Monotone Inclusion Approach to Weakly Monotone Discrete-Time Finite-Horizon Mean-Field Games

Uğur Aydın
Department of Electrical and Computer Engineering
University of Illinois Urbana-Champaign
Urbana, IL 61801, USA
uaydin2@illinois.edu

Tamer Başar
Department of Electrical and Computer Engineering
University of Illinois Urbana-Champaign
Urbana, IL 61801, USA
basar1@illinois.edu

Naci Saldi
Department of Mathematics
Bilkent University
Çankaya, Ankara 06800, Turkey
naci.saldi@bilkent.edu.tr

Abstract

We revisit the problem of computing mean-field equilibria (MFEs) in discrete-time, monotone, finite-horizon mean-field games (MFGs). We show that, when the transition kernel is independent of the state-measure term and the reward function satisfies the usual weak monotonicity condition and is Lipschitz continuous, anchored proximal gradient descent methods can be used to compute a monotone MFE. We also establish last-iterate convergence results for these methods. Our approach relies on formulating the computation problem as an optimization problem over the space of occupation measures. Using this formulation, we show that the problem is equivalent to a class of constrained Lipschitz monotone inclusion problems. We then apply iterative methods for this monotone inclusion formulation to derive a tractable algorithm. The resulting algorithm achieves a convergence rate of $O(1/\sqrt{T})$ after T iterations, without requiring any regularization. This rate holds even in the absence of a uniqueness assumption for the corresponding MFE.

1 Introduction

Mean-field games (MFGs) provide a powerful framework for approximating solutions to multi-agent problems with identical agents. The theory of MFGs was introduced independently by Lasry and Lions in Lasry and Lions [2007], who coined the term "mean-field games," and by Huang, Malhamé, and Caines in Huang et al. [2006], who formulated the problem from the viewpoint of stochastic dynamic games. These foundational works study continuous-time noncooperative differential games with a large but finite number of agents, where the influence of each individual agent on the population becomes asymptotically negligible as the number of agents grows. In this paper, we consider the discrete-time counterpart of this framework, introduced in Saldi et al. [2018].

MFGs reduce multi-agent problems with identical agents to an effective single-agent problem and provide approximate solutions to general Markov games, for which the existence of a solution is not always guaranteed Saldi et al. [2018]. Motivated by the growing interest in multi-agent systems, computational methods for MFGs have also attracted significant attention Guo et al. [2019]. However, the dynamics of MFGs generally do not satisfy any natural contraction property. In fact, computing equilibria in most MFGs is PPAD-complete Yardim et al. [2024], suggesting that the problem is computationally intractable in general.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

In this work, we revisit the learning problem for monotone finite-horizon mean-field games. We assume only that the transition probability is independent of state measures and the reward function is Lipschitz continuous on the space of state measures and satisfies the so-called weak monotonicity condition. Under these assumptions, we reformulate the problem as a minimization problem over the space of occupation measures on $S \times A$, where S denotes the state space and A the action space. By expressing the reward function as an operator of the occupation measure, we show that the problem of finding a mean-field equilibrium (MFE) can be reformulated as a monotone inclusion problem of the form $\mathbf{0} \in F(\mathbf{z}) + A(\mathbf{z})$, where F is a single-valued, monotone, and Lipschitz operator, and A is a maximal monotone operator defined over the space of feasible occupation measures.

The inclusion problem $\mathbf{0} \in F(\boldsymbol{z}) + A(\boldsymbol{z})$ in general cannot be solved via vanilla descent methods Daskalakis and Panageas [2018]. To design iterative an algorithm for this inclusion problem, in the optimization literature, it is well known that one needs further structure, such as optimisim, as in extragradient methods, or accelaration methods, as in anchoring Golowich et al. [2020] Ryu et al. [2019]. For simplicity, we will merely work under anchoring in this work. This will allow us to present a completely tractable algorithm by using the recent result [Cai and Zheng, 2026, Theorem 2] with convergence guarantees without relying on contraction or regularizers.

Two primary structural conditions are commonly used for computational purposes: contractivity and monotonicity. Unfortunately, monotonicity-based approaches often require additional regularization structure and typically apply only in fairly restrictive settings. In the MFG literature, monotonicity conditions are most often used to establish uniqueness of equilibrium Saldi et al. [2020]Perolat et al. [2021]. Beyond uniqueness, such conditions have also found applications in the mean-field reinforcement learning (MFRL) literature Zhang et al. [2023]Isobe et al. [2025]. In this paper, we assume only that the reward function is weakly monotone and Lipschitz continuous on the space of state measures.

Previous work on last-iterate convergence for such methods often relies on proximal techniques, for example, by tracking the proximal operator as the base point changes Isobe et al. [2025]. In contrast, we show that the anchoring mechanism allows us to avoid introducing additional regularization terms, which can otherwise lead to deviations from the true MFE. We also establish a convergence rate for the objective function. To the best of our knowledge, this is the first algorithm in the literature for the non-regularized, non-contractive setting that admits an explicit convergence rate. We further emphasize that our framework allows for the presence of multiple MFEs, and our convergence guarantee remains valid even in this non-unique setting.

1.1 Related Work

The main assumption imposed in this work is the so-called monotonicity condition. Monotonicity conditions are widely used in the study of non-stationary MFEs, as they are among the few structural assumptions known to guarantee uniqueness of MFE (Saldi et al. [2020], Perolat et al. [2021]). Nevertheless, the "weak" monotonicity condition by itself is generally insufficient to ensure uniqueness. In the computational literature, monotonicity has mainly been exploited in algorithms based on policy updates (Isobe et al. [2025], Zhang et al. [2023]). By contrast, we formulate the problem as an optimization problem over the space of measures and develop an algorithm directly on the space of occupation measures.

Representing the underlying MFE problem as an optimization problem over the corresponding space of occupation measures is a standard technique in the study of Markov decision processes (Puterman [2014]). Although the optimization formulation was not explicitly used there, occupation measures were first employed in the context of discrete-time MFGs to establish the existence of MFEs in Saldi et al. [2018]. Later, Guo et al. [2022] used the associated optimization problem to propose a framework for solving finite-horizon MFEs. However, the algorithms developed in that work are intractable and rely on unknown initialization constraints as well as higher-order differentiability assumptions on the system components. For "linear" MFGs, Saldi [2026] also developed optimization-based algorithms for computing MFEs using occupation measures. None of these works provides an algorithm with an explicit convergence rate.

More recently, the works Anahtarci et al. [2025b], Alkır and Saldi [2025], Guo et al. [2022], Anahtarci et al. [2025a] developed reinforcement learning algorithms based on the occupation measure approach. The key distinction between our method and these previous results is that those works formulate the

optimization problem through the space of policies. As a consequence, the required subproblems are generally intractable. In contrast, we formulate the problem directly as an optimization problem over occupation measures and develop an iterative algorithm that operates entirely in the space of occupation measures, without invoking policies even at the level of subroutines. This subtle but important distinction allows us to reformulate the problem as a monotone inclusion problem over occupation measures.

By representing MFE as an optimization problem, we identify an operator F on the space of occupation measures. Since the problem must be restricted to feasible occupation measures, we are naturally led to the constrained monotone inclusion $\mathbf{0} \in F(\mathbf{z}) + A(\mathbf{z})$, where A is a set-valued operator that captures the constraints induced by the occupation measure formulation. The unconstrained case, corresponding to $A \equiv 0$, has recently attracted considerable attention because of its close connection to saddle-point problems arising in the training of generative adversarial networks (GANs) (Golowich et al. [2020];Ryu et al. [2019]). The constrained case $A \not\equiv 0$, however, presents additional technical challenges (Cai and Zheng [2023]). Even in the unconstrained setting, standard proximal gradient methods may fail to converge (Daskalakis and Panageas [2018]). To address this issue, methods based on *optimism* and *acceleration* have been incorporated into proximal gradient schemes (Yoon and Ryu [2021]).

Finally, we briefly note that existing algorithms for weakly monotone MFGs are typically based on policy updates, which are often implemented through KL-divergence regularization or related techniques (Zhang et al. [2023];Isobe et al. [2025];Perolat et al. [2021]). By contrast, the algorithm introduced in this work operates solely on the space of occupation measures and therefore avoids the use of auxiliary policy-update procedures such as proximal methods through KL divergence regularization.

1.2 Contributions

Our main contributions are summarized as follows:

1. We provide a maximal monotone set-inclusion characterization of MFEs under weak monotonicity of the reward function.

2. When the transition probability kernel is independent of the population distribution, we show that the monotone set-inclusion problem corresponding to the MFE problem can be solved in polynomial time using the anchored proximal algorithm introduced by Cai and Zheng [2026]. It appears that this is the first fully tractable and provably convergent algorithm for computing exact finite-horizon mean-field equilibria in discrete-time mean-field games.

3. Tractability of our algorithm relies on solving a convex quadratic program exactly at each iteration, which can be done in polynomial time. Using the interior-point primal–dual path-following algorithm of Monteiro and Adler [1989], we provide an exact arithmetic complexity bound for solving this quadratic program.

4. To the best of our knowledge, the algorithm introduced here is the first in the discrete-time MFG literature to achieve an explicit convergence rate in the non-regularized setting while allowing for multiple mean-field equilibria. In addition, under our assumptions, we construct an example of an MFG showing that the upper bound of our algorithm is tight, in the sense that it is matched by a corresponding lower bound for this algorithm.

A comparison of our contributions we have listed above with (some) of the previous literature can be found in the table below.

Work	Assumption	Type	Rate of Convergence	Uniqueness	Tractable
Guo et al. [2019]	Contraction	Regularized	Exponential	Yes	Yes
Anahtarcı et al. [2020]	Contraction	Nonregularized	Exponential	Yes	Yes
Anahtarci et al. [2023]	Contraction	Regularized	Exponential	Yes	Yes
Cui and Koeppl [2021]	Contraction	Regularized	Exponential	Yes	Yes
Yardım et al. [2023]	Contraction	Regularized	Exponential	Yes	Yes
Zhang et al. [2023]	Strict Mono.	Regularized	$O(\log^2 T/\sqrt{T})$	Yes	Yes
Zhang et al. [2024]	Contraction	Regularized	Exponential	Yes	Yes
Guo et al. [2024]	Differentiable	Nonregularized	N/A	No	No
Dong et al. [2025]	Weak Mono.	Regularized	$O(1/T)$	No	Yes
Isobe et al. [2025]	Weak Mono.	Nonregularized	N/A	No	No
This work	**Weak Mono.**	**Nonregularized**	$O(1/\sqrt{T})$	**No**	**Yes**

Remark 1. *The work Dong et al. [2025] establishes an $O(1/T)$ convergence rate when one has oracle access to the reward function and the transition kernel. In addition, Isobe et al. [2025] shows that exponential convergence is achievable in the regularized setting.*

2 Preliminaries

2.1 Discrete-Time Mean-Field Games

In this subsection, following the framework of Saldi et al. [2018], we introduce the setup for discrete-time mean-field games. A discrete-time finite-horizon mean-field game is specified by a tuple (S, A, r, P, μ_0, H), where

1. S is a finite state space.
2. A is a finite action space.
3. $r : S \times A \times \mathcal{P}(S) \to \mathbb{R}$ is the one-stage reward function, where $\mathcal{P}(S)$ denotes the space of probability measures on S.
4. $P : S \times A \times \mathcal{P}(S) \to \mathcal{P}(S)$ is the one-stage transition kernel.
5. $\mu_0 \in \mathcal{P}(S)$ is the initial state distribution.
6. H is the horizon length.

Each element $\mu \in \mathcal{P}(S)$ will be referred to as a *state measure*, representing the distribution of states across a continuum of identical agents. A *state-measure flow* is a tuple of state measures $\boldsymbol{\mu} = (\mu_t)_{t=0}^{H-1}$. Throughout the paper, whenever an MFG is fixed, we implicitly assume that every state-measure flow starts from the initial state measure μ_0 specified in the definition of the game.

For a given finite horizon $H \in \mathbb{N}$, the objective in an MFG is to find a policy flow $\boldsymbol{\pi}^* = (\pi_h^*)_{h=0}^{H-1}$ and a state-measure flow $\boldsymbol{\mu}^* = (\mu_h^*)_{h=0}^{H-1}$ such that the objective

$$J(\boldsymbol{\pi}^*, \boldsymbol{\mu}^*) := \mathbb{E}\left[\sum_{h=0}^{H-1} r(x_h, a_h, \mu_h^*)\right], \tag{1}$$

where $x_{h+1} \sim P(\cdot \mid x_h, a_h, \mu_h^*)$, $a_h \sim \pi_h^*(\cdot \mid x_h)$, and $x_0 \sim \mu_0$, satisfies

$$J(\boldsymbol{\pi}, \boldsymbol{\mu}^*) \le J(\boldsymbol{\pi}^*, \boldsymbol{\mu}^*) \tag{2}$$

for every policy flow $\boldsymbol{\pi}$, and the state-measure flow $\boldsymbol{\mu}^*$ satisfies the *consistency condition*

$$\mu_{h+1}^*(\cdot) = \sum_{s\in S}\sum_{a\in A} P(\cdot \mid s, a, \mu_h^*)\, \pi_h^*(a \mid s)\, \mu_h^*(s), \qquad h = 0, \dots, H-2. \tag{3}$$

When r and P are independent of the state-measure component in $\mathcal{P}(S)$, the problem reduces to a standard Markov decision process. If the pair $(\boldsymbol{\pi}^*, \boldsymbol{\mu}^*)$ satisfies (2) and (3), then we say that $(\boldsymbol{\pi}^*, \boldsymbol{\mu}^*)$ is a *mean-field equilibrium*.

The main difficulty in computing an MFE arises from the dependence of r and P on the state-measure component $\mathcal{P}(S)$, together with the consistency requirement in (3). Because of this additional dependence and the associated evolution condition, there is generally no natural contractive dynamical system available for computational purposes. We also emphasize that a discrete-time MFG is, in general, neither a standard game nor a standard control problem, but rather lies somewhere in between.

For our purposes, we impose the following assumption on the MFG.

