocation at round h) to a distribution over next states.

To ensure the mechanism can not sell more goods than are available (α_{\max}), we assume that the parameterizations of α_h^θ are such that $\sum_h \alpha_h^\theta(\hat{\nu}_h^{-\perp}) \leq \alpha_{\max}$ almost surely. This can be ensured, for example, by parameterizing α_h^θ as a fraction of *remaining goods* at every h . \mathcal{G}_{auc} allows for complex valuation dynamics, such as single-minded bidders (who stay in \perp), time-dependent or population-dependent evolving valuations, as well as super and subadditive valuations over bundles of goods. Under these dynamics, we denote the expected utility of player i and exploitability as $J_{\text{auc}}^{\tau,i}, \mathcal{E}_{\text{auc}}^{\tau,i}$ respectively, as defined in Definition 1.

We note that parameterizing $\alpha_h^\theta, p_h^\theta$ fully captures the intuition of *reserve prices* in the BA-MFG setting. In many auction formats, a reserve price, i.e., a minimum price that bidders have to bid and pay to win, has been shown to increase revenue [55]. A reserve price $r_h \in \mathcal{A}$ at round h can be implemented for example with $\alpha_h(\nu) = \sum_{a' \geq r_h} \nu(a')$ and $p_h(a, \nu) = a$.

3.1 Auctions at the Mean-Field Regime

From the above description, \mathcal{G}_{auc} clearly has exchangeable agents. However, the corresponding one-step state evolutions are not independent, making Theorem 1 inapplicable here. Motivated by the relevance of large-scale auction design, we show that PMFGs are still relevant and (D1) holds with a refined analysis of batched auctions in the following. We begin by defining the correct MFG that characterizes the batch auction at the limit $N \rightarrow \infty$. While the definition is symbol-laden, we state it for completeness.

Definition 3 (BA-MFG) A *Batched-Auction MFG (BA-MFG)* is the PMFG $\mathcal{M}_{\text{mfa}} := (\mathcal{S}, \mathcal{A}, H, \mu_0, \Theta, \{P_{h,\theta}^{\text{mfa}}\}_{h=0}^{H-1}, \{R_{h,\theta}^{\text{mfa}}\}_{h=0}^{H-1})$, where $R_{\theta,h}^{\text{mfa}}, P_{\theta,h}^{\text{mfa}}$ are given by:

$$\begin{aligned} R_{\theta,h}^{\text{mfa}}(s, a, L) &:= p_{\text{win}}(s, a, L, \alpha_h^\theta(\nu^{-\perp}(L))) u_h(s, p_h(a, \nu^{-\perp}(L))) \\ P_{\theta,h}^{\text{mfa}}(s' | s, a, L) &:= p_{\text{win}}(s, a, L, \alpha_h^\theta(\nu^{-\perp}(L))) w_h(s' | \perp, \xi^{L,\theta}) \\ &\quad + (1 - p_{\text{win}}(s, a, L, \alpha_h^\theta(\nu^{-\perp}(L)))) w_h(s' | s, \xi^{L,\theta}), \end{aligned}$$

where $\nu^{-\perp}(L) := \sum_{s \neq \perp} L(s, \cdot)$, $\xi^{L,\theta} \in \Delta_{\mathcal{S}}$ such that $\xi^{L,\theta}(\perp) = L(\perp) + \langle L, p_{\text{win}}(\cdot, \cdot, L, \alpha_h^\theta(\nu^{-\perp}(L))) \rangle$ and $\xi(s) = \langle L(s, \cdot), p_{\text{win}}(s, \cdot, L, \alpha_h^\theta(\nu^{-\perp}(L))) \rangle$, and p_{win} defined as³

$$p_{\text{win}}(s, a, L, \alpha) := \mathbb{1}_{s \neq \perp} \max \left\{ 0, \min \left\{ 1, \frac{\alpha - \sum_{s' \in \mathcal{V}, a' > a} L(s', a')}{\nu^{-\perp}(L)(a)} \right\} \right\}.$$

We use $V_{\text{mfa}}^\tau, \mathcal{E}_{\text{mfa}}^\tau$ to denote expected reward and exploitability in \mathcal{M}_{mfa} , as in Definition 2.

The intuition behind Definition 3 relies on the fact that the function p_{win} approximately characterizes the marginal winning probability of an agent when N is large. In fact, Theorem 2 below shows that BA-MFG is indeed the correct model for auctions with large N . Existing approximation results (such as [63, 75] as well as Theorem 1) fundamentally are incompatible with this setting due to (1) the fact that transitions in (finite-player) auctions are not independent, and (2) due to the inherent jump discontinuities in both $P_{\theta,h}^{\text{auc}}, R_{\theta,h}^{\text{auc}}$. No zero-dominance (NZD) policies, defined below, identify a subset of policy space Π_H where this difficulty can be circumvented.

Definition 4 (No zero-dominance (NZD)) Let \mathcal{M}_{mfa} be a BA-MFG and $\theta \in \Theta$. $\pi \in \Pi_H$ is said to satisfy the NZD property for θ if at induced $\mathbf{L} = \{L_h\}_{h=0}^{H-1} = \Lambda_{\text{mfa}}(\pi|\theta)$ there exist no $a \in \mathcal{A}, h \in [H]$ such that $\sum_{s \in \mathcal{V}} L_h(s, a) = 0$ and $\sum_{s \in \mathcal{V}, a' > a} L_h(s, a') = \alpha_h^\theta(\nu^{-\perp}(L))$.

While NZD is a technical condition, it is for instance satisfied by any entropy regularized MFG-NE of \mathcal{M}_{mfa} if $\tau > 0$, therefore, contains ε -NE for arbitrarily small $\varepsilon > 0$. With this property, BA-MFGs satisfy (D1) as shown below, making it a relevant model for auction design.

³In this definition we take $\varepsilon/0 = \infty$, for any $\varepsilon > 0$ and $0/0 = 0$, for convenience.

Theorem 2 (Approximation for BA-MFG) Let \mathcal{M}_{mfa} be a BA-MFG with Lipschitz-continuous $u_h, w_h, \alpha_h^\theta, p_h^\theta$ and let $g : \Theta \times \Delta_{\mathcal{S} \times \mathcal{A}}^H \rightarrow \mathbb{R}$ be Lipschitz. Let $\pi \in \Pi_H$ be a policy that satisfies the no zero-dominance property with respect to $\theta \in \Theta$. Then, for $\boldsymbol{\pi} = (\pi, \dots, \pi) \in \Pi_H^N$,

1. $\mathcal{E}_{auc}^\tau(\boldsymbol{\pi}|\theta) \leq \mathcal{E}_{mfa}^\tau(\pi|\theta) + \mathcal{O}(1/\sqrt{N})$, for any $\tau \geq 0$,
2. $|g(\theta, \Lambda_{mfa}(\boldsymbol{\pi}|\theta)) - G(\theta, \boldsymbol{\pi})| \leq \mathcal{O}(1/\sqrt{N})$.

Proof sketch: The proof builds on (1) special handling of the correlated evolution of s_h^i at any round h and (2) showing that for non-zero dominant policies π , the dynamics are locally Lipschitz continuous. The two conclusions are proved separately in Appendices E.3 and E.4. \square

Theorem 2 demonstrates convergence for a broad class of policies and relates to a large strand of work on equilibrium computation for auctions, which we discuss in Appendix B. While a true MF-NE does not necessarily satisfy the no zero-dominance property, an entropy-regularized MF-NE does. In this regard, the results above show that the BA-MFG *essentially* characterizes the limiting behavior of batched auctions.

Remark 2 In general, Theorem 2 incorporates a standard worst-case exponential bound in H , depending on w_h, π . However, in certain cases, such as non-expansive or population-independent w_h and π with full support, the bound becomes polynomial in $H, |\mathcal{S}|, |\mathcal{A}|$ (see Appendix E.3). We later verify the quality of the bound in real-world experiments.

Finally, we state the following differentiability result of F_{omd}^τ , thus satisfying (D2) when combined with the adjoint method described in Section 2.2. This result permits mechanism design by backpropagation for any entropy regularization $\tau > 0$, completing the motivation for BA-MFG.

Lemma 3 (Differentiability on \mathcal{G}_{auc}) Let \mathcal{M}_{mfa} be a BA-MFG on an open parameter space $\Theta \subset \mathbb{R}^d$, with Lipschitz $u_h, w_h, \alpha_h^\theta, p_h^\theta$, then F_{omd}^τ is almost everywhere differentiable on $\mathbb{R}^{[H] \times \mathcal{S} \times \mathcal{A}} \times \Theta$.

Equipped with an algorithmic tool to solve large-scale ID problems, we move to empirical demonstration on applications.

4 Experimental Results

We evaluate our methodology on numerical examples of increasing complexity, using AMID to obtain gradient estimates and ADAM [40] as an update rule on parameters θ . All experiment details, including computational resources, can be found in Appendix F. We also provide reference implementations in JAX and PyTorch⁴.

First, we demonstrate the effectiveness of our approach on the prototypical MFG of the beach bar game [58]. We formulate the PMFG \mathcal{M}_{bb} , where a large population of beachgoers starting from $\mu_0 = \text{Uniform}(\mathcal{S})$ can move left, stay, or move right ($\mathcal{A} : \{-1, 0, 1\}$) on a beach ($\mathcal{S} := [K]$) over H steps, while trying to minimize their distance to the bar located at $s_{\text{bar}} = K/2$ and avoiding busy spots. We parameterize a pricing mechanism $\theta \in [0, 1/2]^{\mathcal{S}}$ for spots on the beach to minimize congestion (the softmax of population flow):

$$F_{\text{bb}}^{h,\theta}(s, a | \mathbf{L}, \theta) := -\frac{d(s, s_{\text{bar}})}{K} + \frac{|a|}{K} - \frac{\log L_h(s)}{3} - \theta_s, \quad g_{\text{bb}}(\theta, \mathbf{L}) := -\sum_{h,s} \exp\{|\mathcal{S}|L_h(s)\}.$$

We report the training curves and induced flows in Figure 1, along with θ^* in Appendix F.

⁴The PyTorch implementation was adapted from MFGLib[31].

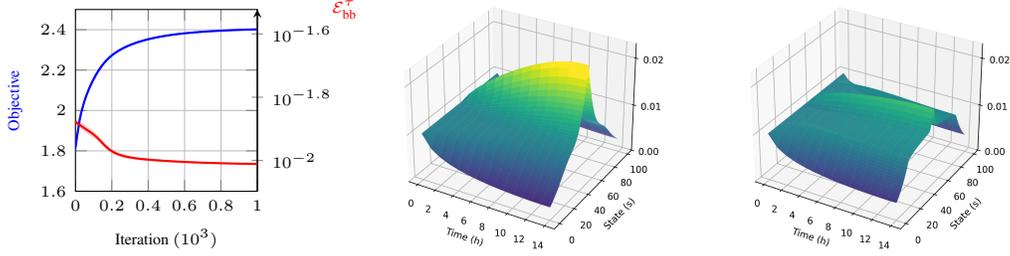


Figure 1: Payment design with AMID in \mathcal{M}_{bb} . **Left:** objective and exploitability throughout training iterations. **Middle-right:** population flow in time before and after learning payments.

Dynamic auction settings. We move to the more challenging but relevant setting of designing neural network mechanisms BA-MFGs. We focus on risk-neutral bidders (u_h linear in payments) and on *direct revelation* mechanisms, i.e. $\mathcal{S} = \mathcal{A}$.⁵ We set $|\mathcal{S}| = 100$, and maximize the revenue objective:

$$g_{\text{rev}}(\theta, \mathbf{L}) := \sum_{h=0}^{H-1} \sum_{(s,a) \in \mathcal{V} \times \mathcal{A}} L_h(s, a) p_{\text{win}}(s, a, L_h, \alpha_h^\theta(\nu^{-\perp}(L_h))) p_h^\theta(a, \nu^{-\perp}(L_h)).$$

The exact settings, labeled (A1)-(A3) are as follows:

- (A1) $H = 4$, $\mu_0 = \text{Uniform}(\mathcal{S})$, $\alpha_{\text{max}} = 0.8$, and single-minded bidders (after winning stay at state \perp) with no evolution in valuations s_h^i otherwise.
- (A2) $H = 4$, $\alpha_{\text{max}} = 0.8$, $\mu_0(s) \propto \gamma^s$ for $\gamma = 0.9$, dynamic values with $w(s'|s) \propto \exp\{- (3s - s')^2 / 2\sigma^2\}$ for $\sigma = 0.2$, bidders are single-minded.
- (A3) $H = 5$, μ_0 uniformly sampled from $\Delta_{\mathcal{S}}$, $\alpha_{\text{max}} \sim \text{Uniform}([0.6, 1.2])$, participants re-enter with probability 0.3 each round.⁶

We parameterize $p_h^\theta, \alpha_h^\theta$ with a residual neural network (architecture clarified in Appendix F) containing $\approx 2 \times 10^5$ parameters, with inputs $\mathbf{e}_h, \nu_h^{-\perp}$, and remaining unsold goods at round h , guaranteeing by parameterization that no more than α_{max} goods are sold in total.

Baselines. We evaluate AMID against several benchmarks. First, we compare against the results of running a simple first-price auction (FIRSTPRICE), i.e., the highest bidders win and pay what they bid, to see how much more revenue we can achieve from optimizing over α_h^θ and p_h^θ . Second, we contrast with various methods without gradient information: two methods using two-point gradient estimators (0-ADAM and 0-SGD respectively), and a 0-order annealing strategy (ANNEAL) using random perturbations of θ . We use $\tau = 10^{-3}$, $\eta = 10$ and $T = 400$ for computing objective G_{approx}^T . We report the training curves in Figure 2, where we evaluate $G_{\text{approx}}^{T_{\text{val}}}$ throughout training $T_{\text{val}} = 500$ for robustness. The results indicate the effectiveness of our method against zeroth-order methods across different auctions. Evaluations on a larger variety of settings and parameterizations (longer horizons H , nonlinear utilities, static mechanisms, other g) are also reported in Appendix F.

Empirically Testing (D1) & (D2). Figure 3 illustrates that we fulfill (D1) & (D2). Notably, the actual revenue in the N player auction is very close to the optimized g_{rev} even for $N \approx 100$. Furthermore, the exploitability curve of OMD iterates at the optimized θ^* also suggests that the iterates are a good approximation of MF-NE, and empirically, the assumptions of Lemma 1 are valid. Namely, OMD iterates produce a valid approximate Nash equilibrium after the end-to-end optimization process with AMID, empirically verifying that the revenue at Nash is indeed optimized.

We further provide empirical evidence supporting (D2) by comparing the computational footprint of AMID against naive backpropagation through the full computational graph induced by OMD. In

⁵The latter choice is motivated by the conceptual simplicity, widespread use in practice, and does not represent a significant restriction given the revelation principle [43].

⁶The setting is more challenging for two reasons. First, the neural mechanism must generalize over α_{max} , which it can observe. Second, it must generalize over all μ_0 , which it does not observe, but potentially infer from $\nu^{-\perp}(L)$. This is also referred to as *prior-free* mechanism design [29, 34]

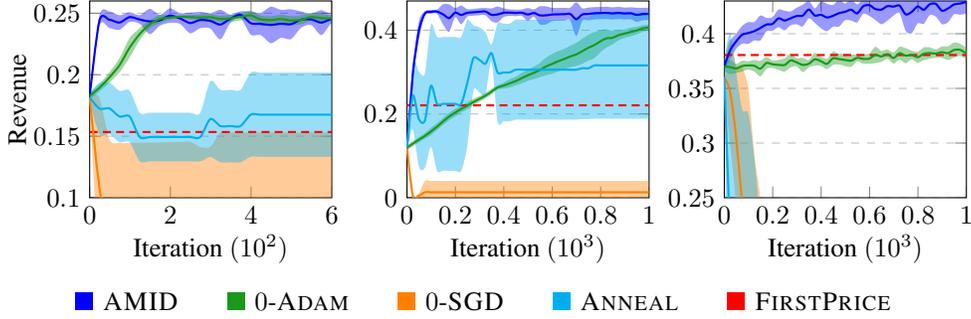


Figure 2: g_{rev} throughout iterations of AMID and baseline algorithms in settings (A1-3), left to right.

Table 2, we report the time and memory usage of the two methods when solving (A1) with increasing time horizons H on a single H100. The results are reported for a single backpropagation step. The modest growth in memory and computation time observed for AMID as H increases highlights the scalability and practical suitability of our methodology for solving large-scale ID problems.

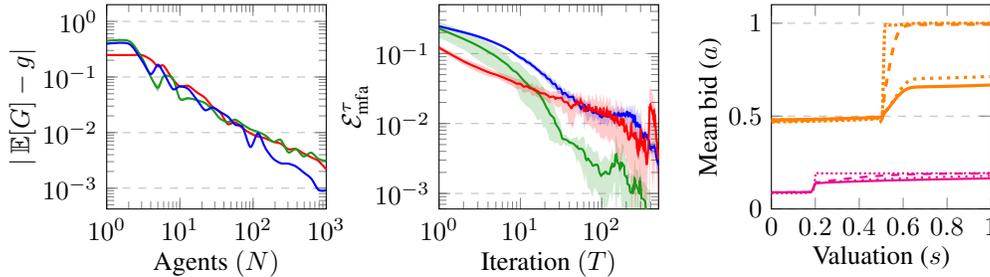


Figure 3: **Left:** deviation in revenue in \mathcal{M}_{mfa} vs N -player \mathcal{G}_{auc} at θ^* as functions of N , and **middle:** exploitability curve of OMD iterations $F_{\text{omd}}^{(T)}(\theta^*, \cdot)$ at optimized θ^* in (A1), (A2), (A3). **Right:** mean bids of NE computed by $F_{\text{omd}}^{(T)}$ for $h \in [4]$ before and after optimization with AMID in (A1).

Table 2: Empirical compute time and memory usage for single-step naive backpropagation vs. single-step AMID across different problem horizons, in setting (A1). The rows with “n/a” indicate the method did not run on a single H100.

Horizon	Naive (time, s)	Naive (memory)	AMID (time, s)	AMID (memory)
$H=5$	0.19 ± 0.02 s	2760 MiB	0.09 ± 0.01 s	560 MiB
$H=10$	0.25 ± 0.10 s	8746 MiB	0.21 ± 0.08 s	586 MiB
$H=25$	0.71 ± 0.15 s	16960 MiB	0.67 ± 0.12 s	826 MiB
$H=50$	n/a	n/a	1.72 ± 0.41 s	1076 MiB

5 Conclusion

In this work, we presented a novel method for ID relying on PMFGs. In particular, we set forth two desiderata in order to use scalable first-order optimization to approximately solve ID problems. Through new analyses, we demonstrated that these conditions hold in both classical Mean Field Game (MFG) settings and batched auction environments. For both settings, we presented a unified algorithm, called AMID, which can solve a span of ID problems, such as congestion pricing or optimal auction design. Overall, the AMID framework offers a flexible foundation for diverse incentive design applications, paving the way for future extensions.

Acknowledgments and Disclosure of Funding

This project was supported by Swiss National Science Foundation (SNSF) under the framework of NCCR Automation and SNSF Starting Grant. V. Thoma acknowledges funding from the Swiss National Science Foundation (SNSF) Project Funding No. 200021-207343 and is supported by an ETH AI Center Doctoral Fellowship.

References

- [1] Alexander Aurell, Rene Carmona, Gokce Dayanikli, and Mathieu Lauriere. Optimal incentives to mitigate epidemics: a stackelberg mean field game approach. *SIAM Journal on Control and Optimization*, 60(2):S294–S322, 2022.
- [2] Santiago R. Balseiro, Omar Besbes, and Gabriel Y. Weintraub. Repeated auctions with budgets in ad exchanges: Approximations and design. *Management Science*, 61(4):864–884, 2015.
- [3] Tamer Basar and Rayadurgam Srikant. Revenue-maximizing pricing and capacity expansion in a many-users regime. In *Proceedings. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies*, volume 1, pages 294–301. IEEE, 2002.
- [4] Alain Bensoussan, Michael HM Chau, and Sheung Chi Phillip Yam. Mean field stackelberg games: Aggregation of delayed instructions. *SIAM Journal on Control and Optimization*, 53(4):2237–2266, 2015.
- [5] Alain Bensoussan, Jens Frehse, Phillip Yam, et al. *Mean field games and mean field type control theory*, volume 101. Springer, 2013.
- [6] Martin Bichler, Maximilian Fichtl, Stefan Heidekrüger, Nils Kohring, and Paul Sutterer. Learning equilibria in symmetric auction games using artificial neural networks. *Nature Machine Intelligence*, 3(8):687–695, August 2021.
- [7] Martin Bichler, Maximilian Fichtl, and Matthias Oberlechner. Computing bayes nash equilibrium strategies in auction games via simultaneous online dual averaging. In *Proceedings of the 24th ACM Conference on Economics and Computation*, EC ’23, page 294, New York, NY, USA, 2023. Association for Computing Machinery.
- [8] Vitor Bosshard, Benedikt Bünz, Benjamin Lubin, and Sven Seuken. Computing Bayes-Nash equilibria in combinatorial auctions with continuous value and action spaces. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, Melbourne, Australia, August 2017.
- [9] Vitor Bosshard, Benedikt Bünz, Benjamin Lubin, and Sven Seuken. Computing bayes-nash equilibria in combinatorial auctions with verification. *Journal of Artificial Intelligence Research*, 69:531–570, 2020.
- [10] Yang Cai and Christos Papadimitriou. Simultaneous bayesian auctions and computational complexity. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, EC ’14, page 895–910, New York, NY, USA, 2014. Association for Computing Machinery.
- [11] Ilayda Canyakmaz, Iosif Sakos, Wayne Lin, Antonios Varvitsiotis, and Georgios Piliouras. Learning and steering game dynamics towards desirable outcomes. In *Proceedings of the 7th Annual Learning for Dynamics & Control Conference*, volume 283 of *Proceedings of Machine Learning Research*, pages 1512–1524. PMLR, 04–06 Jun 2025.
- [12] René Carmona and Gökçe Dayanıklı. Mean field game model for an advertising competition in a duopoly. *International Game Theory Review*, 23(04):2150024, 2021.
- [13] René Carmona, Gökçe Dayanıklı, and Mathieu Laurière. Mean field models to regulate carbon emissions in electricity production. *Dynamic Games and Applications*, 12(3):897–928, 2022.
- [14] René Carmona and François Delarue. Probabilistic analysis of mean-field games. *SIAM Journal on Control and Optimization*, 51(4):2705–2734, 2013.

- [15] René Carmona, Mathieu Laurière, and Zongjun Tan. Model-free mean-field reinforcement learning: mean-field mdp and mean-field q-learning. *The Annals of Applied Probability*, 33(6B):5334–5381, 2023.
- [16] Siyu Chen, Donglin Yang, Jiayang Li, Senmiao Wang, Zhuoran Yang, and Zhaoran Wang. Adaptive model design for Markov decision process. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 3679–3700. PMLR, 17–23 Jul 2022.
- [17] Xi Chen, Xiaotie Deng, and Shang-Hua Teng. Settling the complexity of computing two-player nash equilibria. *Journal of the ACM (JACM)*, 56(3):1–57, 2009.
- [18] Vincent Conitzer and Tuomas Sandholm. Self-interested automated mechanism design and implications for optimal combinatorial auctions. In *Proceedings of the 5th ACM Conference on Electronic Commerce, EC '04*, page 132–141, New York, NY, USA, 2004. Association for Computing Machinery.
- [19] Kai Cui and Heinz Koepl. Approximately solving mean field games via entropy-regularized deep reinforcement learning. In *International Conference on Artificial Intelligence and Statistics*, pages 1909–1917. PMLR, 2021.
- [20] Michael Curry, Vinzenz Thoma, Darshan Chakrabarti, Stephen McAleer, Christian Kroer, Tuomas Sandholm, Niao He, and Sven Seuken. Automated design of affine maximizer mechanisms in dynamic settings. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(9):9626–9635, March 2024.
- [21] Michael Curry, Alexander R Trott, Soham Phade, Yu Bai, and Stephan Zheng. Finding general equilibria in many-agent economic simulations using deep reinforcement learning, 2022.
- [22] Constantinos Daskalakis, Paul W Goldberg, and Christos H Papadimitriou. The complexity of computing a nash equilibrium. *Communications of the ACM*, 52(2):89–97, 2009.
- [23] Gokce Dayanikli and Mathieu Lauriere. A machine learning method for stackelberg mean field games. *arXiv preprint arXiv:2302.10440*, 2023.
- [24] Greg d’Eon, Neil Newman, and Kevin Leyton-Brown. Understanding iterative combinatorial auction designs via multi-agent reinforcement learning. In *Proceedings of the 25th ACM Conference on Economics and Computation, EC '24*, page 1102–1130, New York, NY, USA, 2024. Association for Computing Machinery.
- [25] Stylianos Despotakis, R. Ravi, and Amin Sayedi. First-price auctions in online display advertising. *Journal of Marketing Research*, 58(5):888–907, October 2021.
- [26] Tanner Fiez, Benjamin Chasnov, and Lillian Ratliff. Implicit learning dynamics in stackelberg games: Equilibria characterization, convergence analysis, and empirical study. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 3133–3144. PMLR, 13–18 Jul 2020.
- [27] Ian Gale and Mark Stegeman. Sequential auctions of endogenously valued objects. *Games and Economic Behavior*, 36:74–103, 07 2001.
- [28] Matthias Gerstgrasser and David C. Parkes. Oracles & followers: Stackelberg equilibria in deep multi-agent reinforcement learning. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 11213–11236. PMLR, 23–29 Jul 2023.
- [29] Andrew V. Goldberg, Jason D. Hartline, and Andrew Wright. Competitive auctions and digital goods. page 735 – 744, 2001. Cited by: 199.

- [30] Amy Greenwald, Jiawei Li, and Eric Sodomka. Approximating equilibria in sequential auctions with incomplete information and multi-unit demand. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.
- [31] X. Guo, A. Hu, M. Santamaria, M. Tajrobehkar, and J. Zhang. MFGLib: A library for mean field games. *arXiv preprint arXiv:2304.08630*, 2023.
- [32] Xin Guo, Anran Hu, Renyuan Xu, and Junzi Zhang. Learning mean-field games. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.
- [33] Xin Guo, Lihong Li, Sareh Nabi, Rabih Salhab, and Junzi Zhang. Mesob: Balancing equilibria & social optimality, 2023.
- [34] Jason D. Hartline and Tim Roughgarden. Optimal mechanism design and money burning. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, STOC '08, page 75–84, New York, NY, USA, 2008. Association for Computing Machinery.
- [35] Jiawei Huang, Vinzenz Thoma, Zebang Shen, Heinrich H. Nax, and Niao He. Learning to steer markovian agents under model uncertainty. In *The Thirteenth International Conference on Learning Representations*, 2025.
- [36] Minyi Huang, Roland P Malhamé, and Peter E Caines. Large population stochastic dynamic games: closed-loop mckean-vlasov systems and the nash certainty equivalence principle. *Communications in Information & Systems*, 6(3):221–252, 2006.
- [37] Noboru Isobe, Kenshi Abe, and Kaito Ariu. Last iterate convergence in monotone mean field games. *arXiv preprint arXiv:2410.05127*, 2024.
- [38] Krishnamurthy Iyer, Ramesh Johari, and Mukund Sundararajan. Mean field equilibria of dynamic auctions with learning. *Management Science*, 60(12):2949–2970, dec 2014.
- [39] Robert G. Jeroslow. The polynomial hierarchy and a simple model for competitive analysis. *Mathematical Programming*, 32(2):146–164, June 1985.
- [40] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.
- [41] Paul Klemperer. How (not) to run auctions: The european 3g telecom auctions. *European Economic Review*, 46(4–5):829–845, May 2002.
- [42] Nils Kohring, Fabian Raoul Pieroth, and Martin Bichler. Enabling first-order gradient-based learning for equilibrium computation in markets. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 17327–17342. PMLR, 23–29 Jul 2023.
- [43] Vijay Krishna. *Auction Theory*. Academic Press, 2002.
- [44] Jean-Michel Lasry and Pierre-Louis Lions. Mean field games. *Japanese journal of mathematics*, 2(1):229–260, 2007.
- [45] Mathieu Laurière, Sarah Perrin, Julien Pérolat, Sertan Girgin, Paul Muller, Romuald Élie, Matthieu Geist, and Olivier Pietquin. Learning in mean field games: A survey. *arXiv preprint arXiv:2205.12944*, 2022.
- [46] Ana Ley, Winnie Hu, and Keith Collins. Less traffic, faster buses: Congestion pricing’s first week. *The New York Times*, January 2025. Accessed: 2025-15-01.
- [47] Jiayang Li, Jing Yu, Boyi Liu, Yu Nie, and Zhaoran Wang. Achieving hierarchy-free approximation for bilevel programs with equilibrium constraints. In *Proceedings of the 40th International Conference on Machine Learning, ICML'23, Honolulu, Hawaii, USA, 2023*. JMLR.org.

