Mean-Field Games with Constraints

Anran Hu * Zijiu Lyu †

Abstract

This paper introduces a framework of Constrained Mean-Field Games (CMFGs), where each agent solves a constrained Markov decision process (CMDP). This formulation captures scenarios in which agents' strategies are subject to feasibility, safety, or regulatory restrictions, thereby extending the scope of classical mean field game (MFG) models. We first establish the existence of CMFG equilibria under a strict feasibility assumption, and we further show uniqueness under a classical monotonicity condition. To compute equilibria, we develop Constrained Mean-Field Occupation Measure Optimization (CMFOMO), an optimization-based scheme that parameterizes occupation measures and shows that finding CMFG equilibria is equivalent to solving a single optimization problem with convex constraints and bounded variables. CMFOMO does not rely on uniqueness of the equilibria and can approximate all equilibria with arbitrary accuracy. We further prove that CMFG equilibria induce $O(1/\sqrt{N})$-Nash equilibria in the associated constrained N-player games, thereby extending the classical justification of MFGs as approximations for large but finite systems. Numerical experiments on a modified Susceptible-Infected-Susceptible (SIS) epidemic model with various constraints illustrate the effectiveness and flexibility of the framework.

1 Introduction

In the domain of game theory, the analysis of large-scale strategic interactions poses both profound opportunities and formidable challenges. Multi-agent systems, in which numerous agents interact under shared decision-making rules, are central to fields such as economics, finance, and engineering [1, 2]. Yet the complexity of these systems grows rapidly with the number of participants, making the direct study of N-player stochastic games analytically and computationally intractable.

Introduced by the seminal works of [3] and [4], mean field games (MFGs) provide a powerful framework to address this difficulty. The key idea is to approximate large-population games by considering the limiting regime as the number of agents tends to infinity. In this regime, the influence of any single agent becomes negligible, and each agent interacts only with the collective effect of the population. This simplification transforms an otherwise intractable many-player game into a tractable model where equilibrium analysis can be carried out systematically.

MFG theory has not only offered conceptual clarity but also rigorous approximation results: equilibria of the limiting mean field game can be used to construct approximate Nash equilibria for the corresponding N-player game, with explicit error bounds of order $O(1/\sqrt{N})$ [3, 4, 5, 6]. Over the past decade, this approach has been extensively studied and extended, leading to a rich literature on existence, uniqueness, and numerical methods. Its versatility has enabled applications across domains ranging from financial markets and systemic risk to crowd dynamics, epidemiology, and large-scale social networks.

Despite these successes, classical MFG models typically assume that agents operate without binding restrictions on their states or controls. This unconstrained setting is analytically tractable but often unrealistic. In practice, agents' decisions are almost always shaped by feasibility, safety, or regulatory requirements. The relevance of constraints has long been recognized in related areas such as constrained Markov decision

*Columbia University, IEOR. **Email:** ah4277@columbia.edu

†UC Berkeley, IEOR. **Email:** zijiu.lyu@berkeley.edu

processes (CMDPs), which capture situations where an individual agent face multiple competing objectives. Examples include telecommunication, where throughput must be maximized while keeping delays below a threshold [7]; finance, where risk and return trade-offs are enforced through constraints as in Markowitz's portfolio theory [8]; and resource allocation problems in medical systems, where limited capacity dictates feasible allocations [9]. More recently, the study of safe reinforcement learning has emphasized constraints as a safeguard against catastrophic outcomes [10].

While classical MFG models have yielded many elegant results, they often overlook constraints that naturally arise in real-world systems. Incorporating such constraints can make the framework more realistic and broaden its range of applications. At the same time, constraints introduce significant analytical challenges: the feasible set of agent decisions may depend on the population distribution itself, which prevents a direct application of existing methods and analyses.

Our work and approach. In this paper, we introduce a framework of Constrained Mean Field Games (CMFGs), where both the time horizon and the state space are discrete and finite. Unlike the classical MFG setting, where the representative agent solves an unconstrained Markov decision process (MDP), here each agent faces a constrained MDP (CMDP) that maximizes utility subject to restrictions on states and actions. In the general case, constraints depend on an individual's state and action together with the population distribution, a setting we call agent-population-level constraints. We define CMFG Nash equilibria through a fixed-point argument: given a population distribution, the representative agent solves the induced CMDP to obtain its best response, and this policy must in turn regenerate the same distribution. For clarity, we use the term MFG to refer exclusively to the unconstrained case, and the terms player and agent interchangeably.

For CMFGs with general agent-population-level constraints, we prove the existence of equilibria under suitable assumptions using Kakutani's fixed-point theorem, in line with the classical MFG literature. A key difference is that, with constraints, the feasible domain of each agent depends on the population distribution, which complicates the analysis. To apply Kakutani's theorem, we impose a strict feasibility assumption: for every constraint and any mean-field distribution, there must exist a policy that respects all constraints while satisfying the chosen one with a positive margin. This guarantees that the feasible set for each agent is nonempty and has enough slack to avoid degenerate boundary cases. In addition to existence, we also provide a uniqueness result under stronger assumptions. Specifically, when the transition probabilities and constraints are independent of the mean-field distribution and the reward functions satisfy the classical monotonicity condition of Lasry and Lions, the CMFG admits a unique equilibrium. This complements our general existence result and parallels the uniqueness theory developed for classical MFGs.

To compute equilibria under agent-population-level constraints, we develop a numerical scheme called Constrained Mean Field Occupation Measure Optimization (CMFOMO). The method parameterizes the agent's occupation measure and leverages Karush-Kuhn-Tucker (KKT) conditions to reformulate the equilibrium problem as an optimization program with convex constraints and bounded variables. CMFOMO does not assume uniqueness and can characterize all CMFG equilibria. We show that suboptimal solutions of CMFOMO approximate true equilibria arbitrarily well when the objective is minimized sufficiently.

We then consider a population-level constraint setting, a degenerate case where constraints depend only on the mean-field distribution and not on individual strategies, without imposing the strict feasibility assumption. In this case, the problem reduces to finding standard Nash equilibria that satisfy the population-level constraints, which can be solved using a modified CMFOMO framework by removing the components associated with the strict feasibility assumption, and similar approximation guarantees are obtained.

Finally, we establish a rigorous connection between CMFGs and finite-player constrained games. In particular, we show that CMFG equilibria yield $O(1/\sqrt{N})$-Nash equilibria for the corresponding constrained N-player games. This extends the classical justification of MFGs as approximations for large but finite-player games to the constrained setting.

In the end, numerical experiments on a modified Susceptible–Infected–Susceptible (SIS) model with several different kinds of constraints illustrate the effectiveness of our approach in both the agent-population-level and population-level settings.

Related work. Our approach builds on the linear programming (LP) method for CMDPs, where decision variables are occupation measures. We refer to [7] for a comprehensive treatment of solution approaches and theory, and to [11, 12, 13] for early developments. In the continuous-time setting, the LP method has also been applied to stochastic control problems, potentially with constraints. See [14] and the references therein for a broad survey.

In recent years, a growing body of work has used LP formulations and occupation measures in games without agent constraints. For example, [15, 16] introduce LP formulations for continuous-time MFGs to characterize optimality conditions and establish existence of Nash equilibria with mixed strategies. [17] develops a primal-dual characterization for discrete-time MFGs, leading to an optimization-based formulation (MFOMO), which is later extended to continuous-time MFGs in [18]. Using occupation measures, [19] proposes a globally convergent algorithm for computing equilibria of MFGs under the monotonicity condition, and further extended the analysis to reinforcement learning with regret guarantees in multi-agent systems. Our CMFOMO generalizes the MFOMO framework of [17] by incorporating agent constraints while preserving its favorable properties.

Imposing agent constraints in MFGs has been investigated from several perspectives. In deterministic settings, [20] considers continuous-time MFGs where the agent's state is confined to a compact subset of the Euclidean space, which is referred to as state constraints, and establishes existence and uniqueness of Nash equilibria. Subsequent works such as [21, 22, 23] extend this analysis using PDE-based methods, while [24, 25] further generalize the framework to include mean-field games of controls and mixed state–action constraints. In stochastic formulations, [26] studies portfolio liquidation MFGs with a terminal inventory constraint, requiring each agent's position to be fully liquidated by the terminal time. Building on this, [27] introduces directional trading constraints, where each player is restricted to trade in a single direction depending on their initial position. For finite-player settings, [28] analyzes a stochastic liquidation game with terminal constraints and rigorously connects its equilibria to those of the corresponding MFG.

Compared to these studies, our paper investigates stochastic MFGs with more general constraints, which may depend simultaneously on individual agents' states, actions, and the population distribution. In addition, we propose an optimization-based computational framework for solving such constrained MFGs and provide theoretical guarantees on the approximation error between MFG and N-player equilibria.

Organization of the paper. The remainder of the paper is organized as follows. Section 2 presents the formulation of CMFGs and introduces both agent-population-level and population-level constraints. Section 3 develops existence and uniqueness results as well as our numerical scheme, CMFOMO. Section 4 treats the population-level setting and its approximation results. Section 5 establishes the connection to constrained N-player games, and Section 6 illustrates the framework with numerical experiments on the SIS model.

2 Constrained Mean-Field Games

In this section, we formally introduce (Markov) Constrained Mean-Field Games (CMFGs). The game is defined over a finite time horizon $\mathcal{T} := \{0, 1, \dots, T\}$, where $T < +\infty$, with a finite state space $\mathcal{S} \subset \mathbb{R}^d$ and a finite action space $\mathcal{A} \subset \mathbb{R}^n$. In this game, there is an infinite population of homogeneous players, each independently starting from an initial state $s_0 \sim \mu_0 \in \Delta(\mathcal{S})$. At each time step $t \in \mathcal{T}$, a representative player observes their current state s_t and selects an action $a_t \in \mathcal{A}$ according to a randomized (Markov) policy $\pi_t(\cdot \,|\, s_t) \in \Delta(\mathcal{A})$. The player then receives a reward $r_t(s_t, a_t, L_t)$ and transitions to a new state s_{t+1}

following the transition probability $p_t(\cdot \mid s_t, a_t, L_t)$, where $L_t \in \Delta(\mathcal{S} \times \mathcal{A})$ represents the joint state-action distribution of the population at time t.

This formulation aligns with the standard framework of discrete time Mean-Field Games (MFGs), where the players are free to optimize their policies without additional restrictions. However, in CMFGs, players are subject to certain policy constraints. Alongside receiving rewards, players also incur a cost $c_t(s_t, a_t, L_t)$, defined by a cost function $c : \mathcal{T} \times \mathcal{S} \times \mathcal{A} \times \Delta(\mathcal{S} \times \mathcal{A}) \to \mathbb{R}^k$. Each player must ensure that their expected cumulative cost over the time horizon does not exceed a specified threshold $\gamma_0 \in \mathbb{R}^k$. This constraint enforces that the expected cumulative cost remains within the prescribed limits, introducing an additional layer of complexity to the decision-making process.