Assumption 1 (Weak monotonicity)**.** *For every pair of policy flows $\boldsymbol{\pi} = (\pi_h)_{h=0}^{H-1}$ and $\tilde{\boldsymbol{\pi}} = (\tilde{\pi}_h)_{h=0}^{H-1}$, it holds that*

$$\sum_{h=0}^{H-1} \sum_{s\in S} \sum_{a\in A} \Big(r(s,a,\mu_h^{\boldsymbol{\pi}}) - r(s,a,\mu_h^{\tilde{\boldsymbol{\pi}}}) \Big) \Big(\rho_h^{\boldsymbol{\pi}}(s,a) - \rho_h^{\tilde{\boldsymbol{\pi}}}(s,a) \Big) \leq 0, \tag{4}$$

where $\boldsymbol{\mu}^{\boldsymbol{\pi}}$ and $\boldsymbol{\mu}^{\tilde{\boldsymbol{\pi}}}$ are state-measure flows such that

$$\rho_h^{\boldsymbol{\pi}}(s,a) = \pi_h(a \mid s)\mu_h^{\boldsymbol{\pi}}(s), \qquad \rho_h^{\tilde{\boldsymbol{\pi}}}(s,a) = \tilde{\pi}_h(a \mid s)\mu_h^{\tilde{\boldsymbol{\pi}}}(s).$$

We say that the monotonicity condition is *strict* if (4) holds with equality only when $\boldsymbol{\pi} = \tilde{\boldsymbol{\pi}}$. In general, strict monotonicity [Perolat et al., 2021, Proposition 1], or alternatively additional separability conditions on the reward function Saldi et al. [2020], is needed to ensure uniqueness of the MFE. Therefore, under Assumption 1, the finite-horizon MFE need not be unique in general.

We further note that Assumption 1 guarantees the existence of an MFE; see [Perolat et al., 2021, Proposition 1]. Without monotonicity, existence typically requires additional regularity conditions on the system components r and P, most notably uniform continuity; see Saldi et al. [2018].

Assumption 2. *The transition kernel P is independent of the state-measure component; namely,*

$$P(\cdot \mid s,a,\mu) = P(\cdot \mid s,a) \qquad \textit{for all } \mu \in \mathcal{P}(S).$$

Assumption 2, together with Assumption 1, is frequently imposed to guarantee uniqueness of the MFE Saldi et al. [2020]. It is also adopted in Isobe et al. [2025] to develop an algorithm in the weakly monotone setting. On the other hand, under strict monotonicity, the framework of Zhang et al. [2023] does not require Assumption 2. In our setting, the main role of Assumption 2 is that it renders the constraint set in the optimization formulation a convex polytope, a property that is essential for the development of a fully tractable algorithm. This assumption also commonly occurs in practice; see Weintraub et al. [2008], Subramanian and Mahajan [2019], while excluding applications to congestion Lauriere [2021] and epidemic Olmez et al. [2022] MFGs.

Assumption 3 (Lipschitz continuity)**.** *For all $(x,a) \in S \times A$ and all $\mu, \tilde{\mu} \in \mathcal{P}(S)$, we have*

$$|r(x,a,\mu) - r(x,a,\tilde{\mu})| \leq L_r \|\mu - \tilde{\mu}\|_{\mathrm{TV}},$$

where $\|\cdot\|_{\mathrm{TV}}$ denotes the total variation norm on $\mathcal{P}(S)$.

The Lipschitz continuity assumption imposed above is weaker than the assumptions typically used in contractive approaches for the nonlinear case, where r and P are also required to be Lipschitz continuous with respect to the state and action variables; see Anahtarci et al. [2023], Yardım et al. [2023], Guo et al. [2019], Cui and Koeppl [2021]. In contrast, the work Zhang et al. [2023] does not require any Lipschitz continuity assumption under strict monotonicity. As noted earlier, since we establish convergence rates for first-order methods, it is natural to impose additional regularity conditions on the system components, such as Lipschitz continuity.

Remark 2. *For simplicity, we present our main results in the time-homogeneous setting; that is, for all $h = 0, 1, \ldots, H-1$, $P_h(\cdot \mid \cdot,\cdot) \equiv P(\cdot \mid \cdot,\cdot)$ and $r_h(\cdot,\cdot,\cdot) \equiv r(\cdot,\cdot,\cdot)$. However, our results continue to hold in the time-inhomogeneous setting as well.*

2.2 Monotone Operators and Inclusions

Our computational results rely on representing an MFE as an optimization problem involving monotone operators. The purpose of this subsection is to provide the basic background on monotone operators and monotone inclusions needed to make the presentation as self-contained as possible.

For our purposes, all monotone operators are finite-dimensional. Throughout the paper, unless stated otherwise, every finite-dimensional Euclidean space $\mathbb{R}^d$ is equipped with the Euclidean norm $\|\cdot\|_2$.

Definition 1 (Monotone Operator)**.** *Let $\mathcal{H}$ be a finite-dimensional Hilbert space with inner product $\langle \cdot, \cdot \rangle$, and let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a set-valued operator. The graph of A is* $\mathrm{gra}(A) := \{(\boldsymbol{x}, \boldsymbol{u}) \in \mathcal{H} \times \mathcal{H} : \boldsymbol{u} \in A(\boldsymbol{x})\}$*. The operator A is called* monotone *if*

$$\langle \boldsymbol{u} - \boldsymbol{v}, \boldsymbol{x} - \boldsymbol{y} \rangle \geq 0 \qquad \forall (\boldsymbol{x}, \boldsymbol{u}), (\boldsymbol{y}, \boldsymbol{v}) \in \mathrm{gra}(A).$$

Definition 2 (Maximal Monotone Operator)**.** *The set-valued operator $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is called* maximal monotone *if it is monotone and there is no monotone set-valued operator $B : \mathcal{H} \rightrightarrows \mathcal{H}$ such that* $\mathrm{gra}(A) \subsetneq \mathrm{gra}(B)$.

The following remark says that the normal cone of a compact convex set about a given point is a maximal monotone operator, which will be the case of interest for us.

Remark 3. *[Normal cone is maximal monotone] Let $\mathcal{K} \subset \mathcal{H}$ be a nonempty compact convex set, and define its normal cone operator by*

$$N_{\mathcal{K}}(\boldsymbol{x}) := \begin{cases} \{\, \boldsymbol{g} \in \mathcal{H} : \langle \boldsymbol{g}, \boldsymbol{y} - \boldsymbol{x} \rangle \leq 0 \ \forall \boldsymbol{y} \in \mathcal{K} \,\}, & \boldsymbol{x} \in \mathcal{K}, \\ \varnothing, & \boldsymbol{x} \notin \mathcal{K}. \end{cases}$$

Then, $N_{\mathcal{K}}$ is a maximal monotone operator.

Indeed, since $\mathcal{K}$ is nonempty, compact, and convex, it is in particular closed and convex. Hence the indicator function

$$I_{\mathcal{K}}(\boldsymbol{x}) := \begin{cases} 0, & \boldsymbol{x} \in \mathcal{K}, \\ +\infty, & \boldsymbol{x} \notin \mathcal{K} \end{cases}$$

is proper, lower semicontinuous, and convex Rockafellar [1970]. Moreover,

$$\partial I_{\mathcal{K}}(\boldsymbol{x}) = N_{\mathcal{K}}(\boldsymbol{x}),$$

where $\partial I_{\mathcal{K}}$ denotes the convex subdifferential [Bauschke and Combettes, 2020, Example 16.12]. Since the subdifferential of any proper lower semicontinuous convex function is maximal monotone, it follows that $N_{\mathcal{K}}$ is maximal monotone [Bauschke and Combettes, 2020, Example 20.41].

For a given maximal monotone set-valued operator $A : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ and a single-valued Lipschitz monotone operator $F : \mathbb{R}^d \to \mathbb{R}^d$, in this paper we are interested in finding $\boldsymbol{z} \in \mathbb{R}^d$ such that $\mathbf{0} \in F(\boldsymbol{z}) + A(\boldsymbol{z})$, which is called the (constrained) *monotone inclusion problem*.

3 Discrete-Time Mean-Field Equilibrium as a Monotone Inclusion Problem

Let $\mathrm{MFG}_{\mu_0} = (S, A, r, P, \mu_0, H)$ be a finite-horizon mean-field game. In this section, we represent an MFE of MFG_{μ_0} as a solution to a monotone inclusion problem. To do so, we formulate the MFE problem as an optimization problem over the space of occupation measures. In the next section, we show that this optimization problem can be solved by tractable algorithms, and we present a tractable algorithm together with convergence guarantees.

To represent an MFE as a monotone inclusion problem, we first formulate the MFE problem as an optimization problem. For this purpose, we introduce additional notation and definitions related to occupation measures.

Definition 3. *Let $\boldsymbol{\pi}$ be a policy flow, and let $\boldsymbol{\mu}^{\boldsymbol{\pi}} = (\mu_h^{\boldsymbol{\pi}})_{h=0}^{H-1}$ be a state-measure flow associated with $\boldsymbol{\pi}$. For each $h = 0, \ldots, H-1$, the joint probability measure on $S \times A$ defined by*

$$\rho_h^{\boldsymbol{\pi}}(s, a) := \pi_h(a \mid s)\, \mu_h^{\boldsymbol{\pi}}(s)$$

is called the occupation measure *corresponding to $\boldsymbol{\pi}$ at time h. The collection $\boldsymbol{\rho}^{\boldsymbol{\pi}} = (\rho_h^{\boldsymbol{\pi}})_{h=0}^{H-1}$ is called the corresponding* occupation flow.

For a given horizon length H, we define the ambient space containing all occupation flows corresponding to admissible policies as the finite-horizon occupation-flow polytope

$$\mathcal{K} := \left\{ \boldsymbol{\rho} = (\rho_h)_{h=0}^{H-1} \in \prod_{h=0}^{H-1} \mathbb{R}_+^{S \times A} : \begin{array}{l} \displaystyle\sum_{a \in A} \rho_0(s, a) = \mu_0(s) \quad \forall s \in S, \\ \displaystyle\sum_{a \in A} \rho_{h+1}(s', a) = \sum_{s \in S} \sum_{a \in A} P(s' \mid s, a) \rho_h(s, a) \\ \qquad\qquad \forall s' \in S,\ \forall h = 0, \ldots, H-2 \end{array} \right\}.$$

In general, a policy flow need not determine a unique occupation flow. Conversely, a given occupation flow may correspond to more than one policy flow.

Definition 4. *For a given occupation flow $\boldsymbol{\rho} \in \mathcal{K}$, the* induced state flow *is defined by*

$$\mu_h^{\boldsymbol{\rho}}(s) := \sum_{a \in A} \rho_h(s,a), \qquad h = 0, \dots, H-1.$$

It is straightforward to verify that if $(\boldsymbol{\pi}^*, \boldsymbol{\mu}^*)$ is a finite-horizon MFE, then there exists an occupation flow corresponding to $\boldsymbol{\pi}^*$ whose induced state flow is exactly $\boldsymbol{\mu}^*$.

Lemma 1. *Suppose that Assumption 2 holds. Then, $\mathcal{K}$ is a nonempty, compact, convex polytope in the Euclidean space under consideration.*

Assumption 2 is crucial for showing that $\mathcal{K}$ is convex. For our purposes, the importance of this convexity property is that it implies that the normal cone $N_{\mathcal{K}}(\boldsymbol{x})$ is maximal monotone for every $\boldsymbol{x} \in \mathcal{K}$; see Remark 3. Although Guo et al. [2024] also formulates an optimization problem based on occupation measures, it does not consider the setting of Assumption 2. In the absence of Assumption 2, the corresponding optimization problem for MFE typically becomes nonlinear, and this nonlinearity is difficult to handle in full generality.

Definition 5. *The operator $F : \mathcal{K} \to \mathbb{R}^{H|S||A|}$, defined coordinatewise by*

$$[F(\boldsymbol{\rho})]_{h,s,a} := -\, r\big(s, a, \mu_h^{\boldsymbol{\rho}}\big),$$

is called the mean-field occupation operator.

It is immediate that F is a single-valued operator. For our purposes, we will also need F to be monotone and Lipschitz continuous on $\mathcal{K}$.

Lemma 2. *Suppose Assumptions 1–3 hold. Then, for all $\boldsymbol{\rho}, \tilde{\boldsymbol{\rho}} \in \mathcal{K}$,*

$$\|F(\boldsymbol{\rho}) - F(\tilde{\boldsymbol{\rho}})\|_2 \le |S||A|\, L_r\, \|\boldsymbol{\rho} - \tilde{\boldsymbol{\rho}}\|_2.$$

Lemma 3. *Suppose Assumptions 1–3 hold. Then, the operator F is monotone on $\mathcal{K}$; that is,*

$$\langle F(\boldsymbol{\rho}) - F(\tilde{\boldsymbol{\rho}}),\, \boldsymbol{\rho} - \tilde{\boldsymbol{\rho}} \rangle \ge 0 \qquad \forall \boldsymbol{\rho}, \tilde{\boldsymbol{\rho}} \in \mathcal{K}.$$

The main result of this section is the following.

Theorem 1. *Suppose that Assumptions 1–3 hold. Let $(\boldsymbol{\pi}^\star, \boldsymbol{\mu}^\star)$ be a policy flow and a state-measure flow, respectively, and define*

$$\rho_h^*(s,a) = \pi_h^*(a \mid s)\mu_h^*(s).$$

Then, $(\boldsymbol{\pi}^\star, \boldsymbol{\mu}^\star)$ is an MFE if, and only if,

$$\langle F(\boldsymbol{\rho}^\star),\, \boldsymbol{\rho} - \boldsymbol{\rho}^\star \rangle \ge 0 \qquad \forall \boldsymbol{\rho} \in \mathcal{K}. \tag{5}$$

Outline of the proof. For any policy flow $\boldsymbol{\pi}$ and state-measure flow $\boldsymbol{\mu}$, since the transition kernel is independent of the state-measure flow, we can write

$$J(\boldsymbol{\pi}, \boldsymbol{\mu}) = \sum_{h=0}^{H-1} \sum_{s \in S} \sum_{a \in A} r(s, a, \mu_h)\, \rho_h^{\boldsymbol{\pi}}(s,a) =: L_{\boldsymbol{\mu}}(\boldsymbol{\rho}^{\boldsymbol{\pi}}). \tag{6}$$

For any fixed state-measure flow $\boldsymbol{\mu}$, we extend the functional $L_{\boldsymbol{\mu}}$ to entire $\mathcal{K}$ as

$$L_{\boldsymbol{\mu}}(\boldsymbol{\rho}) = \sum_{h=0}^{H-1} \sum_{s \in S} \sum_{a \in A} r\big(s, a, \mu_h\big)\, \rho_h(s,a).$$

Using the functional $L_{\boldsymbol{\mu}}(\boldsymbol{\rho})$, we show that $(\boldsymbol{\pi}^\star, \boldsymbol{\mu}^\star)$ induces an MFE if, and only if, $L_{\boldsymbol{\mu}^\star}(\boldsymbol{\rho}) \le L_{\boldsymbol{\mu}^\star}(\boldsymbol{\rho}^\star)$ for all $\boldsymbol{\rho} \in \mathcal{K}$, where $\boldsymbol{\rho}^\star$ is the occupation flow corresponding to this MFE. Moreover, using (6), one can readily show that this criterion is equivalent to (5). □

It can be shown that $\mathcal{K}$ is a nonempty, compact, and convex set. Equivalently, (5) can be written as the monotone inclusion [Sedlmayer et al., 2023, Section 1.1]

$$\mathbf{0} \in F(\boldsymbol{\rho}^\star) + N_{\mathcal{K}}(\boldsymbol{\rho}^\star), \tag{7}$$

where $N_{\mathcal{K}}$ denotes the normal-cone operator of $\mathcal{K}$ (Remark 3).