- [48] Jiayang Li, Jing Yu, Yu Nie, and Zhaoran Wang. End-to-end learning and intervention in games. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 16653–16665. Curran Associates, Inc., 2020.
- [49] Anton Likhodedov and Tuomas Sandholm. Methods for boosting revenue in combinatorial auctions. In *Proceedings of the 19th National Conference on Artificial Intelligence, AAAI’04*, page 232–237, San Jose, California, 2004. AAAI Press.
- [50] Boyi Liu, Jiayang Li, Zhuoran Yang, Hoi-To Wai, Mingyi Hong, Yu Nie, and Zhaoran Wang. Inducing equilibria via incentives: Simultaneous design-and-play ensures global convergence. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 29001–29013. Curran Associates, Inc., 2022.
- [51] Zhi-Quan Luo, Jong-Shi Pang, and Daniel Ralph. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, nov 1996.
- [52] Sabita Maharjan, Quanyan Zhu, Yan Zhang, Stein Gjessing, and Tamer Basar. Dependable demand response management in the smart grid: A stackelberg game approach. *IEEE Transactions on Smart Grid*, 4(1):120–132, 2013.
- [53] Carlos Martin and Tuomas Sandholm. Finding mixed-strategy equilibria of continuous-action games without gradients using randomized policy networks. In Edith Elkind, editor, *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, pages 2844–2852. International Joint Conferences on Artificial Intelligence Organization, 8 2023. Main Track.
- [54] Jun Moon and Tamer Başar. Linear quadratic mean field stackelberg differential games. *Automatica*, 97:200–213, 2018.
- [55] Roger B. Myerson. Optimal auction design. *Mathematics of Operations Research*, 6(1):58–73, 1981.
- [56] Julien Pérolat, Sarah Perrin, Romuald Elie, Mathieu Laurière, Georgios Piliouras, Matthieu Geist, Karl Tuyls, and Olivier Pietquin. Scaling mean field games by online mirror descent. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*, pages 1028–1037, 2022.
- [57] Sarah Perrin, Mathieu Laurière, Julien Pérolat, Romuald Élie, Matthieu Geist, and Olivier Pietquin. Generalization in mean field games by learning master policies. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 9413–9421, 2022.
- [58] Sarah Perrin, Julien Pérolat, Mathieu Laurière, Matthieu Geist, Romuald Elie, and Olivier Pietquin. Fictitious play for mean field games: Continuous time analysis and applications. *Advances in Neural Information Processing Systems*, 33:13199–13213, 2020.
- [59] Fabian R. Pieroth, Nils Kohring, and Martin Bichler. Equilibrium computation in multi-stage auctions and contests. *ArXiv*, abs/2312.11751, 2023.
- [60] Zinovi Rabinovich, Victor Naroditskiy, Enrico H. Gerding, and Nicholas R. Jennings. Computing pure bayesian-nash equilibria in games with finite actions and continuous types. *Artificial Intelligence*, 195:106–139, February 2013.
- [61] Daniel M. Reeves and Michael P. Wellman. Computing best-response strategies in infinite games of incomplete information. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence, UAI ’04*, page 470–478, Arlington, Virginia, USA, 2004. AUAI Press.
- [62] Tim Roughgarden. Transaction fee mechanism design. In *Proceedings of the 22nd ACM Conference on Economics and Computation, EC ’21*, page 792, New York, NY, USA, 2021. Association for Computing Machinery.
- [63] Naci Saldi, Tamer Basar, and Maxim Raginsky. Markov–nash equilibria in mean-field games with discounted cost. *SIAM Journal on Control and Optimization*, 56(6):4256–4287, 2018.

- [64] Sina Sanjari, Subhonmesh Bose, and Tamer Başar. Incentive designs for stackelberg games with a large number of followers and their mean-field limits. *Dynamic Games and Applications*, 15(1):238–278, 2025.
- [65] Vinzenz Thoma, Vitor Bosshard, and Sven Seuken. Computing perfect bayesian equilibria in sequential auctions with verification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(13):14158–14166, April 2025.
- [66] Vinzenz Thoma, Michael Curry, Niao He, and Sven Seuken. Learning best response policies in dynamic auctions via deep reinforcement learning, 2023.
- [67] Vinzenz Thoma, Barna Pasztor, Andreas Krause, Giorgia Ramponi, and Yifan Hu. Contextual bilevel reinforcement learning for incentive alignment. In *Advances in Neural Information Processing Systems*, volume 37, pages 127369–127435, 2024.
- [68] Yevgeniy Vorobeychik and Michael P. Wellman. Stochastic search methods for nash equilibrium approximation in simulation-based games. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 2, AAMAS '08*, page 1055–1062, Richland, SC, 2008. International Foundation for Autonomous Agents and Multiagent Systems.
- [69] Jing Wang, Meichen Song, Feng Gao, Boyi Liu, Zhaoran Wang, and Yi Wu. Differentiable arbitrating in zero-sum markov games. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems, AAMAS '23*, page 1034–1043, Richland, SC, 2023. International Foundation for Autonomous Agents and Multiagent Systems.
- [70] Kai Wang, Lily Xu, Andrew Perrault, Michael K. Reiter, and Milind Tambe. Coordinating followers to reach better equilibria: End-to-end gradient descent for stackelberg games. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(5):5219–5227, June 2022.
- [71] Lingxiao Wang, Zhuoran Yang, and Zhaoran Wang. Breaking the curse of many agents: Provable mean embedding q-iteration for mean-field reinforcement learning. In *International conference on machine learning*, pages 10092–10103. PMLR, 2020.
- [72] Leo Widmer, Jiawei Huang, and Niao He. Steering no-regret agents in mfgs under model uncertainty, 2025.
- [73] Batuhan Yardim, Semih Cayci, Matthieu Geist, and Niao He. Policy mirror ascent for efficient and independent learning in mean field games. In *International Conference on Machine Learning*, pages 39722–39754. PMLR, 2023.
- [74] Batuhan Yardim, Semih Cayci, and Niao He. A variational inequality approach to independent learning in static mean-field games. *ACM / IMS Journal of Data Science*, 2(2), July 2025.
- [75] Batuhan Yardim, Artur Goldman, and Niao He. When is mean-field reinforcement learning tractable and relevant? In *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems, AAMAS '24*, page 2038–2046, Richland, SC, 2024. International Foundation for Autonomous Agents and Multiagent Systems.
- [76] Batuhan Yardim and Niao He. Exploiting approximate symmetry for efficient multi-agent reinforcement learning. In *Proceedings of the 7th Annual Learning for Dynamics & Control Conference*, volume 283 of *Proceedings of Machine Learning Research*, pages 31–44. PMLR, 04–06 Jun 2025.
- [77] Muhammad Aneeq Uz Zaman, Alec Koppel, Sujay Bhatt, and Tamer Basar. Oracle-free reinforcement learning in mean-field games along a single sample path. In *International Conference on Artificial Intelligence and Statistics*, pages 10178–10206. PMLR, 2023.
- [78] Brian Hu Zhang, Gabriele Farina, Ioannis Anagnostides, Federico Cacciamani, Stephen McAleer, Andreas Haupt, Andrea Celli, Nicola Gatti, Vincent Conitzer, and Tuomas Sandholm. Steering no-regret learners to a desired equilibrium. In *Proceedings of the 25th ACM Conference on Economics and Computation, EC '24*, page 73–74, New York, NY, USA, 2024. Association for Computing Machinery.

- [79] Fengzhuo Zhang, Vincent YF Tan, Zhaoran Wang, and Zhuoran Yang. Learning regularized monotone graphon mean-field games. *arXiv preprint arXiv:2310.08089*, 2023.
- [80] Stephan Zheng, Alexander Trott, Sunil Srinivasa, David C. Parkes, and Richard Socher. The ai economist: Taxation policy design via two-level deep multiagent reinforcement learning. *Science Advances*, 8(18), May 2022.

NeurIPS Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [Yes]

Justification: The claimed convergence rate is proven in the paper and the experimental results are accurately described.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [Yes]

Justification: We state our assumptions in Section 2.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: See Assumption 2 and Assumption 1 and the extensive appendix with all proofs.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: We introduce and formally define our approximation framework in Definition 2 and describe the algorithmic procedure using pseudocode in Algorithm 1. The experimental setting (Batched Auction MFG) (Definition 3) and objectives are precisely specified. The various experiments where we apply our framework are clearly defined in Section 4, and full details are presented in the appendix.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
 - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: We provide code with both JAX and PyTorch implementations, along with instructions to execute the experiments. All experiments are seeded to ensure reproducibility. The main experiments are implemented using the JAX codebase.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: All experimental settings are documented in Section 4. Details on the hyperparameters and experimental setup are provided in Appendix F, and the accompanying code includes the full configuration used for each experiment.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: All experiments include error bars.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: We provide detailed information about computing resources in Appendix F.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [Yes]

Justification: We use neither human subjects nor sensitive data, and we do not foresee dual use of our work.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [NA]

Justification: Not applicable, no immediate social impact.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.

- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: Not applicable.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: We used JAX, PyTorch, and adapted code from MFGLib. No additional third-party datasets or pretrained models were used. Everything is properly cited.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: Not applicable.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: Not applicable.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: Not applicable.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. Declaration of LLM usage

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: Not applicable.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.

Contents

1	Introduction	1
1.1	Related Works	3
2	Designing Games for Large Populations: Lipschitz Case	3
2.1	Parameterized Mean Field Game Design	4
2.2	Approximating the First Order Derivatives	5
3	Beyond Lipschitz: Mechanism Design for Large-Scale Auctions	6
3.1	Auctions at the Mean-Field Regime	7
4	Experimental Results	8
5	Conclusion	10
A	Frequently-Used Notation	25
B	Extended Related Works	26
C	Automated Mechanism Design as ID	26
D	Mean-Field Mechanism Design	27
D.1	Preliminary Lemmas	27
D.2	Proof of Theorem 1	28
D.3	Proof of Lemma 1	28
D.4	Discussion of Assumptions of Lemma 1	29
D.5	Proof of Lemma 2	29
D.6	Extension of Lemma 2 to General Mirror Descent	31
E	Results on Batched Auctions	31
E.1	Extended Definitions	31
E.2	Preliminary Lemmas	34
E.3	Proof of Theorem 2, part 1 (Approximation in Exploitability)	37
E.4	Proof of Theorem 2, part 2 (Approximation in Objective)	49
E.5	Proof of Lemma 3	50
F	Experiment Details	54
F.1	Additional Results on the Beach Bar Process	55
F.2	Additional Results on Auctions	57

A Frequently-Used Notation

General Notation

$\mathcal{H}(u)$	$:= -\sum_x u(x) \log u(x)$, entropy
$\text{sigmoid}(x)$	$:= \frac{1}{1+e^{-x}}$, sigmoid function
∇f	$\in \mathbb{R}^{d_1 \times d_2}$, Jacobian of function $f : \mathbb{R}^{d_2} \rightarrow \mathbb{R}^{d_1}$
\mathbb{S}_{D-1}	$:= \{x \in \mathbb{R}^D : \ x\ = 1\}$, $(D-1)$ -dimensional unit sphere in \mathbb{R}^D
\mathbb{B}_D	$:= \{x \in \mathbb{R}^D : \ x\ \leq 1\}$, D -dimensional unit closed ball in \mathbb{R}^D
Var	variance of random variable
\mathbf{e}_i	standard unit vector with i -th entry 1.
$\Delta_{\mathcal{X}}$	$:= \{u \in \mathbb{R}^{\mathcal{X}} : \sum_x u_x = 1, u_x \geq 0\}$, probability simplex on \mathcal{X} .
$D_{\text{kl}}(u v)$	$:= \sum_x u(x) \log u(x)/v(x)$, Kullback–Leibler divergence
$\langle \cdot, \cdot \rangle$	dot product.
$\mathbb{1}$	indicator function.
$\ \cdot\ _1$	ℓ_1 norm (on Euclidean space \mathbb{R}^D)
$\ \cdot\ _2$	ℓ_2 norm (on Euclidean space \mathbb{R}^D)
$\otimes_{i \in [N]} m^i$	product measure
$\sigma(\mathbf{x})$	empirical counts of entries of some $\mathbf{x} \in \mathcal{X}^N$, $\sigma(\mathbf{x}) := 1/N \sum_{i=1}^N \mathbf{e}_{x_i}$
$\rho(A)$	spectral radius of matrix $A \in \mathbb{R}^{D,D}$
$\text{Marg}_{\mathcal{Y}}(d)$	$:= \sum_{x \in \mathcal{X}} d(x, \cdot) \in \Delta_{\mathcal{Y}}$, for $d \in \Delta_{\mathcal{X} \times \mathcal{Y}}$.
$\ A\ _{p \rightarrow q}$	$:= \sup_{\ x\ _p \leq 1} \ Ax\ _q$, operator norm of A with norms $\ \cdot\ _p, \ \cdot\ _q$

Generic PMFGs

\mathcal{G}	parameterized MFG
H	horizon (number of rounds) of auction
\mathcal{S}	(finite) state space
\mathcal{A}	(finite) action space
N	number of players/agents
g	ID objective
G	$:= \mathbb{E}[g(\theta, \widehat{\mathbf{L}})]$, true ID objective in N -player game
$\text{Nash}^\tau(\mathcal{G})$	set of τ -regularized Nash equilibria
Θ	parameter space
θ	$\in \Theta$, ID design parameter
Π_H	$\{\pi : [H] \times \mathcal{S} \rightarrow \Delta_{\mathcal{A}}\}$, set of finite-horizon Markovian policies.
μ_0	initial state distribution
$\overline{P}_{h,\theta}$	parameterized state transition dynamics in N -player DG
$\overline{R}_{h,\theta}$	parameterized reward functions in N -player DG
τ	$\in \mathbb{R}$ entropy regularization magnitude
$\boldsymbol{\pi}$	$\in \Pi_H^N$, N -tuples of policies
$J_{\mathcal{G}}^{\tau,i}$	expected reward of player i in dynamic game \mathcal{G} (see Definition 1)
$\mathcal{E}_{\mathcal{G}}^{\tau,i}$	exploitability of player i in dynamic game \mathcal{G} (see Definition 1)
$\mathcal{E}_{\mathcal{G}}^\tau$	maximum exploitability in dynamic game \mathcal{G} (see Definition 1)
\mathcal{M}	parameterized mean-field game (PMFG)
$P_{h,\theta}$	parameterized state transition dynamics in PMFG
$R_{h,\theta}$	parameterized reward functions in PMFG
Γ_h	one-step MFG forward flow (see Definition 2)
Λ	maps policies in Π_H to H -step mean-field population flow in $\Delta_{\mathcal{S} \times \mathcal{A}}^H$ (see Definition 2)

Adjoint method:

V_h^τ	state value function in the PMFG
q_h^τ	q-value value function (on state-action pairs) in the PMFG
F	$:\mathbb{R}^{[H] \times \mathcal{S} \times \mathcal{A}} \times \Theta \rightarrow \mathbb{R}^{[H] \times \mathcal{S} \times \mathcal{A}}$ generic policy update operator for computing NE, defined in log policy space
$\overline{F}_{\text{omd}}$	mirror descent update, in policy space
F_{omd}	mirror descent update, in log-policy space

For auctions:

\mathcal{G}_{auc}	batched Auction
\perp	inactive state for agent not participating in the current round
\mathcal{V}	value space
α_{max}	parameter for maximal amount of goods
α_h^θ	parametrized item allocation function
p_h^θ	parametrized payment function
u_h	utility function of bidders
w_h	valuation dynamics function
$P_{h,\theta}^{\text{auc}}$	transition dynamics in batched auction
$R_{h,\theta}^{\text{auc}}$	reward functions
$J_{\text{auc}}^{\tau,i}$	entropy regularized (τ) sum of rewards for agent $i \in [N]$
$\mathcal{E}_{\text{auc}}^{\tau,i}$	entropy regularized (τ) exploitability for agent $i \in [N]$
\mathcal{M}_{mfa}	batched auction MFG
$P_{h,\theta}^{\text{mfa}}$	BA-MFG transition dynamics
$R_{h,\theta}^{\text{mfa}}$	BA-MFG reward functions
$\nu^{-\perp}$	action marginal of active (non- \perp) states; maps state-action distributions to sub-probability distributions over actions
p_{win}	winning probability function, given a bid a , sold goods α and population bid distribution $\nu^{-\perp}$
$\xi^{L,\theta}$	post allocation state distribution
V_{mfa}^τ	entropy regularized sum of rewards for BA-MFG
$\mathcal{E}_{\text{mfa}}^\tau$	entropy regularized exploitability of rewards for BA-MFG
Λ_{mfa}	population operator for BA-MFG
g_{rev}	revenue objective

B Extended Related Works

Related works on Equilibrium Computation in Auctions. As outlined in Appendix C, many real-world auctions are not strategyproof. It is thus important to evaluate them at equilibrium—both from a predictive (how bidders will likely behave), as well as from a normative (bidding recommendations) standpoint. While some simple formats have been solved analytically [43], existing hardness results for computing exact equilibria in auctions [10] motivate approximate computational approaches. In single-round auctions, a strand of work has used iterated best-response computations to calculate equilibria [61, 68, 60, 8, 9]. Other approaches rely on gradient descent [7, 42] or deep learning [6, 53]. For multi-round auctions, [65] compute ε -perfect Bayesian equilibria, using best response dynamics. Others have used (deep) RL to find approximate Nash equilibria [30, 59, 24, 66]. In our work, by using mean-field approximations, we circumvent the curse of dimensionality inherent to these multi-agent RL approaches. Using mean-field approaches to solve auctions has been explored previously to some extent. [2, 32] study specific repeated ad auctions with budget constraints, and [38] studies dynamic auctions, where bidders iteratively learn about their own type.

Other MFG works. Stackelberg equilibria for MFGs have also been studied in the particular case of linear-quadratic models [54, 4]. Another relevant model in this setting is mean-field incentive design with major and minor players [64], where designing incentives for a leader is studied for the purpose of influencing a population. In continuous time, Stackelberg MFGs have been studied in applications such as regulating carbon markets [13], epidemics [1], and advertising markets [12].

C Automated Mechanism Design as ID

We note that in contrast to our approach many works on automated mechanism design focus on designing strategyproof auctions, i.e. auctions in which bidders bid truthful in equilibrium, relying on the so-called *revelation principle*[20, 18, 49]. The revelation principle states that any non-strategyproof equilibrium of an auction can be implemented as an outcome equivalent strategyproof equilibrium of an adapted auction [43]. Restricting to strategyproof mechanisms bypasses the need

to differentiate through an equilibrium. Instead of a problem like (ID) where the outer objective depends on an inner equilibrium solution, the inner solution is already known—bid truthfully—and instead the problem becomes one of constrained optimization problem, where the so-called *incentive compatibility* (IC) constraints ensure that bidding truthfully is in fact an equilibrium [18].

While restricting to strategyproof mechanisms foregoes the need to differentiate through an equilibrium, many real-world auctions are not strategyproof. In fact, in 2019 Google for example deliberately changed towards non-strategyproof first price auctions for selling ads, citing the increased transparency of simple, non-strategyproof format for the bidders [25]. In such cases without IC constraints, the question is how bidders will respond in equilibrium and in turn designing revenue-optimal auctions becomes an instance of (ID), which we tackle in Sections 3 and 4.

D Mean-Field Mechanism Design

D.1 Preliminary Lemmas

Theorem 3 (Rademacher) *Let $U \subset \mathbb{R}^m$ be an open set and $f : U \rightarrow \mathbb{R}^n$ be a Lipschitz continuous map. Then, f is almost everywhere differentiable on U , that is, the points on which f is not differentiable on U for a set of measure 0.*

In some cases, an explicit differentiability assumption might be useful for PMFG dynamics, which we state below.

Assumption 2 (Differentiability) *For all $s, s' \in \mathcal{S}$, $a \in \mathcal{A}$, the functions $P_\theta(s'|s, a, L)$, $R_\theta(s, a, L)$ are differentiable on θ, L . Furthermore, $g(\theta, \mathbf{L})$ is differentiable on θ, \mathbf{L} with bounded derivatives.*

In particular, the following simple result is useful for the derivation of AMID.

Lemma 4 (Differentiability of operators) *The maps $\Gamma, \Lambda, q_h^\tau, V_h^\tau, \bar{F}_{omd}$ as well as the map $\theta, \pi \rightarrow g(\theta, \Lambda(\pi))$ are almost everywhere differentiable under Assumption 1 and differentiable everywhere under Assumption 2.*

Proof: This result is a straightforward result of the definitions of the mentioned operators, in particular, when Assumption 2 is taken, the mentioned operators are also differentiable as they are the compositions of differentiable functions. In the case where only Assumption 1 holds, the above-mentioned functions are also Lipschitz on every bounded domain, which implies by Theorem 3 that they are almost everywhere differentiable. \square

Finally, we state the following standard lemma from past work on the approximation of MFG dynamics by finite player games.

Lemma 5 *Assume that the conditions of Theorem 1 hold for the PMFG \mathcal{M} and DG \mathcal{G} , let $\pi, \bar{\pi} \in \Pi_H$ be two arbitrary policies, and let $\theta \in \Theta$ be fixed. Let $\mathbf{L}^{\bar{\pi}} = \{L_h^{\bar{\pi}}\}_h = \Lambda(\bar{\pi}|\theta)$ be the population flow induced by $\bar{\pi}$ on the PMFG with fixed parameter θ . Take the trajectories s_h^i, a_h^i, \hat{L}_h induced by the DG with parameter θ and policy profile $(\pi, \bar{\pi}, \dots, \bar{\pi}) \in \Pi_H^N$. Then,*

$$\mathbb{E}[\|\hat{L}_h - L_h^{\bar{\pi}}\|_1] \leq \frac{1 - L_{pop,\mu}^{h+1}}{1 - L_{pop,\mu}} |\mathcal{S}| |\mathcal{A}| \sqrt{\frac{2}{N}} + \frac{1}{N} \sum_{i=0}^{h-1} L_{pop,\mu}^{h-i-1} \Delta_h,$$

where $\Delta_h := \sup_s \|\bar{\pi}_h(\cdot|s) - \pi_h(\cdot|s)\|_1$, and $L_{pop,\mu}$ is a uniform bound on the Lipschitz moduli of Γ_h in L . Furthermore, denoting the random variables s_h, a_h as the distributions of state-action pairs in the PMFG dynamics with population flow $\mathbf{L}^{\bar{\pi}}$ and policy π ,

$$\|\mathbb{P}[s_h = \cdot, a_h = \cdot] - \mathbb{P}[s_h^1 = \cdot, a_h^1 = \cdot]\|_1 \leq K_L \sum_{h'=0}^h \mathbb{E}[\|\hat{L}_h - L_h^{\bar{\pi}}\|_1],$$

where K_L is the Lipschitz modulus of transition dynamics $P_{\theta,h}$ in L .

Proof: The proof is a straightforward extension of the approximation results due to [75], to the case where transition dynamics and rewards also depend on the mean-field flow over actions. See in particular Theorem 3.2 in [75]. \square

D.2 Proof of Theorem 1

First, we present the bound on exploitability. As in the setting of Lemma 5, let $\mathbf{L}^{\bar{\pi}} = \{L_h^{\bar{\pi}}\}_h = \Lambda(\bar{\pi}|\theta)$ be the population flow induced by $\bar{\pi}$ on the PMFG with fixed parameter θ . Take the trajectories $s_h^i, a_h^i, \widehat{L}_h$ induced by the DG with parameter θ and policy profile $\boldsymbol{\pi} := (\pi, \bar{\pi}, \dots, \bar{\pi}) \in \Pi_H^N$. Similarly, denote the random variables s_h, a_h as the distributions of state-action pairs in the PMFG dynamics with population flow $\mathbf{L}^{\bar{\pi}}$ and policy π .

$$\begin{aligned} & |V_{\mathcal{M}}^{\tau}(\mathbf{L}^{\bar{\pi}}, \pi|\theta) - J_{\mathcal{G}}^{\tau,1}(\boldsymbol{\pi}|\theta)| \\ &= \left| \sum_{h,s,a} \mathbb{P}[s_h = s, a_h = a] (R_{h,\theta}(L_h^{\bar{\pi}}, s, a|\theta) + \tau \mathcal{H}(\bar{\pi}_h(s))) \right. \\ &\quad \left. - \sum_{h,s,a,L} \mathbb{P}[s_h = s, a_h = a, \widehat{L}_h = L] (R_{h,\theta}(L, s, a|\theta) + \tau \mathcal{H}(\bar{\pi}_h(s))) \right| \\ &\leq \left| \sum_{h,s,a,L} \mathbb{P}[s_h = s, a_h = a, \widehat{L}_h = L] (R_{h,\theta}(L, s, a|\theta) - R_{h,\theta}(L_h^{\bar{\pi}}, s, a|\theta)) \right| \\ &\quad + \left| \sum_{h,s,a} (\mathbb{P}[s_h^i = s, a_h^i = a] - \mathbb{P}[s_h = s, a_h = a]) (R_{h,\theta}(L_h^{\bar{\pi}}, s, a|\theta) + \tau \mathcal{H}(\bar{\pi}_h(s))) \right|. \end{aligned}$$

Since $|R_{h,\theta}(L_h^{\bar{\pi}}, s, a|\theta) + \tau \mathcal{H}(\bar{\pi}_h(s))| \leq 1 + \tau \log |\mathcal{A}|$, and $R_{h,\theta}$ is Lipschitz in L (say with modulus K), it holds that

$$\begin{aligned} & |V_{\mathcal{M}}^{\tau}(\mathbf{L}^{\bar{\pi}}, \pi|\theta) - J_{\mathcal{G}}^{\tau,1}(\boldsymbol{\pi}|\theta)| \\ &\leq K \sum_h \mathbb{E}[\|\widehat{L}_h - L_h^{\bar{\pi}}\|_1] + \sum_h \|\mathbb{P}[s_h = \cdot, a_h = \cdot] - \mathbb{P}[s_h^1 = \cdot, a_h^1 = \cdot]\|_1 (1 + \tau \log |\mathcal{A}|) \leq \mathcal{O}(1/\sqrt{N}), \end{aligned}$$

by an application of Lemma 5. The corresponding bound on $\mathcal{E}_{\mathcal{G}}^{\tau}(\boldsymbol{\pi}^*|\theta)$ is obtained by maximizing $\bar{\pi}$ as the best response to the population strategy profile in the DG, and setting π to be an MFG-NE π^* for the parameter θ .

We also prove a similar bound for the objective value. Assume now that all N players play policy π^* in the DG. Then, since g is Lipschitz continuous,

$$\begin{aligned} |G(\theta, \boldsymbol{\pi}^*) - g(\theta, \Lambda(\boldsymbol{\pi}^*|\theta))| &\leq \mathbb{E}[|g(\theta, \{\widehat{L}_h\}_{h=0}^{H-1}) - g(\theta, \{L_h^{\bar{\pi}}\}_{h=0}^{H-1})|] \\ &\leq \mathbb{E}[\sum_h K_h \|\widehat{L}_h - L_h^{\bar{\pi}}\|_1], \end{aligned}$$

where we denote the Lipschitz modulus of g with L_h as K_h (with respect to norm $\|\cdot\|_1$). An application of the technical lemma Lemma 5 yields then the $\mathcal{O}(1/\sqrt{N})$ upper bound.