Mathematically, given a mean-field flow $L := L_t \in \Delta(\mathcal{S} \times \mathcal{A})_{t \in \mathcal{T}}$ that characterizes the joint distribution of states and actions in the population at each time step, each player solves the following Constrained Markov Decision Process (CMDP).

Constrained MDP (CMDP).

$$\begin{aligned} \text{maximize}_{\{\pi_t\}_{t\in\mathcal{T}}} \quad & \mathbb{E}\left[\sum_{t\in\mathcal{T}} r_t(s_t, a_t, L_t) \,\middle|\, s_0 \sim \mu_0\right] && \text{(CMDP reward)} \\ \text{subject to} \quad & a_t \sim \pi_t(s_t), \quad s_{t+1} \sim p_t(s_t, a_t, L_t), && \text{(CMDP dynamics)} \\ & \mathbb{E}\left[\sum_{t\in\mathcal{T}} c_t(s_t, a_t, L_t) \,\middle|\, s_0 \sim \mu_0\right] \le \gamma_0, && \text{(CMDP constraints)} \end{aligned}$$

where $\mu_0 \in \Delta(\mathcal{S})$ denotes the initial distribution. For notational convenience, let $\mathcal{M}_c(L)$ denote the constrained MDP problem described above, and let $V_c^*(L)$ denote its optimal expected cumulative reward, whenever it exists.

The extra constraints in (CMDP constraints) show the main difference of our CMFGs compared with the classic unconstrained MFGs and one needs to define a new solution concept that is tailored to our setting. To formally define the Nash Equilibrium (NE) of CMFGs, we first define the mapping $\Psi(\cdot) \in \left[\Delta(\mathcal{S}\times\mathcal{A})\right]^{|\mathcal{T}|}$ which maps any admissible policy π to the flow of population state-action distribution induced by π. The mapping Ψ is defined recursively in the following way. For any $s \in \mathcal{S}$ and $a \in \mathcal{A}$,

$$\begin{cases} \Psi_0(\pi)(s,a) := \pi_0(a \mid s)\mu_0(s), \\ \Psi_{t+1}(\pi)(s,a) := \pi_{t+1}(a \mid s) \sum_{s'\in\mathcal{S}, a'\in\mathcal{A}} p_t(s \mid s', a', \Psi_t(\pi))\Psi_t(\pi)(s', a'). \end{cases}$$

Definition 1 (Nash equilibrium of CMFGs)**.** *We say that (π, L) is a Nash equilibrium of CMFG if it satisfies:*

1) π is an optimal policy of the constrained MDP $\mathcal{M}_c(L)$;

2) $L = \Psi(\pi)$.

Remark 2.1. *In the literature of MFGs, conditions 1) and 2) are often referred to as the optimality/best response and consistency/compatibility conditions, respectively. Here, the condition 1) involves a constrained MDP problem instead of a standard MDP problem, which is the key difference of our framework compared with the standard definition of the NEs for MFGs. The NE of CMFGs indicates that no agent can improve its payoff by unilaterally deviating from current strategy while satisfying constraints.*

Population-level constraints and agent-population-level constraints. When the cost functions take different forms, two types of constraints can be considered. If the cost function c involves the agent's state s and action a, as well as the population distribution L, we say that the constraints are at the *agent-population level*. Because now each agent's strategy does impact the value of the cost function c, when the

mean-field flow L is fixed, some strategies may violate the constraints and therefore the set of admissible strategies is only a subset of that of unconstrained MFGs. As a result, the Nash equilibria of CMFGs (as defined in Definition 1) could be very different from the that of unconstrained MFGs. We shall discuss agent-population-level constraints in detail in Section 3.

On the other hand, for a special case where the cost function c depends only on the mean-field term L and is independent of an agent's state s and action a, the constraints are said to be at the *population level*. Consequently, the strategy of an individual agent does not influence whether the constraints are satisfied. This further implies that CMFG Nash equilibria are equivalent to MFG Nash equilibria with certain patterns. We refer to Section 4 for a more detailed discussion of population-level constraints.

To better illustrate the above ideas and to show the difference between population-level and agent-population-level constraints, we now introduce a concrete example of CMFGs, which is adapted from the Susceptible-Infected-Susceptible (SIS) model in [29, Appendix A.3].

Example 1 (Constrained SIS model)**.** The constrained SIS model evolves over discrete time steps $0, 1, \ldots, T$ with finite horizon $T < \infty$. Each agent has two possible states, S, I, denoting "susceptible" and "infected", and two actions, U, D, representing "going out" and "social distancing". Agents incur a loss when infected but no penalty when susceptible. Social distancing lowers the probability of infection but yields a smaller immediate reward.

The reward function is defined as:

$$r_t(s_t, a_t, L_t) := -\delta_{s_t=I} - 0.5\delta_{a_t=D},$$

where $\delta_\cdot$ is the indicator function, taking a value of 1 if the condition is true and 0 otherwise. The transition probabilities are:

$$\begin{aligned}
&p_t(s_{t+1} = I \,|\, s_t = S, a_t = D, L_t) = 0,\\
&p_t(s_{t+1} = I \,|\, s_t = S, a_t = U, L_t) = 0.9L_t(s_t = I, a_t = U),\\
&p_t(s_{t+1} = S \,|\, s_t = I, a_t = D, L_t) = 0.5,\\
&p_t(s_{t+1} = S \,|\, s_t = I, a_t = U, L_t) = 0.5\Big[1 - 0.9L_t(s_t = I, a_t = U)\Big].
\end{aligned}$$

Here, $L_t(s_t = I, a_t = U) \in [0, 1]$ denotes the proportion of infected agents going out at time t.

To introduce constraints, we propose the following three illustrative examples:

$$\text{Agent-population-level (state):} \quad \frac{1}{|\mathcal{T}|}\mathbb{E}\left(\sum_{t\in\mathcal{T}} \delta_{s_t=I} \,\middle|\, s_0 \sim \mu_0\right) \le \gamma_0, \tag{1}$$

$$\text{Agent-population-level (action):} \quad \frac{1}{|\mathcal{T}|}\mathbb{E}\left(\sum_{t\in\mathcal{T}} \delta_{a_t=U} \,\middle|\, s_0 \sim \mu_0\right) \le \gamma_0, \tag{2}$$

$$\text{Population-level (state):} \quad \frac{1}{|\mathcal{T}|}\sum_{t\in\mathcal{T}} L_t(s_t = I) \le \gamma_0. \tag{3}$$

Constraint (1) operates at the agent-population level, requiring that an individual agent's strategy ensures their infection risk remains below the threshold γ_0. This reflects a personalized safety measure that prioritizes the health of each agent. Similarly, constraint (2) is an agent-population-level restriction, limiting the time an individual agent spends outside. This could represent a government-imposed measure to control exposure during a pandemic. In contrast, constraint (3) is a population-level restriction. It does not enforce specific behavior for any individual agent. Instead, it sets a global limit on the infection rate in the population as a whole. This type of constraint aims to achieve collective safety and health outcomes by controlling the aggregate behavior of the agents.

The above description provides an intuitive explanation of the two types of constraints. In the following, we demonstrate through a concrete calculation that population-level constraints and agent-population-level constraints are indeed mathematically distinct.

For simplicity, let us consider the case where $T = 1$ and the initial distribution is $\mu_0(s_0 = I) = 1$. Denote by $L := L_0(s_0 = I, a_0 = U) \in [0, 1]$ the percentage of infected agents who choose to go out at time 0, and by $\alpha := \mathbb{P}(a_0 = U) \in [0, 1]$ the probability that a specific agent chooses to go out given the population distribution L. A straightforward calculation shows that the expected reward for this specific agent is (we do not consider the action cost at $t = 1$):

$$\mathbb{E}(r_0 + r_1) = 0.5(1 - 0.9L)\alpha - 2.$$

Since $1-0.9L > 0$, the agent maximizes their expected cumulative reward by always going out, *i.e.*, $\alpha^* = 1$. This decision is independent of the population distribution L. Consequently, in the NE of the unconstrained MFG, the percentage of infected agents at $t = 1$ will be:

$$\mathbb{P}(s_1 = I) = 0.95.$$

As a comparison, if all agents instead choose social distancing at the initial time, only half of them will remain infected at $t = 1$.

Now, suppose we impose the following **population-level constraint**:

$$L_1(s_1 = I) \leq \gamma_0. \tag{4}$$

Based on the earlier calculation, there will be no NE in the CMFG if $\gamma_0 < 0.95$. Alternatively, suppose we instead impose the following **agent-population-level constraint**:

$$\mathbb{E}(\delta_{s_1=I}) \leq \gamma_0. \tag{5}$$

Under the constraint (5), there are three cases to discuss:

- Case 1: $\gamma_0 > 0.5$

  In this case, there will be an NE in the CMFG. To spell out, by Definition 1, the NE policy α^{NE} solves the following system of equations:

  $$\begin{cases} 0.5 + 0.45\alpha^{\mathrm{NE}} L^{\mathrm{NE}} = \min\{\gamma_0, 0.95\}, \\ \alpha^{\mathrm{NE}} = L^{\mathrm{NE}}, \end{cases}$$

  which gives:

  $$\alpha^{\mathrm{NE}} = \min\left\{\sqrt{\frac{\gamma_0 - 0.5}{0.45}}, 1\right\}.$$

- Case 2: $\gamma_0 = 0.5$

  In this case, there will be no NE in the CMFG. Since $\mathbb{E}(\delta_{s_1=I}) = 0.5+0.45L\alpha$, the constraint translates into $\alpha L = 0$. When $L = 0$, every $\alpha \in [0, 1]$ is feasible. However, the optimal action $\alpha^* = 1$, which contradicts the consistency condition of the NE. On the other hand, when $L \in (0, 1]$, the only feasible action is $\alpha = 0$, which contradicts the consistency condition, too. As a consequence, there will be no NE.

- Case 3: $\gamma_0 < 0.5$

  In this case, there will be no NE in the CMFG. This is obvious since there will be no feasible action.