Corollary 1. *With the notation of Theorem 1, $(\boldsymbol{\pi}^\star, \boldsymbol{\mu}^\star)$ is a finite-horizon MFE if, and only if,*

$$\mathbf{0} \in F(\boldsymbol{\rho}^\star) + N_{\mathcal{K}}(\boldsymbol{\rho}^\star),$$

where $\rho_h^*(s,a) = \pi_h^*(a|s)\mu_h^*(s)$.

4 Mean-Field Anchored Proximal Gradient over Occupation Measures

In this section, we present an anchored first-order method for computing a finite-horizon MFE in the occupation-measure formulation. In general, for monotone inclusions of the form $\mathbf{0} \in F(\mathbf{z}) + A(\mathbf{z})$, vanilla proximal descent methods are not guaranteed to converge. To address this issue, we employ an acceleration mechanism that ensures convergence.

For any given $\alpha > 0$, since $\mathcal{K}$ is a nonempty, compact, convex polytope, the normal-cone operator $N_{\mathcal{K}}$ is maximal monotone, and its resolvent coincides with the Euclidean projection onto $\mathcal{K}$; that is,

$$J_{\alpha N_{\mathcal{K}}} = (I + \alpha N_{\mathcal{K}})^{-1} = \Pi_{\mathcal{K}}.$$

Therefore, a proximal anchored-gradient step for (7) reduces to a projected anchored step on the occupation variable.

Anchored step. Starting from an initial feasible occupation flow $\boldsymbol{\rho}^{\mathbf{0}} \in \mathcal{K}$, the iteration is given by

$$\boldsymbol{\rho}^{\boldsymbol{k+1}} = \Pi_{\mathcal{K}}\Big((1-\beta_k)\boldsymbol{\rho}^{\boldsymbol{k}} + \beta_k \boldsymbol{\rho}^{\mathbf{0}} - \alpha_k F(\boldsymbol{\rho}^{\boldsymbol{k}})\Big), \qquad k = 0, 1, 2, \dots. \tag{8}$$

Here, $\alpha_k > 0$ denotes the step size, and $\beta_k \in [0, 1]$ is the anchoring parameter. A standard choice is

$$\alpha_k = \frac{1}{L\sqrt{k+\gamma}}, \qquad \beta_k = \frac{\gamma}{k+\gamma}, \qquad \gamma \geq 2, \tag{9}$$

where $L := |S||A|\, L_r$ is the Lipschitz constant of the operator F, as established in Lemma 2.

Remark 4. *The step-size choice in* (9) *was first introduced in Surina et al. [2026] for the unconstrained case, and was later extended to the constrained case in Cai and Zheng [2026].*

The update (8) admits a simple interpretation. The term $-\alpha_k F(\boldsymbol{\rho}^{\boldsymbol{k}})$ is a forward step driven by the current reward under the induced mean-field flow, while the term $\beta_k(\boldsymbol{\rho}^{\mathbf{0}} - \boldsymbol{\rho}^{\boldsymbol{k}})$ pulls the iterate back toward the anchor $\boldsymbol{\rho}^{\mathbf{0}}$, thereby stabilizing the last iterate. The projection $\Pi_{\mathcal{K}}$ then restores feasibility with respect to the occupation-flow constraints. We also note that implementing the update (8) requires access to both the reward function and the transition kernel through suitable oracle queries.

A further remark on the projection step in (8) is in order. Since $\mathcal{K}$ is a compact convex polytope, computing the projection $\Pi_{\mathcal{K}}(\boldsymbol{x})$ is equivalent to solving the quadratic program

$$\min_{\boldsymbol{y} \in \mathcal{K}} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_2^2, \tag{10}$$

which can be solved in polynomial time and is therefore tractable. Indeed, first, note that for some fixed matrix $\boldsymbol{A}$ and vector $\boldsymbol{b}$, the set $\mathcal{K}$ can be rewritten as

$$\mathcal{K} = \{\boldsymbol{\rho} \in \prod_{h=0}^{H-1} \mathbb{R}_+^{S\times A} : \boldsymbol{A}\boldsymbol{\rho} = \boldsymbol{b}\}.$$

In particular, (10) is a convex quadratic program, and when $\boldsymbol{A}$ and $\boldsymbol{b}$ are written in rationals, it can be solved exactly in polynomial time Monteiro and Adler [1989]:

Theorem 2. *Let L be the input size of $\boldsymbol{x}$. Then* (10) *can be solved exactly in $O(H\,|S|\,|A|\,L)$-arithmetic steps.*

Proof. This result directly follows from Monteiro and Adler [1989]. □

Corollary 2. *For each k, let L_k be the input size of each update $(1-\beta_k)\boldsymbol{\rho}^{\boldsymbol{k}} + \beta_k\boldsymbol{\rho}^{\mathbf{0}} - \alpha_k F(\boldsymbol{\rho}^{\boldsymbol{k}})$. Then, each step* (8) *can be calculated in at most $O(H\,|S|\,|A|\,C_r + (H\,|S|\,|A|)^3\,L_k)$-arithmetic steps, where C_r is the computational cost of encoding $r(\cdot,\cdot,\cdot)$.*

Proof. Note that evaluating $F(\boldsymbol{\rho}^{\boldsymbol{k}})$ requires $H\,|S|\,|A|$ scalar reward evaluations. Thus, as C_r is the arithmetic cost of one scalar reward evaluation, $(1-\beta_k)\boldsymbol{\rho}^{\boldsymbol{k}} + \beta_k\boldsymbol{\rho}^{\mathbf{0}} - \alpha_k F(\boldsymbol{\rho}^{\boldsymbol{k}})$ can be evaluated in $O(H\,|S|\,|A|\,C_r)$ steps. Combined with the cost of projection in Theorem 2, we obtain the desired $O(H\,|S|\,|A|\,C_r + (H\,|S|\,|A|)^3\,L_k)$ arithmetic complexity. □

In particular, in contrast to previous work, most notably Isobe et al. [2025], all intermediate steps of our algorithm are fully tractable. A more closely related line of work is Guo et al. [2024], which also uses occupation measures and proposes an iterative algorithm. However, that approach relies on the ability to choose a sufficiently good initialization in order to guarantee merely an asymptotic convergence, whereas our method does not require such an assumption.

Recovery of the policy and state-measure flows. Given an occupation-flow iterate $\boldsymbol{\rho^k}$, recall that the induced state flow is defined by

$$\mu_h^k(s) := \sum_{a \in A} \rho_h^k(s,a), \qquad h = 0, \dots, H-1.$$

A policy flow $\boldsymbol{\pi^k}$ can then be recovered from $\boldsymbol{\rho^k}$ by disintegration:

$$\pi_h^k(a \mid s) = \begin{cases} \dfrac{\rho_h^k(s,a)}{\mu_h^k(s)}, & \text{if } \mu_h^k(s) > 0, \\ \text{any element of } \mathcal{P}(A), & \text{if } \mu_h^k(s) = 0. \end{cases} \tag{11}$$

Thus, each iterate $\boldsymbol{\rho^k}$ naturally defines an approximate equilibrium candidate $(\boldsymbol{\pi^k}, \boldsymbol{\mu^k})$. We note, however, that the recovered policy flow need not be unique, since on states with zero mass one may choose any element of $\mathcal{P}(A)$.

As noted in Cai and Zheng [2026], a natural notion of error metric for the iteration (8) is given by the tangential residual

$$r^{\tan}_{F,N_{\mathcal{K}}}(\boldsymbol{\rho^T}) := \inf_{\boldsymbol{c} \in N_{\mathcal{K}}(\boldsymbol{\rho^T})} \|F(\boldsymbol{\rho^T}) + \boldsymbol{c}\|_2. \tag{12}$$

By Corollary 1, an occupation flow $\boldsymbol{\rho^\star}$ induces an MFE if, and only if, $\mathbf{0} \in F(\boldsymbol{\rho^\star}) + N_{\mathcal{K}}(\boldsymbol{\rho^\star})$. Equivalently, there exists $\boldsymbol{c^\star} \in N_{\mathcal{K}}(\boldsymbol{\rho^\star})$ such that $\mathbf{0} = F(\boldsymbol{\rho^\star}) + \boldsymbol{c^\star}$. Thus, the residual in (12) measures the smallest remaining error after correcting $F(\boldsymbol{\rho^T})$ by a normal-cone direction to $\mathcal{K}$ at $\boldsymbol{\rho^T}$. In particular, the residual vanishes exactly when $\boldsymbol{\rho^T}$ satisfies the MFE condition.

Theorem 3. *Suppose that $G : \mathbb{R}^d \to \mathbb{R}^d$ is monotone and L-Lipschitz continuous with respect to $\|\cdot\|_2$, and that $A : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ is maximally monotone. Further, suppose that there exists $\boldsymbol{z^\star} \in \mathbb{R}^d$ such that $\mathbf{0} \in G(\boldsymbol{z^\star}) + A(\boldsymbol{z^\star})$. Then, the iterates generated by $\boldsymbol{z_{t+1}} = J_{\alpha_t A}\left((1-\beta_t)\boldsymbol{z_t} + \beta_t \boldsymbol{z_0} - \alpha_t G(\boldsymbol{z_t})\right)$ satisfy*

$$r^{\tan}_{G,A}(\boldsymbol{z}_T) \le \frac{25\gamma L \|\boldsymbol{z^\star} - \boldsymbol{z_0}\|_2}{\sqrt{T-1+\gamma}}.$$

Proof. This result is the same as [Cai and Zheng, 2026, Theorem 2]. □

Theorem 4. *Suppose that Assumptions 1–3 hold. Then, the iterates generated by* (8) *with step sizes given by* (9) *satisfy*

$$r^{\tan}_{F,N_{\mathcal{K}}}(\boldsymbol{\rho^T}) \le \frac{25\,\gamma\,\sqrt{2H}\,|S|\,|A|\,L_r}{\sqrt{T-1+\gamma}}.$$

Proof. By Lemma 3, F is monotone on $\mathcal{K}$ and by Lemma 2 F is $|S||A|L_r$-Lipschitz on $\mathcal{K}$. Note that the iterations (8) remain on the set $\mathcal{K}$. Thus, although F is not monotone on the whole $\mathbb{R}^d$, Theorem 3 is still applicable to F. By Remark 3 and Lemma 1, we have that $N_{\mathcal{K}}$ is maximally monotone. Lastly, by [Perolat et al., 2021, Proposition 1] and Assumption 1, there exists $\mathbf{0} \in F(\boldsymbol{\rho}^*) + N_{\mathcal{K}}(\boldsymbol{\rho^\star})$ for some $\boldsymbol{\rho^\star} \in \mathcal{K}$. Thus, since $\|\boldsymbol{z_t} - \boldsymbol{z_0}\|_2 \le \sqrt{2H}$, the desired rate of convergence follows directly from Theorem 3. □

It remains to verify that every accumulation point of the sequence $(\boldsymbol{\rho^T})_T$ induces an MFE. Since this sequence is bounded, all of its accumulation points can be reached via some subsequence.

Lemma 4. *For any accumulation point $\boldsymbol{\rho^\star}$ of $(\boldsymbol{\rho^T})_T$ we have $r^{\tan}_{F,N_{\mathcal{K}}}(\boldsymbol{\rho^\star}) = 0$.*

Let $\boldsymbol{\rho^\star}$ be an accumulation point of $(\boldsymbol{\rho^T})_T$. Thus, there exists a subsequence $(T_n)_n$ such that $\lim_{n\to\infty} \boldsymbol{\rho^{T_n}} = \boldsymbol{\rho^\star}$. By Lemma 4, we have $r^{\tan}_{F,N_{\mathcal{K}}}(\boldsymbol{\rho^\star}) = 0$. Hence, by (12), there exists some $\boldsymbol{c^\star} \in N_{\mathcal{K}}(\boldsymbol{\rho}^*)$ such that $\mathbf{0} = F(\boldsymbol{\rho^\star}) + \boldsymbol{c^\star}$. Equivalently, $\mathbf{0} \in F(\boldsymbol{\rho^\star}) + N_{\mathcal{K}}(\boldsymbol{\rho^\star})$. Therefore, by

Corollary 1, $\boldsymbol{\rho^*}$ induces a finite-horizon MFE. Since $\boldsymbol{\rho^*}$ was chosen as an arbitrary accumulation point of $(\boldsymbol{\rho^T})_T$, every accumulation point of this sequence induces an MFE.

The next result identifies the unique limit of $\boldsymbol{\rho^T}$.

Theorem 5. *Let $\mathcal{S} \subset \mathcal{K}$ denote the set of occupancy measures that induce an MFE. If Assumptions 1–3 hold, then $\mathcal{S}$ is a compact and convex set. Let $\Pi_{\mathcal{S}}(\boldsymbol{\rho^0})$ denote the orthogonal projection of $\boldsymbol{\rho^0}$ onto $\mathcal{S}$. Then, $\lim_{T\to\infty} \boldsymbol{\rho^T} = \Pi_{\mathcal{S}}(\boldsymbol{\rho^0})$ with respect to the Euclidean norm. In particular, when one has complete access to r and P, the iterations in* (8) *are tractable and converge to an occupancy measure that induces an MFE.*

Proof. We show in the appendix that $\lim_{T\to\infty} \boldsymbol{\rho^T} = \Pi_{\mathcal{S}}(\boldsymbol{\rho^0})$. The convergence of the iterations in (8) follows from Theorem 4, Lemma 4, and the fact that $\lim_{T\to\infty} \boldsymbol{\rho^T} = \Pi_{\mathcal{S}}(\boldsymbol{\rho^0})$. When one has access to P, the QP in (10) can be solved in polynomial time by Corollary 2. Thus, the algorithm is fully tractable and converges to an exact MFE. □

Moreover, under our assumptions, one can construct an example for which the rate $\Omega(1/\sqrt{T})$ is also attained as a matching lower bound for this iteration scheme. In this sense, the bound in Theorem 4 is tight for this algorithm.

Proposition 1. *There exists a mean-field game with time-homogeneous reward function and transition kernel for which the convergence rate in Theorem 4 is matched by a corresponding lower bound.*

Proof. We will provide the example here but defer the proof of the lower bound for the rate of convergence of the iterations to the appendix.