D.3 Proof of Lemma 1

We first prove that if $\lim_{T \rightarrow \infty} F_{\text{omd}}^{(T)}(\theta', \zeta)$ exists and $F_{\text{omd}}^{\infty}(\theta', \zeta) = \zeta^*$, the softmax(ζ^*) $\in \text{Nash}_{\mathcal{M}}^{\tau}(\theta')$. Since F_{omd} is a continuous function, it must hold that

$$F_{\text{omd}}(\theta', \zeta^*) = F_{\text{omd}}(\theta', \lim_{T \rightarrow \infty} F_{\text{omd}}^{(T)}(\theta', \zeta)) = F_{\text{omd}}^{\infty}(\theta', \zeta) = \zeta^*,$$

therefore $F_{\text{omd}}(\theta', \zeta^*) = \zeta^*$. Denoting $\pi_h^*(\cdot|s) := \text{softmax}(\zeta^*(h, s, \cdot))$, we then have the relations

$$\begin{aligned} F_{\text{omd}}(\theta, \zeta^*)(h, s, a) &:= (1 - \eta\tau)\zeta^*(h, s, a) + \eta q_h^{\tau}(s, a|\Lambda(\boldsymbol{\pi}^*), \boldsymbol{\pi}^*, \theta), \\ \bar{F}_{\text{omd}}(\theta, \boldsymbol{\pi})(h, s) &:= \arg \max_{u \in \Delta_{\mathcal{A}}} \langle q_h^{\tau}(s, \cdot|\Lambda(\boldsymbol{\pi}), \boldsymbol{\pi}, \theta), u \rangle + \tau \mathcal{H}(u) - \eta^{-1}(1 - \tau\eta) D_{\text{kl}}(u|\boldsymbol{\pi}(s)). \end{aligned}$$

Then, for any h , it holds that $\zeta^*(h, s, a) = \tau^{-1} q_h^{\tau}(s, a|\Lambda(\boldsymbol{\pi}^*), \boldsymbol{\pi}^*, \theta)$. We show that $\boldsymbol{\pi}^*$ is then the best response to $\Lambda(\boldsymbol{\pi}^*)$ by backward induction. Denote for convenience $\mathbf{L}^* := \{L_h^*\}_h = \Lambda(\boldsymbol{\pi}^*)$.

At time $H - 1$, $\zeta^*(H - 1, s, a) = \tau^{-1} R_{h,\theta}(s, a, L_{H-1}^*)$ and $\pi_{H-1}^*(a|s) = \frac{\exp\{\tau^{-1} R_{h,\theta}(s, a, L_{H-1}^*)\}}{\sum_{a'} \exp\{\tau^{-1} R_{h,\theta}(s, a', L_{H-1}^*)\}}$, therefore, by first order optimality conditions, $\pi_{H-1}^*(\cdot|s)$ maximizes (uniquely) the strongly concave function $u \rightarrow \langle u, R_{h,\theta}(s, a, L_{H-1}^*) \rangle + \tau \mathcal{H}(u)$ on the

simplex $\Delta_{\mathcal{A}}$. Hence, at every state s at time $H - 1$, the policy $\pi_{H-1}^*(\cdot|s)$ is optimal. Assume now $\pi_{h'}^*(\cdot|s)$ is optimal for all $h' > h$ for some $h \in \{0, \dots, H - 2\}$. Then, by the inductive assumption, $q_{h'}^r(s, a|\Lambda(\pi^*), \pi^*, \theta)$ is the optimal regularized q function for all $h' \geq h$. Since $\zeta^*(h, s, a) = \tau^{-1}q_h^r(s, a|\Lambda(\pi^*), \pi^*, \theta)$, once again by first order optimality conditions, π_h^* is also the optimal policy at time h .

We move on to the convergence of derivatives. Firstly, since q_h^r is given to be C^1 in a neighborhood of (θ, ζ^*) and $\rho(\partial_\zeta q^r(\cdot, \cdot|\Lambda(\text{softmax}(\zeta^*)), \text{softmax}(\zeta^*), \theta)) < \tau$, there exists an open neighborhood \bar{U} of (θ, ζ^*) where $\rho(\partial_\zeta q^r(\cdot, \cdot|\Lambda(\text{softmax}(\zeta')), \text{softmax}(\zeta'), \theta')) < \tau - \delta_1$ for all $(\theta', \zeta') \in \bar{U}$ for some $\delta_1 > 0$. Then, since

$$F_{\text{omd}}(\theta, \zeta)(h, s, a) := (1 - \eta\tau)\zeta_{h,s,a} + \eta q_h^r(s, a|\Lambda(\text{softmax}(\zeta)), \text{softmax}(\zeta), \theta),$$

on \bar{U} it also holds that $\rho(\partial_\zeta F_{\text{omd}}(\theta, \zeta)) < 1 - \delta_2$ for some $\delta_2 > 0$. On \bar{U} , the map F_{omd}^∞ is implicitly defined by $F_{\text{omd}}(\theta, F_{\text{omd}}^\infty(\theta, \zeta)) = F_{\text{omd}}^\infty(\theta, \zeta)$, therefore by the implicit function theorem F_{omd}^∞ is differentiable in θ and

$$\nabla_\theta F_{\text{omd}}^\infty(\theta, \zeta) = (I - \partial_\zeta F_{\text{omd}}(\theta, F_{\text{omd}}^\infty(\theta, \zeta)))^{-1} \partial_\theta F_{\text{omd}}(\theta, \zeta), \quad \forall (\theta, \zeta) \in \bar{U}.$$

Fixing some direction $v \in \Theta$, define inductively the iterates

$$\begin{aligned} \zeta_0 &:= \zeta', & \zeta_{t+1} &:= F_{\text{omd}}(\theta, \zeta_t) \\ x_0 &:= 0, & x_{t+1} &:= \partial_\zeta F_{\text{omd}}(\theta, \zeta_t)x_t + \partial_\theta F_{\text{omd}}(\theta, \zeta_t)v. \end{aligned}$$

Note that by the chain rule, $x_{t+1} = \nabla_u(F_{\text{omd}}^{(T)}(\theta, \zeta'))v$, therefore, if we show that $\lim_{t \rightarrow \infty} x_t = \nabla_\theta(F_{\text{omd}}^\infty(\theta, \zeta'))v$ for any choice of v , we are done. Firstly, by the assumptions of the lemma, for sufficiently large T_0 , it holds that $\zeta_t \in \bar{U}$ for all $t > T_0$, and $\zeta^* = \lim_{t \rightarrow \infty} \zeta_t$. Defining $x^* := (I - \partial_\zeta F_{\text{omd}}(\theta, \zeta^*))^{-1} \partial_\theta F_{\text{omd}}(\theta, \zeta^*)v$, which satisfies $x^* := \partial_\zeta F_{\text{omd}}(\theta, \zeta^*)x^* + \partial_\theta F_{\text{omd}}(\theta, \zeta^*)v$. Therefore,

$$\begin{aligned} \|x_{t+1} - x^*\| &= \|\partial_\zeta F_{\text{omd}}(\theta, \zeta_t)x_t - \partial_\zeta F_{\text{omd}}(\theta, \zeta^*)x^* + \partial_\theta F_{\text{omd}}(\theta, \zeta_t)v - \partial_\theta F_{\text{omd}}(\theta, \zeta^*)v\| \\ &\leq \|\partial_\zeta F_{\text{omd}}(\theta, \zeta_t)x_t - \partial_\zeta F_{\text{omd}}(\theta, \zeta^*)x_t\| + \|\partial_\zeta F_{\text{omd}}(\theta, \zeta^*)x_t - \partial_\zeta F_{\text{omd}}(\theta, \zeta^*)x^*\| \\ &\quad + \|\partial_\theta F_{\text{omd}}(\theta, \zeta_t)v - \partial_\theta F_{\text{omd}}(\theta, \zeta^*)v\| \\ &\leq \|\partial_\zeta F_{\text{omd}}(\theta, \zeta_t)x_t - \partial_\zeta F_{\text{omd}}(\theta, \zeta^*)x_t\| + (1 - \delta_2)\|x_t - x^*\| \\ &\quad + \|\partial_\theta F_{\text{omd}}(\theta, \zeta_t)v - \partial_\theta F_{\text{omd}}(\theta, \zeta^*)v\|, \end{aligned}$$

which proves that $x_t \rightarrow x^*$ as both $\partial_\theta F_{\text{omd}}, \partial_\zeta$ are continuous in ζ for $\zeta \in \bar{U}$ and $\zeta_t \rightarrow \zeta^*$. By the implicit function theorem, x^* is the gradient of F_{omd}^∞ , concluding the proof.

D.4 Discussion of Assumptions of Lemma 1

We briefly discuss the ramifications of Lemma 1. Firstly, the result is useful only assuming that the OMD iterates converge to an NE of the MFG. NE computation in MFGs is a well-studied research topic on its own, and several positive results are known for various classes of MFGs. We take this for granted in this lemma, as NE computation is not the main goal.

Next, we note that the lemma suggests that the derivative of (T -approx.) is a valid first-order oracle provided that the Jacobian of F_{omd} has bounded spectral radius around the NE induced by a θ . Importantly, for any PMFG, T -step objective approximates the fixed-point gradient provided that τ is sufficiently large. For more structured settings, the result can be strengthened to permit $\tau = 0$, for instance, the derivations in [48] readily extend to the case where $H = 1$, $|\mathcal{S}| = 1$, and the reward function R is monotone in population distribution. We leave as future work to generalize this for monotone PMFGs with dynamics (i.e., for PMFGs with $H > 1$).

D.5 Proof of Lemma 2

Let $\Theta = \mathbb{R}^{D_1}, Z = \mathbb{R}^{D_2}$, and assume the functions $F : \Theta \times Z \rightarrow Z$ and $g : \Theta \times Z \rightarrow \mathbb{R}$ are differentiable. We define $\ell : \Theta \rightarrow \mathbb{R}$

$$\begin{aligned} \ell(\theta) &:= g(\theta, \zeta_{T+1}(\theta)), \\ \zeta_{t+1}(\theta) &:= (1 - \tau\eta)\zeta_t(\theta) + \eta F(\theta, \zeta_t(\theta)), \quad \forall t = 0, \dots, T. \end{aligned} \tag{1}$$

for some ζ_0 constant. Clearly $\ell(\theta)$ is differentiable; our goal is to efficiently compute $\nabla_\theta \ell(\theta)$.

Define for any sequence of (row) vectors $\{\mathbf{a}_t\}_{t=0}^T$ for $\mathbf{a}_t \in \Theta$ the function

$$\begin{aligned} L(\theta, \{\mathbf{a}_t\}_{t=0}^T) &:= g(\theta, \zeta_{T+1}(\theta)) + \sum_{t=0}^T \mathbf{a}_t \cdot (\zeta_{t+1}(\theta) - (1 - \eta\tau)\zeta_t(\theta) - \eta F(\theta, \zeta_t(\theta))) \\ &= g(\theta, \zeta_{T+1}(\theta)) + \mathbf{a}_T \cdot \zeta_{T+1}(\theta) - (1 - \eta\tau)\mathbf{a}_0 \cdot \zeta_0(\theta) - \eta \sum_{t=0}^T \mathbf{a}_t \cdot F(\theta, \zeta_t(\theta)) \\ &\quad + \sum_{t=1}^T \zeta_t \cdot (\mathbf{a}_{t-1} - (1 - \eta\tau)\mathbf{a}_t). \end{aligned}$$

Since for any $\{\mathbf{a}_t\}_{t=0}^T$ it holds that $L(\theta, \{\mathbf{a}_t\}_{t=0}^T) = \ell(\theta)$, we have the identities

$$\partial_\theta L(\theta, \{\mathbf{a}_t\}_{t=0}^T) = \nabla_\theta \ell(\theta), \quad \partial_{\mathbf{a}_t} L(\theta, \{\mathbf{a}_t\}_{t=0}^T) = 0$$

Therefore, for any sequence of functions $\mathbf{a}_t(\theta)$ that depend on θ , we have

$$\nabla_\theta \ell(\theta) = \partial_\theta L(\theta, \{\mathbf{a}_t\}_{t=0}^T) + \sum_t (\nabla_\theta \mathbf{a}_t) \frac{\partial}{\partial \mathbf{a}_t} L(\theta, \{\mathbf{a}_t\}_{t=0}^T) = \partial_\theta L(\theta, \{\mathbf{a}_t\}_{t=0}^T).$$

We will therefore compute $\partial_\theta L(\theta, \{\mathbf{a}_t\}_{t=0}^T)$ with a suitable choice of \mathbf{a}_t . By simple derivation:

$$\begin{aligned} \partial_\theta L(\theta, \{\mathbf{a}_t\}_{t=0}^T) &= \partial_\zeta g(\theta, \zeta_{T+1}(\theta)) \nabla_\theta \zeta_{T+1}(\theta) + \partial_\theta g(\theta, \zeta_{T+1}(\theta)) \\ &\quad + \mathbf{a}_T \nabla_\theta \zeta_{T+1}(\theta) - (1 - \eta\tau)\mathbf{a}_0 \nabla_\theta \zeta_0(\theta) \\ &\quad - \eta \sum_{t=0}^T \mathbf{a}_t \partial_\theta F(\theta, \zeta_t(\theta)) - \eta \sum_{t=0}^T \mathbf{a}_t \partial_\zeta F(\theta, \zeta_t(\theta)) \nabla_\theta \zeta_t(\theta) \\ &\quad + \sum_{t=1}^T (\mathbf{a}_{t-1} - (1 - \eta\tau)\mathbf{a}_t) \nabla_\theta \zeta_t(\theta) \end{aligned}$$

Since $\nabla_\theta \zeta_0 = 0$, as ζ_0 is constant, rearranging the terms yields:

$$\begin{aligned} \partial_\theta L(\theta, \{\mathbf{a}_t\}_{t=0}^T) &= (\nabla_\zeta g(\zeta_{T+1}(\theta)) + \mathbf{a}_T) \nabla_\theta \zeta_{T+1}(\theta) - \eta \sum_{t=0}^T \mathbf{a}_t \partial_\theta F(\theta, \zeta_t(\theta)) + \partial_\theta g(\theta, \zeta_{T+1}(\theta)) \\ &\quad + \sum_{t=1}^T (\mathbf{a}_{t-1} - (1 - \eta\tau)\mathbf{a}_t - \eta \mathbf{a}_t \partial_\zeta F(\theta, \zeta_t(\theta))) \nabla_\theta \zeta_t(\theta). \end{aligned}$$

Therefore, we pick

$$\begin{aligned} \bar{\mathbf{a}}_T &:= -\nabla_\zeta g(\zeta_{T+1}(\theta)) \\ \bar{\mathbf{a}}_{t-1} &:= (1 - \eta\tau)\bar{\mathbf{a}}_t + \eta \bar{\mathbf{a}}_t \partial_\zeta F(\theta, \zeta_t(\theta)), \end{aligned}$$

we obtain the equality

$$\nabla_\theta \ell(\theta) = \partial_\theta L(\theta, \{\bar{\mathbf{a}}_t\}_{t=0}^T) = -\eta \sum_{t=0}^T \bar{\mathbf{a}}_t \partial_\theta F(\theta, \zeta_t(\theta)) + \partial_\theta g(\theta, \zeta_{T+1}(\theta)). \quad (2)$$

Most importantly, Equation (2) permits the computation of $\nabla_\theta \ell$ with T with only caching the variables ζ_t . Namely, the forward system Equation (1) can be used to iteratively compute ζ_{T+1} , and the backward system only requires evaluating Jacobians at the current step, meaning other than ζ_t , the memory requirements are kept constant.

In case memory is a bottleneck, ζ_t can be cached during the forward step every $\beta\sqrt{T}$ steps for some constant β , meaning only $\mathcal{O}(\sqrt{T})$ memory is needed. In this case, during the backward step, ζ_t will need to be recomputed every $\mathcal{O}(\sqrt{T})$ steps for $\mathcal{O}(\sqrt{T})$ OMD iterations, maintaining the $\mathcal{O}(T)$ time complexity. β will introduce a tradeoff between time and space complexity.

D.6 Extension of Lemma 2 to General Mirror Descent

Lemma 2 analyzes the adjoint method under the specific entropy regularization scheme. Let $h : \mathbb{R}^{D_1} \rightarrow \mathbb{R}$ be a strongly convex distance generating function, and $\nabla h : \mathbb{R}^{D_1} \rightarrow \mathbb{R}^{D_1}$ the corresponding mirror map. Then, the generalized mirror descent update rule can be written as (in policy space $\Pi_H \subset \mathbb{R}^{[H] \times \mathcal{S} \times \mathcal{A}}$):

$$\nabla h(\pi_{t+1}) := (1 - \tau\eta)\nabla h(\pi_t) + \eta F(\theta, \pi_t(\theta)), \quad \forall t = 0, \dots, T.$$

Defining the iterates $\zeta_t := \nabla h(\pi_{t+1})$, we obtain the similar update rule

$$\ell(\theta) := g(\theta, (\nabla h)^{-1}(\zeta_{T+1}(\theta))),$$

$$\zeta_{t+1}(\theta) := (1 - \tau\eta)\zeta_t(\theta) + \eta F(\theta, (\nabla h)^{-1}(\zeta_t(\theta))), \quad \forall t = 0, \dots, T,$$

which reduces to the case analyzed in Appendix D.5, by the definitions

$$\bar{F}(\theta, \zeta_t(\theta)) := F(\theta, (\nabla h)^{-1}(\zeta_t(\theta)))$$

$$\bar{g}(\theta, \zeta_{T+1}(\theta)) := g(\theta, (\nabla h)^{-1}(\zeta_{T+1}(\theta))).$$

E Results on Batched Auctions

This section presents the formal analysis of parameterized batched auctions introduced in the main text. We also provide rigorous statements and complete the proofs left out in the main text.

Appendix E.1 provides formal definitions of the settings omitted in the main text and useful auxiliary constructions to assist the proofs. Appendix E.2 presents a collection of auxiliary lemmas that support the subsequent analysis. In Appendix E.3, we prove the first part of Theorem 2, establishing an upper bound on the exploitability of mean field policies under the no zero-dominance condition. The second part of the theorem, which addresses convergence of the mechanism-level objective, is proved in Appendix E.4. Finally, Appendix E.5 provides the proof of Lemma 3, which establishes the Lipschitz continuity of the (entropy regularized) q values under full-support policies.

E.1 Extended Definitions

Additional useful notation. To streamline the proofs, we also introduce some useful notation. For any arbitrary finite set \mathcal{X} and a scalar $\beta \in \mathbb{R}_{\geq 0}$ define the sets:

$$\Delta_{\mathcal{X}}^{\beta} := \{d \in \mathbb{R}^{\mathcal{X}} : \forall x \in \mathcal{X} : d(x) \geq 0, \sum_{x \in \mathcal{X}} d(x) = \beta\},$$

$$\Delta_{\mathcal{X}}^{\geq \beta} := \{d \in \mathbb{R}^{\mathcal{X}} : \forall x \in \mathcal{X} : d(x) \geq 0, \beta \leq \sum_{x \in \mathcal{X}} d(x) \leq 1\},$$

$$\Delta_{\mathcal{X}}^{\leq \beta} := \{d \in \mathbb{R}^{\mathcal{X}} : \forall x \in \mathcal{X} : d(x) \geq 0, \sum_{x \in \mathcal{X}} d(x) \leq \beta\}.$$

For $x \in \mathbb{R}^{\mathcal{X}}$, where \mathcal{X} has a total order, denote the cumulative mass function $S_x(d) := \sum_{x' \geq x} d(x')$. For some $d \in \Delta_{\mathcal{X} \times \mathcal{Y}}$, define the marginal distribution

$$\text{Marg}_{\mathcal{Y}}(d) := \sum_{x \in \mathcal{X}} d(x, \cdot) \in \Delta_{\mathcal{Y}}.$$

For $\alpha \in [0, 1]$, define the *threshold bid operator* $\text{Th}_{\alpha} : \Delta_{\mathcal{A}}^{\geq 0} \rightarrow \mathcal{A}$ as

$$\text{Th}_{\alpha}(d) := \max(\{a \in \mathcal{A} : S_a(d) \geq \alpha, d(a) > 0\} \cup \{a_0\}),$$

where a_0 is the smallest element of \mathcal{A} . For $d \in \Delta_{\mathcal{V} \times \mathcal{A}}^{\geq \alpha}$ we also define $\text{Th}_{\alpha}(d) := \text{Th}_{\alpha}(\text{Marg}_{\mathcal{S}}(d))$.

We define the operator $\Xi_{\alpha} : \Delta_{\mathcal{V} \times \mathcal{A}}^{\geq \alpha} \rightarrow \mathbb{R}^{\mathcal{V} \times \mathcal{A}}$, which maps a matrix state-action distribution L to a matrix L' , where each entry $L'(s, a)$ represents the expected probability mass of agents in state s who chose action a and did not win an item this round, given that α goods are allocated. Formally:

$$\Xi_{\alpha}(d)(s, a) = \begin{cases} 0 & \text{if } a > \text{Th}(d), \alpha > 0 \\ \frac{\sum_{s' \in \mathcal{V}, a' \geq \text{Th}_{\alpha}(d)} d(s', a') - \alpha}{\sum_{s' \in \mathcal{V}} d(s', \text{Th}_{\alpha}(d))} d(s, \text{Th}_{\alpha}(d)) & \text{if } a = \text{Th}(d), \alpha > 0 \\ d(s, a) & \text{otherwise} \end{cases}$$

Ξ_α is well-defined as $\sum_{s' \in \mathcal{V}} d(s', \text{Th}_\alpha(d)) > 0$ whenever $\alpha > 0$ by definition.

Using the operator Ξ_α , we define the post allocation operator $\Upsilon_\alpha : \Delta_{\mathcal{V} \times \mathcal{A}}^{\geq \alpha} \rightarrow \Delta_{\mathcal{V}}$, which maps a sub-probability distribution d over active state-action pairs to the post allocation state distribution. That is,

$$\Upsilon_\alpha(L)(s) := \sum_{a \in \mathcal{A}} \Xi_\alpha(L)(s, a).$$

The operator Γ_α , which governs the transition of the population state distribution after item allocation, can be expressed in terms of Υ_α by explicitly accounting for the inactive state \perp . Specifically, let $d \in \Delta_{\mathcal{S} \times \mathcal{A}}$ be a state-action distribution and assume $\alpha(\nu^{-\perp}(d)) \leq \|\nu^{-\perp}(d)\|_1$. Then $\Gamma_\alpha : \Delta_{\mathcal{S} \times \mathcal{A}} \rightarrow \Delta_{\mathcal{S}}$ is defined as

$$\Gamma_\alpha(d)(s) := \begin{cases} \Upsilon_\alpha(\nu^{-\perp}(d))(d_{-\perp})(s) & \text{if } s \in \mathcal{V}, \\ \sum_{a \in \mathcal{A}} d(\perp, a) + \alpha & \text{if } s = \perp, \end{cases}$$

where $d_{-\perp}$ denotes the restriction of d to active states \mathcal{V} .

The BA-MFG is a parametrized PMFG, where both the payment function and the allocation threshold function are parameterized as α_h^θ and p_h^θ . Throughout this section, we assume these parameters are fixed and omit them, writing without loss of generality α_h, p_h instead.

Definition 5 (Batched Auction MFG (BA-MFG)) A Batched-Auction MFG (BA-MFG) is a MFG defined by the tuple $\mathcal{M}_{mfa} = (\mathcal{S}, \mathcal{A}, H, \mu_0, \{P_h^{mfa}\}_{h=0}^{H-1}, \{R_h^{mfa}\}_{h=0}^{H-1})$ of discrete state space $\mathcal{S} = \mathcal{V} \cup \{\perp\}$, discrete action space \mathcal{A} , horizon $H \in \mathbb{N}_{>0}$, initial distribution $\mu_0 \in \mathcal{S}$, transition dynamics $P_h^{mfa} : \mathcal{S} \times \mathcal{A} \times \Delta_{\mathcal{S} \times \mathcal{A}} \rightarrow \Delta_{\mathcal{S}}$ and reward functions $R_h^{mfa} : \mathcal{S} \times \mathcal{A} \times \Delta_{\mathcal{S} \times \mathcal{A}} \rightarrow [0, 1]$. The transition dynamics P_h^{mfa} and the reward functions R_h^{mfa} depend from the allocation functions $\{\alpha_h\}_{h=0}^{H-1}$, the dynamic functions $\{w_h\}_{h=0}^{H-1}$, the payment functions $\{p_h\}_{h=0}^{H-1}$ and the utility functions $\{u_h\}_{h=0}^{H-1}$. Define the “winning probability” as

$$p_{win}(s, a, L, \alpha) = \mathbb{1}_{s \neq \perp} \begin{cases} 1 & \text{if } \sum_{s'} \sum_{a' \geq a} L(s', a') \leq \alpha \\ 0 & \text{if } \sum_{s'} \sum_{a' > a} L(s', a') \geq \alpha. \\ \frac{\alpha - \sum_{s'} \sum_{a' > a} L(s', a')}{\sum_{s'} L(s, a)} & \text{otherwise} \end{cases}$$

For the allocation α and valuation transition w , define the operators $\Gamma_\alpha : \Delta_{\mathcal{S} \times \mathcal{A}} \rightarrow \Delta_{\mathcal{S}}, \Gamma_w : \Delta_{\mathcal{S}} \times \Delta_{\mathcal{S}}$ as

$$\Gamma_\alpha(L)(s) := \begin{cases} \sum_a L(s, a)(1 - p_{win}(s, a, L, \alpha(\nu^{-\perp}(L)))) & \text{if } s \neq \perp \\ L(\perp) + \sum_{s'} \sum_{a'} L(s', a') p_{win}(s', a', L, \alpha(\nu^{-\perp}(L))) & \text{otherwise} \end{cases},$$

$$\Gamma_w(\xi) := \sum_{z \in \mathcal{S}} \xi(z) w(\cdot | z, \xi),$$

Define also Γ_h, Λ_{mfa} as ⁷

$$\Gamma_h(L, \pi)(s, a) := \Gamma_{w_h}(\Gamma_{\alpha_h}(L))(s) \pi(a | s),$$

$$\Lambda_{mfa}(\pi) := \{\Gamma_{h-1}(\dots \Gamma_1(\Gamma_0(\mu_0 \cdot \pi_0, \pi_1), \pi_2), \dots, \pi_{h-1})\}_{h=0}^{H-1}.$$

The transition probability and rewards can be written as:

$$P_h^{mfa}(s' | s, a, L) := p_{win}(s, a, L, \alpha_h(\nu^{-\perp})) w_h(s' | \perp, \Gamma_{\alpha_h}(L))$$

$$+ (1 - p_{win}(s, a, L, \alpha_h(\nu^{-\perp}))) w_h(s' | s, \Gamma_{\alpha_h}(L))$$

$$R_h^{mfa}(s, a, L) := p_{win}(s, a, L, \alpha_h(\nu^{-\perp})) u_h(s, p_h(a, \nu^{-\perp}(L))).$$

For $\pi \in \Pi_H, \tau \geq 0$ and $\mathbf{L} = \{L_h\}_{h=0}^{H-1}$, the total expected (entropy regularized) reward is

$$V_{mfa}^\tau(\mathbf{L}, \pi) := \mathbb{E} \left[\sum_{h=0}^{H-1} R_h^{mfa}(s, a, L_h) + \tau \mathcal{H}(\pi_h(s_h)) \right]_{s_0 \sim \mu_0, a_h \sim \pi(s_h)} \Bigg|_{s_{h+1} \sim P_h^{mfa}(s_h, a_h, L_h)}$$

⁷Note that this way of writing Γ_h, Λ_{mfa} is consistent with the MFG definition.