This concrete calculation thus demonstrates the difference between population-level constraints and agent-population-level constraints. □

Remark 2.2 (Lagrangian method and penalty method)**.** *Alternative approaches to handling constraints include the Lagrangian method and penalty method. For Lagrangian method, the representative agent solves the following unconstrained MDP problem:*

$$\begin{aligned} \textit{maximize}_{\{\pi_t\}_{t\in\mathcal{T}}} \quad & \mathbb{E}\left[\sum_{t\in\mathcal{T}}\Big(r_t(s_t,a_t,L_t)-\eta_1^\top c_t(s_t,a_t,L_t)\Big)\,\Big|\, s_0\sim\mu_0\right] \\ \textit{subject to} \quad & a_t\sim\pi_t(s_t),\quad s_{t+1}\sim p_t(s_t,a_t,L_t), \end{aligned}$$

where $\eta_1 > 0$ is the Lagrangian multiplier.

For penalty method, the representative agent solves the following unconstrained MDP problem:

$$\begin{aligned} \textit{maximize}_{\{\pi_t\}_{t\in\mathcal{T}}} \quad & \mathbb{E}\left[\sum_{t\in\mathcal{T}}\Big(r_t(s_t,a_t,L_t)-\eta_2^\top (c_t(s_t,a_t,L_t)-\gamma_0)^+\Big)\,\Big|\, s_0\sim\mu_0\right] \\ \textit{subject to} \quad & a_t\sim\pi_t(s_t),\quad s_{t+1}\sim p_t(s_t,a_t,L_t), \end{aligned}$$

where $\eta_2 > 0$ is the penalty coefficient. For these two problems, the unconstrained MDPs always admit solutions, and the corresponding MFGs fall into the standard setting. However, studying the constrained version directly is often preferable or necessary, especially when violating the constraints may lead to detrimental consequences and hard constraints must be enforced. For Lagrangian method, the relationship between γ_0 in (CMDP constraints) *and the Lagrangian multiplier coefficient η_1 is unclear, making it challenging to define an appropriate η_1 in practice. And for penalty method, one needs to increase the value of η_2 to make the problem asymptotically equivalent to the original constrained problem.*

3 Mean-Field Games with Agent-Population-Level Constraints

In this section, we focus on the problem of CMFGs with general agent-population-level constraints, *i.e.*, the cost function c involves agent's state s and action a, as well as population distribution L. We first analyze the existence of CMFG NEs defined by Definition 1. We then provide a numerical scheme to solve the CMFGs.

3.1 Existence and Uniqueness of Nash Equilibrium in CMFGs

Our main result in this section is that a CMFG NE defined by Definition 1 always exists under proper continuity assumptions and a feasibility assumption on the constraints of the CMFG.

The first assumption in the following ensures that for any individual constraint indexed by i, there exists a policy $\pi^{(i)}$ that satisfies all constraints while being strictly feasible for constraint i.

Assumption 1 (Strict feasibility)**.** *There exists $\delta > 0$ such that for any mean-field flow L and any index $1 \le i \le k$, there exists a policy $\pi^{(i)}$ satisfying:*

$$\mathbb{E}_{a\sim\pi^{(i)}}\left[\sum_{t\in\mathcal{T}} c_t(s_t,a_t,L_t)\,\Big|\, s_0\sim\mu_0\right] \le \gamma_0,$$

and

$$\mathbb{E}_{a\sim\pi^{(i)}}\left[\sum_{t\in\mathcal{T}} [c_t]_i(s_t,a_t,L_t)\,\Big|\, s_0\sim\mu_0\right] \le [\gamma_0]_i - \delta,$$

where $[\cdot]_i$ denotes the i-th entry of the vector.

When $\delta = 0$, the assumption above simply ensures the feasibility of the constraints, guaranteeing that for any mean-field flow, at least one policy satisfies the given conditions. This can also be interpreted as stating that each agent always has at least one feasible strategy, regardless of the strategies chosen by competitors. By assuming $\delta > 0$, we strengthen this feasibility condition by ensuring the existence of a *strictly feasible* policy for each constraint. As we will show, this tightening can not be relaxed and plays a crucial role in our subsequent analysis.

We also impose the following continuity assumption. Such continuity conditions on rewards and transition kernels are standard in the existing MFG literature to guarantee the existence of NEs. Here, we extend this assumption to include the continuity of the constraints as well.

Assumption 2 (Continuity)**.** *The functions p, r, and c are continuous with respect to the mean-field flow.*

Under these assumptions, we are able to establish the existence of NE for the constrained MFGs.

Theorem 3.1 (Existence of NE in CMFGs)**.** *Under Assumptions 1 and 2, there exists at least one CMFG NE defined by Definition 1.*

Remark 3.1. *Assumption 1 is critical for the proof of existence. Indeed, the CMFG may not have an NE when Assumption 1 is not satisfied. For instance, let us set $\gamma_0 = 0.5$ in (5) (cf. Example 1). Then there is no positive δ satisfying Assumption 1. By the analysis in Example 1, the resulting CMFG does not have any NE. We further notice that the CMFG also satisfies Assumption 2, which means Assumption 1 cannot be relaxed.*

The proof of Theorem 3.1 relies on the following two lemmas. Lemma 3.2 reformulates the constrained Markov Decision Process (MDP) problem as a linear program (LP). Lemma 3.4 then establishes that the solution to the Karush-Kuhn-Tucker (KKT) conditions, arising from the LP formulation in Lemma 3.2, can always be found within a bounded domain, where the bound is independent of the choice of the mean-field flow.

To aid our later discussion, we first introduce the concept of *occupation measure*, which characterizes the distribution of state-action of a single player. Specifically, when the mean-field flow L is fixed, for any admissible policy π, its associated occupation measure $d \in \left[\Delta(\mathcal{S} \times \mathcal{A})\right]^{|\mathcal{T}|}$ describes the probability of the representative player staying at state s and taking action a at time t. With a slight abuse of notation, for any mean-field flow L, $s \in \mathcal{S}$ and $a \in \mathcal{A}$, define $\Psi(\pi, L) \in \left[\Delta(\mathcal{S} \times \mathcal{A})\right]^{|\mathcal{T}|}$ recursively by:

$$\begin{cases} \Psi_0(\pi, L)(s, a) := \pi_0(a \,|\, s)\mu_0(s), \\ \Psi_{t+1}(\pi, L)(s, a) := \pi_{t+1}(a \,|\, s) \sum_{s' \in \mathcal{S}, a' \in \mathcal{A}} p_t(s \,|\, s', a', L_t)\Psi_t(\pi, L)(s', a'). \end{cases}$$

Then $\Psi(\pi, L)$ is the occupation measure of a representative agent taking policy π under mean-field flow L.

On the other hand, for any $d \in \left[\Delta(\mathcal{S} \times \mathcal{A})\right]^{|\mathcal{T}|}$, we can retrieve the policy which leads to the occupation measure as follows. Specifically, define $\Pi(d)$ by the set of policies such that:

$$\forall t \in \mathcal{T}, s \in \mathcal{S}, a \in \mathcal{A}, \quad \pi_t(a \,|\, s) = \begin{cases} d_t(s, a) \Big/ \sum_{a' \in \mathcal{A}} d_t(s, a'), & \text{if } \sum_{a' \in \mathcal{A}} d_t(s, a') > 0, \\ \text{any vector in } \Delta(\mathcal{A}), & \text{else.} \end{cases}$$

Then any policy $\pi \in \Pi(d)$ induces the occupation measure d.

Utilizing the connection between admissible policies and occupation measures, one can transform the constrained MDP problem into a linear program on occupation measure.

Lemma 3.2 (CMDPs as LPs)**.** *Given the mean-field flow L, solving the constrained MDP problem $\mathcal{M}_c(L)$ is equivalent to solving the following LP problem:*

$$\textit{minimize}_d \quad -r_L^\top d \tag{LP 1}$$

$$\textit{subject to} \quad A_L d = b, \quad c_L d \le \gamma_0, \quad d \ge 0. \tag{LP 2}$$

Here $r_L \in \mathbb{R}^{|\mathcal{S}||\mathcal{A}||\mathcal{T}|}$, $b \in \mathbb{R}^{|\mathcal{S}||\mathcal{T}|}$, $c_L \in \mathbb{R}^{k \times |\mathcal{S}||\mathcal{A}||\mathcal{T}|}$, *and* $A_L \in \mathbb{R}^{|\mathcal{S}||\mathcal{T}| \times |\mathcal{S}||\mathcal{A}||\mathcal{T}|}$ *are defined by:*

$$r_L := \begin{pmatrix} r_0(L_0) \\ r_1(L_1) \\ \vdots \\ r_T(L_T) \end{pmatrix}, \quad r_t(L_t) \in \mathbb{R}^{|\mathcal{S}||\mathcal{A}|}, \quad b := \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \mu_0 \end{pmatrix}, \quad \mu_0 \in \mathbb{R}^{|\mathcal{S}|},$$

$$c_L := \begin{pmatrix} \Big([c_0]_1(L_0)\Big)^\top & \Big([c_1]_1(L_1)\Big)^\top & \cdots & \Big([c_T]_1(L_T)\Big)^\top \\ \Big([c_0]_2(L_0)\Big)^\top & \Big([c_1]_2(L_1)\Big)^\top & \cdots & \Big([c_T]_2(L_T)\Big)^\top \\ \vdots & \vdots & \ddots & \vdots \\ \Big([c_0]_k(L_0)\Big)^\top & \Big([c_1]_k(L_1)\Big)^\top & \cdots & \Big([c_T]_k(L_T)\Big)^\top \end{pmatrix}, \quad [c_t]_i(L_t) \in \mathbb{R}^{|\mathcal{S}||\mathcal{A}|},$$

$$A_L := \begin{pmatrix} A_0(L_0) & -Z & 0 & 0 & \cdots & 0 & 0 \\ 0 & A_1(L_1) & -Z & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & A_{T-1}(L_{T-1}) & -Z \\ Z & 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}, \quad Z := \overbrace{\Big(I_{|\mathcal{S}|}, I_{|\mathcal{S}|}, \cdots, I_{|\mathcal{S}|}\Big)}^{|\mathcal{A}|},$$

where $I_{|\mathcal{S}|}$ *denotes the identity matrix of shape* $|\mathcal{S}| \times |\mathcal{S}|$, *and for each* $t \in \mathcal{T}\backslash\{T\}$, $A_t(L_t) \in \mathbb{R}^{|\mathcal{S}| \times |\mathcal{S}||\mathcal{A}|}$ *is defined by:*

$$\forall 1 \le i \le |\mathcal{S}|, \quad \textit{the } i\textit{-th row of } A_t(L_t) := \Big(p_t(i \,|\, \cdot, \cdot, L_t)\Big)^\top \in \mathbb{R}^{1 \times |\mathcal{S}||\mathcal{A}|}.$$

Specifically, if π^* *is an optimal policy of* $\mathcal{M}_c(L)$, *then there exists* d^* *such that* $\pi^* \in \Pi(d^*)$ *and* d^* *is an optimal solution of the above LP problem. Conversely, if* d^* *is an optimal solution of the above LP problem, then for any* $\pi^* \in \Pi(d^*)$, π^* *is an optimal policy of* $\mathcal{M}_c(L)$.