For the proof we consider the finite-horizon MFG with $H = 2$, state space $S = \{0, 1, 2\}$, action space $A = \{L, R\}$, and initial distribution $\mu_0 = \delta_0$. The transition kernel is independent of the population measure and satisfies $P(1 \mid 1, \cdot) \equiv P(2 \mid 2, \cdot) = 1$ and

$$P(1 \mid 0, L) = P(2 \mid 0, R) = 1, \qquad P(2 \mid 0, L) = P(1 \mid 0, R) = 0,$$

with states 1 and 2 absorbing. The rewards are

$$r(0, \cdot, \cdot) \equiv 0, \qquad r(s, a, \mu) = -\mu(s), \quad s \in \{1, 2\},\ a \in \{L, R\}.$$

It can be easily shown that this example satisfies the assumptions 1–3. □

Several remarks are in order concerning Proposition 1. The upper bound proved in Theorem 4 is not optimal in general for other gradient methods in unconstrained monotone inclusion problems; see Yoon and Ryu [2021]. In our case, however, since $N_{\mathcal{K}} \not\equiv \emptyset$, the problem falls into the constrained setting. For constrained monotone inclusion problems, Cai et al. [2022] shows that, under additional assumptions—which in our framework amount to comonotonicity of the operator $(F, N_{\mathcal{K}})$—one can recover the sharper rate established in Yoon and Ryu [2021].

5 Conclusion

In this work, by formulating the MFE problem as a monotone inclusion problem, we developed algorithms with explicit convergence guarantees for computing MFEs in weakly monotone MFGs. In addition, Proposition 1 shows that the rate $\Omega(1/\sqrt{T})$ appears as a matching lower bound for our algorithm, implying the tightness of the upper bound.

The main limitation of our analysis is Assumption 2. An important direction for future work is therefore to relax this assumption. A second interesting direction is to extend the convergence theory of anchored extragradient and proximal methods to monotone inclusion problems involving merely Lipschitz continuous monotone operators together with maximal monotone perturbations. Such optimistic methods can achieve faster convergence rates of order $O(1/T)$, and are therefore especially attractive for this class of problems.

Acknowledgments and Disclosure of Funding

Research of first and second authors was supported in part by the AFOSR Grant FA9550-24-1-0152.

References

Ş. K. Alkır and N. Saldi. Inverse reinforcement learning for mean-field games with average reward criterion. In *2025 IEEE 64th Conference on Decision and Control (CDC)*, pages 7272–7277. IEEE, 2025.

B. Anahtarcı, C. D. Karıksız, and N. Saldi. Value iteration algorithm for mean-field games. *Systems & Control Letters*, 143:104744, 2020.

B. Anahtarci, C. D. Kariksiz, and N. Saldi. Q-learning in regularized mean-field games. *Dynamic Games and Applications*, 13(1):89–117, 2023.

B. Anahtarci, C. D. Kariksiz, and N. Saldi. Kernel based maximum entropy inverse reinforcement learning for mean-field games. *arXiv preprint arXiv:2507.14529*, 2025a.

B. Anahtarci, C. D. Kariksiz, and N. Saldi. Maximum causal entropy IRL in mean-field games and GNEP framework for forward RL. *Journal of Machine Learning Research*, 26(121):1–40, 2025b.

H. H. Bauschke and P. L. Combettes. Correction to: convex analysis and monotone operator theory in Hilbert spaces. In *Convex Analysis and Monotone OperatorTtheory in Hilbert Spaces*, pages C1–C4. Springer, 2020.

Y. Cai and W. Zheng. Doubly optimal no-regret learning in monotone games. In *International Conference on Machine Learning*, pages 3507–3524. PMLR, 2023.

Y. Cai and W. Zheng. Last-iterate convergence of anchored gradient descent. *arXiv preprint arXiv:2604.12235*, 2026.

Y. Cai, A. Oikonomou, and W. Zheng. Accelerated algorithms for constrained nonconvex-nonconcave min-max optimization and comonotone inclusion. *arXiv preprint arXiv:2206.05248*, 2022.

K. Cui and H. Koeppl. Approximately solving mean field games via entropy-regularized deep reinforcement learning. In *International Conference on Artificial Intelligence and Statistics*, pages 1909–1917. PMLR, 2021.

C. Daskalakis and I. Panageas. The limit points of (optimistic) gradient descent in min-max optimization. *Advances in Neural Information Processing Systems*, 31, 2018.

J. Dong, B. Wang, and Y. Yu. Last-iterate convergence in regularized graphon mean field game. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 13779–13787, 2025.

N. Golowich, S. Pattathil, C. Daskalakis, and A. Ozdaglar. Last iterate is slower than averaged iterate in smooth convex-concave saddle point problems. In *Conference on Learning Theory*, pages 1758–1784. PMLR, 2020.

X. Guo, A. Hu, R. Xu, and J. Zhang. Learning mean-field games. *Advances in Neural Information Processing Systems*, 32, 2019.

X. Guo, A. Hu, and J. Zhang. Optimization frameworks and sensitivity analysis of Stackelberg mean-field games. *arXiv preprint arXiv:2210.04110*, 2022.

X. Guo, A. Hu, and J. Zhang. Mf-omo: An optimization formulation of mean-field games. *SIAM Journal on Control and Optimization*, 62(1):243–270, 2024.

M. Huang, R. P. Malhamé, and P. E. Caines. Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. *Communications in Information and Systems*, 6(3):221–252, 2006.

N. Isobe, K. Abe, and K. Ariu. Last iterate convergence in monotone mean field games. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

J.-M. Lasry and P.-L. Lions. Mean field games. *Jpn. J. Math.*, 2(1):229–260, 2007.

M. Lauriere. Numerical methods for mean field games and mean field type control. *Mean field games*, 78(221-282), 2021.

R. D. Monteiro and I. Adler. Interior path following primal-dual algorithms. part ii: Convex quadratic programming. *Mathematical Programming*, 44(1):43–66, 1989.

S. Y. Olmez, S. Aggarwal, J. W. Kim, E. Miehling, T. Başar, M. West, and P. G. Mehta. Modeling presymptomatic spread in epidemics via mean-field games. In *2022 American control conference (ACC)*, pages 3648–3655. IEEE, 2022.

J. Perolat, S. Perrin, R. Elie, M. Laurière, G. Piliouras, M. Geist, K. Tuyls, and O. Pietquin. Scaling up mean field games with online mirror descent. *arXiv preprint arXiv:2103.00623*, 2021.

M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, 2014.

R. Rockafellar. Convex Analysis. *Princeton Math. Series*, 28, 1970.

E. K. Ryu, K. Yuan, and W. Yin. Ode analysis of stochastic gradient methods with optimism and anchoring for minimax problems. *arXiv preprint arXiv:1905.10899*, 2019.

N. Saldi. Linear mean-field games with discounted cost. *Mathematics of Operations Research*, 51(1): 358–397, 2026.

N. Saldi, T. Başar, and M. Raginsky. Markov–Nash equilibria in mean-field games with discounted cost. *SIAM Journal on Control and Optimization*, 56(6):4256–4287, 2018.

N. Saldi, T. Başar, and M. Raginsky. Approximate Markov-Nash equilibria for discrete-time risk-sensitive mean-field games. *Mathematics of Operations Research*, 45(4):1596–1620, 2020.

M. Sedlmayer, D.-K. Nguyen, and R. I. Bot. A fast optimistic method for monotone variational inequalities. In *International Conference on Machine Learning*, pages 30406–30438. PMLR, 2023.

J. Subramanian and A. Mahajan. Reinforcement learning in stationary mean-field games. In *Proceedings of the 18th international conference on autonomous agents and multiagent systems*, pages 251–259, 2019.

A. Surina, A. Suggala, G. Tsoukalas, A. Kovsharov, S. Shirobokov, F. J. Ruiz, P. Kohli, and S. Chaudhuri. An improved last-iterate convergence rate for anchored gradient descent ascent. *arXiv preprint arXiv:2604.03782*, 2026.

G. Y. Weintraub, C. L. Benkard, and B. Van Roy. Markov perfect industry dynamics with many firms. *Econometrica*, 76(6):1375–1411, 2008.

B. Yardım, S. Cayci, M. Geist, and N. He. Policy mirror ascent for efficient and independent learning in mean field games. In *International Conference on Machine Learning*, pages 39722–39754. PMLR, 2023.

B. Yardim, A. Goldman, and N. He. When is mean-field reinforcement learning tractable and relevant? In *AAMAS'24: Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems*, pages 2038–2046. International Foundation for Autonomous Agents and Multiagent Systems, 2024.

T. Yoon and E. K. Ryu. Accelerated algorithms for smooth convex-concave minimax problems with o (1/kˆ 2) rate on squared gradient norm. In *International Conference on Machine Learning*, pages 12098–12109. PMLR, 2021.

C. Zhang, X. Chen, and X. Di. Stochastic semi-gradient descent for learning mean field games with population-aware function approximation. *arXiv preprint arXiv:2408.08192*, 2024.

F. Zhang, V. Tan, Z. Wang, and Z. Yang. Learning regularized monotone graphon mean-field games. *Advances in Neural Information Processing Systems*, 36:67297–67308, 2023.

A Detailed Proofs of the Statements in Section 3

Lemma 5. *Every policy flow induces a feasible occupation flow.*

Proof. Fix a policy flow $\boldsymbol{\pi} = (\pi_h)_{h=0}^{H-1}$. Let $\boldsymbol{\mu}^{\boldsymbol{\pi}} = (\mu_h^{\boldsymbol{\pi}})_{h=0}^{H-1}$ be the corresponding state flow, defined recursively by

$$\mu_1^{\boldsymbol{\pi}} = \mu_0, \qquad \mu_{h+1}^{\boldsymbol{\pi}}(s') = \sum_{s\in S}\sum_{a\in A} P(s' \mid s,a)\,\pi_h(a \mid s)\,\mu_h^{\boldsymbol{\pi}}(s) \quad (h = 0,\dots,H-2).$$

Define an occupation flow associated with $\boldsymbol{\pi}$ (starting from $\pi_0(a|s)\mu_0(s)$) by

$$\rho_h^{\boldsymbol{\pi}}(s,a) := \pi_h(a \mid s)\,\mu_h^{\boldsymbol{\pi}}(s).$$

Then for every $s \in S$,

$$\sum_{a\in A} \rho_0^{\boldsymbol{\pi}}(s,a) = \sum_{a\in A} \pi_0(a \mid s)\,\mu_0^{\boldsymbol{\pi}}(s) = \mu_0^{\boldsymbol{\pi}}(s) = \mu_0(s).$$

Also, for each $h = 1,\dots,H-1$ and $s' \in S$,

$$\begin{aligned}\sum_{a\in A} \rho_{h+1}^{\boldsymbol{\pi}}(s',a) = \mu_{h+1}^{\boldsymbol{\pi}}(s') &= \sum_{s\in S}\sum_{a\in A} P(s' \mid s,a)\,\pi_h(a \mid s)\,\mu_h^{\boldsymbol{\pi}}(s)\\ &= \sum_{s\in S}\sum_{a\in A} P(s' \mid s,a)\,\rho_h^{\boldsymbol{\pi}}(s,a).\end{aligned}$$

Hence $\boldsymbol{\rho}^{\boldsymbol{\pi}} \in \mathcal{K}$. □

Lemma 6. *Every feasible occupation flow comes from some policy flow.*

Proof. Fix $\boldsymbol{\rho} \in \mathcal{K}$ and define

$$\mu_h^{\boldsymbol{\rho}}(s) := \sum_{a\in A} \rho_h(s,a).$$

We define a policy $\boldsymbol{\pi}^{\boldsymbol{\rho}} = (\pi_h^{\boldsymbol{\rho}})_{h=0}^{H-1}$ by

$$\pi_h^{\boldsymbol{\rho}}(a \mid s) := \begin{cases} \dfrac{\rho_h(s,a)}{\mu_h^{\boldsymbol{\rho}}(s)}, & \text{if } \mu_h^{\boldsymbol{\rho}}(s) > 0,\\ \text{an arbitrary element of } \mathcal{P}(A), & \text{if } \mu_h^{\boldsymbol{\rho}}(s) = 0.\end{cases}$$

We claim that the state flow generated by the policy flow $\boldsymbol{\pi}^{\boldsymbol{\rho}}$ is exactly $\boldsymbol{\mu}^{\boldsymbol{\rho}}$ and that its occupation flow is exactly $\boldsymbol{\rho}$.

We prove by induction on h that $\mu_h^{\boldsymbol{\pi}^{\boldsymbol{\rho}}} = \mu_h^{\boldsymbol{\rho}}$. For $h = 0$,

$$\mu_0^{\boldsymbol{\pi}^{\boldsymbol{\rho}}}(s) = \mu_0(s) = \sum_{a\in A} \rho_0(s,a) = \mu_0^{\boldsymbol{\rho}}(s).$$

Assume that $\mu_k^{\boldsymbol{\pi}^{\boldsymbol{\rho}}} = \mu_k^{\boldsymbol{\rho}}$ for all $k = 1,\cdots,h$. Then, for each $s' \in S$,

$$\begin{aligned}\mu_{h+1}^{\boldsymbol{\pi}^{\boldsymbol{\rho}}}(s') &= \sum_{s\in S}\sum_{a\in A} P(s' \mid s,a)\,\pi_h^{\boldsymbol{\rho}}(a \mid s)\,\mu_h^{\boldsymbol{\pi}^{\boldsymbol{\rho}}}(s)\\ &= \sum_{s\in S}\sum_{a\in A} P(s' \mid s,a)\,\pi_h^{\boldsymbol{\rho}}(a \mid s)\,\mu_h^{\boldsymbol{\rho}}(s)\\ &= \sum_{s\in S}\sum_{a\in A} P(s' \mid s,a)\,\rho_h(s,a)\\ &= \sum_{a\in A} \rho_{h+1}(s',a)\\ &= \mu_{h+1}^{\boldsymbol{\rho}}(s').\end{aligned}$$

By induction, it follows that $\boldsymbol{\mu}^{\boldsymbol{\pi}^{\boldsymbol{\rho}}} = \boldsymbol{\mu}^{\boldsymbol{\rho}}$.

Finally,

$$\rho_h^{\boldsymbol{\pi}^{\boldsymbol{\rho}}}(s,a) = \pi_h^{\boldsymbol{\rho}}(a \mid s)\mu_h^{\boldsymbol{\pi}^{\boldsymbol{\rho}}}(s) = \pi_h^{\boldsymbol{\rho}}(a \mid s)\mu_h^{\boldsymbol{\rho}}(s) = \rho_h(s,a),$$

with the convention that if $\mu_h^{\boldsymbol{\rho}}(s) = 0$, then both sides are zero. Thus every $\boldsymbol{\rho} \in \mathcal{K}$ is the occupation flow of some policy flow. □

Lemma 7. *$\mathcal{K}$ is a nonempty compact convex polytope.*

Proof. By definition, $\mathcal{K}$ is the intersection of finitely many affine hyperplanes and the nonnegative orthant, hence it is a convex polyhedron. By Lemma 5 it is nonempty.