For a policy $\pi \in \Pi_H$ we denote with $L^\pi = \Lambda_{\text{mfa}}(\pi)$ the induced state-action distribution, with μ^π the induced state distribution and with ξ^π the induced hidden state distribution, with $\xi_h^\pi = \Gamma_{\alpha_h}(L_h^\pi)$. Additionally, we define with P_α the transition probabilities associated with the item allocation dynamics

$$P_\alpha(s'|s, a, L) := \begin{cases} p_{\text{win}}(s, a, L, \alpha(\nu^{-\perp}(L))) & \text{if } s' = \perp \\ (1 - p_{\text{win}}(s, a, L, \alpha(\nu^{-\perp}(L)))) & \text{if } s' = s \\ 0 & \text{otherwise} \end{cases}.$$

Then, the total expected (entropy regularized) reward can also be expressed as

$$V_{\text{mfa}}^\tau(\mathbf{L}, \pi) := \mathbb{E} \left[\sum_{h=0}^{H-1} R_h^{\text{mfa}}(s, a, L_h) + \tau \mathcal{H}(\pi_h(s_h)) \middle| \begin{array}{l} s_0 \sim \mu_0, a_h \sim \pi(s_h), z_h \sim P_{\alpha_h}(s_h, a_h, L_h) \\ s_{h+1} \sim w_h(z_h, \Gamma_{\alpha_h}(L_h)) \end{array} \right]$$

Definition 6 (N-player Batched Auction (N-BA)) An N -player batched auction (N-BA) is a dynamic game $\mathcal{G}_{\text{auc}} = (N, \mathcal{S}, \mathcal{A}, H, \mu_0, \{P_h^{\text{auc}}\}_{h=0}^{H-1}, \{R_h^{\text{auc}}\}_{h=0}^{H-1})$ of discrete state space $\mathcal{S} = \mathcal{V} \cup \{\perp\}$, discrete action space \mathcal{A} , horizon $H \in \mathbb{N}_{>0}$, starting distribution $\mu_0 \in \mathcal{S}$, transition dynamics $P_h^{\text{auc}} : \mathcal{S}^N \times \mathcal{A}^N \rightarrow \Delta_{\mathcal{S}}^N$ and rewards functions $R_h^{\text{auc}} : \mathcal{S}^N \times \mathcal{A}^N \rightarrow [0, 1]^N$. The transition dynamics P_h^{auc} and the reward functions R_h^{auc} depend from the allocation functions $\{\alpha_h\}_{h=0}^{H-1}$, the dynamic functions $\{w_h\}_{h=0}^{H-1}$, the payment functions $\{p_h\}_{h=0}^{H-1}$ and the utility functions $\{u_h\}_{h=0}^{H-1}$. Let $\mathbf{s} = (s^1, \dots, s^N) \in \mathcal{S}^N$ denote the joint state of all agents, and $\mathbf{a} = (a^1, \dots, a^N) \in \mathcal{A}^N$ the joint action profile. Assume items are allocated to the top bidders according to the submitted actions, with uniform random tie-breaking at the allocation threshold. This induces a joint allocation probability kernel $P_{N, \alpha} : \mathcal{S}^N \times \mathcal{A}^N \rightarrow \Delta_{\mathcal{S}^N}$ formally defined as

$$P_{N, \alpha}(\mathbf{z}|\mathbf{s}, \mathbf{a}) = \begin{cases} 1 & \forall i \in \mathcal{W} \cup \{i \in [N] : s^j = \perp\}, z^i = \perp, \\ \frac{1}{\binom{|\mathcal{T}|}{\lfloor \alpha(\nu) \cdot N \rfloor - |\mathcal{W}|}} & \text{if } |\{j \in \mathcal{T} : z^j = \perp\}| = \lfloor \alpha(\nu) \cdot N \rfloor - |\mathcal{W}|, \\ \binom{\lfloor \alpha N \rfloor - |\mathcal{W}|}{|\mathcal{T}|} & \forall i \in \mathcal{L}, z^i = s^i \\ 0 & \text{otherwise} \end{cases},$$

where $\nu = \sum_j \mathbf{e}_{a^j} \mathbb{1}_{s^j \neq \perp}$, $a^* = \max\{a \in \mathcal{A} : \sum_{j \in [N]} \mathbb{1}_{a^j \geq a, s^j \neq \perp} \geq \lfloor \alpha(\nu) N \rfloor\}$, $\mathcal{T} = \{j \in [N] : a^j = a^*, s^j \neq \perp\}$, $\mathcal{W} = \{j \in [N] : a^j > a^*, s^j \neq \perp\}$ and $\mathcal{L} = \{j \in [N] : a^j < a^*, s^j \neq \perp\}$. The marginal winning probability and the rewards, similar as its corresponding MFG can be expressed as

$$P_{N, \alpha}^i(\perp | \mathbf{s}, \mathbf{a}) = \sum_{\mathbf{z}} P_{N, \alpha}(\mathbf{z}|\mathbf{s}, \mathbf{a}) \mathbb{1}_{z^i = \perp} = p_{\text{win}}(s^i, a^i, \sigma(\mathbf{s}, \mathbf{a}), \frac{\lfloor N\alpha(\nu^{-\perp}(\sigma(\mathbf{s}, \mathbf{a}))) \rfloor}{N})$$

$$R_h^{\text{auc}, i}(\mathbf{s}, \mathbf{a}) = p_{\text{win}}(s^i, a^i, \sigma(\mathbf{s}, \mathbf{a}), \frac{\lfloor N\alpha_h(\nu^{-\perp}(\sigma(\mathbf{s}, \mathbf{a}))) \rfloor}{N}) u_h(s^i, p_h(a^i, \nu^{-\perp}(\sigma(\mathbf{s}, \mathbf{a}))))$$

We define with $R_{N, h}^{\text{mfa}}$ the N -player discretization of R_h^{mfa} :

$$R_{N, h}^{\text{mfa}}(s, a, L) := p_{\text{win}}(s, a, L, \frac{\lfloor N\alpha_h(\nu^{-\perp}(L)) \rfloor}{N}) u_h(s, p_h(a, \nu^{-\perp}(L)))$$

Note that if $L(s, a) = 0$, the reward $R_{N, h}^{\text{mfa}}(s, a, L)$ is not defined. Additionally, observe that

$$R_h^{\text{auc}, i}(\mathbf{s}, \mathbf{a}) = R_{N, h}^{\text{mfa}}(s^i, a^i, \sigma(\mathbf{s}, \mathbf{a})).$$

For a strategy profile $\boldsymbol{\pi} \in \Pi_H^N$, $\tau \geq 0$ the (entropy regularized) sum of rewards of player $i \in [N]$ is defined as

$$J_{\text{auc}}^{\tau, i}(\boldsymbol{\pi}) := \mathbb{E} \left[\sum_{h=0}^{H-1} R_{N, h}^{\text{mfa}}(s_h^i, a_h^i, \widehat{L}_h) + \tau \mathcal{H}(\pi_h^i(s_h^i)) \middle| \begin{array}{l} \forall j \in [N] : s_0^j \sim \mu_0, a_h^j \sim \pi_h^j(s_h^j) \\ \mathbf{z}_h \sim P_{N, \alpha_h}(\mathbf{s}_h, \mathbf{a}_h), \widehat{L}_h = \frac{1}{N} \sum_{j \in [N]} \mathbf{e}_{s_h^j, a_h^j} \\ s_{h+1}^j \sim w(z_h^j, \widehat{\xi}_h), \widehat{\xi}_h = \frac{1}{N} \sum_{j \in [N]} \mathbf{e}_{z_h^j} \end{array} \right],$$

exploitability $\mathcal{E}_{\text{auc}}^{\tau, i}$ as $\mathcal{E}_{\text{auc}}^{\tau, i}(\boldsymbol{\pi}) := \max_{\boldsymbol{\pi}' \in \Pi_H} J_{\text{auc}}^{\tau, i}(\boldsymbol{\pi}', \boldsymbol{\pi}^{-i}) - J_{\text{auc}}^{\tau, i}(\boldsymbol{\pi})$.

E.2 Preliminary Lemmas

We present several important lemmas that will be used later to prove the main convergence theorem.

Lemma 6 (Sensitivity of Ξ_α to the Population) *Let $\alpha \in [0, 1]$ and let $d, d' \in \Delta_{\mathcal{V} \times \mathcal{A}}^{\geq \alpha}$. Then Ξ_α is Lipschitz-continuous with respect to the ℓ_1 -norm, with constant 1:*

$$\|\Xi_\alpha(d) - \Xi_\alpha(d')\|_1 \leq \|d - d'\|_1.$$

Proof: It is straightforward to verify that Ξ_α is continuous. We verify that it is also Lipschitz continuous with modulus 1. Assume $\alpha > 0$, as otherwise $\Xi_\alpha(d)$ is the identity map and the claim is trivial. Denote the bids by $\mathcal{A} := \{1, \dots, A\}$, and let $S_a(d) := \sum_{s'} \sum_{a' \geq a} d(s', a')$ with $S_{A+1}(d) := 0$, and $r_a(d) := \sum_{s' \in \mathcal{V}} d(s', a)$. For $\bar{a} \in \{1, \dots, A+1\}$, define the regions

$$\mathcal{R}_{\bar{a}} := \{d \in \Delta_{\mathcal{V} \times \mathcal{A}}^{\geq \alpha} : S_{\bar{a}}(d) \geq \alpha > S_{\bar{a}+1}(d)\}.$$

On the region $\mathcal{R}_{\bar{a}}$, the map $\Xi_\alpha(d)$ is differentiable, in fact, for $d \in \mathcal{R}_{\bar{a}}$, it holds that $\bar{a} = \text{Th}_\alpha(d)$ and

$$\Xi_\alpha(d)(s, a) = \begin{cases} 0 & \text{if } a > \bar{a} \\ \frac{S_{\bar{a}}(d) - \alpha}{r_{\bar{a}}(d)} d(s, \bar{a}) & \text{if } a = \bar{a} \\ d(s, a) & \text{otherwise} \end{cases}$$

We calculate the Jacobian of Ξ_α and upper bound its operator norm $\|\nabla \Xi_\alpha\|_{1 \rightarrow 1}$ given by the max column sum $\|\nabla \Xi_\alpha\|_{1 \rightarrow 1} = \max_{s, a} \sum_{s', a'} |(\nabla \Xi_\alpha)_{s' a', s a}|$. We upper bound the column sums corresponding to s, a .

Case 1. If $a < \bar{a}$, then $\frac{\partial \Xi_\alpha(d)(s', a')}{\partial d_{sa}} = 0$ for any $(s', a') \neq (s, a)$ and $\frac{\partial \Xi_\alpha(d)(s, a)}{\partial d_{sa}} = 1$, therefore the column sum is

$$\sum_{s', a'} \left| \frac{\partial \Xi_\alpha(d)(s', a')}{\partial d_{sa}} \right| = \left| \frac{\partial \Xi_\alpha(d)(s, a)}{\partial d_{sa}} \right| = 1.$$

Case 2. If $a = \bar{a}$, then $\frac{\partial \Xi_\alpha(d)(s', a')}{\partial d_{sa}} = 0$ if $a' \neq \bar{a}$, therefore

$$\begin{aligned} \sum_{s', a'} \left| \frac{\partial \Xi_\alpha(d)(s', a')}{\partial d_{sa}} \right| &= \sum_{s'} \left| \frac{\partial \Xi_\alpha(d)(s', \bar{a})}{\partial d_{s\bar{a}}} \right| = \left| \frac{\partial \Xi_\alpha(d)(s, \bar{a})}{\partial d_{s\bar{a}}} \right| + \sum_{s' \neq s} \left| \frac{\partial \Xi_\alpha(d)(s', \bar{a})}{\partial d_{s\bar{a}}} \right| \\ &= \left| \frac{S_{\bar{a}}(d) - \alpha}{r_{\bar{a}}(d)} + d(s, \bar{a}) \frac{r_{\bar{a}}(d) - S_{\bar{a}}(d) + \alpha}{r_{\bar{a}}(d)^2} \right| \\ &\quad + \sum_{s' \neq s} \left| d(s', \bar{a}) \frac{r_{\bar{a}}(d) - S_{\bar{a}}(d) + \alpha}{r_{\bar{a}}(d)^2} \right| = 1, \end{aligned}$$

since all terms in the absolute values are nonnegative if $\bar{a} = \text{Th}_\alpha(d)$.

Case 3. If $a > \bar{a}$, then only the rows corresponding to the active action \bar{a} has nonzero gradient, and

$$\sum_{s', a'} \left| \frac{\partial \Xi_\alpha(d)(s', a')}{\partial d_{sa}} \right| = \sum_{s'} \left| \frac{\partial \Xi_\alpha(d)(s', \bar{a})}{\partial d_{s\bar{a}}} \right| = \sum_{s'} \frac{1}{\sum_{s''} d(s'', \bar{a})} d(s, \bar{a}) = 1.$$

To conclude, it holds that on any arbitrary region $\mathcal{R}_{\bar{a}}$,

$$\|\nabla \Xi_\alpha\|_{1 \rightarrow 1} = \max_{s, a} \sum_{s', a'} \left| \frac{\partial \Xi_\alpha(d)(s', \bar{a})}{\partial d_{s\bar{a}}} \right| = 1.$$

Therefore, $\|\nabla \Xi_\alpha\|_{1 \rightarrow 1} \leq 1$ on all regions $\mathcal{R}_{\bar{a}}$, and Ξ_α is non-expansive in the ℓ_1 norm. \square

Lemma 7 (Sensitivity of Ξ_α to the Allocation Parameter) *Let $d \in \Delta_{\mathcal{V} \times \mathcal{A}}^{\leq 1}$ arbitrary, let $\alpha_1, \alpha_2 \leq \|d\|_1$ arbitrary. Then*

$$\|\Xi_{\alpha_1}(d) - \Xi_{\alpha_2}(d)\|_1 \leq |\alpha_1 - \alpha_2|.$$

Proof: As in Lemma 6, let $\mathcal{A} := \{1, \dots, A\}$, and let $S_a(d) := \sum_{s'} \sum_{a' \geq a} d(s', a')$ with $S_{A+1}(d) := 0$, and $r_a(d) := \sum_{s' \in \mathcal{V}} d(s', a)$.

For $\bar{a} \in \{1, \dots, A+1\}$ such that $r_{\bar{a}}(d) > 0$, define the partition of the interval $[0, \|d\|_1]$ into the intervals

$$\mathcal{R}_{\bar{a}} := \{\alpha \in [0, \|d\|_1] : S_{\bar{a}}(d) \geq \alpha > S_{\bar{a}+1}(d)\}.$$

On the interval $\mathcal{R}_{\bar{a}}$, the map $\Xi_\alpha(d)$ is differentiable in α , and it holds that $\bar{a} = \text{Th}_\alpha(d)$ and

$$\Xi_\alpha(d)(s, a) = \begin{cases} 0 & \text{if } a > \bar{a} \\ \frac{S_{\bar{a}}(d) - \alpha}{r_{\bar{a}}(d)} d(s, \bar{a}) & \text{if } a = \bar{a} \\ d(s, a) & \text{otherwise} \end{cases}$$

Then, again, we upper bound the operator norm of the Jacobian (in this case, gradient):

$$\|\nabla_\alpha \Xi_\alpha(d)\|_{1 \rightarrow 1} = \sum_{s', a'} \left| \frac{\partial \Xi_\alpha(d)(s', a')}{\partial \alpha} \right| = \sum_{s'} \left| \frac{-1}{r_{\bar{a}}(d)} d(s', \bar{a}) \right| = 1,$$

So $\|\nabla_\alpha \Xi_\alpha(d)\|_{1 \rightarrow 1} \leq 1$, implying the claim of the lemma. \square

For completeness, we state the simple corollaries of the above two sensitivity analyses in the following.

Corollary 1 (Lipschitz Continuity of Ξ) Let $d, d' \in \Delta_{\mathcal{V} \times \mathcal{A}}^{\leq 1}$, $\alpha \in [0, \|d\|_1]$, and $\alpha' \in [0, \|d'\|_1]$ arbitrary, then

$$\|\Xi_\alpha(d) - \Xi_{\alpha'}(d')\|_1 \leq \|d - d'\|_1 + |\alpha - \alpha'|.$$

Proof: Let $d, d' \in \Delta_{\mathcal{V} \times \mathcal{A}}^{\leq 1}$, $\alpha \in [0, \|d\|_1]$, and $\alpha' \in [0, \|d'\|_1]$ arbitrary. Without loss of generality assume $\|d\|_1 \geq \|d'\|_1$, then by applying triangular inequality we have

$$\|\Xi_\alpha(d) - \Xi_{\alpha'}(d')\|_1 \leq \|\Xi_\alpha(d) - \Xi_{\alpha'}(d)\|_1 + \|\Xi_{\alpha'}(d) - \Xi_{\alpha'}(d')\|_1 \leq |\alpha - \alpha'| + \|d - d'\|_1,$$

where the last step follows from Lemmas 6 and 7. \square

Corollary 2 (Lipschitz Continuity of Υ_α) Let $d, d' \in \Delta_{\mathcal{V} \times \mathcal{A}}^{\leq 1}$, $\alpha \in [0, \|d\|_1]$, and $\alpha' \in [0, \|d'\|_1]$ arbitrary, then Υ_α is non-expansive in the ℓ_1 norm, that is

$$\|\Upsilon_\alpha(d) - \Upsilon_{\alpha'}(d')\|_1 \leq |\alpha - \alpha'| + \|d - d'\|_1.$$

Proof: Let $d, d' \in \Delta^{\leq 1}$, $\alpha \in [0, \|d\|_1]$, and $\alpha' \in [0, \|d'\|_1]$ arbitrary, then

$$\begin{aligned} \|\Upsilon_\alpha(d) - \Upsilon_{\alpha'}(d')\|_1 &= \left\| \sum_{a \in \mathcal{A}} \Xi_\alpha(d)(\cdot, a) - \sum_{a \in \mathcal{A}} \Xi_{\alpha'}(d')(\cdot, a) \right\|_1 \\ &\leq \sum_{a \in \mathcal{A}} \|\Xi_\alpha(d)(\cdot, a) - \Xi_{\alpha'}(d')(\cdot, a)\|_1 = \|\Xi_\alpha(d) - \Xi_{\alpha'}(d')\|_1. \end{aligned}$$

The upper bound follows by the result of Corollary 1. \square

Corollary 3 (Lipschitz Continuity of Γ_α) Let $d, d' \in \Delta_{\mathcal{S} \times \mathcal{A}}$, and suppose the threshold function $\alpha : \Delta_{\mathcal{A}}^{\leq 1} \rightarrow [0, 1]$ is Lipschitz continuous with constant K_α , and satisfies the feasibility condition $\alpha(\nu) \leq \|\nu\|_1$ for all $\nu \in \Delta_{\mathcal{A}}^{\leq 1}$. Then,

$$\|\Gamma_\alpha(d) - \Gamma_\alpha(d')\|_1 \leq (K_\alpha + 1) \|d - d'\|_1.$$

Proof: By Corollary 2 and the Lipschitz continuity of α , it follows that:

$$\begin{aligned} \|\Gamma_\alpha(d) - \Gamma_\alpha(d')\|_1 &\leq |\alpha(\nu^{-\perp}(d)) - \alpha(\nu^{-\perp}(d'))| + \|d - d'\|_1 \\ &\leq K_\alpha \|d - d'\|_1 + \|d - d'\|_1. \end{aligned}$$

\square

In the next sequence of results, we deal with the stability of winning probabilities given by the function p_{win} . In general, p_{win} is easily seen to have discontinuous jumps, however, a local stability result can be shown if the NZD condition holds.

Lemma 8 (Stability of Winning Probabilities) Let $\mathcal{A} = \{a_1, a_2, \dots, a_K\}$ be a finite set of actions with total order $a_1 < a_2 < \dots < a_K$, and let $\nu \in \Delta_{\mathcal{A}}^{\leq 1}$. For any $\alpha \in [0, 1]$, define the winning probability for action $a \in \mathcal{A}$ as

$$p_{\text{win}}(a, \nu, \alpha) := \begin{cases} 0 & \text{if } \sum_{a' > a} \nu(a') \geq \alpha \text{ or } \alpha = 0, \\ 1 & \text{if } \sum_{a' \geq a} \nu(a') \leq \alpha \text{ and } \alpha > 0, \\ \frac{\alpha - \sum_{a' > a} \nu(a')}{\nu(a)} & \text{otherwise.} \end{cases}$$

Assume that α, ν satisfy the no zero-dominance property (i.e., for any $a \in \mathcal{A}$, $\nu(a) = 0 \Rightarrow \sum_{a' > a} \nu(a') < \alpha(\nu)$) and $\|\nu\|_1 > 0$. Then for all $\nu' \in \Delta_{\mathcal{A}}^{\leq 1}$, $\alpha' \in [0, 1]$, $a \in \mathcal{A}$, p_{win} satisfies

$$|p_{\text{win}}(a, \nu, \alpha) - p_{\text{win}}(a, \nu', \alpha')| \leq C_{\nu, \alpha} \|\nu - \nu'\|_1 + C_{\nu, \alpha} |\alpha - \alpha'|,$$

where:

$$C_{\nu, \alpha} = \begin{cases} \frac{1}{\min\{\nu(a^*)\}} & , \text{ if } \alpha = 0 \\ \frac{1}{\min\{\nu(a^*), \nu(a^-), \Delta_{\nu, \alpha}\}} & , \text{ if } \Delta_{\nu, \alpha} > 0, a^- \neq \perp, \alpha > 0 \\ \frac{1}{\min\{\nu(a^*), \Delta_{\nu, \alpha}\}} & , \text{ if } \Delta_{\nu, \alpha} > 0, a^- = \perp, \alpha > 0 \\ \frac{1}{\min\{\nu(a^*), \nu(a^-)\}} & , \text{ if } \Delta_{\nu, \alpha} = 0, a^- \neq \perp, \alpha > 0 \\ \frac{1}{\min\{\nu(a^*)\}} & , \text{ if } \Delta_{\nu, \alpha} = 0, a^- = \perp, \alpha > 0 \end{cases}$$

and where $a^* := \text{Th}_{\alpha}(\nu)$ is the threshold action, a^- is the action just below the threshold ($a^- := a_{k-1}$ if $a^* = a_k$ for some $k > 1$ and $a^- = \perp$ if $a^* = a_1$), and $\Delta_{\nu, \alpha} := \min_{a \in \mathcal{A}} \left| \alpha - \sum_{a' \geq a} \nu(a') \right|$.

Proof: Importantly, $C_{\nu, \alpha}$ is finite in each case if ν, L satisfies the NZD condition. The proof, while notionally dense, works on a simple idea: in general, p_{win} incorporates discontinuities where the winning probability of an action below the threshold might jump from 0 to 1. However, this does not happen *locally* when there is some probability mass on a^- just below the threshold action. Note that when $\|\nu\|_1 \neq 0$ and NZD holds, it holds that $\nu(a^*) > 0$, and $\nu(a^-) > 0$ if $a^* \neq a_1$ and $\alpha > 0$. Furthermore, by NZD, if $\alpha = 0$, it must hold that $\nu(a_K) > 0$ and $a^* = a_K$ by definition. Define the useful constant

$$\delta := \begin{cases} \Delta_{\nu, \alpha}, & \text{if } \Delta_{\nu, \alpha} > 0, \\ \max\{1/\nu(a^*), 1/\nu(a^-)\}, & \text{if } \Delta_{\nu, \alpha} = 0, a^- \neq \perp \\ 1/\nu(a^*), & \text{if } \Delta_{\nu, \alpha} = 0, a^- = \perp \end{cases}$$

which will be the radius of the open set around which there are no discontinuities of p_{win} .

First, we show that $|p_{\text{win}}(a, \nu, \alpha) - p_{\text{win}}(a, \nu, \alpha')| \leq C_{\nu, \alpha} |\alpha - \alpha'|$ for any α' . Without loss of generality, we can assume that $\alpha \leq \|\nu\|_1, \alpha' \leq \|\nu\|_1$, as $p_{\text{win}}(a, \nu, \alpha') = p_{\text{win}}(a, \nu, \min\{\alpha', \|\nu\|_1\})$ for any α' by definition, and $|\min\{\alpha, \|\nu\|_1\} - \min\{\alpha', \|\nu\|_1\}| \leq |\alpha - \alpha'|$. If $|\alpha - \alpha'| < \delta$, then $\alpha' \in (\alpha - \delta, \alpha + \delta) \cap \mathbb{R}_{\geq 0}$. On the interval $(\alpha - \delta, \alpha + \delta) \cap \mathbb{R}_{\geq 0}$, p_{win} is continuous for any a since

$$\begin{aligned} \text{For } a < a^*, a \neq a^- : \quad & p_{\text{win}}(a, \nu, \alpha') = 0, \\ \text{For } a > a^* : \quad & p_{\text{win}}(a, \nu, \alpha') = 1, \\ \text{For } a = a^* : \quad & p_{\text{win}}(a^*, \nu, \alpha') = \begin{cases} 1 & \text{if } \sum_{a' \geq a^*} \nu(a') \leq \alpha', \\ \frac{\alpha' - \sum_{a' > a^*} \nu(a')}{\nu(a^*)} & \text{otherwise.} \end{cases} \\ \text{For } a = a^-, \text{ if } a^- \neq \perp : \quad & p_{\text{win}}(a^-, \nu, \alpha') = \begin{cases} 0 & \text{if } \sum_{a' > a^-} \nu(a') \geq \alpha', \\ \frac{\alpha' - \sum_{a' > a^-} \nu(a')}{\nu(a^-)} & \text{otherwise.} \end{cases} \end{aligned}$$

Moreover, from above, p_{win} is almost everywhere differentiable in α' in this interval with

$$\left| \frac{\partial p_{\text{win}}(a, \nu, \alpha)}{\partial \alpha} \right| \leq \begin{cases} \max\{1/\nu(a^-), 1/\nu(a^*)\}, & \text{if } a^* \neq a_1, \alpha > 0 \\ 1/\nu(a^*), & \text{otherwise} \end{cases}$$

as in this interval the ‘‘active’’ threshold action will always be either a^* or a^- . Therefore,

$$|p_{\text{win}}(a, \nu, \alpha) - p_{\text{win}}(a, \nu, \alpha')| \leq C_{\nu, \alpha} |\alpha - \alpha'|.$$

If on the other hand $|\alpha - \alpha'| \geq \delta$, then

$$|p_{\text{win}}(a, \nu, \alpha) - p_{\text{win}}(a, \nu, \alpha')| \leq 1 \leq \delta^{-1} |\alpha - \alpha'| \leq C_{\nu, \alpha} |\alpha - \alpha'|.$$

So in both cases it holds that $|p_{\text{win}}(a, \nu, \alpha) - p_{\text{win}}(a, \nu, \alpha')| \leq C_{\nu, \alpha} |\alpha - \alpha'|$.

Next, we show the stability in ν . Once again, assume that $\|\nu' - \nu\|_1 < \delta$. On the open set $\{\nu' : \|\nu' - \nu\|_1 < \delta\}$ the function $p_{\text{win}}(a, \cdot, \alpha)$ is once again continuous in ν' for any a , and almost everywhere differentiable with

$$\left| \frac{\partial p_{\text{win}}(a, \nu', \alpha)}{\partial \nu'(a')} \right| \leq \begin{cases} \frac{1}{\nu(a^*)}, & \text{if } a^- = \perp \text{ or } \alpha = 0 \\ \frac{1}{\min\{\nu(a^*), \nu(a^-)\}}, & \text{if } a^- \neq \perp, \alpha > 0 \end{cases}$$

for almost every ν' . Therefore, it holds that

$$|p_{\text{win}}(a, \nu', \alpha) - p_{\text{win}}(a, \nu, \alpha)| \leq C_{\nu, \alpha} \|\nu - \nu'\|_1.$$

On the other hand, if $\|\nu' - \nu\|_1 \geq \delta$, then

$$|p_{\text{win}}(a, \nu', \alpha) - p_{\text{win}}(a, \nu, \alpha)| \leq 1 \leq \delta^{-1} \|\nu - \nu'\|_1.$$

To complete the proof, we use the triangle inequality:

$$|p_{\text{win}}(a, \nu', \alpha') - p_{\text{win}}(a, \nu, \alpha)| \leq |p_{\text{win}}(a, \nu, \alpha') - p_{\text{win}}(a, \nu, \alpha)| + |p_{\text{win}}(a, \nu, \alpha') - p_{\text{win}}(a, \nu', \alpha')|$$

□

Remark 3 *The proof of Lemma 8 can be adapted to handle action distributions with full support. In this case, a refined version of the first part of the argument shows that the winning probability is Lipschitz-continuous with a constant bounded by $\frac{1}{\min\{\nu(a^-), \nu(a^*), \nu(a^+)\}}$, where a^*, a^-, a^+ are actions around the threshold. Consequently, for policies with full support, i.e., $\pi_h(a|s) \geq \epsilon$ for all $s \in \mathcal{V}$, $a \in \mathcal{A}$, and some $\epsilon > 0$, the deviation in winning probability is bounded by $\frac{1}{(1-\alpha_{\max})\epsilon} \|\nu - \nu'\|_1$.*

Lemma 9 *Let $\mu, \mu' \in \Delta_{\mathcal{S}}$ be two arbitrary state distributions, and let $\pi, \pi' \in \Pi$ be two arbitrary policies. Define the corresponding state-action distributions $L, L' \in \Delta_{\mathcal{S} \times \mathcal{A}}$ as*

$$L(s, a) := \mu(s)\pi(a|s), \quad L'(s, a) := \mu'(s)\pi'(a|s).$$

Then, it holds that

$$\|L - L'\|_1 \leq \sup_s \|\pi(\cdot|s) - \pi'(\cdot|s)\|_1 + \|\mu - \mu'\|_1.$$

Proof: We compute the total variation distance:

$$\begin{aligned} \|L - L'\|_1 &= \sum_{s, a} |\mu(s)\pi(a|s) - \mu'(s)\pi'(a|s)| \\ &= \sum_{s, a} |\mu(s)\pi(a|s) - \mu(s)\pi'(a|s) + \mu(s)\pi'(a|s) - \mu'(s)\pi'(a|s)| \\ &\leq \sum_s \mu(s) \sum_a |\pi(a|s) - \pi'(a|s)| + \sum_s |\mu(s) - \mu'(s)| \sum_a \pi'(a|s) \\ &= \sup_s \|\pi(\cdot|s) - \pi'(\cdot|s)\|_1 + \|\mu - \mu'\|_1, \end{aligned}$$

where the last line uses that $\sum_a \pi'(a|s) = 1$ for all s . □

E.3 Proof of Theorem 2, part 1 (Approximation in Exploitability)

The theorem considers BA-MFGs with Lipschitz-continuous reward, dynamics, and utility functions. Let $K_w \in [0, 2]$ denote the Lipschitz modulus of the state dynamics w_h , K_p the Lipschitz modulus of the payment function p_h , K_u the Lipschitz modulus of the utility function u_h with respect to its second argument, and K_α be the Lipschitz modulus of the allocation function α_h . That is, for any $s \in \mathcal{V}$, $a \in \mathcal{A}$, $\nu, \nu' \in \Delta_{\mathcal{A}}$, $p, p' \in \mathbb{R}_{\geq 0}$,

$$\begin{aligned} |u_h(s, p) - u_h(s, p')| &\leq K_u \cdot |p - p'|, & |p_h(a, \nu) - p_h(a, \nu')| &\leq K_p \cdot \|\nu - \nu'\|_1, \\ |\alpha_h(\nu) - \alpha_h(\nu')| &\leq K_\alpha \cdot \|\nu - \nu'\|_1. \end{aligned}$$

The composed function $u_h(s, p_h(a, \nu))$ then is also Lipschitz with modulus $K_u K_p$ as

$$|u_h(s, p_h(a, \nu)) - u_h(s, p_h(a, \nu'))| \leq K_u \cdot |p_h(a, \nu) - p_h(a, \nu')| \leq K_u K_p \cdot \|\nu - \nu'\|_1,$$

for all $\nu, \nu' \in \Delta_{\mathcal{A}}^{\leq 1}$.