The LP formulation of MDPs has been exploited in [17], and the proof of Lemma 3.2 follows from that of [17, Lemma 2, Lemma 3] and thus is omitted here.

By Lemma 3.2 and the strong duality of LP (see [30], for example), solving $\mathcal{M}_c(L)$ is then equivalent to finding the solution that satisfies the following KKT conditions.

KKT Conditions.

$$\begin{aligned} \text{find} \quad & \Big\{ d \in \mathbb{R}^{|\mathcal{S}||\mathcal{A}||\mathcal{T}|}, y \in \mathbb{R}^{|\mathcal{S}||\mathcal{T}|}, z \in \mathbb{R}^{|\mathcal{S}||\mathcal{A}||\mathcal{T}|}, \lambda \in \mathbb{R}^k \Big\} \\ \text{such that} \quad & -r_L = A_L^\top y + z - c_L^\top \lambda, \quad A_L d = b, \quad c_L d \le \gamma_0, \quad d \ge 0, \\ & z \ge 0, \quad z^\top d = 0, \quad \lambda \ge 0, \quad \lambda^\top (c_L d - \gamma_0) = 0. \end{aligned}$$

Since the KKT conditions are uniquely determined by the mean-field flow L, we denote them by $\mathcal{K}(L)$ in the rest of our paper. It is obvious that finding an NE in a CMFG is equivalent to finding a fixed point of the KKT conditions, in the following sense.

Proposition 3.3. (π, L) *is a CMFG NE defined by Definition 1 if and only if* $\pi \in \Pi(L)$ *and there exists* (y, z, λ) *such that* (L, y, z, λ) *is a solution of* $\mathcal{K}(L)$.

The following lemma guarantees that one can always find solutions y, z, λ in a compact set.

Lemma 3.4 (A bounded solution of $\mathcal{K}(L)$)**.** *Given the mean-field flow L, under Assumption 1, let (d, y, z, λ) be a solution of $\mathcal{K}(L)$. Then, $||\lambda||_\infty \leq |\mathcal{T}| r_{\max}/\delta$, and there exists (y', z') in a bounded domain such that (d, y', z', λ) is a solution of $\mathcal{K}(L)$. Specifically, the bounds are given by:*

$$\left|\left|y'\right|\right|_1 \leq \frac{|\mathcal{S}||\mathcal{T}|(|\mathcal{T}|+1)}{2} r_{\max}\left(1 + \frac{|\mathcal{T}|}{\delta} c_{\max}\right),$$
$$\left|\left|z'\right|\right|_1 \leq |\mathcal{S}||\mathcal{A}|(|\mathcal{T}|^2 - |\mathcal{T}| + 2) r_{\max}\left(1 + \frac{|\mathcal{T}|}{\delta} c_{\max}\right),$$

where $r_{\max}$ and $c_{\max}$ denote the maximum values of r and c, respectively, i.e.*,*

$$r_{\max} := \max_{t\in\mathcal{T}, s\in\mathcal{S}, a\in\mathcal{A}, L\in\Delta(\mathcal{S}\times\mathcal{A})} r_t(s, a, L),$$
$$c_{\max} := \max_{t\in\mathcal{T}, s\in\mathcal{S}, a\in\mathcal{A}, L\in\Delta(\mathcal{S}\times\mathcal{A})} ||c_t(s, a, L)||_\infty.$$

Proof. Suppose (d, y, z, λ) is a solution of $\mathcal{K}(L)$. We first prove that $||\lambda||_\infty \leq |\mathcal{T}| r_{\max}/\delta$. To see this, notice that d is a minimizer of the LP problem:

$$\begin{aligned} \text{minimize}_d \quad & -\left(r_L - c_L^\top \lambda\right)^\top d \\ \text{subject to} \quad & A_L d = b, \quad d \geq 0. \end{aligned}$$

Therefore, for any index $1 \leq i \leq k$, choosing policy $\pi^{(i)}$ defined in Assumption 1 and defining the associated occupation measure $d^{(i)} := \Psi\left(\pi^{(i)}, L\right)$, we will have:

$$\begin{aligned} \left(r_L - c_L^\top \lambda\right)^\top d \geq \left(r_L - c_L^\top \lambda\right)^\top d^{(i)}, \\ r_L^\top \left(d - d^{(i)}\right) \geq \lambda^\top c_L \left(d - d^{(i)}\right). \end{aligned} \tag{6}$$

In (6), LHS $\leq |\mathcal{T}| r_{\max}$, and RHS $= \lambda^\top\left(\gamma_0 - c_L d^{(i)}\right) \geq [\lambda]_i \left[\gamma_0 - c_L d^{(i)}\right]_i \geq \delta [\lambda]_i$. Thus, we see that $[\lambda]_i \leq |\mathcal{T}| r_{\max}/\delta$, which yields the desired bound of λ.

The boundedness of y' and z' is directly implied by combining the proof of [17, Proposition 6] and . The only difference with [17, Proposition 6] is that r_L is replaced by $r_L - c_L^\top \lambda$. This finalizes our proof. □

Finally, we present the proof of Theorem 3.1 below.

Proof of Theorem 3.1. We start with defining the set-valued mapping Γ, which maps a mean-field flow L to the set of optimal solutions of the LP introduced in Lemma 3.2. By Proposition 3.3, it suffices to establish that Γ admits a fixed point. In the following, we prove this by aplying Kakutani's fixed-point theorem.

To proceed, we verify the conditions of Kakutani's fixed-point theorem. First, notice that the set of mean-field flows can be viewed as a concatenation of $|\mathcal{T}|$ probability simplices, each of dimension $|\mathcal{S}||\mathcal{A}| - 1$, making it a non-empty, compact and convex subset of $\mathbb{R}^{|\mathcal{T}|(|\mathcal{S}||\mathcal{A}|-1)}$. Second, the convexity of $\Gamma(L)$ for any mean-field flow L follows from the linearity of (LP 1) and (LP 2). Finally, we verify that Γ has a closed graph. Suppose $\{L_n\}_{n\geq 0}$ and $\{d_n\}_{n\geq 0}$ are sequences of mean-field flows and occupation measures such that

1. $d_n \in \Gamma(L_n)$ for all n, and
2. there exist $\hat{L}$ and $\hat{d}$ such that $\lim_{n\to\infty} L_n = \hat{L}$ and $\lim_{n\to\infty} d_n = \hat{d}$.

To show that $\hat{d} \in \Gamma(\hat{L})$, we note that for each n, there exists (y_n, z_n, λ_n) such that $(d_n, y_n, z_n, \lambda_n)$ is a solution of $\mathcal{K}(L_n)$. By Lemma 3.4, the sequence $\{(d_n, y_n, z_n, \lambda_n)\}_{n\geq 0}$ is uniformly bounded, ensuring the existence of a convergent subsequence. Let $(\hat{d}, \hat{y}, \hat{z}, \hat{\lambda})$ be its limit. Then by Assumption 2, this limit is a solution to $\mathcal{K}(\hat{L})$, which implies $\hat{d} \in \Gamma(\hat{L})$. This completes the proof. □

Remark 3.2. *The proof of Theorem 3.1 differs from existing methods to prove the existence of NEs, and highly relies on the primal-dual formulation. Specifically, the fixed point mapping in our approach is defined in terms of mean-field flows (and occupation measures), with the analysis depending on the boundedness of the dual variables (Lemma 3.4). In contrast, existing approaches, such as those in [29], define the fixed point mapping via optimal policies and analyze it using value functions and Q-functions. However, this approach is difficult to apply to CMFGs due to the absence of Bellman equations.*

Beyond existence, we also establish a uniqueness result under the classical Lasry-Lions monotonicity condition.

Assumption 3 (Monotonicity condition)**.** *The transition probabilities and the constraints are independent of the population distribution. In addition, the reward functions satisfy the following monotonicity condition: for any two mean-field flows L^1 and L^2,*

$$\sum_{t\in\mathcal{T}}\sum_{s\in\mathcal{S}}\sum_{a\in\mathcal{A}}\big(r_t(s,a,L_t^1)-r_t(s,a,L_t^2)\big)\big(L_t^1(s,a)-L_t^2(s,a)\big)\le 0,$$

with equality if and only if $L^1\equiv L^2$.

Theorem 3.5 (Uniqueness of CMFG NE)**.** *Under Assumptions 1, 2, and 3, the CMFG admits a unique Nash equilibrium mean-field flow as defined in Definition 1.*

Proof. By Theorem 3.1, there exists at least one CMFG NE. Suppose for contradiction that there are two distinct NE mean-field flows, L^1 and L^2. Then, by the definition of CMFG NE and Lemma 3.2,

$$r_{L^1}^\top L^1\ge r_{L^1}^\top L^2,\quad r_{L^2}^\top L^2\ge r_{L^2}^\top L^1.$$

This yields

$$0>(r_{L^1}-r_{L^2})^\top(L^1-L^2)=r_{L^1}^\top L^1-r_{L^1}^\top L^2+r_{L^2}^\top L^2-r_{L^2}^\top L^1\ge 0,$$

a contradiction. Hence the equilibrium is unique in terms of the mean-field flow. □

3.2 Optimization Framework for Solving NE in CMFGs

In this section, we provide a framework of *Constrained Mean-Field Occupation Measure Optimization* (CMFOMO), which transforms the problem of solving NEs into an optimization problem with bounded variables and convex constraints. This framework extends [17, Theorem 5] to incorporate constraints in the representative agent's problem. The following Theorem 3.6 is a direct implication of Definition 1 and Lemmas 3.2 and 3.4.