To show boundedness, let $\boldsymbol{\rho} \in \mathcal{K}$. Summing the initial constraints over s gives

$$\sum_{s\in S}\sum_{a\in A} \rho_0(s,a) = \sum_{s\in S} \mu_0(s) = 1.$$

Now assume $\sum_{s,a} \rho_h(s,a) = 1$. Then,

$$\begin{aligned}
\sum_{s'\in S}\sum_{a\in A} \rho_{h+1}(s',a) &= \sum_{s'\in S}\sum_{s\in S}\sum_{a\in A} P(s' \mid s,a)\rho_h(s,a) \\
&= \sum_{s\in S}\sum_{a\in A} \rho_h(s,a) \sum_{s'\in S} P(s' \mid s,a) \\
&= \sum_{s\in S}\sum_{a\in A} \rho_h(s,a) \\
&= 1.
\end{aligned}$$

Hence each stage occupation vector has total mass 1, and thus each coordinate belongs to $[0,1]$. Therefore $\mathcal{K}$ is bounded. Being a closed and bounded polyhedron, it is compact. Thus $\mathcal{K}$ is a nonempty compact convex polytope. □

Lemma 8. *Best response at fixed mean field is a linear program on $\mathcal{K}$*

Proof. Fix an exogenous mean-field flow $\boldsymbol{\mu} = (\mu_h)_{h=0}^{H-1}$. For a policy flow $\boldsymbol{\pi} = (\pi_h)_{h=0}^{H-1}$, let $\boldsymbol{\mu}^{\boldsymbol{\pi}} = (\mu_h^{\boldsymbol{\pi}})_{h=0}^{H-1}$ denote the state-flow induced by $\boldsymbol{\pi}$ under the transition kernel $P(\cdot \mid s,a)$ and the initial distribution μ_0. Define a corresponding occupation flow $\boldsymbol{\rho}^{\boldsymbol{\pi}} = (\rho_h^{\boldsymbol{\pi}})_{h=0}^{H-1}$ by

$$\rho_h^{\boldsymbol{\pi}}(s,a) := \pi_h(a \mid s)\mu_h^{\boldsymbol{\pi}}(s).$$

Then the best-response objective at the exogenous flow $\boldsymbol{\mu}$ is

$$J(\boldsymbol{\mu},\boldsymbol{\pi}) := \mathbb{E}^{\boldsymbol{\pi}}\left[\sum_{h=0}^{H-1} r(x_h,a_h,\mu_h)\right],$$

and therefore by Assumption 2 we obtain

$$J(\boldsymbol{\mu},\boldsymbol{\pi}) = \sum_{h=0}^{H-1}\sum_{s\in S}\sum_{a\in A} \rho_h^{\boldsymbol{\pi}}(s,a)\, r(s,a,\mu_h).$$

Therefore, for a fixed exogenous mean-field flow $\boldsymbol{\mu} = (\mu_h)_{h=0}^{H-1}$, maximizing $J(\boldsymbol{\mu},\boldsymbol{\pi})$ over policy flows $\boldsymbol{\pi}$ is equivalent to maximizing the linear functional

$$L_{\boldsymbol{\mu}}(\boldsymbol{\rho}) := \sum_{h=0}^{H-1}\sum_{s\in S}\sum_{a\in A} \rho_h(s,a)\, r(s,a,\mu_h)$$

over $\boldsymbol{\rho} = (\rho_h)_{h=0}^{H-1} \in \mathcal{K}$. Indeed, by Lemma 5, every policy flow $\boldsymbol{\pi}$ induces an occupation flow $\boldsymbol{\rho}^{\boldsymbol{\pi}} \in \mathcal{K}$, and by Lemma 6, every $\boldsymbol{\rho} \in \mathcal{K}$ is the occupation flow of some policy flow. Moreover, if $\boldsymbol{\rho}^{\boldsymbol{\pi}}$ denotes the occupation flow induced by $\boldsymbol{\pi}$, then

$$J(\boldsymbol{\mu},\boldsymbol{\pi}) = \sum_{h=0}^{H-1}\sum_{s\in S}\sum_{a\in A} \rho_h^{\boldsymbol{\pi}}(s,a)\, r(s,a,\mu_h) = L_{\boldsymbol{\mu}}(\boldsymbol{\rho}^{\boldsymbol{\pi}}).$$

Hence maximizing $J(\boldsymbol{\mu},\boldsymbol{\pi})$ over policy flows is equivalent to maximizing $L_{\boldsymbol{\mu}}(\boldsymbol{\rho})$ over $\boldsymbol{\rho} \in \mathcal{K}$. □

Lemma 9. *Mean-field equilibria are exactly solutions of the variational inequality* (5)*.*

Proof. Suppose first that $(\boldsymbol{\mu}^\star, \boldsymbol{\pi}^\star)$ is an MFE, and let $\boldsymbol{\rho}^\star := \boldsymbol{\rho}^{\boldsymbol{\pi}^\star}$ be the occupation flow induced by $\boldsymbol{\pi}^\star$. Since $(\boldsymbol{\mu}^\star, \boldsymbol{\pi}^\star)$ is an MFE, $\boldsymbol{\pi}^\star$ is a best response to the exogenous mean-field flow $\boldsymbol{\mu}^\star$, that is,

$$J(\boldsymbol{\mu}^\star, \boldsymbol{\pi}) \le J(\boldsymbol{\mu}^\star, \boldsymbol{\pi}^\star) \qquad \forall \boldsymbol{\pi}.$$

Let $\boldsymbol{\rho} \in \mathcal{K}$ be arbitrary. By Lemma 6, there exists a policy flow $\boldsymbol{\pi}^{\boldsymbol{\rho}}$ whose occupation flow is exactly $\boldsymbol{\rho}$. By Lemma 7,

$$J(\boldsymbol{\mu}^\star, \boldsymbol{\pi}^{\boldsymbol{\rho}}) = L_{\boldsymbol{\mu}^\star}(\boldsymbol{\rho}), \qquad J(\boldsymbol{\mu}^\star, \boldsymbol{\pi}^\star) = L_{\boldsymbol{\mu}^\star}(\boldsymbol{\rho}^\star),$$

where

$$L_{\boldsymbol{\mu}^\star}(\boldsymbol{\rho}) := \sum_{h=0}^{H-1} \sum_{s \in S} \sum_{a \in A} \rho_h(s,a)\, r(s,a,\mu_h^\star).$$

Since $(\boldsymbol{\mu}^\star, \boldsymbol{\pi}^\star)$ is an MFE,

$$L_{\boldsymbol{\mu}^\star}(\boldsymbol{\rho}) = J(\boldsymbol{\mu}^\star, \boldsymbol{\pi}^{\boldsymbol{\rho}}) \le J(\boldsymbol{\mu}^\star, \boldsymbol{\pi}^\star) = L_{\boldsymbol{\mu}^\star}(\boldsymbol{\rho}^\star) \qquad \forall \boldsymbol{\rho} \in \mathcal{K}.$$

Equivalently,

$$\sum_{h=0}^{H-1} \sum_{s \in S} \sum_{a \in A} (\rho_h(s,a) - \rho_h^\star(s,a))\, r(s,a,\mu_h^\star) \le 0 \qquad \forall \boldsymbol{\rho} \in \mathcal{K}.$$

Since $(\boldsymbol{\mu}^\star, \boldsymbol{\pi}^\star)$ is an MFE, the consistency condition gives

$$\boldsymbol{\mu}^\star = \boldsymbol{\mu}^{\boldsymbol{\pi}^\star}.$$

Moreover, since $\boldsymbol{\rho}^\star$ is the occupation flow of $\boldsymbol{\pi}^\star$, we also have

$$\boldsymbol{\mu}^{\boldsymbol{\pi}^\star} = \boldsymbol{\mu}^{\boldsymbol{\rho}^\star}.$$

Hence

$$\boldsymbol{\mu}^\star = \boldsymbol{\mu}^{\boldsymbol{\rho}^\star},$$

and thus the previous inequality becomes

$$\sum_{h=0}^{H-1} \sum_{s \in S} \sum_{a \in A} (\rho_h(s,a) - \rho_h^\star(s,a))\, r(s,a,\mu_h^{\boldsymbol{\rho}^\star}) \le 0 \qquad \forall \boldsymbol{\rho} \in \mathcal{K}.$$

Rearranging,

$$\sum_{h=0}^{H-1} \sum_{s \in S} \sum_{a \in A} \big[-r(s,a,\mu_h^{\boldsymbol{\rho}^\star})\big](\rho_h(s,a) - \rho_h^\star(s,a)) \ge 0 \qquad \forall \boldsymbol{\rho} \in \mathcal{K}.$$

By the definition of F, this is exactly

$$\langle F(\boldsymbol{\rho}^\star), \boldsymbol{\rho} - \boldsymbol{\rho}^\star \rangle \ge 0 \qquad \forall \boldsymbol{\rho} \in \mathcal{K}.$$

Hence $\boldsymbol{\rho}^\star$ solves the variational inequality (5).

Conversely, suppose that $\boldsymbol{\rho}^\star \in \mathcal{K}$ solves the variational inequality

$$\langle F(\boldsymbol{\rho}^\star), \boldsymbol{\rho} - \boldsymbol{\rho}^\star \rangle \ge 0 \qquad \forall \boldsymbol{\rho} \in \mathcal{K}.$$

By the definition of F, this means that

$$\sum_{h=0}^{H-1} \sum_{s \in S} \sum_{a \in A} (\rho_h(s,a) - \rho_h^\star(s,a))\, r(s,a,\mu_h^{\boldsymbol{\rho}^\star}) \le 0 \qquad \forall \boldsymbol{\rho} \in \mathcal{K}.$$

Define

$$\boldsymbol{\mu}^\star := \boldsymbol{\mu}^{\boldsymbol{\rho}^\star}, \qquad \boldsymbol{\pi}^\star := \boldsymbol{\pi}^{\boldsymbol{\rho}^\star},$$

where $\boldsymbol{\pi}^{\boldsymbol{\rho}^\star}$ is given by Lemma 6 (the exact choice does not matter). Then

$$\sum_{h=0}^{H-1} \sum_{s \in S} \sum_{a \in A} (\rho_h(s,a) - \rho_h^\star(s,a))\, r(s,a,\mu_h^\star) \le 0 \qquad \forall \boldsymbol{\rho} \in \mathcal{K},$$

that is,

$$L_{\boldsymbol{\mu}^\star}(\boldsymbol{\rho}) \le L_{\boldsymbol{\mu}^\star}(\boldsymbol{\rho}^\star) \qquad \forall \boldsymbol{\rho} \in \mathcal{K}.$$

Now let $\boldsymbol{\pi}$ be any policy flow, and let $\boldsymbol{\rho}^{\boldsymbol{\pi}}$ be a corresponding occupation flow. By Lemma 5, we have $\boldsymbol{\rho}^{\boldsymbol{\pi}} \in \mathcal{K}$. Therefore,

$$L_{\boldsymbol{\mu}^\star}(\boldsymbol{\rho}^{\boldsymbol{\pi}}) \le L_{\boldsymbol{\mu}^\star}(\boldsymbol{\rho}^\star).$$

Using Lemma 7 again,

$$J(\boldsymbol{\mu}^\star, \boldsymbol{\pi}) = L_{\boldsymbol{\mu}^\star}(\boldsymbol{\rho}^{\boldsymbol{\pi}}) \le L_{\boldsymbol{\mu}^\star}(\boldsymbol{\rho}^\star) = J(\boldsymbol{\mu}^\star, \boldsymbol{\pi}^\star).$$

Hence $\boldsymbol{\pi}^\star$ is a best response to $\boldsymbol{\mu}^\star$.

Finally, by construction,

$$\boldsymbol{\mu}^\star = \boldsymbol{\mu}^{\boldsymbol{\rho}^\star} \qquad \text{and} \qquad \boldsymbol{\rho}^\star = \boldsymbol{\rho}^{\boldsymbol{\pi}^\star},$$

and thus

$$\boldsymbol{\mu}^\star = \boldsymbol{\mu}^{\boldsymbol{\pi}^\star}.$$

Thus the consistency condition also holds, and therefore $(\boldsymbol{\pi}^\star, \boldsymbol{\mu}^\star)$ is a mean-field equilibrium. □

Lemma 10. *F is monotone on $\mathcal{K}$.*

Proof. Take any $\boldsymbol{\rho}, \tilde{\boldsymbol{\rho}} \in \mathcal{K}$. By the definition of F,

$$\begin{aligned}\langle F(\boldsymbol{\rho}) - F(\tilde{\boldsymbol{\rho}}), \boldsymbol{\rho} - \tilde{\boldsymbol{\rho}} \rangle &= \sum_{h=0}^{H-1} \sum_{s\in S} \sum_{a\in A} \Big(-r(s,a,\mu_h^{\boldsymbol{\rho}}) + r(s,a,\mu_h^{\tilde{\boldsymbol{\rho}}})\Big)\big(\rho_h(s,a) - \tilde{\rho}_h(s,a)\big) \\ &= -\sum_{h=0}^{H-1} \sum_{s\in S} \sum_{a\in A} \Big(r(s,a,\mu_h^{\boldsymbol{\rho}}) - r(s,a,\mu_h^{\tilde{\boldsymbol{\rho}}})\Big)\big(\rho_h(s,a) - \tilde{\rho}_h(s,a)\big).\end{aligned}$$

By Lemma 6, there exist policy flows $\boldsymbol{\pi}^{\boldsymbol{\rho}}$ and $\boldsymbol{\pi}^{\tilde{\boldsymbol{\rho}}}$ whose occupation flows are exactly $\boldsymbol{\rho}$ and $\tilde{\boldsymbol{\rho}}$, respectively. Moreover, by construction, the induced state flows of these policies are $\boldsymbol{\mu}^{\boldsymbol{\rho}}$ and $\boldsymbol{\mu}^{\tilde{\boldsymbol{\rho}}}$.