Additionally, let B_p be an upper bound on the absolute value of payments, i.e., $|p_h(a, \nu)| \leq B_p$, and let B_u be an upper bound on the absolute value of utilities, i.e., $|u_h(s, p)| \leq B_u$, for all s, a, ν, h . Such an upper bound always exists if u_h, p_h are continuous in ν , since $\Delta_{\mathcal{A}}^{\leq 1}$ is a compact set.

The argument proceeds in three steps:

1. First, we bound the expected deviation between the empirical distributions and their mean field counterparts. That is, we show an upper bound on the deviation $\mathbb{E} \left[\|L_h^{\bar{\pi}} - \widehat{L}_h\|_1 \right]$.
2. Second, we show that for policies satisfying the no zero-dominance property, the expected transition probabilities associated with item allocation, under both the mean field and finite-population settings, differ proportionally to the deviation in population distributions.
3. Finally, we bound the exploitability of a single agent when all other agents follow a policy that satisfies the no zero-dominance property with respect to the mechanism.

In our analysis, we also make use of the population distribution ξ_h after item allocation at round h . In the mean field setting, this is defined as $\xi_h := \Gamma_{\alpha_h}(L_h)$, where $L_h \in \Delta_{\mathcal{S} \times \mathcal{A}}$ is the state-action distribution at round h , and the operator Γ_{α_h} captures the expected post-allocation state distribution under the mechanism (e.g., by marking winners as inactive). This quantity serves as the input to the state transition function w_h in the MFG dynamics. In the finite-agent setting, we denote the analogous empirical quantity by $\widehat{\xi}_h$, representing the empirical distribution over states immediately after allocation. We also define the random variables $\{z_h^i\}$ as $z_h^i = \perp$ if agent i was not active in round h (i.e. $s_h^i = \perp$) or agent i won the the auction in round h , and $z_h^i = s_h^i$ otherwise. With this definition,

$$\widehat{\xi}_h = \frac{1}{N} \sum_{i=1}^N \mathbf{e}_{z_h^i} \in \Delta_{\mathcal{S}}.$$

Finally, we define the constants used in our convergence analysis as follows. Define

$$K_s := \sup_{s, s', \xi} \|w(s, \xi) - w(s', \xi)\|_1, \quad K_\xi := K_w + \frac{1}{2} K_s.$$

E.3.1 Step 1: Expected Deviation of Empirical Distributions

We derive explicit bounds on the expected deviation between the empirical distributions and their mean field counterparts. In particular, we bound the deviations for the state distribution and the state-action distribution.

Lemma 10 *Let $\mathcal{M}_{mfa} = (\mathcal{S}, \mathcal{A}, H, \mu_0, \{P_h^{mfa}\}_{h=0}^{H-1}, \{R_h^{mfa}\}_{h=0}^{H-1})$ define a BA-MFG. Consider the corresponding finite-agent Batched Auction model with N agents, \mathcal{G}_{auc} , which is approximated by \mathcal{M}_{mfa} . Let $\boldsymbol{\pi} = \{\pi_h^i\}_{h=0, \dots, H-1, i \in [N]} \in \Pi_H^N$ denote the joint policy of the population. Denote by $\widehat{\mu}_h \in \Delta_{\mathcal{S}}$ the empirical state distribution and by $\widehat{L}_h \in \Delta_{\mathcal{S} \times \mathcal{A}}$ the empirical state-action distribution at round h .*

Let $\bar{\pi} \in \Pi_H$ arbitrary, and define the associated mean field state-action distribution flow $L^{\bar{\pi}} := \Lambda_{mfa}(\bar{\pi})$, with corresponding marginal state distribution $\mu_h^{\bar{\pi}} := \sum_a L_h^{\bar{\pi}}(\cdot, a)$. Then, for all $h \in \{0, \dots, H-1\}$, the following bound holds:

$$\mathbb{E} \left[\|L_h^{\bar{\pi}} - \widehat{L}_h\|_1 \right] \leq \mathbb{E} \left[\|\mu_h^{\bar{\pi}} - \widehat{\mu}_h\|_1 \right] + \frac{1}{N} \sum_{i \in [N]} \|\bar{\pi}_h - \pi_h^i\|_1 + \sqrt{\frac{|\mathcal{A}||\mathcal{S}|}{N}}.$$

Proof: We decompose the deviation:

$$\mathbb{E} \left[\|\widehat{L}_h^\pi - \widehat{L}_h\|_1 \middle| \{s_h^i\}_{i,h} \right] \leq \underbrace{\|\widehat{L}_h^\pi - \mathbb{E}[\widehat{L}_h | \{s_h^i\}_{i,h}]\|_1}_{(\square)} + \underbrace{\mathbb{E} \left[\left\| \widehat{L}_h - \mathbb{E}[\widehat{L}_h | \{s_h^i\}_{i,h}] \right\|_1 \middle| \{s_h^i\}_{i,h} \right]}_{(\triangle)}.$$

We bound the two terms separately. For (\square) , define $L \in \Delta_{\mathcal{S} \times \mathcal{A}}$ as $L(s, a) = \widehat{\mu}_h(s) \widehat{\pi}(a|s) = \frac{1}{N} \sum_{i \in [N]} \mathbb{1}_{s_h^i = s} \widehat{\pi}(a|s)$. We (almost surely) have:

$$\begin{aligned} \|\widehat{L}_h^\pi - \mathbb{E}[\widehat{L}_h | \{s_h^i\}_{i,h}]\|_1 &\leq \|\widehat{L}_h^\pi - L\|_1 + \|\mathbb{E}[\widehat{L}_h | \widehat{\mu}_h] - L\|_1 \\ &\leq \sum_{s \in \mathcal{S}} \sum_{a \in \mathcal{A}} |\mu_h^\pi(s) \widehat{\pi}(a|s) - \widehat{\mu}_h(s) \widehat{\pi}(a|s)| + \|\mathbb{E}[\widehat{L}_h | \{s_h^i\}_{i,h}] - L\|_1 \\ &\leq \|\mu_h^\pi - \widehat{\mu}_h\|_1 + \frac{1}{N} \sum_{i \in [N]} \sum_{a \in \mathcal{A}} |\widehat{\pi}(a|s_h^i) - \pi_h^i(a|s_h^i)| \\ &\leq \|\mu_h^\pi - \widehat{\mu}_h\|_1 + \frac{1}{N} \sum_{i \in [N]} \|\widehat{\pi}_h - \pi_h^i\|_1. \end{aligned}$$

For the second term (\triangle) , by applying Jensen's inequality, we have:

$$\begin{aligned} \mathbb{E} \left[\left\| \widehat{L}_h - \mathbb{E}[\widehat{L}_h | \{s_h^i\}_{i,h}] \right\|_1 \middle| \{s_h^i\}_{i,h} \right] &= \sum_{s \in \mathcal{S}} \sum_{a \in \mathcal{A}} \mathbb{E} \left[\left| \widehat{L}_h(s, a) - \mathbb{E}[\widehat{L}_h(s, a) | \{s_h^i\}_{i,h}] \right| \middle| \{s_h^i\}_{i,h} \right] \\ &\leq \sum_{s \in \mathcal{S}} \sum_{a \in \mathcal{A}} \sqrt{\text{Var} \left[\widehat{L}_h(s, a) \middle| \{s_h^i\}_{i,h} \right]} \\ &= \frac{1}{N} \sum_{s \in \mathcal{S}} \sum_{a \in \mathcal{A}} \sqrt{\sum_{i \in [N], s_h^i = s} \pi_h^i(a|s) (1 - \pi_h^i(a|s))}. \end{aligned}$$

Applying Cauchy-Schwarz's inequality, we get for any $s \in \mathcal{S}$:

$$\begin{aligned} \sum_{a \in \mathcal{A}} \sqrt{\sum_{i \in [N], s_h^i = s} \pi_h^i(a|s) (1 - \pi_h^i(a|s))} &\leq \sqrt{|\mathcal{A}| \sum_{a \in \mathcal{A}} \sum_{i \in [N], s_h^i = s} \pi_h^i(a|s) (1 - \pi_h^i(a|s))} \\ &\leq \sqrt{|\mathcal{A}| \sum_{a \in \mathcal{A}} \sum_{i \in [N], s_h^i = s} \pi_h^i(a|s)} \\ &\leq \sqrt{N |\mathcal{A}| \widehat{\mu}_h(s)}. \end{aligned}$$

By integrating this result into the previous computation and using Cauchy-Schwarz's inequality, we get:

$$\begin{aligned} \mathbb{E} \left[\left\| \widehat{L}_h - \mathbb{E}[\widehat{L}_h | \{s_h^i\}_{i,h}] \right\|_1 \middle| \{s_h^i\}_{i,h} \right] &= \frac{1}{N} \sum_{s \in \mathcal{S}} \sum_{a \in \mathcal{A}} \sqrt{\sum_{i \in [N], s_h^i = s} \pi_h^i(a|s) (1 - \pi_h^i(a|s))} \\ &\leq \frac{\sqrt{|\mathcal{A}|}}{\sqrt{N}} \sum_{s \in \mathcal{S}} \sqrt{\widehat{\mu}_h(s)} \leq \sqrt{\frac{|\mathcal{A}| |\mathcal{S}|}{N}} \sum_{s \in \mathcal{S}} \widehat{\mu}_h(s) \leq \sqrt{\frac{|\mathcal{A}| |\mathcal{S}|}{N}}. \end{aligned}$$

Combining the upper bounds derived for terms (\square) and (\triangle) , we obtain the desired result, as $\mathbb{E} \left[\|\widehat{L}_h^\pi - \widehat{L}_h\|_1 \right] = \mathbb{E} \left[\mathbb{E} \left[\|\widehat{L}_h^\pi - \widehat{L}_h\|_1 \middle| \{s_h^i\}_{i,h} \right] \right]$. \square

Lemma 11 (Deviation Between Empirical and Mean Field Population) Let $\mathcal{M}_{mfa} = (\mathcal{S}, \mathcal{A}, H, \mu_0, \{P_h^{mfa}\}_{h=0}^{H-1}, \{R_h^{mfa}\}_{h=0}^{H-1})$ be a BA-MFG. Let $\{\alpha_h\}_{h=0}^{H-1}$, $\{w_h\}_{h=0}^{H-1}$, $\{p_h\}_{h=0}^{H-1}$, and $\{u_h\}_{h=0}^{H-1}$ denote the allocation thresholds, transition dynamics, payment, and utility functions, respectively, from which $\{P_h^{mfa}\}_{h=0}^{H-1}$ and $\{R_h^{mfa}\}_{h=0}^{H-1}$ are derived. Assume these functions are Lipschitz continuous, with respective Lipschitz constants K_α , K_w , K_p , and K_u .

Consider the corresponding finite-agent Batched Auction model with N agents \mathcal{G}_{auc} , which is approximated by \mathcal{M}_{mfa} . Let $\boldsymbol{\pi} = \{\pi_h^i\}_{h=0, \dots, H-1, i \in [N]} \in \Pi_H^N$ denote the joint policy of the population. For

each round h , denote by $\widehat{L}_h \in \Delta_{\mathcal{S} \times \mathcal{A}}$ the empirical state-action distribution and by $\widehat{\mu}_h \in \Delta_{\mathcal{S}}$ the corresponding empirical state distribution.

Let $\bar{\pi} \in \Pi_H$ be an arbitrary policy, and define the associated mean field state-action distribution flow $L^{\bar{\pi}} := \Lambda_{\text{mfa}}(\bar{\pi})$, with corresponding marginal state distribution $\mu_h^{\bar{\pi}} := \sum_a L_h^{\bar{\pi}}(\cdot, a)$. Then, for all $h \in \{0, \dots, H-1\}$, it holds that

$$\begin{aligned} \mathbb{E} [\|\mu_h^{\bar{\pi}} - \widehat{\mu}_h\|_1] &\leq \frac{1 - (K_\xi(1 + K_\alpha))^{h+1}}{1 - K_\xi(1 + K_\alpha)} \frac{\sqrt{|\mathcal{S}|}}{\sqrt{N}} \\ &\quad + K_\xi \frac{1 - (K_\xi(1 + K_\alpha))^h}{1 - K_\xi(1 + K_\alpha)} \left(\frac{\sqrt{|\mathcal{S}|}}{2\sqrt{N}} + \frac{1}{N} + (1 + K_\alpha) \frac{\sqrt{|\mathcal{S}||\mathcal{A}|}}{\sqrt{N}} \right) \\ &\quad + \sum_{h' < h} (K_\xi \cdot (K_\alpha + 1))^{h-h'} \cdot \frac{1}{N} \sum_{i \in [N]} \|\bar{\pi}_{h'} - \pi_{h'}^i\|_1, \end{aligned}$$

if $K_\xi(1 + K_\alpha) \neq 1$, and

$$\begin{aligned} \mathbb{E} [\|\mu_h^{\bar{\pi}} - \widehat{\mu}_h\|_1] &\leq (h+1) \frac{\sqrt{|\mathcal{S}|}}{\sqrt{N}} + h \frac{\sqrt{|\mathcal{S}||\mathcal{A}|}}{\sqrt{N}} + hK_\xi \left(\frac{\sqrt{|\mathcal{S}|}}{2\sqrt{N}} + \frac{1}{N} \right) \\ &\quad + \sum_{h' < h} \frac{1}{N} \sum_{i \in [N]} \|\bar{\pi}_{h'} - \pi_{h'}^i\|_1, \end{aligned}$$

if $K_\xi(1 + K_\alpha) = 1$.

Proof: Let $\{s_h^i\}_{h=0, \dots, H-1, i \in [N]}$, $\{a_h^i\}_{h=0, \dots, H-1, i \in [N]}$ and $\{z_h^i\}_{h=0, \dots, H-1, i \in [N]}$ as in Definition 6. We prove the lemma inductively over h . For $h = 0$ we have $\mu_0^{\bar{\pi}} = \mu_0 = \mathbb{E}[\widehat{\mu}_0]$. Let $X_s := \sum_{i \in [N]} \mathbb{1}_{\{s_0^i = s\}}$, which is by definition X_s a binomial random variable with parameters N and $\mu_0(s)$. Since it is a sum of independent Bernoulli random variables, its variance is $\text{Var}[X_s] = N\mu_0(s)(1 - \mu_0(s))$. By using Jensen's, we can upper bound the expected absolute deviation for each state $s \in \mathcal{V}$

$$\mathbb{E}[|\mu_0(s) - \widehat{\mu}_0(s)|] \leq \sqrt{\text{Var}[\widehat{\mu}_0(s)]} = \sqrt{\frac{\mu_0(s)(1 - \mu_0(s))}{N}} \leq \sqrt{\frac{\mu_0(s)}{N}}$$

By summing over all states $s \in \mathcal{V}$ and applying Cauchy-Schwarz's inequality, we get:

$$\mathbb{E}[\|\mu_0 - \widehat{\mu}_0\|_1] = \sum_{s \in \mathcal{V}} \mathbb{E}[|\mu_0(s) - \widehat{\mu}_0(s)|] \leq \sum_{s \in \mathcal{V}} \sqrt{\frac{\mu_0(s)}{N}} \stackrel{C.S.}{\leq} \frac{\sqrt{|\mathcal{V}|}}{\sqrt{N}}.$$

Next, for $h \geq 0$, we compute an upper bound for the deviation at step $h+1$. In particular, we analyze the conditional expectation

$$\begin{aligned} &\mathbb{E} [\|\mu_{h+1}^{\bar{\pi}} - \widehat{\mu}_{h+1}\|_1 | \{z_h^i\}_{i=1}^N] \\ &\leq \underbrace{\mathbb{E} [\|\mu_{h+1}^{\bar{\pi}} - \mathbb{E}[\widehat{\mu}_{h+1} | \{z_h^i\}_{i=1}^N]\|_1 | \{z_h^i\}_{i=1}^N]}_{:= (\Delta)} + \underbrace{\mathbb{E} [\|\widehat{\mu}_{h+1} - \mathbb{E}[\widehat{\mu}_{h+1} | \{z_h^i\}_{i=1}^N]\|_1 | \{z_h^i\}_{i=1}^N]}_{:= (\square)}, \end{aligned}$$

almost surely, where $\{z_h^i\}_{i=1}^N$ are the states of agents after the item allocation in round h as before. We upper bound the two terms (\square) and (Δ) separately. For (\square) we have:

$$(\square) = \sum_{s \in \mathcal{S}} \mathbb{E} [|\widehat{\mu}_{h+1}(s) - \mathbb{E}[\widehat{\mu}_{h+1}(s) | \{z_h^i\}_{i=1}^N]|_1 | \{z_h^i\}_{i=1}^N] \leq \sum_{s \in \mathcal{S}} \sqrt{\text{Var}[\widehat{\mu}_{h+1}(s) | \{z_h^i\}_{i=1}^N]}.$$

For arbitrary $s \in \mathcal{S}$, noting that $\widehat{\xi}_h$ is $\{z_h^i\}$ -measurable,

$$\text{Var}[\widehat{\mu}_{h+1}(s) | \{z_h^i\}] = \frac{1}{N^2} \sum_{s' \in \mathcal{S}} N \widehat{\xi}_h(s') w(s|s', \widehat{\xi}_h) (1 - w(s|s', \widehat{\xi}_h)) \leq \frac{1}{N} \sum_{s' \in \mathcal{S}} \widehat{\xi}_h(s') w(s|s', \widehat{\xi}_h).$$

By the Cauchy-Schwarz inequality, we have

$$\left(\sum_{s \in \mathcal{S}} \sqrt{\sum_{s' \in \mathcal{S}} \widehat{\xi}_h(s') w(s|s', \widehat{\xi}_h)} \right)^2 \leq \left(\sum_{s \in \mathcal{S}} \sum_{s' \in \mathcal{S}} \widehat{\xi}_h(s') w(s|s', \widehat{\xi}_h) \right) |\mathcal{S}| = |\mathcal{S}|.$$

Therefore, we (almost surely) have

$$(\square) \leq \frac{1}{\sqrt{N}} \sum_{s \in \mathcal{S}} \sqrt{\sum_{s' \in \mathcal{S}} \widehat{\xi}_h(s') w(s|s', \widehat{\xi}_h)} \leq \frac{\sqrt{|\mathcal{S}|}}{\sqrt{N}}.$$

For (\triangle) it holds that

$$\begin{aligned} (\triangle) &= \mathbb{E} \left[\|\mu_{h+1}^{\bar{\pi}} - \mathbb{E}[\widehat{\mu}_{h+1} | \{z_h^i\}]\|_1 | \{z_h^i\} \right] = \|\mu_{h+1}^{\bar{\pi}} - \mathbb{E}[\widehat{\mu}_{h+1} | \{z_h^i\}]\|_1 \\ &= \left\| \Gamma_g(\xi_h^{\bar{\pi}}) - \Gamma_g(\widehat{\xi}_h) \right\|_1 \leq \left(K_w + \frac{K_s}{2} \right) \|\xi_h^{\bar{\pi}} - \widehat{\xi}_h\|_1, \end{aligned}$$

where in the last step we applied Lemma 2.2 from [75]. Merging the upper bounds for (\triangle) and (\square) and taking expectations, we obtain

$$\mathbb{E}[\|\mu_{h+1}^{\bar{\pi}} - \widehat{\mu}_{h+1}\|_1] \leq \frac{\sqrt{|\mathcal{S}|}}{\sqrt{N}} + \left(K_w + \frac{K_s}{2} \right) \mathbb{E}[\|\xi_h^{\bar{\pi}} - \widehat{\xi}_h\|_1].$$

Next, we bound the expected deviation between the post-allocation state distribution in the mean field model and its empirical counterpart in the finite-agent system. Specifically, we consider $\mathbb{E}[\|\xi_h^{\bar{\pi}} - \widehat{\xi}_h\|_1]$, where $\xi_h^{\bar{\pi}} := \Gamma_{\alpha_h}(L_h^{\bar{\pi}})$ is the mean field post-allocation distribution induced by policy $\bar{\pi}$, and $\widehat{\xi}_h$ is the corresponding empirical distribution in the finite-agent auction, computed from realized allocations rather than via the operator Γ_{α_h} . Denote the σ -algebra induced by $\{s_h^i, a_h^i\}_i$ as \mathcal{F}_h for simplicity, then

$$\mathbb{E}[\|\xi_h^{\bar{\pi}} - \widehat{\xi}_h\|_1 | \mathcal{F}_h] \leq \underbrace{\mathbb{E}[\|\xi_h^{\bar{\pi}} - \mathbb{E}[\widehat{\xi}_h | \mathcal{F}_h]\|_1 | \mathcal{F}_h]}_{:= (\diamond)} + \underbrace{\mathbb{E}[\|\widehat{\xi}_h - \mathbb{E}[\widehat{\xi}_h | \mathcal{F}_h]\|_1 | \mathcal{F}_h]}_{:= (\heartsuit)}.$$

We upper-bound the two terms separately once again. For (\heartsuit) we have:

$$\begin{aligned} (\heartsuit) &= \mathbb{E}[\|\widehat{\xi}_h - \mathbb{E}[\widehat{\xi}_h | \mathcal{F}_h]\|_1 | \mathcal{F}_h] = \mathbb{E} \left[\sum_{s \in \mathcal{V}} \left| \widehat{\xi}_h(s) - \mathbb{E}[\widehat{\xi}_h(s) | \mathcal{F}_h] \right| \middle| \mathcal{F}_h \right] \\ &= \sum_{s \in \mathcal{V}} \mathbb{E} \left[\left| \widehat{\xi}_h(s) - \mathbb{E}[\widehat{\xi}_h(s) | \mathcal{F}_h] \right| \middle| \mathcal{F}_h \right]. \end{aligned}$$

We establish an upper bound on the absolute deviation for each state $s \in \mathcal{V}$. Let $s \in \mathcal{V}$ be arbitrary. Given the empirical state-action distribution $\widehat{L}_h \in \Delta_{\mathcal{S} \times \mathcal{A}}$, the corresponding empirical state distribution is given by marginalizing over actions: $\widehat{\mu}_h = \sum_{a \in \mathcal{A}} \widehat{L}_h(\cdot, a)$. By definition of $\widehat{\mu}_h$, there are $N\widehat{\mu}_h(s)$ agents in state s at round h . Denoting these agents as $i_1, \dots, i_{N\widehat{\mu}_h(s)}$, we can express $\widehat{\xi}_h(s)$ as:

$$\widehat{\xi}_h(s) = \frac{1}{N} \sum_{j \in [N\widehat{\mu}_h(s)]} \mathbb{1}_{z_h^{i_j} = s}$$

Two key observations can be made regarding these indicator variables:

1. The indicators are negatively correlated due to the structure of the auction. To illustrate this, assume without loss of generality that the first M agents have not won yet, i.e., $s_h^i \neq \perp$ for all $i \in [M]$. Since the number of items in each round is fixed, when conditioning on \widehat{L}_h , we have

$$\left[\alpha_h(\nu^{-\perp}(\widehat{L}_h))N \right] = \sum_{i=1}^M \mathbb{1}_{z_h^i = \perp},$$

which implies that the indicator variables $\mathbb{1}_{z_h^i = \perp}$ are negatively correlated. Consequently, their complements $\mathbb{1}_{z_h^i = s_h^i} = 1 - \mathbb{1}_{z_h^i = \perp}$ are also negatively correlated. This implies that any subset of these indicator variables retains this negative correlation property. Specifically, for every state s , the random variables $\mathbb{1}_{z_h^{i_j} = s}$, $j \in [N\widehat{\mu}_h(s)]$, are negatively correlated.

2. Since these are Bernoulli random variables, their variance is at most $1/4$.

It follows from the two observations above that the variance of $\widehat{\xi}_h$ conditioned on \mathcal{F}_h can be upper bounded almost surely as follows:

$$\begin{aligned}\mathbb{V}\text{ar}\left[\widehat{\xi}_h(s)\middle|\mathcal{F}_h\right] &= \mathbb{V}\text{ar}\left[\frac{1}{N}\sum_{j\in[N\widehat{\mu}_h(s)]}\mathbb{1}_{z_h^{ij}=s}\middle|\mathcal{F}_h\right] = \frac{1}{N^2}\mathbb{V}\text{ar}\left[\sum_{j\in[N\widehat{\mu}_h(s)]}\mathbb{1}_{z_h^{ij}=s}\middle|\mathcal{F}_h\right] \\ &\leq \frac{1}{N^2}\sum_{j\in[N\widehat{\mu}_h(s)]}\mathbb{V}\text{ar}\left[\mathbb{1}_{z_h^{ij}=s}\middle|\mathcal{F}_h\right] \leq \frac{\widehat{\mu}_h(s)}{4N}.\end{aligned}$$

Using Jensen's inequality, we can bound the absolute deviation using the variance:

$$\mathbb{E}\left[\left|\widehat{\xi}_h(s) - \mathbb{E}[\widehat{\xi}_h(s)\middle|\mathcal{F}_h]\right|\middle|\mathcal{F}_h\right] \leq \sqrt{\mathbb{V}\text{ar}\left[\widehat{\xi}_h(s)\middle|\mathcal{F}_h\right]} \leq \sqrt{\frac{\widehat{\mu}_h(s)}{4N}}$$

Using this result, together with Cauchy-Schwarz's inequality, we can further bound (\heartsuit) :

$$(\heartsuit) = \sum_{s\in\mathcal{V}}\mathbb{E}\left[\left|\widehat{\xi}_h(s) - \mathbb{E}[\widehat{\xi}_h(s)\middle|\mathcal{F}_h]\right|\middle|\mathcal{F}_h\right] \leq \sum_{s\in\mathcal{V}}\sqrt{\frac{\widehat{\mu}_h(s)}{4N}} \stackrel{C.S.}{\leq} \frac{\sqrt{|\mathcal{V}|}}{2\sqrt{N}}$$

For the term (\diamond) , applying the result of Corollary 3 yields:

$$\begin{aligned}(\diamond) &= \mathbb{E}\left[\left\|\widehat{\xi}_h^\pi - \mathbb{E}[\widehat{\xi}_h\middle|\mathcal{F}_h]\right\|_1\middle|\mathcal{F}_h\right] = \|\Gamma_{\alpha_h}(L_h^\pi) - \Gamma_{\alpha_h}(\widehat{L}_h)\|_1 + \|\mathbb{E}[\widehat{\xi}_h\middle|\mathcal{F}_h] - \Gamma_{\alpha_h}(\widehat{L}_h)\|_1 \\ &\leq |\alpha_h(\nu^\perp(L_h^\pi)) - \alpha_h(\nu^\perp(\widehat{L}_h))| + \|L_h^\pi - \widehat{L}_h\|_1 + \frac{1}{N} \\ &\leq K_\alpha\|(\nu^\perp(L_h^\pi) - \nu^\perp(\widehat{L}_h))\|_1 + \|L_h^\pi - \widehat{L}_h\|_1 + \frac{1}{N} \\ &\leq (K_\alpha + 1)\|L_h^\pi - \widehat{L}_h\|_1 + \frac{1}{N},\end{aligned}$$

where the second to last step follows from the Lipschitz continuity of α_h , while the last step follows from $\|(\nu^\perp(L_h^\pi) - \nu^\perp(\widehat{L}_h))\|_1 \leq \|L_h^\pi - \widehat{L}_h\|_1$. Additionally, $\|\mathbb{E}[\widehat{\xi}_h\middle|\mathcal{F}_h] - \Gamma_{\alpha_h}(\widehat{L}_h)\|_1 \leq \frac{1}{N}$ comes from Lemma 7, as $|\lfloor N\alpha_h(\nu^\perp(L_h^\pi)) \rfloor - N\alpha_h(\nu^\perp(\widehat{L}_h))| \leq 1$.