Theorem 3.6 (Constrained MF-OMO)**.** *Let $c_i>0$ $(i=1,2,3,4,5)$. Under Assumption 1, (π^*,L^*) is a CMFG NE defined by Definition 1 if and only if there exists (y^*,z^*,λ^*,w^*) such that $(L^*,y^*,z^*,\lambda^*,w^*)$ is an optimal solution of the following optimization problem with the optimal objective value being 0:*

$$\begin{aligned}
\textit{minimize}_{L,y,z,\lambda,w}\quad & c_1\left|\left|A_L^\top y+z+r_L-c_L^\top\lambda\right|\right|_2+c_2\left|\left|A_LL-b\right|\right|_2+c_3(z^\top L)\\
&+c_4\left|\left|\gamma_0-c_LL-w\right|\right|_2+c_5\left|\lambda^\top(\gamma_0-c_LL)\right| \qquad \text{(CMFOMO)}\\
\textit{subject to}\quad & L\ge 0,\quad \mathbf{1}^\top L_t=1,\quad \forall t\in\mathcal{T},\\
& ||y||_1\le\frac{|\mathcal{S}||\mathcal{T}|(|\mathcal{T}|+1)}{2}r_{\max}\left(1+\frac{|\mathcal{T}|}{\delta}c_{\max}\right),\\
& z\ge 0,\quad ||z||_1\le|\mathcal{S}||\mathcal{A}|\left(|\mathcal{T}|^2-|\mathcal{T}|+2\right)r_{\max}\left(1+\frac{|\mathcal{T}|}{\delta}c_{\max}\right),
\end{aligned}$$

$$0 \le \lambda \le \frac{|\mathcal{T}| r_{\max}}{\delta},$$
$$0 \le w \le \gamma_0,$$

where the mean-field flow L is treated as a vector in $\mathbb{R}^{|\mathcal{S}||\mathcal{A}||\mathcal{T}|}$, L_t is defined by the slice of L whose index ranges from $(|\mathcal{S}||\mathcal{A}|t+1)$ to $|\mathcal{S}||\mathcal{A}|(t+1)$, and $\mathbf{1}$ is defined by a vector of length $|\mathcal{S}||\mathcal{A}|$ whose entries are equal to 1. We refer to Assumption 1 and Lemma 3.2 for the definition of the rest of the notations.

Remark 3.3. *In the objective of CMFOMO, c_2 and c_4 control the consistency (cf. Remark 2.1) and feasibility of L, respectively; c_1 characterizes the optimality condition (cf. Remark 2.1); c_3 and c_5 are the complementarity terms connecting the previous conditions into a single optimization problem. When $c_4 = c_5 = 0$ and fixing $\lambda = 0$, the objective of CMFOMO reduces to MF-OMO introduced in [17].*

When allowing optimization errors, as we shall show next, ϵ-optimal solutions of CMFOMO are approximations of the NEs of the CMFG. To quantify the gaps between approximated NEs and true NEs, we introduce the following metrics.

Definition 2 (Optimality gap)**.** *For any policy π, we define its optimality gap by:*

$$G_{opt}(\pi) := V_c^*(\Psi(\pi)) - V^\pi(\Psi(\pi)),$$

where $V_c^*(\Psi(\pi))$ denotes the maximum expected cumulative reward of $\mathcal{M}_c(\Psi(\pi))$ (which always exists under Assumption 1), and $V^\pi(\Psi(\pi))$ denotes the expected cumulative reward of policy π under the mean-field flow $\Psi(\pi)$, i.e.*,*

$$V^\pi(\Psi(\pi)) := \mathbb{E}_{a\sim\pi}\left[\sum_{t\in\mathcal{T}} r_t(s_t, a_t, \Psi_t(\pi)) \,\middle|\, s_0 \sim \mu_0\right].$$

Definition 3 (Feasibility gap)**.** *For any policy π and mean-field flow L, we define the feasibility gap of π under L by:*

$$G_{fea}(\pi, L) := \left|\left|\min\Big(0, \gamma_0 - \gamma_L(\pi)\Big)\right|\right|_2,$$

where the min *operator is applied entry-wise, and*

$$\gamma_L(\pi) := \mathbb{E}_{a\sim\pi}\left[\sum_{t\in\mathcal{T}} c_t(s_t, a_t, L_t) \,\middle|\, s_0 \sim \mu_0\right].$$

With an abuse of notation, we define the feasibility gap of π by:

$$G_{fea}(\pi) := G_{fea}(\pi, \Psi(\pi)).$$

Remark 3.4. *For any given policy π, $G_{opt}(\pi)$ measures its performance compared with the optimal policy when the population distribution is induced by π, while $G_{fea}(\pi)$ measures the extent to which π violates the constraints.*

When $G_{\text{fea}}(\pi, L) > 0$, we say that π is *infeasible* under L. Otherwise, we say π is *feasible* under L. If π is feasible under $\Psi(\pi)$, then by definition, $G_{\text{opt}}(\pi)$ will always be non-negative. However, if π is infeasible under $\Psi(\pi)$, $G_{\text{opt}}(\pi)$ may be negative.

In order to obtain the non-asymptotic analysis, we impose a stronger assumption in the subsequent analysis, which requires the rewards, transition kernels and constraints to be Lipschitz continuous in the mean-field flow.

Assumption 4 (Lipschitz continuity)**.** *The functions p, r, and c are Lipschitz continuous with respect to the mean-field flow. That is, there exist constants $C_p, C_r, C_c > 0$, such that for any $t \in \mathcal{T}$ and any $L^1, L^2 \in \Delta(\mathcal{S} \times \mathcal{A})$, we have the inequalities:*

$$\begin{aligned}
\left|\left| \sum_{s\in\mathcal{S},a\in\mathcal{A}} \left| p_t\left(s,a,L^1\right) - p_t\left(s,a,L^2\right) \right| \right|\right|_1 &\le C_p \left|\left|L^1 - L^2\right|\right|_1, \\
\left|\left| r_t\left(L^1\right) - r_t\left(L^2\right) \right|\right|_1 &\le C_r \left|\left|L^1 - L^2\right|\right|_1, \\
\left|\left| \sum_{s\in\mathcal{S},a\in\mathcal{A}} \left| c_t\left(s,a,L^1\right) - c_t\left(s,a,L^2\right) \right| \right|\right|_1 &\le C_c \left|\left|L^1 - L^2\right|\right|_1,
\end{aligned}$$

where $p_t(s,a,L_t)$ is treated as a vector in $\mathbb{R}^{|\mathcal{S}|}$, $r_t(L_t)$ and L_t are treated as vectors in $\mathbb{R}^{|\mathcal{S}||\mathcal{A}|}$, and $c_t(s,a,L_t)$ is treated as a vector in $\mathbb{R}^k$.

The following theorem shows that any ϵ-sub-optimal solution of CMFOMO leads to an approximated NE of the CMFG with explicit approximation error bounded by a linear function of ϵ. Its proof can be found in Appendix A.

Theorem 3.7 (Suboptimality of CMFOMO)**.** *Let $\epsilon \ge 0$. Under Assumption 1 and Assumption 4, suppose (L, y, z, λ, w) is a feasible solution of CMFOMO defined in Theorem 3.6 with its objective* (CMFOMO) $\le \epsilon$. *Then, for any $\pi \in \Pi(L)$, its optimality gap and feasibility gap are bounded by:*

$$-\epsilon\sqrt{k}\frac{|\mathcal{T}|r_{\max}}{\delta}\zeta_3 \le G_{opt}(\pi) \le \epsilon\Big[(|\mathcal{T}|+1)\zeta_1 + \zeta_2 + \zeta_4\Big], \tag{7}$$

$$0 \le G_{fea}(\pi) \le \epsilon\zeta_3, \tag{8}$$

where

$$\begin{aligned}
\beta_y &:= \frac{|\mathcal{S}||\mathcal{T}|(|\mathcal{T}|+1)}{2} r_{\max}\left(1 + \frac{|\mathcal{T}|}{\delta}c_{\max}\right), \\
\beta_z &:= |\mathcal{S}||\mathcal{A}|(|\mathcal{T}|^2 - |\mathcal{T}| + 2) r_{\max}\left(1 + \frac{|\mathcal{T}|}{\delta}c_{\max}\right), \\
\alpha(c_2) &:= \frac{1}{c_2}\frac{(C_p+1)^{|\mathcal{T}|+1} - 2C_p - 1}{C_p^2}\sqrt{|\mathcal{S}|}, \qquad (9) \\
\zeta_1(c_1,c_2) &:= \frac{1}{c_1} + \alpha(c_2)\left(C_p\beta_y + C_r + kC_c\frac{|\mathcal{T}|r_{\max}}{\delta}\right), \\
\zeta_2(c_2,c_3) &:= \frac{1}{c_3} + \beta_z\alpha(c_2), \\
\zeta_3(c_2,c_4) &:= \frac{1}{c_4} + \alpha(c_2)\Big(kc_{\max} + C_c|\mathcal{S}||\mathcal{A}||\mathcal{T}|\Big), \qquad (10) \\
\zeta_4(c_2,c_5) &:= \frac{1}{c_5} + k\alpha(c_2)\frac{|\mathcal{T}|r_{\max}}{\delta}\Big(kc_{\max} + C_c|\mathcal{S}||\mathcal{A}||\mathcal{T}|\Big).
\end{aligned}$$

In practice, as optimization error is inevitable, the optimized policy may violate the constraints, *i.e.*, infeasible under the threshold γ_0. To guarantee the feasibility, one approach is to optimize CMFOMO with γ_0 replaced by $\gamma_0 - \epsilon_0$ where $\epsilon_0 > 0$ is a small constant. However, the improved feasibility comes at the expense of a worse optimality, as specified in the following theorem.

Theorem 3.8. *Under Assumption 1 and Assumption 4, let $0 < \epsilon_0 < \delta$ be a constant such that Assumption 1 remains satisfied with γ_0 (*resp. *δ) replaced by $\gamma_0 - \epsilon_0$ (*resp. *$\delta - \epsilon_0$). Suppose (L, y, z, λ, w) is a feasible*

solution of CMFOMO defined in Theorem 3.6, with γ_0 (resp. δ) *replaced by* $\gamma_0 - \epsilon_0$ (resp. $\delta - \epsilon_0$), *and its objective* (CMFOMO) $\leq \epsilon_0/\zeta_3$ *where* ζ_3 *is defined by* (10). *Then, for any* $\pi \in \Pi(L)$, *its feasibility gap (under* γ_0*) is 0, and its optimality gap (under* γ_0*) is bounded by:*

$$0 \leq G_{opt}(\pi) \leq \frac{\epsilon_0}{\zeta_3}\left[(|\mathcal{T}|+1)\tilde{\zeta}_1 + \tilde{\zeta}_2 + \tilde{\zeta}_4 + \frac{\zeta_3 k|\mathcal{T}|r_{\max}}{\delta - \epsilon_0}\right],$$

where

$$\begin{aligned}
\tilde{\beta}_y &:= \frac{|\mathcal{S}||\mathcal{T}|(|\mathcal{T}|+1)}{2} r_{\max}\left(1 + \frac{|\mathcal{T}|}{\delta - \epsilon_0} c_{\max}\right), \\
\tilde{\beta}_z &:= |\mathcal{S}||\mathcal{A}|(|\mathcal{T}|^2 - |\mathcal{T}| + 2) r_{\max}\left(1 + \frac{|\mathcal{T}|}{\delta - \epsilon_0} c_{\max}\right), \\
\tilde{\zeta}_1(c_1, c_2) &:= \frac{1}{c_1} + \alpha(c_2)\left(C_p\tilde{\beta}_y + C_r + kC_c\frac{|\mathcal{T}|r_{\max}}{\delta - \epsilon_0}\right), \\
\tilde{\zeta}_2(c_2, c_3) &:= \frac{1}{c_3} + \tilde{\beta}_z\alpha(c_2), \\
\tilde{\zeta}_4(c_2, c_5) &:= \frac{1}{c_5} + k\alpha(c_2)\frac{|\mathcal{T}|r_{\max}}{\delta - \epsilon_0}\Big(kc_{\max} + C_c|\mathcal{S}||\mathcal{A}||\mathcal{T}|\Big),
\end{aligned}$$

and $\alpha(c_2)$ *is defined by* (9).