Applying Assumption 1 to the two policy flows $\boldsymbol{\pi}^{\boldsymbol{\rho}}$ and $\boldsymbol{\pi}^{\tilde{\boldsymbol{\rho}}}$, we obtain

$$\sum_{h=0}^{H-1} \sum_{s\in S} \sum_{a\in A} \Big(r(s,a,\mu_h^{\boldsymbol{\rho}}) - r(s,a,\mu_h^{\tilde{\boldsymbol{\rho}}})\Big)\big(\rho_h(s,a) - \tilde{\rho}_h(s,a)\big) \le 0.$$

Therefore,

$$\langle F(\boldsymbol{\rho}) - F(\tilde{\boldsymbol{\rho}}), \boldsymbol{\rho} - \tilde{\boldsymbol{\rho}} \rangle \ge 0.$$

Hence F is monotone on $\mathcal{K}$. □

Lemma 11. *F is Lipschitz continuous.*

Proof. Let $N :=$. For any given h, s, a, Assumption 3 implies

$$|F_{h,s,a}(\boldsymbol{\rho}) - F_{h,s,a}(\tilde{\boldsymbol{\rho}})| = |r(s,a,\mu_h^{\boldsymbol{\rho}}) - r(s,a,\mu_h^{\tilde{\boldsymbol{\rho}}})| \le L_r \|\mu_h^{\boldsymbol{\rho}} - \mu_h^{\tilde{\boldsymbol{\rho}}}\|_1.$$

Hence

$$\sum_{s\in S} \sum_{a\in A} |F_{h,s,a}(\boldsymbol{\rho}) - F_{h,s,a}(\tilde{\boldsymbol{\rho}})|^2 \le |S|^2 |A|^2 L_r^2 \|\mu_h^{\boldsymbol{\rho}} - \mu_h^{\tilde{\boldsymbol{\rho}}}\|_1^2.$$

Note that

$$\begin{aligned}\|\mu_h^{\boldsymbol{\rho}} - \mu_h^{\tilde{\boldsymbol{\rho}}}\|_1 &= \sum_{s\in S} \left| \sum_{a\in A} \big(\rho_h(s,a) - \tilde{\rho}_h(s,a)\big) \right| \\ &\le \sum_{s\in S} \sum_{a\in A} |\rho_h(s,a) - \tilde{\rho}_h(s,a)| \\ &= \|\rho_h - \tilde{\rho}_h\|_1 \\ &\le |S|\,|A|\,\|\rho_h - \tilde{\rho}_h\|_2.\end{aligned}$$

Therefore

$$\sum_{s\in S}\sum_{a\in A}|F_{h,s,a}(\boldsymbol{\rho})-F_{h,s,a}(\tilde{\boldsymbol{\rho}})|^2\le|S|^2\,|A|^2\,L_r^2\|\rho_h-\tilde{\rho}_h\|_2^2.$$

Summing over $h=0,\dots,H-1$, we obtain

$$\|F(\boldsymbol{\rho})-F(\tilde{\boldsymbol{\rho}})\|_2^2\le|S|^2|A|^2L_r^2\|\boldsymbol{\rho}-\tilde{\boldsymbol{\rho}}\|_2^2.$$

Taking square roots yields

$$\|F(\boldsymbol{\rho})-F(\tilde{\boldsymbol{\rho}})\|_2\le|S||A|\,L_r\,\|\boldsymbol{\rho}-\tilde{\boldsymbol{\rho}}\|_2.$$

Thus, F is a Lipschitz continuous functional, as desired. □

Lemma 12. (5) *and monotone inclusion are equivalent.*

Proof. Recall that for a closed convex set $\mathcal{K}$, the normal cone at $\boldsymbol{\rho}^\star\in\mathcal{K}$ is

$$N_{\mathcal{K}}(\boldsymbol{\rho}^\star):=\left\{\boldsymbol{g}\in\mathbb{R}^{H|S||A|}:\langle\boldsymbol{g},\boldsymbol{\rho}-\boldsymbol{\rho}^\star\rangle\le0\quad\forall\boldsymbol{\rho}\in\mathcal{K}\right\}.$$

Thus,

$$\langle F(\boldsymbol{\rho}^\star),\boldsymbol{\rho}-\boldsymbol{\rho}^\star\rangle\ge0\quad\forall\boldsymbol{\rho}\in\mathcal{K}$$

is equivalent to

$$-F(\boldsymbol{\rho}^\star)\in N_{\mathcal{K}}(\boldsymbol{\rho}^\star),$$

which is equivalent to

$$\mathbf{0}\in F(\boldsymbol{\rho}^\star)+N_{\mathcal{K}}(\boldsymbol{\rho}^\star).$$

This proves (7) and completes the proof. □

Corollary 3. *Under the assumptions of the theorem, the finite-horizon MFE problem can be solved in the occupation variable by algorithms for monotone inclusions. In particular, since F is monotone and Lipschitz and $N_{\mathcal{K}}$ is maximal monotone, the proximal anchored gradient method applies to*

$$\mathbf{0}\in F(\boldsymbol{\rho})+N_{\mathcal{K}}(\boldsymbol{\rho}).$$

B Further Details of Section 4

Lemma 13. *The graph of $N_{\mathcal{K}}$ is closed.*

Proof. By Lemma 1, $\mathcal{K}$ is compact and convex. It follows that the graph of $N_{\mathcal{K}}$ is closed by [Rockafellar, 1970, Theorem 24.4] and Remark 3. □

In what follows, we focus on the proof of Lemma 4.

Proof of Lemma 4. It remains to verify that every accumulation point of the sequence $(\boldsymbol{\rho}^{\boldsymbol{T}})_T$ induces an MFE. The only missing detail from Section 4 is that any accumulation point $\boldsymbol{\rho}^\star$ of $(\boldsymbol{\rho}^{\boldsymbol{T}})_T$ satisfies

$$r^{\tan}_{F,N_{\mathcal{K}}}(\boldsymbol{\rho}^\star)=0.$$

Let $(\boldsymbol{\rho}^{\boldsymbol{T_n}})_n$ be a convergent subsequence such that

$$\boldsymbol{\rho}^{\boldsymbol{T_n}}\to\boldsymbol{\rho}^\star.$$

Since $\mathcal{K}$ is compact, it is closed. Hence $\boldsymbol{\rho}^\star\in\mathcal{K}$.

By Theorem 4 we have

$$r^{\tan}_{F,N_{\mathcal{K}}}(\boldsymbol{\rho}^{\boldsymbol{T_n}})\to0.$$

By the definition of the tangent residual, for each sufficiently large n there exists

$$\boldsymbol{c_n}\in N_{\mathcal{K}}(\boldsymbol{\rho}^{\boldsymbol{T_n}})$$

such that

$$\|F(\boldsymbol{\rho}^{\boldsymbol{T_n}})+\boldsymbol{c_n}\|_2\le r^{\tan}_{F,N_{\mathcal{K}}}(\boldsymbol{\rho}^{\boldsymbol{T_n}})+\frac{1}{n}.$$

Therefore,

$$F(\boldsymbol{\rho}^{\boldsymbol{T_n}}) + \boldsymbol{c_n} \to \mathbf{0}.$$

Equivalently,

$$\boldsymbol{c_n} = -F(\boldsymbol{\rho}^{\boldsymbol{T_n}}) + o(1).$$

Since F is continuous and $\boldsymbol{\rho}^{\boldsymbol{T_n}} \to \boldsymbol{\rho}^{\boldsymbol{*}}$, the sequence $(F(\boldsymbol{\rho}^{\boldsymbol{T_n}}))_n$ is bounded. Hence $(\boldsymbol{c_n})_n$ is also bounded. Passing to a further subsequence if necessary, we may assume that

$$\boldsymbol{c_n} \to \boldsymbol{c^*}$$

for some vector $\boldsymbol{c^*}$.

The normal cone mapping $N_{\mathcal{K}}$ has a closed graph because $\mathcal{K}$ is closed and convex. Since

$$\boldsymbol{\rho}^{\boldsymbol{T_n}} \to \boldsymbol{\rho^*}, \qquad \boldsymbol{c_n} \to \boldsymbol{c^*}, \qquad \boldsymbol{c_n} \in N_{\mathcal{K}}(\boldsymbol{\rho}^{\boldsymbol{T_n}}),$$

we obtain

$$\boldsymbol{c^*} \in N_{\mathcal{K}}(\boldsymbol{\rho}^*).$$

Moreover, by continuity of F,

$$F(\boldsymbol{\rho}^{\boldsymbol{T_n}}) \to F(\boldsymbol{\rho^*}).$$

Taking limits in

$$F(\boldsymbol{\rho}^{\boldsymbol{T_n}}) + \boldsymbol{c_n} \to \mathbf{0}$$

gives

$$F(\boldsymbol{\rho^*}) + \boldsymbol{c^*} = \mathbf{0}.$$

Since $\boldsymbol{c^*} \in N_{\mathcal{K}}(\boldsymbol{\rho}^*)$, it follows that

$$\mathbf{0} \in F(\boldsymbol{\rho^*}) + N_{\mathcal{K}}(\boldsymbol{\rho^*}).$$

Therefore,

$$r^{\tan}_{F,N_{\mathcal{K}}}(\boldsymbol{\rho^*}) = \inf_{\boldsymbol{c} \in N_{\mathcal{K}}(\boldsymbol{\rho^*})} \|F(\boldsymbol{\rho}^*) + \boldsymbol{c}\|_2 = 0,$$

as desired. □

Next, we show that $\boldsymbol{\rho}^{\boldsymbol{T}} \to \Pi_S(\boldsymbol{\rho^0})$ under the Euclidean norm. First, we demonstrate that occupancy measures that induce an MFE form a convex and compact set.

Lemma 14. *The following statements hold:*

- *The set of occupation flows that induce an MFE,*

$$\mathcal{S} := \{\boldsymbol{\rho} \in \mathcal{K} : \langle F(\boldsymbol{\rho}), \boldsymbol{z} - \boldsymbol{\rho}\rangle \geq 0 \quad \forall \boldsymbol{z} \in \mathcal{K}\},$$

  is closed and convex.

- *For every anchor term $\boldsymbol{\rho^0} \in \mathcal{K}$, there exists a unique orthogonal projection $\Pi_{\mathcal{S}}(\boldsymbol{\rho^0})$.*

Proof. Note that by Corollary 1, the occupancy measures that induce an MFE belong to $\mathcal{S}$. Furthermore, every element of $\mathcal{S}$ induces an MFE too.

To prove that $\mathcal{S}$ is closed and convex, we provide an equivalent formulation for $\mathcal{S}$ using monotonicity. We claim that $\mathcal{S}$ is equivalent to

$$\overline{\mathcal{S}} = \{\boldsymbol{\rho} \in \mathcal{K} : \langle F(\boldsymbol{z}), \boldsymbol{z} - \boldsymbol{\rho}\rangle \geq 0 \quad \forall \boldsymbol{z} \in \mathcal{K}\}. \tag{13}$$

First, suppose that $\boldsymbol{\rho} \in \mathcal{S}$. Then, for every $\boldsymbol{z} \in \mathcal{K}$,

$$\langle F(\boldsymbol{\rho}), \boldsymbol{z} - \boldsymbol{\rho}\rangle \geq 0.$$

By monotonicity of F,

$$\langle F(\boldsymbol{z}) - F(\boldsymbol{\rho}), \boldsymbol{z} - \boldsymbol{\rho}\rangle \geq 0.$$

Adding these inequalities gives

$$\langle F(\boldsymbol{z}), \boldsymbol{z} - \boldsymbol{\rho}\rangle \geq 0 \qquad \forall \boldsymbol{z} \in \mathcal{K}.$$

Conversely, suppose that $\boldsymbol{\rho} \in \mathcal{K}$ satisfies

$$\langle F(\boldsymbol{z}), \boldsymbol{z} - \boldsymbol{\rho} \rangle \geq 0 \qquad \forall \boldsymbol{z} \in \mathcal{K}.$$

Fix an arbitrary $\boldsymbol{y} \in \mathcal{K}$, and for $t \in (0, 1]$, define

$$\boldsymbol{z}_t := (1 - t)\boldsymbol{\rho} + t\boldsymbol{y}.$$

Since $\mathcal{K}$ is convex, $\boldsymbol{z}_t \in \mathcal{K}$. Therefore,

$$\langle F(\boldsymbol{z}_t), \boldsymbol{z}_t - \boldsymbol{\rho} \rangle \geq 0.$$

Since

$$\boldsymbol{z}_t - \boldsymbol{\rho} = t(\boldsymbol{y} - \boldsymbol{\rho}),$$

we obtain

$$\langle F(\boldsymbol{z}_t), \boldsymbol{y} - \boldsymbol{\rho} \rangle \geq 0.$$

Letting $t \downarrow 0$, continuity of F yields

$$\langle F(\boldsymbol{\rho}), \boldsymbol{y} - \boldsymbol{\rho} \rangle \geq 0.$$

Since $\boldsymbol{y} \in \mathcal{K}$ was arbitrary, it follows that $\boldsymbol{\rho} \in \mathcal{S}$. This proves (13).

For each fixed $\boldsymbol{z} \in \mathcal{K}$, define

$$H_{\boldsymbol{z}} := \left\{ \boldsymbol{\rho} \in \mathbb{R}^{H|S||A|} : \langle F(\boldsymbol{z}), \boldsymbol{z} - \boldsymbol{\rho} \rangle \geq 0 \right\}.$$

The set $H_{\boldsymbol{z}}$ is a closed half-space in the variable $\boldsymbol{\rho}$, and is therefore closed and convex. By (13),

$$\mathcal{S} = \mathcal{K} \cap \bigcap_{\boldsymbol{z} \in \mathcal{K}} H_{\boldsymbol{z}}.$$

Since $\mathcal{K}$ is closed and convex, and arbitrary intersections of closed convex sets are closed and convex, it follows that $\mathcal{S}$ is closed and convex.

Now, under monotonicity and Lipschitz continuity, we have $\mathcal{S} \neq \varnothing$. Since $\mathcal{S}$ is closed and is contained in the compact set $\mathcal{K}$, it is compact. Thus, existence and uniqueness of $\Pi_{\mathcal{S}}(\boldsymbol{\rho^0})$ follows from the Hilbert projection theorem. □

Theorem 6. *Let $(\boldsymbol{\rho^k})_{k \geq 0}$ be generated by*

$$\boldsymbol{\rho^{k+1}} = \Pi_{\mathcal{K}} \left((1 - \beta_k)\boldsymbol{\rho^k} + \beta_k \boldsymbol{\rho^0} - \alpha_k F(\boldsymbol{\rho^k}) \right),$$

where

$$\alpha_k = \frac{1}{L\sqrt{k + \gamma}}, \qquad \beta_k = \frac{\gamma}{k + \gamma}, \qquad \gamma \geq 2.$$

Then

$$\boldsymbol{\rho^k} \longrightarrow \Pi_{\mathcal{S}}(\boldsymbol{\rho^0}).$$

Consequently, the sequence has a unique accumulation point.

Proof. By [Cai and Zheng, 2026, Theorem 2], we have

$$r^{\tan}_{F, N_{\mathcal{K}}}(\boldsymbol{\rho^k}) \longrightarrow 0.$$

We also demonstrated that every accumulation point of $(\boldsymbol{\rho^k})_{k \geq 0}$ therefore belongs to $\mathcal{S}$. By Lemma 14, the set $\mathcal{S}$ is nonempty, closed, and convex. Hence $\boldsymbol{p} := \Pi_{\mathcal{S}}(\boldsymbol{\rho^0})$ exists and is unique. We show that $\boldsymbol{\rho^k} \to \boldsymbol{p}$.