Combining the upper bounds for (\heartsuit) and (\diamond) we get:

$$\begin{aligned}\mathbb{E}\left[\|\widehat{\xi}_h^\pi - \widehat{\xi}_h\|_1\right] &\leq (K_\alpha + 1)\mathbb{E}\left[\|L_h^\pi - \widehat{L}_h\|_1\right] + \frac{\sqrt{|\mathcal{V}|}}{2\sqrt{N}} + \frac{1}{N} \\ &\leq (K_\alpha + 1)\left(\mathbb{E}\left[\|\mu_h^\pi - \widehat{\mu}_h\|_1\right] + \frac{1}{N}\sum_{i\in[N]}\|\pi_h - \pi_h^i\|_1 + \sqrt{\frac{|\mathcal{A}|\cdot|\mathcal{V}|}{N}}\right) + \frac{\sqrt{|\mathcal{V}|}}{2\sqrt{N}} + \frac{1}{N},\end{aligned}$$

where the last step follows from Corollary 4.

Combining this result with the bound on $\mathbb{E}[\|\mu_{h+1}^\pi - \widehat{\mu}_{h+1}\|_1]$, we apply induction on h to conclude the proof of the lemma. \square

Corollary 4 (Deviation Between Empirical and Mean Field State-Action Population) *Under the conditions of Lemma 11, for all $h \in \{0, \dots, H-1\}$, it holds that:*

$$\begin{aligned}\mathbb{E}\left[\|L_h^\pi - \widehat{L}_h\|_1\right] &\leq \frac{1 - (K_\xi(1 + K_\alpha))^{h+1}}{1 - K_\xi(1 + K_\alpha)}\left(\frac{\sqrt{|\mathcal{S}|}}{\sqrt{N}} + \frac{\sqrt{|\mathcal{S}||\mathcal{A}|}}{\sqrt{N}}\right) \\ &\quad + K_\xi\frac{1 - (K_\xi(1 + K_\alpha))^h}{1 - K_\xi(1 + K_\alpha)}\left(\frac{\sqrt{|\mathcal{S}|}}{2\sqrt{N}} + \frac{1}{N}\right) \\ &\quad + \sum_{h'\leq h}(K_\xi \cdot (K_\alpha + 1))^{h-h'} \cdot \frac{1}{N}\sum_{i\in[N]}\|\pi_{h'} - \pi_{h'}^i\|_1,\end{aligned}$$

if $K_\xi(1 + K_\alpha) \neq 1$, and

$$\begin{aligned} \mathbb{E} \left[\|L_h^\pi - \widehat{L}_h\|_1 \right] &\leq (h+1) \left(\frac{\sqrt{|\mathcal{S}|}}{\sqrt{N}} + \frac{\sqrt{|\mathcal{S}||\mathcal{A}|}}{\sqrt{N}} \right) + hK_\xi \left(\frac{\sqrt{|\mathcal{S}|}}{2\sqrt{N}} + \frac{1}{N} \right) \\ &\quad + \sum_{h' \leq h} \frac{1}{N} \sum_{i \in [N]} \|\bar{\pi}_{h'}^i - \pi_{h'}^i\|_1, \end{aligned}$$

if $K_\xi(1 + K_\alpha) = 1$.

Proof: The upper bound is obtained easily from Lemmas 10 and 11. \square

E.3.2 Step 2: Expected Deviation of Winning probabilities

We derive an explicit upper bound on the expected deviation in winning distributions between the mean field auction and its finite-agent counterpart, under a single-agent deviation from a common policy.

Lemma 12 (Expected Deviation in Allocation Dynamics) Let $\mathcal{M}_{mfa} = (\mathcal{S}, \mathcal{A}, H, \mu_0, \{P_h^{mfa}\}_{h=0}^{H-1}, \{R_h^{mfa}\}_{h=0}^{H-1})$ be a BA-MFG. Let $\{\alpha_h\}_{h=0}^{H-1}$, $\{w_h\}_{h=0}^{H-1}$, $\{p_h\}_{h=0}^{H-1}$, and $\{u_h\}_{h=0}^{H-1}$ denote the allocation thresholds, transition dynamics, payment, and utility functions, respectively, from which $\{P_h^{mfa}\}_{h=0}^{H-1}$ and $\{R_h^{mfa}\}_{h=0}^{H-1}$ are derived. Assume these functions are Lipschitz continuous, with respective Lipschitz moduli K_α , K_w , K_p , and K_u . Consider the corresponding finite-agent Batched Auction model with N agents approximated by \mathcal{M}_{mfa} . Let $\boldsymbol{\pi} = \{\pi_h^i\}_{h=0, \dots, H-1, i \in [N]} \in \Pi_H^N$ denote the joint policy of the population, and let $\widehat{L}_h \in \Delta_{\mathcal{S}}$ be the empirical state-action distribution at round h .

Let $\bar{\pi} \in \Pi_H$ arbitrary, and define the associated mean field state-action distribution flow $L^\pi := \Lambda_{mfa}(\bar{\pi})$. Then, for all $h \in \{0, \dots, H-1\}$, the following bounds hold:

$$\begin{aligned} &\mathbb{E} \left[\left| p_{\text{win}}(s_h^i, a_h^i, L_h^\pi, \nu^{-\perp}(L_h^\pi)) - p_{\text{win}}(s_h^i, a_h^i, \widehat{L}_h, \frac{\lfloor N\nu^{-\perp}(\widehat{L}_h) \rfloor}{N}) \right| \right] \\ &\leq C_{L_h^\pi, \alpha_h} \cdot \mathbb{E} \left[\|L_h^\pi - \widehat{L}_h\|_1 \right] + \frac{C_{L_h^\pi, \alpha_h}}{K_\alpha + 1} \cdot \frac{1}{N} \end{aligned}$$

and

$$\mathbb{E} \left[\|P_{\alpha_h}(s_h^i, a_h^i, L_h^\pi) - P_{N, \alpha_h}^i(\mathbf{s}_h, \mathbf{a}_h)\|_1 \right] \leq 2C_{L_h^\pi, \alpha_h} \mathbb{E} \left[\|L_h^\pi - \widehat{L}_h\|_1 \right] + 2\frac{C_{L_h^\pi, \alpha_h}}{K_\alpha + 1} \frac{1}{N},$$

where for an arbitrary state-action distribution $L \in \Delta_{\mathcal{S} \times \mathcal{A}}$ and $\alpha : \Delta_{\mathcal{S}}^{\leq 1} \rightarrow [0, 1]$, the constant $C_{L, \alpha}$ is defined as in Lemma 8.

Proof: For the first inequality, we have

$$\begin{aligned} &\mathbb{E} \left[\left| p_{\text{win}}(s_h^i, a_h^i, L_h^\pi, \nu^{-\perp}(L_h^\pi)) - p_{\text{win}}(s_h^i, a_h^i, \widehat{L}_h, \frac{\lfloor N\nu^{-\perp}(\widehat{L}_h) \rfloor}{N}) \right| \right] \\ &\leq \sum_{s, a, L} \mathbb{P}[s_h^i = s, a_h^i = a, \widehat{L}_h = L] \left| p_{\text{win}}(s_h^i, a_h^i, L_h^\pi, \nu^{-\perp}(L_h^\pi)) - p_{\text{win}}(s_h^i, a_h^i, \widehat{L}_h, \frac{\lfloor N\nu^{-\perp}(\widehat{L}_h) \rfloor}{N}) \right| \\ &\leq \sum_{s, a, L} \mathbb{P}[s_h^i = s, a_h^i = a, \widehat{L}_h = L] C_{L_h^\pi, \alpha_h} \cdot \|L_h^\pi - L\|_1 + \frac{C_{L_h^\pi, \alpha_h}}{K_\alpha + 1} \cdot \frac{1}{N} \\ &\leq C_{L_h^\pi, \alpha_h} \mathbb{E} \left[\|L_h^\pi - \widehat{L}_h\|_1 \right] + \frac{C_{L_h^\pi, \alpha_h}}{K_\alpha + 1} \frac{1}{N}, \end{aligned}$$

where in the second-to-last step we used Lemma 8.

Additionally, for the allocation dynamics P_{N,α_h} , since each state $s \in \mathcal{V}$ can transition only either to itself or to the inactive state \perp , it follows that:

$$\begin{aligned} \mathbb{E} [\|P_{\alpha_h}(s_h^i, a_h^i, L_h^\pi) - P_{N,\alpha_h}^i(s_h, \mathbf{a}_h)\|_1] &= \sum_{s \in \mathcal{S}} \mathbb{E} [\|P_{\alpha_h}(s|s_h^i, a_h^i, L_h^\pi) - P_{N,\alpha_h}^i(s|s_h, \mathbf{a}_h)\|_1] \\ &= \mathbb{E} [\|P_{\alpha_h}(s_h^i|s_h^i, a_h^i, L_h^\pi) - P_{N,\alpha_h}^i(s_h^i|s_h, \mathbf{a}_h)\|_1] \\ &\quad + \mathbb{E} [\|P_{\alpha_h}(\perp|s_h^i, a_h^i, L_h^\pi) - P_{N,\alpha_h}^i(\perp|s_h, \mathbf{a}_h)\|_1] \\ &= 2\mathbb{E} [\|P_{\alpha_h}(\perp|s_h^i, a_h^i, L_h^\pi) - P_{N,\alpha_h}^i(\perp|s_h, \mathbf{a}_h)\|_1]. \end{aligned}$$

Since the marginal probability of transitioning to the state \perp corresponds to the winning probability, the bound follows directly from the first inequality. \square

E.3.3 Step 3: Exploitability Deviation for BA-MFG

Finally we prove the absolute difference in expected reward due to a single-side policy deviation.

Theorem 4 Let $\mathcal{M}_{mfa} = (\mathcal{S}, \mathcal{A}, H, \mu_0, \{P_h^{mfa}\}_{h=0}^{H-1}, \{R_h^{mfa}\}_{h=0}^{H-1})$ be a Batched Auction Mean Field Game (BA-MFG). Let $\{\alpha_h\}_{h=0}^{H-1}$, $\{w_h\}_{h=0}^{H-1}$, $\{p_h\}_{h=0}^{H-1}$, and $\{u_h\}_{h=0}^{H-1}$ denote the allocation thresholds, transition dynamics, payment, and utility functions, respectively, from which $\{P_h^{mfa}\}_{h=0}^{H-1}$ and $\{R_h^{mfa}\}_{h=0}^{H-1}$ are derived. Assume these functions are Lipschitz continuous, with respective Lipschitz constants K_α , K_w , K_p , and K_u .

Consider the corresponding finite-agent Batched Auction \mathcal{G}_{auc} with N agents, which is approximated by \mathcal{M}_{mfa} . Let $\bar{\pi} \in \Pi_H$ an arbitrary policy satisfying the no-zero dominance property. Then, for any policy $\pi \in \Pi_H$, $\tau \geq 0$ it holds

$$\left| V_{mfa}^\tau(\Lambda_{mfa}(\bar{\pi}), \pi) - J_{auc}^{\tau,1}(\pi, \underbrace{\bar{\pi}, \dots, \bar{\pi}}_{N-1 \text{ times}}) \right| = \mathcal{O}\left(\frac{1}{\sqrt{N}}\right)$$

Proof: Define the random variables $\{s_h^i, a_h^i, z_h^i\}_{i \in [N], h \in \{0, \dots, H-1\}}$, along with $\{\widehat{L}_h\}_{h=0}^{H-1}$, $\{\widehat{\mu}_h\}_{h=0}^{H-1}$, and $\{\widehat{\xi}_h\}_{h=0}^{H-1}$, as in the definition of the N -player Batched Auction (see Definition 6). Here, s_h^i denotes the state of agent i at round h , a_h^i its action, and z_h^i its hidden state following the allocation step. The random variables \widehat{L}_h , $\widehat{\mu}_h$, and $\widehat{\xi}_h$ represent, respectively, the empirical state-action distribution, the empirical state distribution, and the empirical post-allocation state distribution at round h .

For the Mean-Field Batched Auction, define $\{s_h, a_h, z_h\}_{h=0}^{H-1}$, where s_h and a_h represent the state and action of a representative agent at round h , and z_h denotes its post-allocation hidden state. These evolve deterministically according to the mean-field population flows L^π, μ^π, ξ^π induced by the population policy $\bar{\pi}$.

We divide the proof into three steps:

1. We show that for every $h \in \{0, \dots, H-1\}$ we have:

$$\begin{aligned} &\|\mathbb{P}[s_h = \cdot] - \mathbb{P}[s_h^1 = \cdot]\|_1 \\ &\leq \sum_{h' < h} \mathbb{E} [\|L_{h'}^\pi - \widehat{L}_{h'}\|_1] (K_w(K_\alpha + 1) + 2C_{L_{h'}, \alpha_{h'}}) + \frac{2C_{L_{h'}, \alpha_{h'}}}{(K_\alpha + 1)N} \\ &\quad + hK_w \left(\frac{\sqrt{|\mathcal{S}|}}{2\sqrt{N}} + \frac{1}{N} \right) \end{aligned}$$

2. We show that for every $h \in \{0, \dots, H-1\}$ we have:

$$\begin{aligned} &|\mathbb{E}[R_h^{mfa}(s_h, a_h, L_h^\pi) - R_{N,h}^{mfa}(s_h^1, a_h^1, \widehat{L}_h)]| \\ &\leq (K_u K_p + B_u C_{L_h^\pi, \alpha_h}) \mathbb{E} [\|L_h^\pi - \widehat{L}_h\|_1] + \frac{C_{L_h^\pi, \alpha_h} B_u}{K_\alpha + 1} \frac{B_u}{N} \\ &\quad + B_u \|\mathbb{P}[s_h = \cdot] - \mathbb{P}[s_h^1 = \cdot]\|_1. \end{aligned}$$

3. In the last step we combine the results of the previous steps to prove the main claim of the theorem.

Step 1: We prove the inequality by induction on h . For the base case $h = 0$, we have $\|\mathbb{P}[s_0 = \cdot] - \mathbb{P}[s_0^1 = \cdot]\|_1 = 0$, since both distributions are equal to the initial distribution μ_0 . Now, assuming the bound holds for some $h \geq 0$, we show that it also holds for round $h + 1$.

$$\|\mathbb{P}[s_{h+1} = \cdot] - \mathbb{P}[s_{h+1}^1 = \cdot]\|_1 = \left\| \sum_{z, \xi} \mathbb{P}[z_h^1 = z, \hat{\xi}_h = \xi] w_h(z, \xi) - \sum_z \mathbb{P}[z_h = z] w_h(z, \xi^{\bar{\pi}}) \right\|_1.$$

By adding and subtracting $\sum_{z, \xi} \mathbb{P}[z_h^1 = z, \hat{\xi}_h = \xi] w_h(z, \xi^{\bar{\pi}}) = \sum_z \mathbb{P}[z_h^1 = z] w_h(z, \xi^{\bar{\pi}})$, and applying triangular inequality, we get

$$\begin{aligned} \|\mathbb{P}[s_{h+1} = \cdot] - \mathbb{P}[s_{h+1}^1 = \cdot]\|_1 &\leq \left\| \sum_{z, \xi} \mathbb{P}[z_h^1 = z, \hat{\xi}_h = \xi] (w_h(z, \xi) - w_h(z, \xi^{\bar{\pi}})) \right\|_1 \\ &\quad + \left\| \sum_z (\mathbb{P}[z_h = z] - \mathbb{P}[z_h^1 = z]) w_h(z, \xi^{\bar{\pi}}) \right\|_1 \\ &\leq K_w \mathbb{E}[\|\xi_h^{\bar{\pi}} - \hat{\xi}_h\|_1] + \|\mathbb{P}[z_h = \cdot] - \mathbb{P}[z_h^1 = \cdot]\|_1. \end{aligned}$$

The term $\mathbb{E}[\|\xi_h^{\bar{\pi}} - \hat{\xi}_h\|_1]$, using the same derivation as in the inductive step of Lemma 11, can be further upper bounded as

$$\mathbb{E}[\|\xi_h^{\bar{\pi}} - \hat{\xi}_h\|_1] \leq (K_\alpha + 1) \mathbb{E}[\|L_h^{\bar{\pi}} - \hat{L}_h\|_1] + \frac{\sqrt{|\mathcal{V}|}}{2\sqrt{N}} + \frac{1}{N}.$$

Finally applying a similar reasoning we can upper bound $\|\mathbb{P}[z_h = \cdot] - \mathbb{P}[z_h^1 = \cdot]\|_1$.

$$\begin{aligned} \|\mathbb{P}[z_h = \cdot] - \mathbb{P}[z_h^1 = \cdot]\|_1 &= \left\| \sum_{\mathbf{s}, \mathbf{a}} \mathbb{P}[s_h = \mathbf{s}, \mathbf{a}_h = \mathbf{a}] P_{N, \alpha_h}^1(\mathbf{s}, \mathbf{a}) - \sum_{s, a} \mathbb{P}[s_h = s, a_h = a] P_{\alpha_h}(s, a, L_h^{\bar{\pi}}) \right\|_1. \end{aligned}$$

By adding and subtracting $\sum_{\mathbf{s}, \mathbf{a}} \mathbb{P}[s_h = \mathbf{s}, \mathbf{a}_h = \mathbf{a}] P_{\alpha_h}(s^1, a^1, L_h^{\bar{\pi}}) = \sum_{s, a} \mathbb{P}[s_h^1 = s, a_h^1 = a] P_{\alpha_h}(s, a, L_h^{\bar{\pi}})$, and applying triangular inequality, we get

$$\begin{aligned} \|\mathbb{P}[z_h = \cdot] - \mathbb{P}[z_h^1 = \cdot]\|_1 &\leq \sum_{\mathbf{s}, \mathbf{a}} \mathbb{P}[s_h = \mathbf{s}, \mathbf{a}_h = \mathbf{a}] \|P_{N, \alpha_h}^1(\mathbf{s}, \mathbf{a}) - P_{\alpha_h}(s^1, a^1, L_h^{\bar{\pi}})\|_1 \\ &\quad + \sum_{s, a} \|P_{\alpha_h}(s, a, L_h^{\bar{\pi}})\|_1 |\mathbb{P}[s_h = s, a_h = a] - \mathbb{P}[s_h^1 = s, a_h^1 = a]| \\ &\leq \mathbb{E}[\|P_{N, \alpha_h}^1(\mathbf{s}_h, \mathbf{a}_h) - P_{\alpha_h}(s_h^1, a_h^1, L_h^{\bar{\pi}})\|_1] \\ &\quad + \sum_s |\mathbb{P}[s_h = s] - \mathbb{P}[s_h^1 = s]| \sum_a \pi_h(a|s) \|P_{\alpha_h}(s, a, L_h^{\bar{\pi}})\|_1. \end{aligned}$$

Applying Lemma 12 it follows

$$\begin{aligned} \|\mathbb{P}[z_h = \cdot] - \mathbb{P}[z_h^1 = \cdot]\|_1 &\leq 2C_{L_h^{\bar{\pi}}, \alpha_h} \mathbb{E}[\|L_h^{\bar{\pi}} - \hat{L}_h\|_1] + 2 \frac{C_{L_h^{\bar{\pi}}, \alpha_h}}{K_\alpha + 1} \frac{1}{N} + \|\mathbb{P}[s_h = \cdot] - \mathbb{P}[s_h^1 = \cdot]\|_1. \end{aligned}$$

By combining the two bounds, we obtain

$$\begin{aligned} \|\mathbb{P}[s_{h+1} = \cdot] - \mathbb{P}[s_{h+1}^1 = \cdot]\|_1 &\leq K_w \left((K_\alpha + 1) \mathbb{E}[\|L_h^{\bar{\pi}} - \hat{L}_h\|_1] + \frac{\sqrt{|\mathcal{V}|}}{2\sqrt{N}} + \frac{1}{N} \right) \\ &\quad + 2C_{L_h^{\bar{\pi}}, \alpha_h} \mathbb{E}[\|L_h^{\bar{\pi}} - \hat{L}_h\|_1] + \frac{2C_{L_h^{\bar{\pi}}, \alpha_h}}{(K_\alpha + 1)N} \\ &\quad + \|\mathbb{P}[s_h = \cdot] - \mathbb{P}[s_h^1 = \cdot]\|_1. \end{aligned}$$

Applying the induction hypothesis completes the proof for round $h + 1$.

Step 2: We prove the inequality by using the result obtained in step 1. Let $h \in \{0, \dots, H - 1\}$ arbitrary, then

$$\begin{aligned} & \left| \mathbb{E}[R_h^{\text{mfa}}(s_h, a_h, L_h^{\bar{\pi}}) - R_{N,h}^{\text{mfa}}(s_h^1, a_h^1, \widehat{L}_h)] \right| \\ & \leq \underbrace{\left| \mathbb{E}[R_h^{\text{mfa}}(s_h, a_h, L_h^{\bar{\pi}}) - R_h^{\text{mfa}}(s_h^1, a_h^1, L_h^{\bar{\pi}})] \right|}_{(\square)} + \underbrace{\left| \mathbb{E}[R_h^{\text{mfa}}(s_h^1, a_h^1, L_h^{\bar{\pi}}) - R_{N,h}^{\text{mfa}}(s_h^1, a_h^1, \widehat{L}_h)] \right|}_{(\triangle)}. \end{aligned}$$

The first term (\square) can be upper bounded by

$$(\square) = \left| \sum_s (\mathbb{P}[s_h = s] - \mathbb{P}[s_h^1 = s]) \sum_a \pi_a(a|s) R_h^{\text{mfa}}(s, a, L_h^{\bar{\pi}}) \right| \leq B_u \|\mathbb{P}[s_h = \cdot] - \mathbb{P}[s_h^1 = \cdot]\|_1.$$

For the second term (\triangle) we have:

$$\begin{aligned} (\triangle) & \leq \sum_{s,a,L} \mathbb{P}[s_h^1 = s, a_h^1 = a, \widehat{L} = L] \left| R_h^{\text{mfa}}(s, a, L_h^{\bar{\pi}}) - R_{N,h}^{\text{mfa}}(s, a, L) \right| \\ & \leq \sum_{s,a,L} \mathbb{P}[s_h^1 = s, a_h^1 = a, \widehat{L} = L] \left| u_h(s, p_h(a, \nu^{-\perp}(L_h^{\bar{\pi}}))) - u_h(s, p_h(a, \nu^{-\perp}(L))) \right| \\ & \quad + B_u \mathbb{E} \left[\left| p_{\text{win}}(s_h^1, a_h^1, L_h^{\bar{\pi}}, \nu^{-\perp}(L_h^{\bar{\pi}})) - p_{\text{win}}(s_h^1, a_h^1, \widehat{L}_h, \frac{\lfloor N \nu^{-\perp}(\widehat{L}_h) \rfloor}{N}) \right| \right] \\ & \leq K_u K_p \sum_{s,a,L} \mathbb{P}[s_h^1 = s, a_h^1 = a, \widehat{L} = L] \|L_h^{\bar{\pi}} - L\|_1 \\ & \quad + B_u C_{L_h^{\bar{\pi}}, \alpha_h} \mathbb{E} \left[\|L_h^{\bar{\pi}} - \widehat{L}_h\|_1 \right] + \frac{C_{L_h^{\bar{\pi}}, \alpha_h} B_u}{K_\alpha + 1} \frac{B_u}{N} \\ & \leq (K_u K_p + B_u C_{L_h^{\bar{\pi}}, \alpha_h}) \mathbb{E} \left[\|L_h^{\bar{\pi}} - \widehat{L}_h\|_1 \right] + \frac{C_{L_h^{\bar{\pi}}, \alpha_h} B_u}{K_\alpha + 1} \frac{B_u}{N}, \end{aligned}$$

where in the second to last step we used the Lipschitz continuity of $u \circ p$ and Lemma 12. Combining both results we get:

$$\begin{aligned} \left| \mathbb{E}[R_h^{\text{mfa}}(s_h, a_h, L_h^{\bar{\pi}}) - R_{N,h}^{\text{mfa}}(s_h^1, a_h^1, \widehat{L}_h)] \right| & \leq (K_u K_p + B_u C_{L_h^{\bar{\pi}}, \alpha_h}) \mathbb{E} \left[\|L_h^{\bar{\pi}} - \widehat{L}_h\|_1 \right] + \frac{C_{L_h^{\bar{\pi}}, \alpha_h} B_u}{K_\alpha + 1} \frac{B_u}{N} \\ & \quad + B_u \|\mathbb{P}[s_h = \cdot] - \mathbb{P}[s_h^1 = \cdot]\|_1. \end{aligned}$$

Step 3: We now combine the results from the previous two steps to establish the final bound stated in the theorem.

$$\begin{aligned} & \left| V_{\text{mfa}}^\tau(\Lambda_{\text{mfa}}(\bar{\pi}), \pi) - J_{\text{auc}}^{\tau,1}(\pi, \bar{\pi}, \dots, \bar{\pi}) \right| \\ & = \left| \mathbb{E} \left[\sum_{h=0}^{H-1} R_h^{\text{mfa}}(s_h, a_h, L_h^{\bar{\pi}}) + \tau \mathcal{H}(\pi_h(s_h)) - R_{N,h}^{\text{mfa}}(s_h^1, a_h^1, \widehat{L}_h) - \tau \mathcal{H}(\pi_h(s_h^1)) \right] \right| \\ & \leq \sum_{h=0}^{H-1} \left| \mathbb{E} \left[R_h^{\text{mfa}}(s_h, a_h, L_h^{\bar{\pi}}) + \tau \mathcal{H}(\pi_h(s_h)) - R_{N,h}^{\text{mfa}}(s_h^1, a_h^1, \widehat{L}_h) - \tau \mathcal{H}(\pi_h(s_h^1)) \right] \right|. \end{aligned}$$

We proceed by bounding each term individually for every round h . Let $h \in \{0, \dots, H - 1\}$ arbitrary, then

$$\begin{aligned} & \left| \mathbb{E} \left[R_h^{\text{mfa}}(s_h, a_h, L_h^{\bar{\pi}}) + \tau \mathcal{H}(\pi_h(s_h)) - R_{N,h}^{\text{mfa}}(s_h^1, a_h^1, \widehat{L}_h) - \tau \mathcal{H}(\pi_h(s_h^1)) \right] \right| \\ & \leq \left| \mathbb{E} \left[R_h^{\text{mfa}}(s_h, a_h, L_h^{\bar{\pi}}) - R_{N,h}^{\text{mfa}}(s_h^1, a_h^1, \widehat{L}_h) \right] \right| + \tau \left| \mathbb{E} \left[\mathcal{H}(\pi_h(s_h)) - \mathcal{H}(\pi_h(s_h^1)) \right] \right|. \end{aligned}$$

The first term is bounded using the result from Step 2, while the second term can be handled as follows:

$$\begin{aligned} \left| \mathbb{E} \left[\mathcal{H}(\pi_h(s_h)) - \mathcal{H}(\pi_h(s_h^1)) \right] \right| & \leq \sum_s |\mathbb{P}[s_h = s] - \mathbb{P}[s_h^1 = s]| \mathcal{H}(\pi_h(s)) \\ & \leq \|\mathbb{P}[s_h = \cdot] - \mathbb{P}[s_h^1 = \cdot]\|_1 \log(|\mathcal{A}|). \end{aligned}$$