The proof of Theorem 3.8 utilizes similar techniques as the proof of Theorem 3.7, which can also be found in Appendix A. In addition, as a direct corollary of Theorem 3.7, we get the following asymptotic results.

Corollary 3.9. *Under Assumptions 1 and 4. Suppose* $\left\{\left(c_1^{(i)}, c_2^{(i)}, c_3^{(i)}, c_4^{(i)}, c_5^{(i)}\right)\right\}_{i\geq 1}$ *is a sequence of coefficients in CMFOMO's objective function* (CMFOMO). *Denote by* $\left\{\left(L^{(i)}, y^{(i)}, z^{(i)}, \lambda^{(i)}, w^{(i)}\right)\right\}_{i\geq 1}$ *a sequence of feasible variables (*i.e., *satisfying all the constraints in Theorem 3.6) such that:*

$$\begin{aligned}
\exists \epsilon > 0,\ s.t.\ \forall i \geq 1,\ & c_1^{(i)}\left|\left|A_{L^{(i)}}^\top y^{(i)} + z^{(i)} + r_{L^{(i)}} - c_{L^{(i)}}^\top \lambda^{(i)}\right|\right|_2 + c_2^{(i)}\left|\left|A_{L^{(i)}}L^{(i)} - b\right|\right|_2 \\
& + c_3^{(i)}\left(\left(z^{(i)}\right)^\top L^{(i)}\right) + c_4^{(i)}\left|\left|\gamma_0 - c_{L^{(i)}}L^{(i)} - w^{(i)}\right|\right|_2 + c_5^{(i)}\left|\left(\lambda^{(i)}\right)^\top\left(\gamma_0 - c_{L^{(i)}}L^{(i)}\right)\right| \leq \epsilon.
\end{aligned}$$

We have:

1) *if* $\overline{\lim}_{i\to\infty} c_2^{(i)} = \overline{\lim}_{i\to\infty} c_4^{(i)} = \infty$, *then there exists a sub-sequence of* $\left\{\left(L^{(i)}\right)\right\}_{i\geq 1}$, *whose limiting point is denoted by* L, *such that* $\forall \pi \in \Pi(L)$, $G_{fea}(\pi) = 0$.

2) *if* $\overline{\lim}_{i\to\infty} c_1^{(i)} = \overline{\lim}_{i\to\infty} c_2^{(i)} = \overline{\lim}_{i\to\infty} c_3^{(i)} = \overline{\lim}_{i\to\infty} c_4^{(i)} = \overline{\lim}_{i\to\infty} c_5^{(i)} = \infty$, *then there exists a sub-sequence of* $\left\{\left(L^{(i)}\right)\right\}_{i\geq 1}$, *whose limiting point is denoted by* L, *such that* $\forall \pi \in \Pi(L)$, (π, L) *is a CMFG NE defined by Definition 1.*

Finally, we mention that the inverse of Theorem 3.7 is also true, *i.e.*, for any policy π whose optimality and feasibility gaps are sufficiently close to 0, one can always find a feasible solution of CMFOMO whose objective is also arbitrarily close to 0. We shall call this property the *characteristicity* of CMFOMO. The proof of the following theorem is also deferred to Appendix A.

Theorem 3.10 (Characteristicity of CMFOMO)**.** *Let* $\epsilon_1, \epsilon_2 \geq 0$ *and fix* $c_i > 0$ $(i = 1, 2, 3, 4, 5)$. *Under Assumption 1 and Assumption 2, let* π *be a policy such that* $|G_{opt}(\pi)| \leq \epsilon_1$ *and* $0 \leq G_{fea}(\pi) \leq \epsilon_2$. *Then, by choosing* $L = \Psi(\pi)$, $w = \max(0, \gamma_0 - c_L L)$ *and* (d, y, z, λ) *to be a solution of* $\mathcal{K}(L)$ (i.e., *the KKT Conditions induced by* L), (L, y, z, λ, w) *will be a feasible solution of CMFOMO defined in Theorem 3.6 with:*

$$0 \leq \textit{objective}\ (\text{CMFOMO}) \leq \epsilon_1(c_3 + 2c_5) + \epsilon_2\left[\sqrt{k}\frac{|\mathcal{T}|r_{\max}}{\delta}(c_3 + c_5) + c_4\right].$$

4 Population-Level Constrained Mean-Field Games

In this section, we consider a special class of CMFGs where the cost function is independent of agent state and action, without the strictly feasibility assumption. As we shall see in the next, such CMFGs are degenerate since the representative agent now solves an unconstrained MDP problem. Following a similar procedure as Section 3, we show that there is a one-to-one correspondence between the CMFG NEs and the optimal solutions of an optimization problem. Finally, we close this section by showing that, roughly speaking, population-level CMFG NEs form a subset of agent-population-level CMFG NEs.

Despite their similarity, we emphasize that Section 4 is not a special case of Section 3, as Assumption 1 is no longer imposed. Applying this assumption in the population-level setting would render the constraints redundant, since they are automatically satisfied for any admissible policy.

4.1 Optimization Framework for Solving NE in CMFGs

Firstly, under population-level constraints, the CMDP problem $\mathcal{M}_c(L)$ reduces to an unconstrained MDP problem (provided that the constraints are satisfied under L). We denote by $\mathcal{M}(L)$ the following unconstrained version of $\mathcal{M}_c(L)$:

Unconstrained MDP.

$$\begin{aligned} \text{maximize}_{\{\pi_t\}_{t\in\mathcal{T}}} \quad & \mathbb{E}\left[\sum_{t\in\mathcal{T}} r_t(s_t, a_t, L_t) \,\middle|\, s_0 \sim \mu_0\right] \\ \text{subject to} \quad & a_t \sim \pi_t(s_t), \quad s_{t+1} \sim p_t(s_t, a_t, L_t). \end{aligned}$$

Similar to Lemma 3.2, $\mathcal{M}(L)$ can be transformed into an LP problem (*i.e.*, one only needs to remove the constraints $c_L d \le \gamma_0$ with the rest remaining the same), and it has the following KKT conditions:

KKT Conditions (Population-level).

$$\begin{aligned} \text{find} \quad & \left\{ d \in \mathbb{R}^{|\mathcal{S}||\mathcal{A}||\mathcal{T}|}, y \in \mathbb{R}^{|\mathcal{S}||\mathcal{T}|}, z \in \mathbb{R}^{|\mathcal{S}||\mathcal{A}||\mathcal{T}|} \right\} \\ \text{such that} \quad & -r_L = A_L^\top y + z, \quad A_L d = b, \quad d \ge 0, \quad z \ge 0, \quad z^\top d = 0. \end{aligned}$$

In the rest, we denote the above KKT conditions by $\mathcal{K}^p(L)$.

Then, a careful check reveals that, under population-level constraints, (π, L) is a CMFG NE defined by Definition 1 if and only if (π, L) is an MFG NE and the constraints are satisfied under L:

Proposition 4.1. *Suppose the constraints are population-level. (π, L) is a CMFG NE defined by Definition 1 if and only if:*

1) π is an optimal policy of $\mathcal{M}(L)$;

2) $L = \Psi(\pi)$;

3) L satisfies (CMDP constraints).

Remark 4.1. *By Proposition 4.1, a CMFG NE exists if and only if there exists an MFG NE (π^0, L^0) such that $\gamma_0 \ge \sum_{t\in\mathcal{T}} c_t(L_t^0)$.*

Next, we present a result parallel to Theorem 3.6, which characterizes the CMFG NEs as the optimal solutions to a single optimization problem. We note that Assumption 1 is not imposed in the rest of this section.

Theorem 4.2 (Population-level CMFOMO)**.** *Suppose the constraints are population-level. Let $c_i > 0$ $(i = 1, 2, 3, 4)$. (π^*, L^*) is a CMFG NE defined by Definition 1 if and only if there exists (y^*, z^*, w^*) such that (L^*, y^*, z^*, w^*) is an optimal solution of the following optimization problem with the optimal objective value being 0:*

$$\begin{aligned} \textit{minimize}_{L,y,z,w} \quad & c_1 \left|\left|A_L^\top y + z + r_L\right|\right|_2 + c_2 \left|\left|A_L L - b\right|\right|_2 + c_3(z^\top L) + c_4 \left|\left|\gamma_0 - c_L L - w\right|\right|_2 \\ & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(CMFOMO population)} \\ \textit{subject to} \quad & L \geq 0, \quad \mathbf{1}^\top L_t = 1, \quad \forall t \in \mathcal{T}, \\ & ||y||_1 \leq \frac{|\mathcal{S}||\mathcal{T}|(|\mathcal{T}|+1)}{2} r_{\max}, \qquad (11) \\ & z \geq 0, \quad ||z||_1 \leq |\mathcal{S}||\mathcal{A}| \left(|\mathcal{T}|^2 - |\mathcal{T}| + 2\right) r_{\max}, \qquad (12) \\ & 0 \leq w \leq \gamma_0. \end{aligned}$$

Remark 4.2. *Theorem 4.2 provides an alternative way to identify the existence of CMFG NEs without Assumption 1, that is, a CMFG NE exists if and only if the optimal objective value of the above optimization problem is 0.*

The proof of Theorem 4.2 follows directly by combining Proposition 4.1 and [17, Theorem 5]. And we mention that the upper bounds on the RHS of (11) and (12) come from [17, Corollary 7].

Let π be an arbitrary policy. Its optimality gap (*cf.* Definition 2) is only well-defined when $\Psi(\pi)$ satisfies the population-level constraints, *i.e.*, $G_{\text{fea}}(\pi) = 0$. Therefore, the bounds (7) on G_{opt} in Theorem 3.7 are no longer valid. Alternatively, we have the following modified version of Theorem 3.7, whose proof is similar to that of Theorem 3.7.