Since $\boldsymbol{p} \in \mathcal{S}$, we have $-F(\boldsymbol{p}) \in N_{\mathcal{K}}(\boldsymbol{p})$, and hence, for every k, $\boldsymbol{p} = \Pi_{\mathcal{K}}\big(\boldsymbol{p} - \alpha_k F(\boldsymbol{p})\big)$. Using this identity, the nonexpansiveness of $\Pi_{\mathcal{K}}$, and the same monotonicity–Lipschitz estimate used in the proof of [Cai and Zheng, 2026, Lemma 1], we show that

$$\|\boldsymbol{\rho^{k+1}} - \boldsymbol{p}\|_2^2 \leq (1 - \theta_k)\|\boldsymbol{\rho^k} - \boldsymbol{p}\|_2^2 + 2\beta_k(1 - \beta_k) \left\langle \boldsymbol{\rho^0} - \boldsymbol{p}, \boldsymbol{\rho^k} - \boldsymbol{p} \right\rangle + \varepsilon_k,$$

where

$$\theta_k := 2\beta_k - \beta_k^2 - \alpha_k^2 L^2$$

and

$$0 \le \varepsilon_k \le O\big(\beta_k^2 + \alpha_k\beta_k\big).$$

Indeed, since $\boldsymbol{p} \in \mathcal{S}$, we have $\mathbf{0} \in F(\boldsymbol{p}) + N_{\mathcal{K}}(\boldsymbol{p})$, and therefore $-F(\boldsymbol{p}) \in N_{\mathcal{K}}(\boldsymbol{p})$. Since $N_{\mathcal{K}}(\boldsymbol{p})$ is a cone, it follows that $-\alpha_k F(\boldsymbol{p}) \in N_{\mathcal{K}}(\boldsymbol{p})$. By the characterization of the Euclidean projection, $\boldsymbol{p} = \Pi_{\mathcal{K}}\big(\boldsymbol{p} - \alpha_k F(\boldsymbol{p})\big)$.

Using the nonexpansiveness of $\Pi_{\mathcal{K}}$, we obtain

$$\begin{aligned}
\|\boldsymbol{\rho^{k+1}} - \boldsymbol{p}\|_2^2 &\le \Big\|(1-\beta_k)(\boldsymbol{\rho^k} - \boldsymbol{p}) + \beta_k(\boldsymbol{\rho^0} - \boldsymbol{p}) - \alpha_k(F(\boldsymbol{\rho^k}) - F(\boldsymbol{p}))\Big\|_2^2 \\
&= (1-\beta_k)^2\|\boldsymbol{\rho^k} - \boldsymbol{p}\|_2^2 + \beta_k^2\|\boldsymbol{\rho^0} - \boldsymbol{p}\|_2^2 + \alpha_k^2\|F(\boldsymbol{\rho^k}) - F(\boldsymbol{p})\|_2^2 \\
&\quad + 2\beta_k(1-\beta_k)\langle \boldsymbol{\rho^0} - \boldsymbol{p}, \boldsymbol{\rho^k} - \boldsymbol{p}\rangle \\
&\quad - 2\alpha_k(1-\beta_k)\langle (F(\boldsymbol{\rho^k}) - F(\boldsymbol{p})), \boldsymbol{\rho^k} - \boldsymbol{p}\rangle - 2\alpha_k\beta_k\langle (F(\boldsymbol{\rho^k}) - F(\boldsymbol{p})), \boldsymbol{\rho^0} - \boldsymbol{p}\rangle.
\end{aligned}$$

By monotonicity of F,

$$\langle (F(\boldsymbol{\rho^k}) - F(\boldsymbol{p})), (\boldsymbol{\rho^k} - \boldsymbol{p})\rangle = \Big\langle F(\boldsymbol{\rho^k}) - F(\boldsymbol{p}), \boldsymbol{\rho^k} - \boldsymbol{p}\Big\rangle \ge 0 \text{ and hence } -2\alpha_k(1-\beta_k)\langle (F(\boldsymbol{\rho^k}) - F(\boldsymbol{p})), (\boldsymbol{\rho^k} - \boldsymbol{p})\rangle \le 0.$$

Moreover, since F is L-Lipschitz, $\|(F(\boldsymbol{\rho^k}) - F(\boldsymbol{p}))\|_2 \le L\|(\boldsymbol{\rho^k} - \boldsymbol{p})\|_2$. Since $\mathcal{K}$ is compact and $\boldsymbol{\rho^k}, \boldsymbol{p} \in \mathcal{K}$, there exists $D < \infty$ such that

$$\|(\boldsymbol{\rho^k} - \boldsymbol{p})\|_2 \le D \qquad \forall k \ge 0.$$

Therefore,

$$\begin{aligned}
-2\alpha_k\beta_k\langle (F(\boldsymbol{\rho^k}) - F(\boldsymbol{p})), (\boldsymbol{\rho^0} - \boldsymbol{p})\rangle &\le 2\alpha_k\beta_k\|(F(\boldsymbol{\rho^k}) - F(\boldsymbol{p}))\|_2\|(\boldsymbol{\rho^0} - \boldsymbol{p})\|_2 \\
&\le 2\alpha_k\beta_k LD\|(\boldsymbol{\rho^0} - \boldsymbol{p})\|_2.
\end{aligned}$$

Combining the preceding estimates gives

$$\|\boldsymbol{\rho^{k+1}} - \boldsymbol{p}\|_2^2 \le \big((1-\beta_k)^2 + \alpha_k^2 L^2\big)\,\|\boldsymbol{\rho^k} - \boldsymbol{p}\|_2^2 + 2\beta_k(1-\beta_k)\Big\langle \boldsymbol{\rho^0} - \boldsymbol{p}, \boldsymbol{\rho^k} - \boldsymbol{p}\Big\rangle + \varepsilon_k, \tag{14}$$

where

$$\varepsilon_k := \beta_k^2\|\boldsymbol{\rho^0} - \boldsymbol{p}\|_2^2 + 2\alpha_k\beta_k LD\|\boldsymbol{\rho^0} - \boldsymbol{p}\|_2.$$

Finally, defining

$$\theta_k := 1 - \big((1-\beta_k)^2 + \alpha_k^2 L^2\big) = 2\beta_k - \beta_k^2 - \alpha_k^2 L^2,$$

we may rewrite (14) as

$$\|\boldsymbol{\rho^{k+1}} - \boldsymbol{p}\|_2^2 \le (1-\theta_k)\|\boldsymbol{\rho^k} - \boldsymbol{p}\|_2^2 + 2\beta_k(1-\beta_k)\Big\langle \boldsymbol{\rho^0} - \boldsymbol{p}, \boldsymbol{\rho^k} - \boldsymbol{p}\Big\rangle + \varepsilon_k. \tag{15}$$

For completeness, expanding

$$\Big\|(1-\beta_k)(\boldsymbol{\rho^k} - \boldsymbol{p}) + \beta_k(\boldsymbol{\rho^0} - \boldsymbol{p}) - \alpha_k(F(\boldsymbol{\rho^k}) - F(\boldsymbol{p}))\Big\|_2^2.$$

The term

$$-2\alpha_k(1-\beta_k)\langle (F(\boldsymbol{\rho^k}) - F(\boldsymbol{p})), (\boldsymbol{\rho^k} - \boldsymbol{p})\rangle$$

is nonpositive by monotonicity, while Lipschitz continuity and the compactness of $\mathcal{K}$ bound all remaining error terms by $O(\beta_k^2 + \alpha_k\beta_k)$.

For the prescribed step sizes,

$$\theta_k = \frac{2\gamma - 1}{k+\gamma} - \frac{\gamma^2}{(k+\gamma)^2}.$$

Hence

$$0 < \theta_k < 1, \qquad \sum_{k=0}^{\infty} \theta_k = \infty,$$

and

$$\frac{\varepsilon_k}{\theta_k} \longrightarrow 0, \qquad \frac{2\beta_k(1-\beta_k)}{\theta_k} \longrightarrow \frac{2\gamma}{2\gamma - 1}.$$

We next claim that

$$\limsup_{k\to\infty} \left\langle \boldsymbol{\rho^0} - \boldsymbol{p}, \boldsymbol{\rho^k} - \boldsymbol{p} \right\rangle \le 0.$$

Indeed, choose a subsequence attaining this limsup. Since $\mathcal{K}$ is compact, it has a further subsequence of $\boldsymbol{\rho^k}$ converging to some accumulation point $\bar{\boldsymbol{\rho}}$. As shown above, $\bar{\boldsymbol{\rho}} \in \mathcal{S}$. The characterization of the projection

$$\boldsymbol{p} = \Pi_{\mathcal{S}}(\boldsymbol{\rho^0})$$

then gives

$$\left\langle \boldsymbol{\rho^0} - \boldsymbol{p}, \bar{\boldsymbol{\rho}} - \boldsymbol{p} \right\rangle \le 0,$$

which proves the claim.

The preceding estimates allow (14) to be written as

$$\|\boldsymbol{\rho^{k+1}} - \boldsymbol{p}\|_2^2 \le (1-\theta_k)\|\boldsymbol{\rho^k} - \boldsymbol{p}\|_2^2 + \theta_k b_k,$$

where $\limsup_{k\to\infty} b_k \le 0$. Since $a_k \ge 0$, $0 < \theta_k < 1$, and $\sum_k \theta_k = \infty$.

The preceding estimates allow (14) to be written as

$$a_{k+1} \le (1-\theta_k)a_k + \theta_k b_k, \qquad a_k := \|\boldsymbol{\rho^k} - \boldsymbol{p}\|_2^2,$$

where

$$\limsup_{k\to\infty} b_k \le 0.$$

We now show directly that $a_k \to 0$. Fix $\varepsilon > 0$. Since $\limsup_{k\to\infty} b_k \le 0$, there exists K_ε such that

$$b_k \le \varepsilon \qquad \text{for every } k \ge K_\varepsilon.$$

Consequently,

$$a_{k+1} - \varepsilon \le (1-\theta_k)(a_k - \varepsilon), \qquad k \ge K_\varepsilon.$$

Taking positive parts gives

$$(a_{k+1} - \varepsilon)_+ \le (1-\theta_k)(a_k - \varepsilon)_+.$$

Therefore, for every $n \ge K_\varepsilon$,

$$(a_{n+1} - \varepsilon)_+ \le (a_{K_\varepsilon} - \varepsilon)_+ \prod_{k=K_\varepsilon}^{n} (1-\theta_k).$$

Since

$$\prod_{k=K_\varepsilon}^{n} (1-\theta_k) \le \exp\left(- \sum_{k=K_\varepsilon}^{n} \theta_k \right) \longrightarrow 0,$$

we obtain

$$\limsup_{k\to\infty} a_k \le \varepsilon.$$

As $\varepsilon > 0$ is arbitrary and $a_k \ge 0$, it follows that

$$a_k \longrightarrow 0.$$

Hence,

$$\boldsymbol{\rho^k} \longrightarrow \boldsymbol{p} = \Pi_{\mathcal{S}}(\boldsymbol{\rho^0}).$$

Hence every subsequence converges to the same point $\boldsymbol{p}$, and $\boldsymbol{p}$ is the unique accumulation point.

□

C Proof of Proposition 1

In this section we will provide a proof of Proposition 1. Proposition 1 follows directly from the following proposition.

Proposition 2 (A matching $\Omega(T^{-1/2})$ lower-bound example)**.** *Consider the finite-horizon mean-field game with horizon $H = 2$, state space*

$$S = \{0, 1, 2\},$$

action space

$$A = \{L, R\},$$

and initial state distribution

$$\mu_0 = \delta_0.$$

Let the transition kernel be independent of the state-measure and defined by

$$P_0(s' \mid s, a) = \begin{cases} 1, & s = 0,\ a = L,\ s' = 1, \\ 1, & s = 0,\ a = R,\ s' = 2, \\ 1, & s \in \{1, 2\},\ s' = s, \\ 0, & \textit{otherwise}. \end{cases}$$

Let the rewards be

$$r(0, \cdot, \mu) \equiv 0,$$

and

$$r_1(1, \cdot, \mu) \equiv -\mu(1), \qquad r_1(2, \cdot, \mu) \equiv -\mu(2).$$

Then, the following hold:

1. *The game satisfies Assumptions* (1)–(3)*.*
2. *Let $\mathcal{K}$ and F be the occupation-flow polytope and the mean-field occupation operator associated with this MFG. Let $\boldsymbol{\rho^0} \in \mathcal{K}$ be the occupation flow defined by*

$$\rho_0^0(0, L) = 1, \qquad \rho_0^0(0, R) = 0, \qquad \rho_1^0(1, L) = \rho_1^0(1, R) = \frac{1}{2},$$
$$\rho_1^0(2, L) = \rho_1^0(2, R) = 0,$$

   with all other coordinates equal to 0*.*

3. *Consider the anchored iteration*

$$\boldsymbol{\rho^{k+1}} = \Pi_{\mathcal{K}}\Big((1 - \beta_k)\boldsymbol{\rho^k} + \beta_k \boldsymbol{\rho^0} - \alpha_k F(\boldsymbol{\rho^k})\Big),$$

   with

$$\alpha_k = \frac{1}{6\sqrt{k+2}}, \qquad \beta_k = \frac{2}{k+2}.$$

   Then, for every $k \geq 1$,

$$r^{\tan}_{F, N_{\mathcal{K}}}(\boldsymbol{\rho^k}) \geq \frac{1}{\sqrt{3(k+2)}}.$$

   In particular,

$$r^{\tan}_{F, N_{\mathcal{K}}}(\boldsymbol{\rho^k}) = \Omega(k^{-1/2}).$$

Proof. We divide the proof into Steps.

By the definition of the mean-field occupation operator,

$$F_h(\boldsymbol{\rho})(s, a) = -r\big(s, a, \mu_h^{\boldsymbol{\rho}}\big), \qquad h \in \{0, 1\}.$$

We first verify Assumptions 1–3. Let $\boldsymbol{\rho}, \boldsymbol{\eta} \in \mathcal{K}$. Since

$$r(0, a, \mu) = 0, \qquad r(s, a, \mu) = -\mu(s), \quad s \in \{1, 2\},$$

we have

$$\begin{aligned}
&\sum_{h=0}^{1}\sum_{s\in S}\sum_{a\in A}\left(r\big(s,a,\mu_h^{\boldsymbol{\rho}}\big)-r\big(s,a,\mu_h^{\boldsymbol{\eta}}\big)\right)\left(\rho_h(s,a)-\eta_h(s,a)\right)\\
&=-\sum_{h=0}^{1}\sum_{s=1}^{2}\left(\mu_h^{\boldsymbol{\rho}}(s)-\mu_h^{\boldsymbol{\eta}}(s)\right)\sum_{a\in A}\left(\rho_h(s,a)-\eta_h(s,a)\right)\\
&=-\sum_{h=0}^{1}\sum_{s=1}^{2}\left(\mu_h^{\boldsymbol{\rho}}(s)-\mu_h^{\boldsymbol{\eta}}(s)\right)^2\le 0.
\end{aligned}$$

Thus, the reward function is weakly monotone. Notice that the time-0 contribution is zero because

$$\mu_0^{\boldsymbol{\rho}}=\mu_0^{\boldsymbol{\eta}}=\delta_0.$$

Moreover, for every $s\in\{1,2\}$ and $a\in A$,

$$|r(s,a,\mu)-r(s,a,\nu)|=|\mu(s)-\nu(s)|\le\|\mu-\nu\|_2,$$

whereas

$$|r(0,a,\mu)-r(0,a,\nu)|=0.$$

Hence, Assumption 3 holds with Lipschitz constant 1.