Therefore the (entropy regularized) absolute difference in rewards at round h can be upper bounded by

$$\begin{aligned} & \left| \mathbb{E} \left[R_h^{\text{mfa}}(s_h, a_h, L_h^{\bar{\pi}}) + \tau \mathcal{H}(\pi_h(s_h)) - R_{N,h}^{\text{mfa}}(s_h^1, a_h^1, \hat{L}_h) - \tau \mathcal{H}(\pi_h(s_h^1)) \right] \right| \\ & \leq (K_u K_p + B_u C_{L_h^{\bar{\pi}}, \alpha_h}) \mathbb{E} \left[\|L_h^{\bar{\pi}} - \hat{L}_h\|_1 \right] + \frac{C_{L_h^{\bar{\pi}}, \alpha_h} B_u}{K_\alpha + 1} \frac{B_u}{N} \\ & \quad + (B_u + \tau \log(|\mathcal{A}|)) \|\mathbb{P}[s_h = \cdot] - \mathbb{P}[s_h^1 = \cdot]\|_1. \end{aligned}$$

By applying the bound on $\|\mathbb{P}[s_h = \cdot] - \mathbb{P}[s_h^1 = \cdot]\|_1$ derived in Step 1, and summing the per-round deviations over all h , it follows:

$$\begin{aligned} & \left| V_{\text{mfa}}^\tau(\Lambda_{\text{mfa}}(\bar{\pi}), \pi) - J_{\text{auc}}^{\tau,1}(\pi, \underbrace{\bar{\pi}, \dots, \bar{\pi}}_{N-1 \text{ times}}) \right| \\ & \leq \sum_{h=0}^{H-1} (K_u K_p + B_u C_{L_h^{\bar{\pi}}, \alpha_h}) \mathbb{E} \left[\|L_h^{\bar{\pi}} - \hat{L}_h\|_1 \right] + \frac{C_{L_h^{\bar{\pi}}, \alpha_h} B_u}{K_\alpha + 1} \frac{B_u}{N} \\ & \quad + (B_u + \tau \log(|\mathcal{A}|)) \sum_{h=0}^{H-1} \sum_{h' < h} \mathbb{E} \left[\|L_{h'}^{\bar{\pi}} - \hat{L}_{h'}\|_1 \right] (K_w (K_\alpha + 1) + 2C_{L_{h'}^{\bar{\pi}}, \alpha_{h'}}) + \frac{2C_{L_{h'}^{\bar{\pi}}, \alpha_{h'}}}{(K_\alpha + 1)N} \\ & \quad + (B_u + \tau \log(|\mathcal{A}|)) \frac{H(H-1)}{2} K_w \left(\frac{\sqrt{|\mathcal{S}|}}{2\sqrt{N}} + \frac{1}{N} \right) \end{aligned}$$

In particular, Corollary 4 implies that the total (entropy-regularized) reward difference is of order $\mathcal{O}(\frac{1}{\sqrt{N}})$. \square

Conclusion and Statement of Result. Let \mathcal{M}_{mfa} be a BA-MFG with Lipschitz-continuous $\{u_h\}_{h=0}^{H-1}, \{w_h\}_{h=0}^{H-1}, \{\alpha_h\}_{h=0}^{H-1}, \{p_h\}_{h=0}^{H-1}$. Let $\pi_\delta \in \Pi_H$ be a policy that satisfy the no zero-dominance property. Let further assume π_δ is a δ -MFG-NE, namely

$$\delta \geq \max_{\pi' \in \Pi_H} V_{\text{mfa}}^\tau(\Lambda_{\text{mfa}}(\pi_\delta), \pi') - V_{\text{mfa}}^\tau(\Lambda_{\text{mfa}}(\pi_\delta), \pi_\delta).$$

Then, for $\boldsymbol{\pi} = (\pi_\delta, \dots, \pi_\delta)$, we have:

$$\begin{aligned} & \max_{\pi' \in \Pi_H} J_{\text{auc}}^{\tau,i}(\pi', \boldsymbol{\pi}^{-i}) - J_{\text{auc}}^{\tau,i}(\boldsymbol{\pi}) \\ & = \max_{\pi' \in \Pi_H} J_{\text{auc}}^{\tau,i}(\pi', \boldsymbol{\pi}^{-i}) - J_{\text{auc}}^{\tau,i}(\boldsymbol{\pi}) \\ & \quad + V_{\text{mfa}}^\tau(\Lambda_{\text{mfa}}(\pi_\delta), \pi') - V_{\text{mfa}}^\tau(\Lambda_{\text{mfa}}(\pi_\delta), \pi') + V_{\text{mfa}}^\tau(\Lambda_{\text{mfa}}(\pi_\delta), \pi_\delta) - V_{\text{mfa}}^\tau(\Lambda_{\text{mfa}}(\pi_\delta), \pi_\delta) \\ & \leq \max_{\pi' \in \Pi_H} V_{\text{mfa}}^\tau(\Lambda_{\text{mfa}}(\pi_\delta), \pi') - V_{\text{mfa}}^\tau(\Lambda_{\text{mfa}}(\pi_\delta), \pi_\delta) \\ & \quad + |V_{\text{mfa}}^\tau(\Lambda_{\text{mfa}}(\pi_\delta), \pi') - J_{\text{auc}}^{\tau,i}(\pi', \boldsymbol{\pi}^{-i})| + |V_{\text{mfa}}^\tau(\Lambda_{\text{mfa}}(\pi_\delta), \pi_\delta) - J_{\text{auc}}^{\tau,i}(\boldsymbol{\pi})| \\ & \leq \delta + \frac{C_1}{\sqrt{N}} + \frac{C_2}{N} = \delta + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right). \end{aligned}$$

In case $K_\xi(K_\alpha + 1) \neq 1$, the constants C_1 and C_2 are given by

$$\begin{aligned}
C_1 &= 2 \sum_{h=0}^{H-1} (K_u K_p + B_u C_{L_h^{\pi_\delta, \alpha_h}}) \frac{1 - (K_\xi(1 + K_\alpha))^{h+1}}{1 - K_\xi(1 + K_\alpha)} \left(\sqrt{|\mathcal{S}|} + \sqrt{|\mathcal{S}||\mathcal{A}|} \right) \\
&\quad + \sum_{h=0}^{H-1} (K_u K_p + B_u C_{L_h^{\pi_\delta, \alpha_h}}) K_\xi \frac{1 - (K_\xi(1 + K_\alpha))^h}{1 - K_\xi(1 + K_\alpha)} \sqrt{|\mathcal{S}|} \\
&\quad + \sum_{h=0}^{H-1} \sum_{h' < h} \left(K_w(K_\alpha + 1) + 2C_{L_{h'}^{\pi_\delta, \alpha_{h'}}} \right) \\
&\quad \cdot \left[2(B_u + \tau \log |\mathcal{A}|) \frac{1 - (K_\xi(1 + K_\alpha))^{h'+1}}{1 - K_\xi(1 + K_\alpha)} \left(\sqrt{|\mathcal{S}|} + \sqrt{|\mathcal{S}||\mathcal{A}|} \right) \right. \\
&\quad \quad \left. + (B_u + \tau \log |\mathcal{A}|) K_\xi \frac{1 - (K_\xi(1 + K_\alpha))^{h'}}{1 - K_\xi(1 + K_\alpha)} \sqrt{|\mathcal{S}|} \right] \\
&\quad + (B_u + \tau \log(|\mathcal{A}|)) H(H-1) \frac{K_w \sqrt{|\mathcal{S}|}}{2},
\end{aligned}$$

$$\begin{aligned}
C_2 &= 2 \sum_{h=0}^{H-1} (K_u K_p + B_u C_{L_h^{\pi_\delta, \alpha_h}}) \left(\frac{1 - (K_\xi(1 + K_\alpha))^{h+1}}{1 - (K_\xi(1 + K_\alpha))} + K_\xi \frac{1 - (K_\xi(1 + K_\alpha))^h}{1 - (K_\xi(1 + K_\alpha))} \right) \\
&\quad + 2 \sum_{h=0}^{H-1} B_u \frac{C_{L_h^{\pi_\delta, \alpha_h}}}{K_\alpha + 1} \\
&\quad + 2(B_u + \tau \log |\mathcal{A}|) \sum_{h=0}^{H-1} \sum_{h' < h} \left(K_w(K_\alpha + 1) + 2C_{L_{h'}^{\pi_\delta, \alpha_{h'}}} \right) \frac{1 - (K_\xi(1 + K_\alpha))^{h'+1}}{1 - (K_\xi(1 + K_\alpha))} \\
&\quad + 2(B_u + \tau \log |\mathcal{A}|) \sum_{h=0}^{H-1} \sum_{h' < h} \left(K_w(K_\alpha + 1) + 2C_{L_{h'}^{\pi_\delta, \alpha_{h'}}} \right) K_\xi \frac{1 - (K_\xi(1 + K_\alpha))^{h'}}{1 - (K_\xi(1 + K_\alpha))} \\
&\quad + 2(B_u + \tau \log |\mathcal{A}|) \sum_{h=0}^{H-1} \sum_{h' < h} \frac{2C_{L_{h'}^{\pi_\delta, \alpha_{h'}}}}{K_\alpha + 1} \\
&\quad + (B_u + \tau \log |\mathcal{A}|) K_w H(H-1).
\end{aligned}$$

In case $K_\xi(K_\alpha + 1) = 1$, the constants C_1 and C_2 are given by

$$\begin{aligned}
C_1 &= 2 \sum_{h=0}^{H-1} (h+1) \cdot (K_u K_p + B_u C_{L_h^{\pi_\delta, \alpha_h}}) \left(\sqrt{|\mathcal{S}|} + \sqrt{|\mathcal{S}||\mathcal{A}|} \right) \\
&\quad + \sum_{h=0}^{H-1} h \cdot (K_u K_p + B_u C_{L_h^{\pi_\delta, \alpha_h}}) K_\xi \sqrt{|\mathcal{S}|} \\
&\quad + 2(B_u + \tau \log |\mathcal{A}|) \sum_{h=0}^{H-1} \sum_{h' < h} (h'+1) \cdot \left(K_w(K_\alpha + 1) + 2C_{L_{h'}^{\pi_\delta, \alpha_{h'}}} \right) \left(\sqrt{|\mathcal{S}|} + \sqrt{|\mathcal{S}||\mathcal{A}|} \right) \\
&\quad + (B_u + \tau \log |\mathcal{A}|) \sum_{h=0}^{H-1} \sum_{h' < h} h' \cdot \left(K_w(K_\alpha + 1) + 2C_{L_{h'}^{\pi_\delta, \alpha_{h'}}} \right) K_\xi \sqrt{|\mathcal{S}|} \\
&\quad + (B_u + \tau \log(|\mathcal{A}|)) H(H-1) \frac{K_w \sqrt{|\mathcal{S}|}}{2},
\end{aligned}$$

$$\begin{aligned}
C_2 &= 2 \sum_{h=0}^{H-1} (K_u K_p + B_u C_{L_h^{\pi_\delta, \alpha_h}})(h+1 + K_\xi h) + B_u \frac{C_{L_h^{\pi_\delta, \alpha_h}}}{K_\alpha + 1} \\
&\quad + 2(B_u + \tau \log(|\mathcal{A}|)) \sum_{h=0}^{H-1} \sum_{h' < h} (K_w(K_\alpha + 1) + 2C_{L_{h'}^{\pi_\delta, \alpha_{h'}}})(h'+1 + K_\xi h') \\
&\quad + 2(B_u + \tau \log(|\mathcal{A}|)) \sum_{h=0}^{H-1} \sum_{h' < h} \frac{2C_{L_{h'}^{\pi_\delta, \alpha_{h'}}}}{K_\alpha + 1} \\
&\quad + (B_u + \tau \log(|\mathcal{A}|)) K_w H(H-1).
\end{aligned}$$

Remark 4 For policies with full support, the Lipschitz constant associated with the winning probability is of order $\frac{1}{\epsilon}$, as discussed in Remark 3. Consequently, the constants C_1 and C_2 scale as $\mathcal{O}(\frac{1}{\epsilon} H^2 \cdot \frac{1 - (K_\xi(K_\alpha + 1))^H}{1 - K_\xi(K_\alpha + 1)})$ when $K_\xi(K_\alpha + 1) \neq 1$, and as $\mathcal{O}(\frac{1}{\epsilon} H^3)$ when $K_\xi(K_\alpha + 1) = 1$.

Explanation of Constants. In the above expression, the constants represent key components of the BA-MFG dynamics:

- B_u and K_u are the bound and Lipschitz constant of the utility function u respectively.
- B_p and K_p are the bound and Lipschitz constant of the payment function p respectively.
- K_w denotes the Lipschitz constant of the transition function w .
- K_α is the Lipschitz constant of the allocation threshold function α .
- $K_s = \sup_{s, s', \xi} \|w(s, \xi) - w(s', \xi)\|_1$.
- $K_\xi = K_w + \frac{1}{2} K_s$.
- $C_{L, \alpha}$ is the Lipschitz constant of the winning probability function evaluated at the distribution L , assuming L satisfies the no zero-dominance property. For its precise definition, see Lemma 8.
- τ is the entropy regularization parameter.

E.4 Proof of Theorem 2, part 2 (Approximation in Objective)

We show that, under Lipschitz conditions, the objective computed under the mean field approximation closely matches its expected value under a finite population of agents.

Theorem 5 (Convergence of the Mechanism Objective) Let $g : \Delta_{\mathcal{S} \times \mathcal{A}}^H \rightarrow \mathbb{R}$ be a Lipschitz-continuous objective defined over the class of Batched Auction Mean Field Games (BA-MFGs). Let \mathcal{M}_{mfa} be a BA-MFG with Lipschitz-continuous $\{u_h\}_{h=0}^{H-1}$, $\{w_h\}_{h=0}^{H-1}$, $\{\alpha_h\}_{h=0}^{H-1}$, $\{p_h\}_{h=0}^{H-1}$. Let $\boldsymbol{\pi} = (\pi, \dots, \pi)$ be the joint population policy for some $\pi \in \Pi_H$. Then:

$$|g(\Lambda_{mfa}(\boldsymbol{\pi})) - G(\boldsymbol{\pi})| = \mathcal{O}\left(\frac{1}{\sqrt{N}}\right),$$

where $G(\boldsymbol{\pi}) = \mathbb{E}[g(\{\widehat{L}_h\}_{h=0}^{H-1}) | \boldsymbol{\pi}]$.

Proof: We use a decomposition over the support of $\widehat{\mathbf{L}}$ and apply the triangle inequality:

$$\begin{aligned}
|g(\Lambda_{mfa}(\boldsymbol{\pi})) - G(\boldsymbol{\pi})| &= \left| \sum_{\mathbf{L}} \mathbb{P}[\widehat{\mathbf{L}} = \mathbf{L} | \boldsymbol{\pi}] (g(\Lambda_{mfa}(\boldsymbol{\pi})) - g(\mathbf{L})) \right| \\
&\leq \sum_{\mathbf{L}} \mathbb{P}[\widehat{\mathbf{L}} = \mathbf{L} | \boldsymbol{\pi}] |g(\Lambda_{mfa}(\boldsymbol{\pi})) - g(\mathbf{L})| \\
&\leq \sum_{\mathbf{L}} \mathbb{P}[\widehat{\mathbf{L}} = \mathbf{L} | \boldsymbol{\pi}] K_g \|\Lambda_{mfa}(\boldsymbol{\pi}) - \mathbf{L}\|_1 \\
&= K_g \mathbb{E} \left[\|\Lambda_{mfa}(\boldsymbol{\pi}) - \widehat{\mathbf{L}}\|_1 \right],
\end{aligned}$$

where K_g is the Lipschitz constant of g . The result follows by applying the bound from Corollary 4. \square

Lemma 13 (Lipschitz Continuity of Expected Revenue) *Let g_{rev} denote the expected revenue objective. Let \mathcal{M}_{mfa} be a BA-MFG with Lipschitz-continuous payment functions $\{p_h\}_{h=0}^{H-1}$ and allocation functions $\{\alpha_h\}_{h=0}^{H-1}$. Let $\mathbf{L} = \{L_h\}_{h=0}^{H-1}$ and $\mathbf{L}' = \{L'_h\}_{h=0}^{H-1}$ be two arbitrary state-action distribution trajectories over H rounds. Then,*

$$|g_{\text{rev}}(\mathbf{L}) - g_{\text{rev}}(\mathbf{L}')| \leq (B_p(2 + K_\alpha) + K_p)\|\mathbf{L} - \mathbf{L}'\|_1,$$

where B_p is a uniform bound on the absolute value of the payment functions, K_p is their Lipschitz constant, and K_α is the Lipschitz constant of the allocation threshold functions.

Proof: The expected revenue, for $\mathbf{L} = \{L_h\}_{h=0}^{H-1}$, can be rewritten using the operator Ξ

$$g_{\text{rev}}(\mathbf{L}) := \sum_{h=0}^{H-1} \sum_{s,a} (L_h(s, a) - \Xi_{\alpha_h(\nu^{-\perp}(L_h))}(L_h)(s, a)) p_h(a, \nu^{-\perp}(L_h)).$$

To simplify notation, we define

$$\bar{\Xi}_\alpha(L) := L - \Xi_\alpha(L),$$

representing the residual (unallocated) mass at each state-action pair. Applying the triangle inequality:

$$\begin{aligned} |g_{\text{rev}}(\mathbf{L}) - g_{\text{rev}}(\mathbf{L}')| &\leq \sum_{h=0}^{H-1} \left| \sum_{s,a} (\bar{\Xi}_{\alpha_h(\nu^{-\perp}(L_h))}(L_h)(s, a)) p_h(a, \nu^{-\perp}(L_h)) \right. \\ &\quad \left. - (\bar{\Xi}_{\alpha_h(\nu^{-\perp}(L'_h))}(L'_h)(s, a)) p_h(a, \nu^{-\perp}(L'_h)) \right| \\ &\leq \sum_{h=0}^{H-1} \sum_{s,a} \left| \bar{\Xi}_{\alpha_h(\nu^{-\perp}(L_h))}(L_h)(s, a) - \bar{\Xi}_{\alpha_h(\nu^{-\perp}(L'_h))}(L'_h)(s, a) \right| |p_h(a, \nu^{-\perp}(L_h))| \\ &\quad + \sum_{h=0}^{H-1} \sum_{s,a} \bar{\Xi}_{\alpha_h(\nu^{-\perp}(L'_h))}(L'_h)(s, a) |p_h(a, \nu^{-\perp}(L_h)) - p_h(a, \nu^{-\perp}(L'_h))|. \end{aligned}$$

Using the boundedness of the payment function p and the Lipschitz continuity of the allocation operator Ξ_α , the first term can be bounded by $B_p(2 + K_\alpha) \sum_{h=0}^{H-1} \|L_h - L'_h\|_1$. For the second term, the Lipschitz property of p implies a bound of $\sum_{h=0}^{H-1} K_p \|L_h - L'_h\|_1$. Combining these, we conclude that the revenue objective g_{rev} is Lipschitz continuous with constant $B_p(2 + K_\alpha) + K_p$, and satisfies the bound

$$|g_{\text{rev}}(\mathbf{L}) - g_{\text{rev}}(\mathbf{L}')| \leq (B_p(2 + K_\alpha) + K_p)\|\mathbf{L} - \mathbf{L}'\|_1$$

□

Corollary 5 (Convergence of Expected Revenue) *Let g_{rev} the expected revenue objective. Let \mathcal{M}_{mfa} be a BA-MFG with Lipschitz-continuous $\{u_h\}_{h=0}^{H-1}, \{w_h\}_{h=0}^{H-1}, \{\alpha_h\}_{h=0}^{H-1}, \{p_h\}_{h=0}^{H-1}$. Let $\boldsymbol{\pi} = (\pi, \dots, \pi)$ be the joint population policy for some $\pi \in \Pi_H$. Then:*

$$|g_{\text{rev}}(\Lambda_{\text{mfa}}(\boldsymbol{\pi})) - G_{\text{rev}}(\boldsymbol{\pi})| = \mathcal{O}\left(\frac{1}{\sqrt{N}}\right).$$

Proof: The result follows by combining Theorem 5 and lemma 13. □

E.5 Proof of Lemma 3

In this section, we prove that the (entropy regularized) q-functions are Lipschitz continuous with respect to the population policy, assuming full support. We begin by showing that the population flow is Lipschitz in the policy. Next, we establish that both the transition dynamics and the reward function are Lipschitz continuous with respect to the population distribution. Finally, we combine these results to derive a bound on the Lipschitz constant of the (entropy regularized) q-functions.

Lemma 14 (Lipschitz Continuity of Population Operator) Let \mathcal{M}_{mfa} be a Batched Auction Mean Field Game (BA-MFG). Let $\{\alpha_h\}_{h=0}^{H-1}$, $\{P_h^{mfa}\}_{h=0}^{H-1}$, $\{R_h^{mfa}\}_{h=0}^{H-1}$, $\{p_h\}_{h=0}^{H-1}$, and $\{u_h\}_{h=0}^{H-1}$ denote the allocation thresholds, transition dynamics, payment, and utility functions, respectively, from which $\{P_h^{mfa}\}_{h=0}^{H-1}$ and $\{R_h^{mfa}\}_{h=0}^{H-1}$ are derived. Assume these functions are Lipschitz continuous. Consider two arbitrary policies $\pi, \pi' \in \Pi_H$, then

$$\|\Lambda_{mfa}(\pi) - \Lambda_{mfa}(\pi')\|_1 \leq \sum_{h' \leq h} (K_\xi(1 + K_\alpha))^{h-h'} \|\pi_{h'} - \pi'_{h'}\|_1$$

Proof: We prove the bound inductively. Let $L^\pi := \Lambda_{mfa}(\pi)$ and $L^{\pi'} := \Lambda_{mfa}(\pi')$ denote the population flows induced by policies π and π' , respectively. Similarly, let μ^π and $\mu^{\pi'}$ denote the corresponding marginal state distributions.

For $h = 0$ we have $\mu_0^\pi = \mu_0^{\pi'} = \mu_0$. Therefore by Lemma 9 we have $\|L_0^\pi - L_0^{\pi'}\|_1 \leq \|\pi_0 - \pi'_0\|_1$.

For $h + 1 > 0$ by applying the result of Lemma 9 we have

$$\|L_{h+1}^\pi - L_{h+1}^{\pi'}\|_1 \leq \|\pi_{h+1} - \pi'_{h+1}\|_1 + \|\mu_{h+1}^\pi - \mu_{h+1}^{\pi'}\|_1.$$

We then bound the variational difference in state distribution:

$$\begin{aligned} \|\mu_{h+1}^\pi - \mu_{h+1}^{\pi'}\|_1 &= \|\Gamma_w(\xi_h^\pi) - \Gamma_w(\xi_h^{\pi'})\|_1 \leq K_\xi \|\xi_h^\pi - \xi_h^{\pi'}\|_1 \\ &= K_\xi \|\Gamma_{\alpha_h}(L_h^\pi) - \Gamma_{\alpha_h}(L_h^{\pi'})\|_1 \leq K_\xi(1 + K_\alpha) \|L_h^\pi - L_h^{\pi'}\|_1, \end{aligned}$$

where in the last step we used Corollary 3. By induction over h the claim follows. \square

Lemma 15 (Lipschitz Continuity of Transitions and Rewards under Full-Support Policies)

Let \mathcal{M}_{mfa} be a BA-MFG with utility functions $\{u_h\}_{h=0}^{H-1}$, transition dynamics $\{w_h\}_{h=0}^{H-1}$, payment functions $\{p_h\}_{h=0}^{H-1}$, and allocation thresholds $\{\alpha_h\}_{h=0}^{H-1}$, all of which are Lipschitz-continuous with constants K_u , K_w , K_p , and K_α , respectively. Consider two policies $\pi, \pi' \in \Pi_H$ with full support; that is, for all $s \in \mathcal{S}$, $a \in \mathcal{A}$, and $h \in 0, \dots, H-1$, we have $\pi_h(a|s) > \epsilon$ and $\pi'_h(a|s) > \epsilon$ for some constant $\epsilon > 0$. Then

$$\begin{aligned} &|R_h^{mfa}(s, a, L_h^\pi) - R_h^{mfa}(s, a, L_h^{\pi'})| \\ &\leq \left(\frac{B_u}{(1 - \alpha_{max})\epsilon} + K_u K_p \right) \sum_{h' \leq h} (K_\xi(1 + K_\alpha))^{h-h'} \|\pi_{h'} - \pi'_{h'}\|_1, \end{aligned}$$

and

$$\begin{aligned} &\|P_h^{mfa}(s, a, L_h^\pi) - P_h^{mfa}(s, a, L_h^{\pi'})\|_1 \\ &\leq \left(\frac{2}{(1 - \alpha_{max})\epsilon} + K_\alpha + 1 \right) \sum_{h' \leq h} (K_\xi(1 + K_\alpha))^{h-h'} \|\pi_{h'} - \pi'_{h'}\|_1. \end{aligned}$$

Proof: Let $\pi, \pi' \in \Pi_H$ be two arbitrary policies with full support. Let $L_h^\pi = \Lambda_{mfa}(\pi)$ and $L_h^{\pi'} = \Lambda_{mfa}(\pi')$. We then bound separately the rewards and the transition probabilities.

We first bound the absolute difference in rewards, let $s \in \mathcal{S}$, $a \in \mathcal{A}$ arbitrary, assume $s \neq \perp$, else the claim holds trivially, then, by applying triangular inequality we have:

$$\begin{aligned} &\left| R_h^{mfa}(s, a, L_h^{\pi'}) - R_h^{mfa}(s, a, L_h^\pi) \right| \\ &\leq B_u \left| p_{win}(s, a, L_h^\pi, \alpha_h(\nu^\perp(L_h^\pi))) - p_{win}(s, a, L_h^{\pi'}, \alpha_h(\nu^\perp(L_h^{\pi'}))) \right| \\ &\quad + \left| u_h(s, p_h(a, \nu^\perp(L_h^\pi))) - u_h(s, p_h(a, \nu^\perp(L_h^{\pi'}))) \right| \\ &\leq \frac{B_u}{(1 - \alpha_{max})\epsilon} \|L_h^\pi - L_h^{\pi'}\|_1 + K_u K_p \|L_h^\pi - L_h^{\pi'}\|_1, \end{aligned}$$

where the last step follows by Remark 3 and the Lipschitz continuity of p and u . By Lemma 14 the bound for the rewards follows.

Similarly for the transition probabilities we have:

$$\begin{aligned}
& \|P_h^{\text{mfa}}(s, a, L_h^\pi) - P_h^{\text{mfa}}(s, a, L_h^{\pi'})\|_1 \\
& \leq \sum_z \|w_h(z | \Gamma_{\alpha_h}(L_h^{\pi'}))\|_1 \left| P_{\alpha_h}(z | s, a, L_h^{\pi'}) - P_{\alpha_h}(z | s, a, L_h^\pi) \right| \\
& \quad + \sum_z P_{\alpha_h}(z | s, a, L_h^\pi) \|w_h(z, \Gamma_{\alpha_h}(L_h^\pi)) - w_h(z, \Gamma_{\alpha_h}(L_h^{\pi'}))\|_1 \\
& \leq \|P_{\alpha_h}(s, a, L_h^{\pi'}) - P_{\alpha_h}(s, a, L_h^\pi)\|_1 + K_w \|\Gamma_{\alpha_h}(L_h^\pi) - \Gamma_{\alpha_h}(L_h^{\pi'})\|_1.
\end{aligned}$$

From Remark 3 and Corollary 3 it follows

$$\|P_h^{\text{mfa}}(s, a, L_h^\pi) - P_h^{\text{mfa}}(s, a, L_h^{\pi'})\|_1 \leq \left(\frac{2}{(1 - \alpha_{\max})\epsilon} + K_\alpha + 1 \right) \|L_h^\pi - L_h^{\pi'}\|_1,$$

by Lemma 14 the bound follows. \square

Lemma 16 (Lipschitz Continuity of Regularized Value Functions) *Let \mathcal{M}_{mfa} be a BA-MFG with utility functions $\{u_h\}_{h=0}^{H-1}$, transition dynamics $\{w_h\}_{h=0}^{H-1}$, payment functions $\{p_h\}_{h=0}^{H-1}$, and allocation thresholds $\{\alpha_h\}_{h=0}^{H-1}$, all of which are Lipschitz-continuous with constants K_u , K_w , K_p , and K_α , respectively.*

Let $\pi, \pi' \in \Pi_H$ be two policies with full support, i.e., for some $\epsilon > 0$, $\pi_h(a|s), \pi'_h(a|s) > \epsilon$, for all h, s, a . Let $V_h^\tau(s|L^\pi, \pi)$ and $V_h^\tau(s|L^{\pi'}, \pi')$ denote the entropy-regularized value functions under policies π and π' , defined recursively as

$$V_h^\tau(s|L^\pi, \pi) := \tau H(\pi_h(s)) + \sum_{a \in \mathcal{A}} \pi_h(a|s) \left[R_h^{\text{mfa}}(s, a, L_h^\pi) + \sum_{s'} P_h^{\text{mfa}}(s'|s, a, L_h^\pi) V_{h+1}^\tau(s'|L^\pi, \pi) \right],$$

where $\{L_h^\pi\}_{h=0}^{H-1} := \Lambda_{\text{mfa}}(\pi)$ is the population distribution induced by policy π .