Theorem 4.3 (Suboptimality of population-level CMFOMO)**.** *Let $\epsilon \geq 0$. Under Assumption 2, suppose (L, y, z, w) is a feasible solution of CMFOMO defined in Theorem 4.2 with its objective*

$$0 \leq \text{(CMFOMO population)} \leq \epsilon.$$

Then, for any $\pi \in \Pi(L)$, its feasibility gap is bounded by:

$$0 \leq G_{fea}(\pi) \leq \epsilon\zeta_3,$$

where ζ_3 is defined in (10)*. If specially $G_{fea}(\pi) = 0$, then the optimality gap will be well-defined and bounded by:*

$$0 \leq G_{opt}(\pi) \leq \epsilon\Big[(|\mathcal{T}|+1)\zeta_1' + \zeta_2'\Big],$$

where

$$\begin{aligned} \beta_y' &:= \frac{|\mathcal{S}||\mathcal{T}|(|\mathcal{T}|+1)}{2} r_{\max}, \\ \beta_z' &:= |\mathcal{S}||\mathcal{A}|(|\mathcal{T}|^2 - |\mathcal{T}| + 2) r_{\max}, \\ \zeta_1'(c_1, c_2) &:= \frac{1}{c_1} + \alpha(c_2)\left(C_p\beta_y' + C_r\right), \\ \zeta_2'(c_2, c_3) &:= \frac{1}{c_3} + \beta_z'\alpha(c_2), \end{aligned}$$

and α is defined in (9)*.*

A similar result to Theorem 3.10 can also be established by requiring $\epsilon_2 = 0$, as formally stated below.

Theorem 4.4 (Characteristicity under population-level constraints)**.** *Suppose the constraints are population-level. Let $\epsilon_1 \geq 0$ and fix $c_i > 0$ $(i = 1, 2, 3, 4)$. Let π be a policy such that $G_{fea}(\pi) = 0$ and $0 \leq G_{opt}(\pi) \leq \epsilon_1$. Then, by choosing $L = \Psi(\pi)$, $w = 0$ and (d, y, z) to be a solution of $\mathcal{K}^p(L)$ (*i.e.*, the KKT Conditions induced by L), (L, y, z, w) will be a feasible solution of CMFOMO defined in Theorem 4.2 with its objective:*

$$0 \leq \text{(CMFOMO population)} \leq \epsilon_1 c_3.$$

4.2 Connection between Population-Level and Agent-Population-Level Constraints

We end this section with a discussion on the connection between population-level and agent-population-level constraints. Our main result is that population-level CMFG NEs comprise of a subset of agent-population-level CMFG NEs.

Definition 4. *Let c^p (*resp. *c^a) be the cost function of population-level (*resp. *agent-population-level) constraints. c^p and c^a are said to be* twinned *if for any mean-field flow L, it holds that:*

$$\sum_{t\in\mathcal{T}} c_t^p\big(L_t\big) = \sum_{t\in\mathcal{T}} \sum_{s\in\mathcal{S},a\in\mathcal{A}} c_t^a\big(s,a,L_t\big)L_t(s,a).$$

The constraints (4) and (5) in Example 1 are an example of twinned cost functions. The point of Definition 4 is that for any policy π, if π is feasible in $\mathcal{M}_{c^p}\big(\Psi(\pi)\big)$, then π will also be feasible in $\mathcal{M}_{c^a}\big(\Psi(\pi)\big)$. Furthermore, if π is optimal in $\mathcal{M}_{c^p}\big(\Psi(\pi)\big)$ (which is equivalent to π being optimal in $\mathcal{M}\big(\Psi(\pi)\big)$), then π will also be optimal in $\mathcal{M}_{c^a}\big(\Psi(\pi)\big)$. This argument proves the following Theorem 4.5. And we mention that the calculation results in Example 1 conform to Theorem 4.5.

Theorem 4.5. *Let c^p and c^a be a pair of twinned cost functions at the population level and agent population level, respectively. If (π, L) is a CMFG NE defined by Definition 1 under c^p, then (π, L) will also be a CMFG NE under c^a.*

5 Approximation of NE in Finite-player Games with Agent-population-level Constraints

It is well-known that NEs of MFGs are approximated NEs of N-player games. In this section, we provide the corresponding analysis and show that NEs of CMFGs are also approximated NEs of the counterpart constrained N-player games, with an explicit dependency of the approximation error on the number of players N. Throughout this section, we always assume that the constraints are agent-population-level (*cf.* Section 3).

Let $0 < N < +\infty$. We assume that the N players are homogeneous and have mean-field-type interactions. In the following, we prove that the NE of CMFGs (*cf.* Definition 1) is an ϵ-NE of N-player games, and that ϵ can be made arbitrarily close to 0 as N tends to infinity. In the rest, we inherit the notations in §2, and we use the superscript $^{(i)}$ to indicate the private variables of the i-th player.

To begin with, for each $1 \le i \le N$, the constrained control problem of the i-th player writes:

Constrained Control Problem of the i-th Player.

$$\begin{aligned}
\text{maximize}_{\left\{\pi_t^{(i)}\right\}_{t\in\mathcal{T}}} \quad & \mathbb{E}\left[\sum_{t\in\mathcal{T}} r_t\left(s_t^{(i)}, a_t^{(i)}, L_t^{(N)}\right) \,\middle|\, s_0^{(i)} \sim \mu_0\right] && (N \text{ reward})\\
\text{subject to} \quad & a_t^{(i)} \sim \pi_t^{(i)}\left(s_t^{(i)}\right), \quad s_{t+1}^{(i)} \sim p_t\left(s_t^{(i)}, a_t^{(i)}, L_t^{(N)}\right), && (N \text{ dynamics})\\
& \mathbb{E}\left[\sum_{t\in\mathcal{T}} c_t\left(s_t^{(i)}, a_t^{(i)}, L_t^{(N)}\right) \,\middle|\, s_0^{(i)} \sim \mu_0\right] \le \gamma_0, && (N \text{ constraints})
\end{aligned}$$

where $L_t^{(N)}$ denotes the empirical distribution of the N players at time step t, *i.e.*,

$$\forall t \in \mathcal{T}, s \in \mathcal{S}, a \in \mathcal{A}, \quad L_t^{(N)}(s,a) := \frac{1}{N}\sum_{i=1}^{N} \delta_{\left(s_t^{(i)}, a_t^{(i)}\right)=(s,a)}.$$

For convenience, with an abuse of notation, we use $\mathcal{M}_c^{(i)}$ to denote the above constrained control problem of the i-th player, and we denote by $V^{(i)}\left(\pi^{(1)}, \cdots, \pi^{(i)}, \cdots, \pi^{(N)}\right)$ the expected cumulative reward of the i-th player when the players take policies $\left\{\pi^{(i)}\right\}_{1\le i\le N}$.

Remark 5.1. *Unlike that the players are independent in the constrained MDP problem $\mathcal{M}_c(L)$ defined in §2, the N players in $\left\{\mathcal{M}_c^{(i)}\right\}_{1\le i\le N}$ are correlated. This is because the decision of a specific player does affect the law of the empirical distribution of the N players, which in turn affects the evolution of other players. Another consequence is that $\mathcal{M}_c^{(i)}$ is no longer a traditional MDP even when the policies of the rest players are fixed.*

Next, we generalize the concept of feasibility gap (*cf.* Definition 3) to the N-player setting.

Definition 5 (Feasibility gap of the i-th player)**.** *Denote by $\left\{\pi^{(i)}\right\}_{1\le i\le N}$ the policies of the N players. For any $1\le i\le N$, we define the feasibility gap of the i-th player by:*

$$G_{fea}^{(i)}\left(\pi^{(1)},\cdots,\pi^{(i)},\cdots,\pi^{(N)}\right):=\left|\left|\min\left(0,\gamma_0-\gamma^{(i)}\left(\pi^{(1)},\cdots,\pi^{(i)},\cdots,\pi^{(N)}\right)\right)\right|\right|_2,$$

where min *is taken entry-wise, and*

$$\gamma^{(i)}\left(\pi^{(1)},\cdots,\pi^{(i)},\cdots,\pi^{(N)}\right):=\mathbb{E}_{\forall 1\le j\le N,\,a^{(j)}\sim\pi^{(j)}}\left[\sum_{t\in\mathcal{T}}c_t\left(s_t^{(i)},a_t^{(i)},L_t^{(N)}\right)\,\middle|\,s_0^{(i)}\sim\mu_0\right].$$

Given the policies of all but the i-th player, we denote by $\mathcal{F}^{(i)}\left(\pi^{(1)},\cdots,\pi^{(i-1)},\pi^{(i+1)},\cdots,\pi^{(N)}\right)$ the set of policies of the i-th player whose feasibility gap is 0 (*i.e.*, the set of feasible policies of the i-th player provided the policies of the rest players). To ease the exposure of notations, we abbreviate it by $\mathcal{F}^{(i)}$ if the policies of the rest players are clearly indicated in the context, similarly for $G_{\text{fea}}^{(i)}$.

Definition 6 (ϵ-NE of N-player games)**.** *Let $\epsilon_1,\epsilon_2\ge 0$. The set of policies $\left\{\pi^{(i)}\right\}_{1\le i\le N}$ is called an (ϵ_1,ϵ_2)-NE of the N-player game $\left\{\mathcal{M}_c^{(i)}\right\}_{1\le i\le N}$ if for any $1\le i\le N$:*

$$\sup_{\pi\in\mathcal{F}^{(i)}}V^{(i)}\left(\pi^{(1)},\cdots,\pi^{(i-1)},\pi,\pi^{(i+1)},\cdots,\pi^{(N)}\right)\le V^{(i)}\left(\pi^{(1)},\cdots,\pi^{(i-1)},\pi^{(i)},\pi^{(i+1)},\cdots,\pi^{(N)}\right)+\epsilon_1.$$

and $G_{fea}^{(i)}\le\epsilon_2$. Specially, $\left\{\pi^{(i)}\right\}_{1\le i\le N}$ is called an NE if $\epsilon_1=\epsilon_2=0$.

However, Definition 6 is not well-defined when $\mathcal{F}^{(i)}$ is empty. To ensure that $\mathcal{F}^{(i)}$ is non-empty (especially when N is large enough), we need to slightly strengthen Assumption 1 by the following assumption.