We next study the anchored iterations. After removing coordinates that are identically zero for every element of $\mathcal{K}$, we use the coordinate ordering

$$(\rho_0(0,L),\rho_0(0,R),\rho_1(1,L),\rho_1(1,R),\rho_1(2,L),\rho_1(2,R))\,.$$

In these coordinates,

$$\mathcal{K}=\left\{(x,1-x,u,v,w,z):\ \begin{array}{l}0\le x\le 1,\\ u,v,w,z\ge 0,\\ u+v=x,\\ w+z=1-x\end{array}\right\}.$$

For $p\in[0,1]$, define

$$\boldsymbol{\rho}(p):=\left(p,1-p,\frac{p}{2},\frac{p}{2},\frac{1-p}{2},\frac{1-p}{2}\right).$$

The occupation measure specified in the statement satisfies

$$\boldsymbol{\rho^0}=\boldsymbol{\rho}(1).$$

The state distribution induced at time 1 by $\boldsymbol{\rho}(p)$ is

$$\mu_1^{\boldsymbol{\rho}(p)}(1)=p,\qquad \mu_1^{\boldsymbol{\rho}(p)}(2)=1-p.$$

Since the reward function is time homogeneous and $\mu_0=\delta_0$, we have

$$F_0\big(\boldsymbol{\rho}(p)\big)(0,L)=F_0\big(\boldsymbol{\rho}(p)\big)(0,R)=0.$$

At time 1,

$$F_1\big(\boldsymbol{\rho}(p)\big)(1,a)=p,\qquad F_1\big(\boldsymbol{\rho}(p)\big)(2,a)=1-p,\qquad a\in A.$$

Consequently,

$$F\big(\boldsymbol{\rho}(p)\big)=(0,0,p,p,1-p,1-p).$$

We now show that the iterates remain in the one-dimensional set $\mathcal{M}:=\{\boldsymbol{\rho}(p):p\in[0,1]\}$. Suppose that $\boldsymbol{\rho^k}=\boldsymbol{\rho}(p_k)$ for some $p_k\in[0,1]$, and define

$$a_k:=(1-\beta_k)p_k+\beta_k.$$

Then

$$\begin{aligned}
\boldsymbol{y}^k&:=(1-\beta_k)\boldsymbol{\rho^k}+\beta_k\boldsymbol{\rho^0}-\alpha_kF(\boldsymbol{\rho^k})\\
&=(a_k,1-a_k,c_k,c_k,d_k,d_k)\,,
\end{aligned}$$

where

$$c_k = \frac{a_k}{2} - \alpha_k p_k$$

and

$$d_k = \frac{1 - a_k}{2} - \alpha_k (1 - p_k).$$

For a fixed $x \in [0, 1]$, the minimizers of

$$(u - c_k)^2 + (v - c_k)^2$$

subject to $u + v = x$ are

$$u = v = \frac{x}{2}.$$

Similarly, the minimizers of

$$(w - d_k)^2 + (z - d_k)^2$$

subject to $w + z = 1 - x$ are

$$w = z = \frac{1 - x}{2}.$$

Therefore, the projection of $\boldsymbol{y}^k$ onto $\mathcal{K}$ is of the form $\boldsymbol{\rho}(x)$, where x minimizes

$$\varphi_k(x) = 2(x - a_k)^2 + 2\left(\frac{x}{2} - c_k\right)^2 + 2\left(\frac{1 - x}{2} - d_k\right)^2$$

over $x \in [0, 1]$.

Differentiating gives

$$\varphi_k'(x) = 6x - 6a_k + 4\alpha_k p_k - 2\alpha_k.$$

Thus, the unconstrained minimizer is

$$x = a_k - \frac{2}{3}\alpha_k p_k + \frac{\alpha_k}{3}.$$

It follows that

$$p_{k+1} = \left(1 - \beta_k - \frac{2}{3}\alpha_k\right) p_k + \beta_k + \frac{\alpha_k}{3}. \tag{16}$$

We next verify that the unconstrained minimizer belongs to $(0, 1)$. For $k = 0$, since $\beta_0 = 1$ and $p_0 = 1$,

$$p_1 = 1 - \frac{\alpha_0}{3} = 1 - \frac{1}{18\sqrt{2}} \in \left(\frac{1}{2}, 1\right).$$

For every $k \geq 1$,

$$1 - \beta_k - \frac{2}{3}\alpha_k = 1 - \frac{2}{k+2} - \frac{1}{9\sqrt{k+2}} > 0.$$

Suppose that $p_k \in (1/2, 1)$. Then

$$\begin{aligned} p_{k+1} &> \frac{1}{2}\left(1 - \beta_k - \frac{2}{3}\alpha_k\right) + \beta_k + \frac{\alpha_k}{3} \\ &= \frac{1}{2} + \frac{\beta_k}{2} > \frac{1}{2}. \end{aligned}$$

Moreover,

$$\begin{aligned} p_{k+1} &< 1 - \beta_k - \frac{2}{3}\alpha_k + \beta_k + \frac{\alpha_k}{3} \\ &= 1 - \frac{\alpha_k}{3} < 1. \end{aligned}$$

Therefore,

$$p_k \in \left(\frac{1}{2}, 1\right) \qquad \text{for every } k \geq 1.$$

In particular, the projection does not reach the boundary of the interval, and (16) holds for every $k \geq 0$.

Define

$$e_k := p_k - \frac{1}{2}.$$

Subtracting $1/2$ from both sides of (16) gives

$$e_{k+1} = \left(1 - \beta_k - \frac{2}{3}\alpha_k\right) e_k + \frac{\beta_k}{2}.$$

Using

$$\alpha_k = \frac{1}{6\sqrt{k+2}}, \qquad \beta_k = \frac{2}{k+2},$$

we obtain

$$e_{k+1} = \left(1 - \frac{2}{k+2} - \frac{1}{9\sqrt{k+2}}\right) e_k + \frac{1}{k+2}. \tag{17}$$

We claim that

$$e_k \geq \frac{1}{2\sqrt{k+2}} \qquad \text{for every } k \geq 1. \tag{18}$$

Indeed, for $k = 1$,

$$e_1 = \frac{1}{2} - \frac{1}{18\sqrt{2}} > \frac{1}{2\sqrt{3}}.$$

Suppose that (18) holds for some $k \geq 1$, and set

$$n := k + 2.$$

Since $n \geq 3$,

$$1 - \frac{2}{n} - \frac{1}{9\sqrt{n}} > 0.$$

Hence, by (17),

$$\begin{aligned} e_{k+1} &\geq \left(1 - \frac{2}{n} - \frac{1}{9\sqrt{n}}\right) \frac{1}{2\sqrt{n}} + \frac{1}{n} \\ &= \frac{1}{2\sqrt{n}} - \frac{1}{n\sqrt{n}} + \frac{17}{18n}. \end{aligned}$$

Furthermore,

$$\begin{aligned} &\frac{1}{2\sqrt{n}} - \frac{1}{n\sqrt{n}} + \frac{17}{18n} - \frac{1}{2\sqrt{n+1}} \\ &= \left(\frac{1}{2\sqrt{n}} - \frac{1}{2\sqrt{n+1}}\right) + \frac{1}{n}\left(\frac{17}{18} - \frac{1}{\sqrt{n}}\right) \geq 0. \end{aligned}$$

Therefore,

$$e_{k+1} \geq \frac{1}{2\sqrt{n+1}} = \frac{1}{2\sqrt{k+3}}.$$

This proves (18) by induction.

It remains to compute the tangent residual. Fix $p \in (0,1)$. Since all coordinates of $\boldsymbol{\rho}(p)$ that are not identically zero on $\mathcal{K}$ are strictly positive,

$$\boldsymbol{\rho}(p) \in \mathrm{ri}(\mathcal{K}),$$

where $\mathrm{ri}(\mathcal{K})$ denotes the relative interior. The tangent space to $\mathcal{K}$ at $\boldsymbol{\rho}(p)$ is

$$T_{\mathcal{K}} = \{(\xi, -\xi, \zeta_1, \zeta_2, \zeta_3, \zeta_4) : \zeta_1 + \zeta_2 = \xi,\ \zeta_3 + \zeta_4 = -\xi\}.$$

An orthogonal basis of $T_{\mathcal{K}}$ is

$$\boldsymbol{t} = \left(1, -1, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}\right),$$

together with

$$\boldsymbol{d}_1 = (0, 0, 1, -1, 0, 0) \text{ and } \boldsymbol{d}_2 = (0, 0, 0, 0, 1, -1).$$

Since

$$F\big(\boldsymbol{\rho}(p)\big) = (0, 0, p, p, 1-p, 1-p),$$

we have

$$\big\langle F\big(\boldsymbol{\rho}(p)\big), \boldsymbol{d}_1 \big\rangle = \big\langle F\big(\boldsymbol{\rho}(p)\big), \boldsymbol{d}_2 \big\rangle = 0.$$

Moreover,

$$\big\langle F\big(\boldsymbol{\rho}(p)\big), \boldsymbol{t} \big\rangle = 2p - 1$$

and

$$\|\boldsymbol{t}\|_2^2 = 3.$$

Consequently,

$$\big\|\Pi_{T_{\mathcal{K}}} F\big(\boldsymbol{\rho}(p)\big)\big\|_2 = \frac{|2p-1|}{\sqrt{3}}.$$

Since

$$N_{\mathcal{K}}\big(\boldsymbol{\rho}(p)\big) = T_{\mathcal{K}}^{\perp},$$

the tangent residual satisfies

$$\begin{aligned} r^{\tan}_{F,N_{\mathcal{K}}}\big(\boldsymbol{\rho}(p)\big) &= \operatorname{dist}\big(\mathbf{0}, F\big(\boldsymbol{\rho}(p)\big) + N_{\mathcal{K}}\big(\boldsymbol{\rho}(p)\big)\big) \\ &= \big\|\Pi_{T_{\mathcal{K}}} F\big(\boldsymbol{\rho}(p)\big)\big\|_2 \\ &= \frac{|2p-1|}{\sqrt{3}}. \end{aligned}$$

Finally, for every $k \geq 1$,

$$p_k = \frac{1}{2} + e_k$$

with $e_k > 0$. Therefore,

$$\begin{aligned} r^{\tan}_{F,N_{\mathcal{K}}}(\boldsymbol{\rho^k}) &= \frac{|2p_k - 1|}{\sqrt{3}} \\ &= \frac{2e_k}{\sqrt{3}} \\ &\geq \frac{1}{\sqrt{3(k+2)}}. \end{aligned}$$

Hence,

$$r^{\tan}_{F,N_{\mathcal{K}}}(\boldsymbol{\rho^k}) = \Omega(k^{-1/2}),$$

as claimed. □

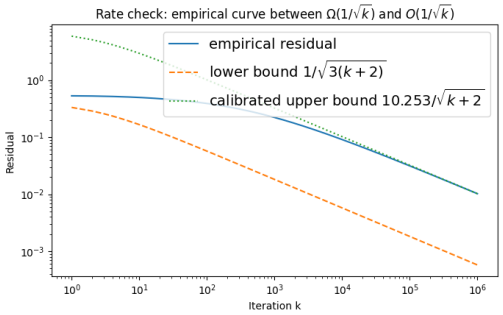


Figure 1: Empirical residual error rate for the example we use for Proposition 1.

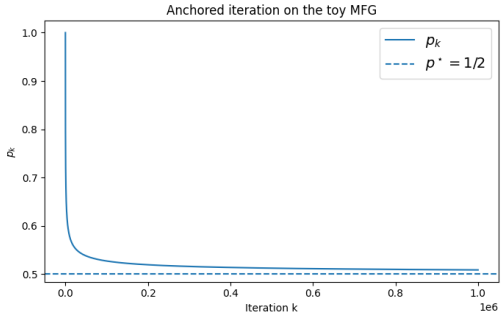


Figure 2: Comparison between the iterates produced by our algorithm for the example in Proposition 1 and the corresponding MFE.

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: Subsection 1.2 in the introduction clearly outlines the paper's contributions. We also include a table comparing our results with prior work in the literature, and the abstract accurately reflects the scope of the paper.

   Guidelines:

   - The answer [N/A] 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 [N/A] 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: When introducing the assumptions in Subsection 2.1, we briefly compare them with those used in prior literature. This discussion also highlights their limitations relative to existing work.

   Guidelines:

   - The answer [N/A] 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: We clearly state all assumptions used in the proofs in Sections 3 and 4. Where appropriate, we include proof sketches of the main results in the main text, and provide complete proofs for all claimed results in the appendix.

Guidelines:

- The answer [N/A] 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. **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: [N/A]

Justification: The paper does not include experiments requiring code.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.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://neurips.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.

5. **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) necessary to understand the results?

Answer: [N/A]

Justification: The paper does not include experiments.

Guidelines:

- The answer [N/A] 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.

6. **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: [N/A]

   Justification: The paper does not include experiments.

   Guidelines:

   - The answer [N/A] 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.

7. **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: [N/A]

   Justification: The main body of the paper does not include any numerical experiments. However, at the end of Section C, we include a graph illustrating the performance of our algorithm on the example studied in that section. The implementation of the algorithm is relatively straightforward, and the graph is intended only to show its behavior across iterations. We do not make any claims about the algorithm's runtime performance.

   Guidelines:

   - The answer [N/A] 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).

8. **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 have reviewed the NeurIPS Code of Ethics, and we confirm that our paper conforms to the NeurIPS Code of Ethics.

Guidelines:

- The answer [N/A] 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).

9. **Broader impacts**

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

Answer: [N/A]

Justification: This paper presents a method for computing mean-field equilibria and establishes theoretical guarantees for the convergence of the iterative procedure. However, these theoretical contributions do not have immediate societal implications.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] 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).

10. **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., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: No data or model is released with this paper.

Guidelines:

- The answer [N/A] 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.

11. **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: [N/A]

    Justification: This paper does not use existing assets.

    Guidelines:

    - The answer [N/A] 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.

12. **New assets**

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

    Answer: [N/A]

    Justification: This paper does not release new assets.

    Guidelines:

    - The answer [N/A] 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.

13. **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: [N/A]

    Justification: This paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] 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.

14. **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: [N/A]

    Justification: This paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] 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.

15. **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 rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components. However, we have used LLMs in the writing.

    Guidelines:

    - The answer [N/A] 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 in the NeurIPS handbook for what should or should not be described.