Then, there exists a constant $C > 0$ such that for all h and all $s \in \mathcal{S}$, the following bound holds:

$$\left| V_h^\tau(s|L^\pi, \pi) - V_h^\tau(s|L^{\pi'}, \pi') \right| \leq C \|\pi - \pi'\|_1,$$

where $C = \mathcal{O}(\frac{1}{\epsilon} H^2)$ when $K_\xi(1 + K_\alpha) = 1$, and $C = \mathcal{O}(\frac{1}{\epsilon} H \frac{1 - (K_\xi(1 + K_\alpha))^H}{1 - K_\xi(1 + K_\alpha)})$ when $K_\xi(1 + K_\alpha) \neq 1$.

Proof: We prove a stronger inductive bound that implies the result of the lemma. To simplify notation, we introduce the following constants:

- $\beta := K_\xi(1 + K_\alpha)$,
- $B := B_u + \tau \log(|\mathcal{A}|)$,
- $g_0 := \frac{B_u}{(1 - \alpha_{\max})\epsilon} + K_u K_p$,
- $g_1 := (B_u + \tau \log(|\mathcal{A}|)) \left(\frac{2}{(1 - \alpha_{\max})\epsilon} + K_\alpha + 1 \right)$.

We also define the following h -dependent quantities, which will be used in the inductive argument:

- $A_h := \tau(\log(\frac{1}{\epsilon}) + 1) + (H - h)B$,
- $G_h := g_0 + g_1(H - h - 1)$,
- $\Delta_h := \|\pi_h - \pi'_h\|_1$.

We proceed by backward induction on the round h , and show that the following bound holds:

$$\left| V_h^\tau(s | L^\pi, \pi) - V_h^\tau(s | L^{\pi'}, \pi') \right| \leq \sum_{h'=h}^{H-1} A_{h'} \Delta_{h'} + \sum_{h'=h}^{H-1} G_{h'} \sum_{h'' \leq h'} \beta^{h' - h''} \Delta_{h''}.$$

For $h = H$ by definition, the value function at round H is zero for all states, i.e.,

$$V_H^\tau(s|L^\pi, \pi) = V_H^\tau(s|L^{\pi'}, \pi') = 0$$

Assume that for value function at time $h + 1$, V_{h+1}^τ the upper bound holds. We now prove the same for V_h^τ . The regularized value function is:

$$V_h^\tau(s|L^\pi, \pi) = \tau \mathcal{H}(\pi_h(s)) + \sum_a \pi_h(a|s) q_h^\tau(s, a|L^\pi, \pi).$$

Then,

$$\begin{aligned} & |V_h^\tau(s|L^\pi, \pi) - V_h^\tau(s|L^{\pi'}, \pi')| \\ & \leq \underbrace{\tau |\mathcal{H}(\pi_h(s)) - \mathcal{H}(\pi_h'(s))|}_{:= (\square)} + \underbrace{\left| \sum_a \pi_h(a|s) q_h^\tau(s, a|L^\pi, \pi) - \sum_a \pi_h'(a|s) q_h^\tau(s, a|L^{\pi'}, \pi') \right|}_{:= (\triangle)}. \end{aligned}$$

For the entropy term (\square) we have

$$\tau |\mathcal{H}(\pi_h(s)) - \mathcal{H}(\pi_h'(s))| \leq \tau (\log(\frac{1}{\epsilon}) + 1) \|\pi_h(s) - \pi_h'(s)\|_1$$

By applying triangular inequality we can upper bound (\triangle) as follows:

$$(\triangle) \leq \sum_a |\pi_h(a|s) - \pi_h'(a|s)| q_h^\tau(s, a|L^\pi, \pi) + \sum_a \pi_h'(a|s) |q_h^\tau(s, a|L^\pi, \pi) - q_h^\tau(s, a|L^{\pi'}, \pi')|$$

The first term can be bounded as a function of H and the maximal absolute utility B_u :

$$\sum_a |\pi(a|s) - \pi'(a|s)| q_h^\tau(s, a|L^\pi, \pi) \leq (H - h) (B_u + \tau \log(|\mathcal{A}|)) \|\pi_h(s) - \pi_h'(s)\|_1,$$

where the term $\tau \log(|\mathcal{A}|)$ comes from the maximal additional reward from the entropy regularizer.

For the second term we bound the absolute difference in q functions:

$$\begin{aligned} & |q_h^\tau(s, a|L^\pi, \pi) - q_h^\tau(s, a|L^{\pi'}, \pi')| \\ & \leq |R_h^{\text{mfa}}(s, a, L_h^\pi) - R_h^{\text{mfa}}(s, a, L_h^{\pi'})| \\ & \quad + \sum_{s'} |P_h^{\text{mfa}}(s'|s, a, L_h^\pi) - P_h^{\text{mfa}}(s'|s, a, L_h^{\pi'})| |V_{h+1}^\tau(s, a|L^\pi, \pi)| \\ & \quad + \sum_s P_h^{\text{mfa}}(s'|s, a, L_h^{\pi'}) |V_{h+1}^\tau(s'|L^\pi, \pi) - V_{h+1}^\tau(s'|L^{\pi'}, \pi')| \\ & \leq |R_h^{\text{mfa}}(s, a, L_h^\pi) - R_h^{\text{mfa}}(s, a, L_h^{\pi'})| \\ & \quad + (H - h - 1) (B_u + \tau \log(|\mathcal{A}|)) \|P_h^{\text{mfa}}(s, a, L_h^\pi) - P_h^{\text{mfa}}(s, a, L_h^{\pi'})\|_1 \\ & \quad \max_s |V_{h+1}^\tau(s'|L^\pi, \pi) - V_{h+1}^\tau(s'|L^{\pi'}, \pi')|. \end{aligned}$$

By combining all intermediate bounds and applying Lemma 15 we have

$$\begin{aligned} & \left| V_h^\tau(s | L^\pi, \pi) - V_h^\tau(s | L^{\pi'}, \pi') \right| \leq A_h \Delta_h + G_h \sum_{h' \leq h} \beta^{h-h'} \Delta_{h'} \\ & \quad + \max_s |V_{h+1}^\tau(s'|L^\pi, \pi) - V_{h+1}^\tau(s'|L^{\pi'}, \pi')|, \end{aligned}$$

by induction the bound holds. As next we compute the global Lipschitz constant.

$$\left| V_h^\tau(s | L^\pi, \pi) - V_h^\tau(s | L^{\pi'}, \pi') \right| \leq \underbrace{\sum_{h'=h}^{H-1} A_{h'} \Delta_{h'}}_{(\square)} + \underbrace{\sum_{h'=h}^{H-1} G_{h'} \sum_{h'' \leq h'} \beta^{h'-h''} \Delta_{h''}}_{(\triangle)}.$$

For (\square) we have:

$$\sum_{h'=h}^{H-1} A_{h'} \Delta_{h'} \leq \sum_{h=0}^{H-1} A_h \Delta_h \leq A_0 \sum_{h=0}^{H-1} \Delta_h$$

For (\triangle) we have:

$$\begin{aligned} \sum_{h'=h}^{H-1} G_{h'} \sum_{h'' \leq h'} \beta^{h'-h''} \Delta_{h''} &\leq \sum_{h=0}^{H-1} G_h \sum_{h' \leq h} \beta^{h-h'} \Delta_{h'} \\ &= \sum_{h'=0}^{H-1} \Delta_{h'} \sum_{j=0}^{H-1-h'} G_{h'+j} \beta^j \\ &= \sum_{h'=0}^{H-1} \Delta_{h'} \left(G_{h'} \sum_{j=0}^{H-1-h'} \beta^j - g_1 \sum_{j=0}^{H-1-h'} j \beta^j \right) \\ &\leq \sum_{h'=0}^{H-1} \Delta_{h'} G_{h'} \sum_{j=0}^{H-1-h'} \beta^j \end{aligned}$$

Using the definition of the geometric sum, and observing that the constants $G_{h'}$ as well as the nested geometric sum decrease as h' increases, we can further simplify the bound by pulling out the leading terms. For $\beta \neq 1$:

$$\sum_{h'=0}^{H-1} \Delta_{h'} G_{h'} \sum_{j=0}^{H-1-h'} \beta^j \leq \frac{1 - \beta^H}{1 - \beta} (g_0 + g_1(H-1)) \sum_{h=0}^{H-1} \Delta_h,$$

while for $\beta = 1$:

$$\sum_{h'=0}^{H-1} \Delta_{h'} G_{h'} \sum_{j=0}^{H-1-h'} \beta^j \leq H(g_0 + g_1(H-1)) \sum_{h=0}^{H-1} \Delta_h.$$

Combining the results (for $\beta \neq 1$) of (\square) and (\triangle) we have:

$$\left| V_h^\tau(s | L^\pi, \pi) - V_h^\tau(s | L^{\pi'}, \pi') \right| \left(\tau(\log(\frac{1}{\epsilon}) + 1) + HB + \frac{1 - \beta^H}{1 - \beta} (g_0 + g_1(H-1)) \right) \sum_{h=0}^{H-1} \Delta_h.$$

For $\beta = 1$ the geometric term is replaced by H . By the definitions of B , g_0 and g_1 the claim follows. \square

F Experiment Details

Implementation Details. The experiments were implemented in JAX and PyTorch, the code is provided in the supplementary material. We implement the adjoint method in JAX. For the PyTorch implementation, some code was adapted from [31]. All error bars in experiments are one standard deviation away from the mean.

Hardware and Compute Time. We run our experiments on a single NVIDIA H100 GPU with an AMD EPYC 16-core CPU. One run of AMID for 1000 iterations takes 6 minutes, apart from the experiment (A8) described below with a long time horizon $H = 100$, which takes 20 minutes for 1000 iterations. The beach bar process experiments take roughly 3 minutes for 1000 iterations.

Parameterizing θ in \mathcal{M}_{bb} . For the beach bar process, we parameterize the per state payments as $\theta_s = p_{\max} \text{sigmoid}(\xi_s)$, where p_{\max} is the maximum per state payment and the unconstrained parameters $\xi \in \mathbb{R}^S$ are learned via AMID.

Parameterizing θ in \mathcal{M}_{mfa} . We parameterize p_h^θ and sold goods α_h^θ as residual neural networks sharing a base. The base network, f_{base}^θ , has $d_{\text{in}} = H + |\mathcal{A}| + 1$ inputs consisting of one-hot encoded time vector \mathbf{e}_h , $|\mathcal{A}|$ -dimensional vector of bid distribution $\nu^{-\perp}$, and remaining goods at round h denoted r_h , (given by $\alpha_{\text{max}} - \sum_{h'=0}^{h-1} \alpha_{h'}$). For the input vector $x_{\text{in}} \in \mathbb{R}^{d_{\text{in}}}$, the base residual network is defined as

$$\begin{aligned} x_{\text{in}} &= \begin{bmatrix} \mathbf{e}_h \\ \nu^{-\perp} \\ r_h \end{bmatrix} \in \mathbb{R}^{H+d+1}, \\ h^{(1)} &= \text{ReLU}(W^{(1)} x_{\text{in}} + b^{(1)}), \\ y_{\text{base}} := h^{(2)} &= \text{ReLU}(W^{(2)} h^{(1)} + b^{(2)} + V^{(2)} x_{\text{in}} + c^{(2)}) \in \mathbb{R}^{d_{\text{hidden}}}. \end{aligned}$$

The goods to be sold this round are then computed by:

$$\alpha_h = r \times \sigma(w_g^\top y_{\text{base}} + b_g), \in \mathbb{R}.$$

and the payments functions for bids is computed then by:

$$\begin{aligned} h^{(3)} &= \text{ReLU}(W^{(3)} y_{\text{base}} + b^{(3)}) + y_{\text{base}}, \quad h^{(4)} = \frac{1}{A-1} \sigma(W^{(4)} t + b^{(4)}) \in \mathbb{R}^{A-1}, \\ x_{\text{payment},1} &= 0, \quad x_{\text{payment},i} = \sum_{j=1}^{i-1} h_j^{(4)}, \quad i = 2, \dots, A. \end{aligned}$$

Note that this parameterization ensures that the payment rule p_h^θ is a monotonic increasing function of bids a . The parameters θ of the mechanism overall are $W^{(1)}, V^{(2)} \in \mathbb{R}^{d_{\text{hidden}} \times (H+d+1)}$, $W^{(2)}, W^{(3)} \in \mathbb{R}^{d_{\text{hidden}} \times d_{\text{hidden}}}$, $W^{(4)} \in \mathbb{R}^{(A-1) \times d_{\text{hidden}}}$ and $b^{(1)}, b^{(2)}, c^{(2)}, b^{(3)}, w_g \in \mathbb{R}^{d_{\text{hidden}}}$, $b^{(4)} \in \mathbb{R}^{A-1}$, $b_g \in \mathbb{R}$.

Baseline algorithms. For the zeroth-order baseline algorithms 0-SGD and 0-Adam, we use the standard 2-point (biased) gradient estimator

$$\widehat{\nabla}_\theta := \frac{G_{\text{approx}}^T(\theta + u_{\text{zero}}z) - G_{\text{approx}}^T(\theta - u_{\text{zero}}z)}{2u_{\text{zero}}} Dz,$$

where $\theta \in \mathbb{R}^D$, z is uniformly distributed on the sphere \mathbb{S}_{D-1} , and u_{zero} is a tunable hyperparameter. This estimator satisfies the well-known property

$$\mathbb{E}[\widehat{\nabla}_\theta] = \nabla \mathbb{E}[G_{\text{approx}}^T(\theta + u_{\text{zero}}\bar{z})], \quad \bar{z} \propto \text{Uniform}(\mathbb{B}_D).$$

That is, $\widehat{\nabla}_\theta$ is an unbiased estimator of the gradient of a smoothed version of the function G_{approx}^T . The bias is tunable by the parameter u_{zero} , with smaller values corresponding to less bias but potentially higher variance in estimates. Since a single evaluation of this gradient estimator takes 2 forward passes over G_{approx}^T , its run time is comparable to that of AMID per iteration. For the baseline ANNEAL, each iteration, we sample a perturbation n from the D -dimensional standard normal distribution. After evaluating $G_{\text{approx}}^T(\theta), G_{\text{approx}}^T(\theta + \sigma_{\text{anneal}}n), G_{\text{approx}}^T(\theta - \sigma_{\text{anneal}}n)$ for the tunable hyperparameter $\sigma_{\text{anneal}} > 0$, ANNEAL updates θ to be the best among $\theta, \theta - \sigma_{\text{anneal}}n, \theta + \sigma_{\text{anneal}}n$.

Hyperparameters. All hyperparameters for the baselines as well as AMID are presented in Table 4. For a fair comparison, we perform a grid search on a range of values for the parameters for all baselines and take the best run after 10 repetitions. In our experiments, AMID is robust to hyperparameter choices while zeroth order methods require some tuning.

F.1 Additional Results on the Beach Bar Process

As mentioned in the main body of the paper, we first present the payment function θ_s learnt AMID after 1000 iterations. As before, we report these by using a slightly higher OMD iteration step $T_{\text{val}} = 500$ than used for training, to demonstrate the robustness of our method.

Parameter	Explanation	Values
η	Adam/SGD learning rate	$3e-5, 1e-4, 3e-4, 1e-3, 1e-3, 1e-2$
u_{zero}	Noise magnitude for evaluating zeroth-order gradient estimator $\widehat{\nabla}_{\theta}$, for 0-SGD and 0-ADAM	$1e-3, 1e-2, 3e-2$
σ_{anneal}	Perturbation magnitude for ANNEAL	$1e-6, 1e-5, 1e-4, 1e-3, 1e-2, 3e-2$
τ	Entropy regularization	$1e-3$
η_{OMD}	OMD learning rate	10
T	OMD iterations in (T -approx.)	400
T_{val}	OMD iterations for validation	500
d_{hidden}	Hidden dimension of residual network parameterizing payments and sold goods	256

Table 4: Hyperparameters for the experiments on auctions.

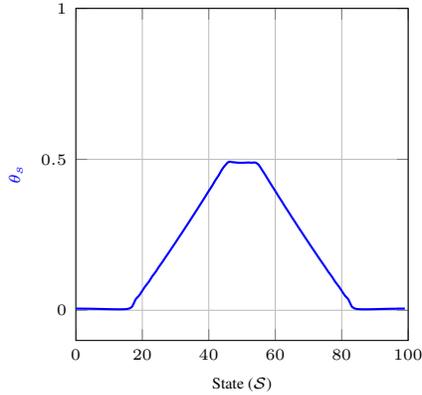


Figure 4: Payment function $s \rightarrow \theta_s$ learned after training with AMID, where payments are bounded on $[0, 1/2]$.

A bottleneck in the beach bar experiment is the magnitude of payments θ_s , which is restricted to be bounded on $[0, 1/2]$. We also report the experiment when $\theta \in [0, 4/5]^S$ below (i.e., when $p_{\max} = 0.8$), in Figure 5. As expected, the population distributions L_h are smoother and closer to uniform in this case. In both cases, AMID behaves as expected: the exploitability of T iterates of OMD is consistently low throughout training, suggesting that the T step approximation objective remains close to a NE throughout training iterations (with exploitability < 0.02). The fact that the exploitability of the policy induced by the T repeated iterations of OMD is due to the fact that in the beach bar experiments, tuning payments yields a NE distribution closer to uniform, which is also the initialization of OMD iterates.

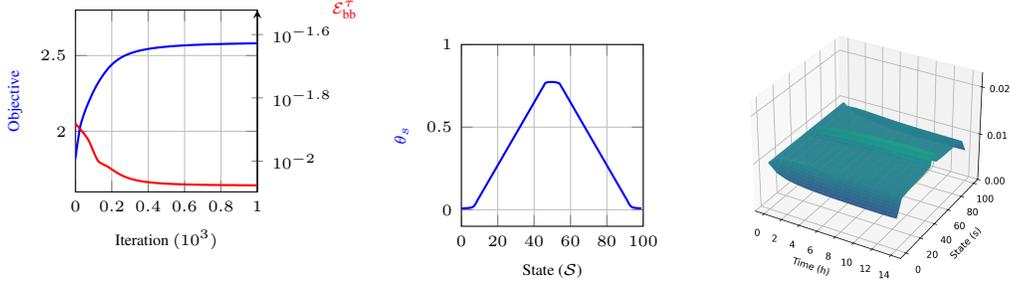


Figure 5: Payment design with AMID in \mathcal{M}_{bb} with larger payments. **Left:** objective and exploitability throughout training iterations. **Middle:** learned payment rule after training with AMID. **Right:** population flow in time after learning payments.

F.2 Additional Results on Auctions

We first analyze 3 further auctions with nonlinear utility functions for bidders. Namely, we take (A1) presented in the main body of the paper where $H = 4$, $\mu_0 = \text{Uniform}(\mathcal{S})$, $\alpha_{\max} = 0.8$, $|\mathcal{S}| = |\mathcal{A}| = 100$ and bidders are single-minded with no evolution in valuations s_h^i other than to transitions to \perp .

(A4) Risk-averse utilities formulated by

$$u_h(s, p) = \frac{1 - \exp\{-\beta(s - p)\}}{1 - \exp^{-\beta}},$$

where we take $\beta = 1$.

(A5) Risk-seeking utilities formulated by

$$u_h(s, p) = \frac{\exp\{\beta(s - p)\} - 1}{\exp^{\beta} - 1},$$

where we take $\beta = 1$.

(A6) Hyperbolic time discounting, which discounts future rewards, where the utility at time h is given by:

$$u_h(s, p) = \frac{s - p}{1 + \lambda h},$$

where we take $\lambda = 1$ as the time discount factor.

Across utility functions, AMID manages to beat all baselines. In our experiments, we also observed significant qualitative changes in both the mean-field Nash equilibrium and the payment rule when nonlinear utility functions are used. We show NE and payment rules suggested by the neural mechanism as NE in Figure 7.

Experiments regarding the impact of horizon. We report the impact of agent regeneration and time horizon by evaluating AMID on the following auction setups:

(A7) $H = 6$, agents regenerate with probability 1 (that is, they never transition to \perp even when they win a round), linear utilities for bidders, $\alpha_{\max} = 0.8$, $\mu_0(s) \propto \gamma^s$ for $\gamma = 0.9$, dynamic values with $w(s'|s) \propto \exp\{-(1.2s - s')^2/2\sigma^2\}$ for $\sigma = 0.2$.

(A8) Long horizon $H = 100$, $\alpha_{\max} = 5$, linear utilities for bidders, agents regenerate with probability 0.015, $\mu_0(s) \propto \gamma^s$ for $\gamma = 0.9$, dynamic values with $w(s'|s) \propto \exp\{-(1.2s - s')^2/2\sigma^2\}$ for $\sigma = 0.01$.

The results are reported on Figure 8.

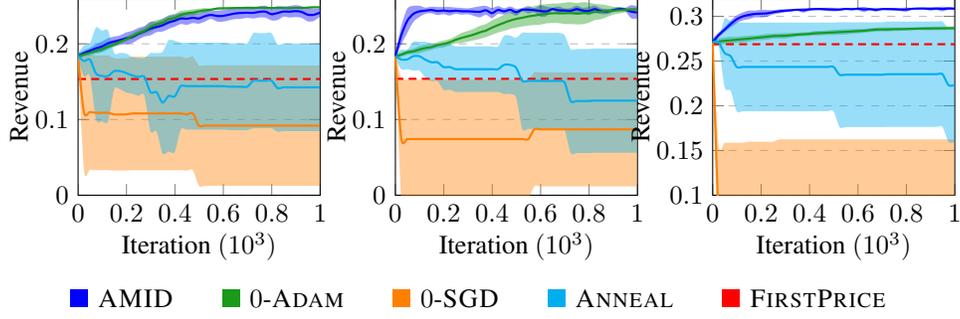


Figure 6: g_{rev} throughout iterations of AMID and baseline algorithms in settings with nonlinear utilities for bidders (A4-6), left to right.

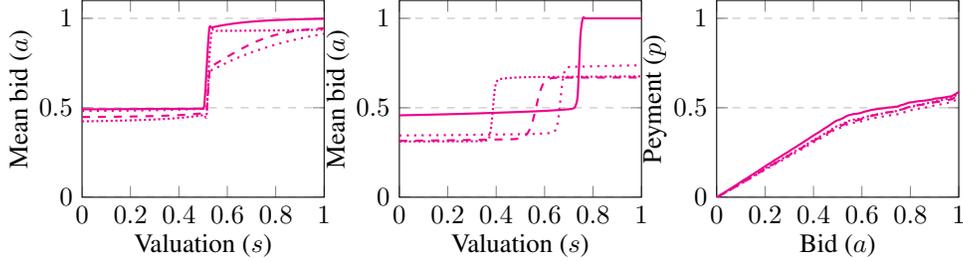


Figure 7: **Left:** mean bids at NE at θ^* after training with AMID on the risk-seeking utility experiment (A5) for $h \in \{0, 1, 2, 3\}$. **Middle:** mean bids at NE at θ^* after training with AMID on the hyperbolic time discounting utility experiment (A6) for $h \in \{0, 1, 2, 3\}$. **Right:** payment function in setting (A5) at the bids induced by the NE policy at θ^* for $h \in \{0, 1, 2, 3\}$.

General objectives. Finally, we explore the impact of optimizing over more general objectives other than revenue. We define the objective

$$g_{\text{mix}}(\theta, \mathbf{L}) := g_{\text{rev}}(\theta, \mathbf{L}) + g_{\text{efficiency}}(\theta, \mathbf{L}),$$

where we define $g_{\text{efficiency}}$ as:

$$g_{\text{efficiency}}(\theta, \mathbf{L}) := \sum_{h=0}^{H-1} \sum_{(s,a) \in \mathcal{V} \times \mathcal{A}} L_h(s, a) p_{\text{win}}(s, a, L_h, \alpha_h^\theta(\nu^{-\perp}(L_h))) u_h(s, p_h^\theta(a, \nu^{-\perp}(L_h))).$$

We modify experiment (A1) and evaluate our AMID on the following setting.

(A9) $H = 4$, $\mu_0 = \text{Uniform}(\mathcal{S})$, $\alpha_{\text{max}} = 0.8$, and single-minded bidders (after winning stay at state \perp) with no evolution in valuations s_h^i otherwise. The objective function is g_{mix} .

The results are reported on Figure 8.

Experiments with static α_h and payment rules. Finally, to verify the impact of the mechanism having access to bids $\nu_h^{-\perp}$ at round h , we run a final experiment where the mechanism is independent of bids, which we call a *static mechanism*. In this case, we simply parameterize $p_h^\theta(a) = \text{sigmoid}(\theta_{h,a}^{(1)}) \cdot a$ and $\alpha_h^\theta(a) = \alpha_{\text{max}} \frac{\exp\{\theta_h^{(2)}\}}{\sum_{h'} \exp\{\theta_{h'}^{(2)}\}}$, to ensure no more than α_{max} hoods are sold and the payments never exceed the bid. The parameter space is then $\theta := [\theta^{(1)}, \theta^{(2)}]$, for $\theta^{(1)} \in \mathbb{R}^{[H] \times \mathcal{A}}$, $\theta^{(2)} \in \mathbb{R}^{[H]}$. For static mechanisms, we observe much less significant improvement over the first price mechanism in general, which most likely originates from better allocation of goods over time when there are dynamics such as regeneration.

(A10) $H = 4$, static mechanism parameterization (independent of $\nu_h^{-\perp}$), each agent regenerates with probability 0.3 at the end of every round, and has a linear utility function.

The results in this setting are reported in Figure 9.

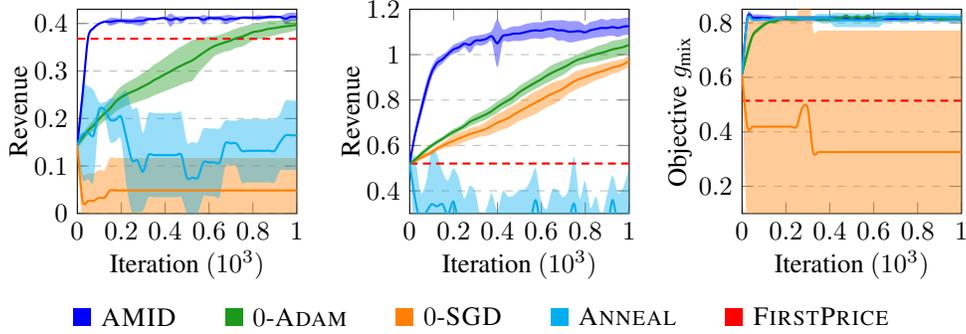


Figure 8: **Left-middle:** g_{rev} throughout iterations of AMID and baseline algorithms in settings (A7-8). **Right:** g_{mix} throughout iterations of AMID and baseline algorithms, in setting (A9).

Experiments on approximating exploitability. We also report our attempts to measure the true N player exploitability gap suggested by Theorem 2 when N is finite. For the first price mechanism, we compute the MFG-NE on the auction setting (A1) introduced in the main paper. Then, fixing $N = 1000$, we simulate trajectories in the batched auction by setting the policies of 999 agents to the MFG-NE, and train PPO on the last bidder. In this setting, we were not able to achieve a better mean reward for the last bidder than the MFG-NE. This suggests the exploitability is close to 0, we report the expected reward achieved by PPO throughout training in Figure 9.

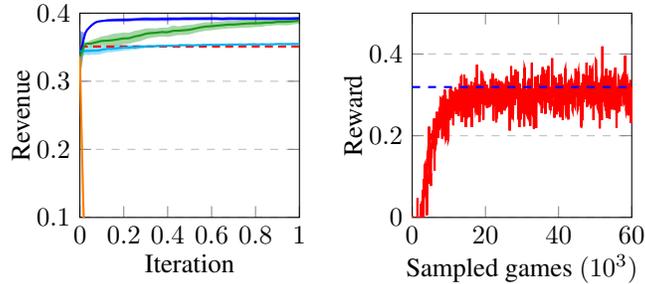


Figure 9: **Left:** Revenue throughout training in the static mechanism setting (A10). **Right:** PPO episodic rewards trained on the batched auction (A1) with $N = 1000$ agents, all but one playing NE. Blue is the MFG best response expected reward, red is PPO's expected reward throughout training.