Assumption 5 (Strengthened strict feasibility)**.** *There exists $\delta>0$, such that for any mean-field flow L, there exists a policy π that satisfies:*

$$\mathbb{E}_{a\sim\pi}\left[\sum_{t\in\mathcal{T}}c_t(s_t,a_t,L_t)\,\middle|\,s_0\sim\mu_0\right]\le\gamma_0-\delta.$$

Remark 5.2. *Assumption 1 only requires the existence of strictly feasible policy for each of the constraints. The feasible policies can be different across different constraints. However, in Assumption 5, we require the existence of a single policy that is strictly feasible for all constraints. When there is only one constraint, i.e., when $k=1$, the two assumptions are equivalent.*

With the new assumption, we present the main result of this section, whose proof can be found in Appendix B.

Theorem 5.1 (ϵ-NE approximated by CMFGs)**.** *Under Assumptions 4 and 5, suppose (π^*,L^*) is a CMFG NE defined by Definition 1. Fix $0<\epsilon\le\delta$ and let $N>0$ be such that:*

$$\left(\frac{1}{2\sqrt{N}}+\frac{2}{N}\right)\tilde{C}\le\epsilon,$$

where

$$\tilde{C} := C_p|\mathcal{S}||\mathcal{A}|(|\mathcal{T}|-1)^2\frac{C_{p,\mathcal{S},\mathcal{A}}^{|\mathcal{T}|}-1}{C_{p,\mathcal{S},\mathcal{A}}-1}\max\Big\{C_r+r_{\max},\, C_c+c_{\max}\Big\},$$
$$C_{p,\mathcal{S},\mathcal{A}} := (C_p+1)\,|\mathcal{S}||\mathcal{A}|.$$

Then, $\{\pi^*\}_{1\le i\le N}$ *(*i.e.*, N copies of π^*) is an (ϵ_1,ϵ_2)-NE of the N-player game defined by Definition 6, where*

$$\epsilon_1 = \epsilon\left(2+\frac{k|\mathcal{T}|r_{\max}}{\delta}\right),$$
$$\epsilon_2 = \epsilon\sqrt{k}.$$

Let $\epsilon_0 > 0$ be a small constant. As a corollary of Theorem 5.1, any CMFG NE under $\gamma_0-\epsilon_0$ will be strictly feasible in the finite-player game under γ_0 when the number of players is large enough. Same as Theorem 3.8, the improved feasibility comes at the expense of a worse optimality.

Corollary 5.2. *Under Assumptions 4 and 5, let $0<\epsilon_0<\delta$ be a constant. Suppose (π^*,L^*) is a CMFG NE defined by Definition 1 by replacing γ_0 with $\gamma_0-\epsilon_0$. Fix $0<\epsilon\le\epsilon_0$ and let $N>0$ be such that:*

$$\left(\frac{1}{2\sqrt{N}}+\frac{2}{N}\right)\tilde{C}\le\epsilon,$$

where $\tilde{C}$ is defined in Theorem 5.1. Then, $\{\pi^*\}_{1\le i\le N}$ *(*i.e.*, N copies of π^*) is an $(\tilde{\epsilon}_1,0)$-NE of the N-player game (under γ_0) defined by Definition 6, where*

$$\tilde{\epsilon}_1 = \epsilon_0\frac{k|\mathcal{T}|r_{\max}}{\delta-\epsilon_0}+\epsilon\left(2+\frac{k|\mathcal{T}|r_{\max}}{\delta}\right).$$

Finally, by connecting Theorem 5.1 with Theorem 3.7, we show that CMFOMO can be used to obtain ϵ-NEs of the finite-player game.

Corollary 5.3 (ϵ-NE approximated by CMFOMO)**.** *Under Assumptions 4 and 5, let $\epsilon'>0$ and suppose (L,y,z,λ,w) is a feasible solution of CMFOMO defined in Theorem 3.6 with its objective* (CMFOMO) $\le\epsilon'$. *Fix $0<\epsilon\le\delta$ and let N be such that:*

$$\left(\frac{1}{2\sqrt{N}}+\frac{2}{N}\right)\tilde{C}\le\epsilon,$$

where $\tilde{C}$ is defined in Theorem 5.1. Then, for any $\pi\in\Pi(L)$, one can show that $\{\pi\}_{1\le i\le N}$ *(*i.e.*, N copies of π) is an $(\epsilon_1',\epsilon_2')$-NE of the N-player game defined by Definition 6, where*

$$\epsilon_1' := \epsilon'\Big[(|\mathcal{T}|+1)\zeta_1+\zeta_2+\zeta_4\Big]+\epsilon\left(2+\frac{k|\mathcal{T}|r_{\max}}{\delta}\right),$$
$$\epsilon_2' := \epsilon'\zeta_3+\epsilon\sqrt{K}.$$

We refer to Theorem 3.7 for the definition of $\zeta_1,\zeta_2,\zeta_3,\zeta_4$.

6 Numerical Experiments

In this section, we illustrate our CMFOMO framework developed before on the modified SIS problem introduced in Example 1. We will discuss different types of constraints including both agent-population-level constraints and population-level constraints and how they affect the solutions of the MFGs. Our implementation is based on PyTorch and MFGLib proposed in [31].

In the following numerical experiments, we set $T=10$ and initialize the population with $\mu_0(s_0=I)=0.5$. The reward functions and transition probabilities follow those in Example 1. Note that, because of the different choices of T and μ_0, the threshold value of γ_0 ensuring the existence of a Nash equilibrium differs from that in Example 1.

6.1 Agent-population-level constraint on state

In this section, we impose a condition on the agents by restricting the probability of each agent getting infected to be upper bounded by a certain threshold. Mathematically speaking, we consider the following agent-population-level constraint on state:

$$\frac{1}{|\mathcal{T}|}\mathbb{E}\left(\sum_{t\in\mathcal{T}}\delta_{s_t=I}\,\middle|\, s_0\sim\mu_0\right)\leq\gamma_0. \tag{13}$$

Note that this setting satisfies Assumptions 1 and 2, so there exists a CMFG NE defined by Definition 1 according to Theorem 3.1.

We use Projected Gradient Descent (PGD) to optimize the objective function (CMFOMO) of CMFOMO defined in Theorem 3.6, combined with the Adam optimizer embedded in PyTorch. Inspired by Theorem 3.10, in each iteration, once L and λ are determined, we solve y and z from the following modified KKT conditions (d is discarded):

$$\begin{aligned}\text{find}\quad & \left\{d\in\mathbb{R}^{|\mathcal{S}||\mathcal{A}||\mathcal{T}|}, y\in\mathbb{R}^{|\mathcal{S}||\mathcal{T}|}, z\in\mathbb{R}^{|\mathcal{S}||\mathcal{A}||\mathcal{T}|}\right\}\\ \text{such that}\quad & -r_L+c_L^\top\lambda=A_L^\top y+z,\quad A_Ld=b,\quad d\geq 0,\quad z\geq 0,\quad z^\top d=0,\end{aligned}$$

and determine w the same way as Theorem 3.10, *i.e.*, $w=\max(0,\gamma_0-c_LL)$. This helps our algorithm to converge faster. In a word, the only variables to be optimized are L and λ, and our optimization procedure is summarized below:

Algorithm 1 Optimization of CMFOMO defined in Theorem 3.6

Require: Threshold γ_0; initial values for (L^0,λ^0); loss weights c_i $(i=1,2,3,4,5)$ for CMFOMO; gradient optimizer; maximum number of iterations N.

$n\leftarrow 0$
while $n<N$ **do**
  Compute (y^n,z^n,w^n) by the modified KKT conditions and Theorem 3.10
  $\text{loss}^n\leftarrow\text{CMFOMO}(L^n,y^n,z^n,\lambda^n,w^n)$ ▷ Objective (CMFOMO) of CMFOMO
  loss^n.backward()
  $\hat{L}^{n+1}\leftarrow\text{optimizer}(n,L^n,L^n.\text{grad}),\quad\hat{\lambda}^{n+1}\leftarrow\text{optimizer}(n,\lambda^n,\lambda^n.\text{grad})$
  $L^{n+1}\leftarrow\text{proj}_L(\hat{L}^{n+1}),\quad\lambda^{n+1}\leftarrow\text{proj}_\lambda(\hat{\lambda}^{n+1})$ ▷ Ensure L^{n+1} a mean-field flow and $\lambda^{n+1}\geq 0$
  $n\leftarrow n+1$
end while
return L^N

As for other hyper-parameters, the initial value L^0 is induced by the uniform policy, *i.e.*, the policy that chooses all actions with an equal probability. The initial value λ^0 is chosen to be 0. We choose an equal weight for CMFOMO, *i.e.*, $c_1=c_2=c_3=c_4=c_5=1$. Finally, we set the learning rate of the Adam optimizer to be 5×10^{-3}.

When γ_0 is set 0.25, in Figure 1 (Left), we plot the evolution of G_{opt} and G_{fea} with respect to optimization iteration. In the figure legends, "G_{opt}" and "G_{fea}" represent the optimality gap and feasibility gap of the policy at the current iteration, respectively, and "G^0_{opt}" refers to the optimality gap of the initial policy. We note that G_{fea}/γ_0 (see the orange line) is initially very small because the starting policy is feasible. As the iterations proceed, the policy quickly moves toward minimizing G_{opt}, temporarily sacrificing feasibility. Nevertheless, G_{fea} eventually converges to an acceptable region. In Figure 1 (Right), we plot the values of $\left\{\mathbb{E}\left(\delta_{s_t=I}\right)\right\}_{t\in\mathcal{T}}$ under the optimized policy, *i.e.*, the percentage of infected agents at each time step. Under the optimized policy, the LHS of (13) is 0.2521, slightly exceeding the threshold $\gamma_0=0.25$.

Remark 6.1. *Here the optimized policy is not feasible under $\gamma_0 = 0.25$. By Theorem 3.8, in order to produce a feasible policy under $\gamma_0 = 0.25$, one practical way is to optimize CMFOMO with γ_0 replaced by $\gamma_0 - \epsilon_0$ for some small $\epsilon_0 > 0$. For illustration, in Table 1, we vary $\epsilon_0 \in [0.01, 0.05]$ and list the optimality gap (under $\gamma_0 = 0.25$) and average infected agents (*i.e.*, *the LHS of* (13)*) of the optimized policies. As observed, with the increasing of ϵ_0, the feasibility is improved at the expense of a worse optimality gap.*

ϵ_0	$G_{\text{opt}}(\pi)$ under $\gamma_0 = 0.25$	Average infected agents
0	0.1625	0.2521
0.01	0.1642	0.2513
0.02	0.1663	0.2493
0.03	0.1664	0.2486
0.04	0.1675	0.2473
0.05	0.1880	0.2469

Table 1: Optimality gap (under $\gamma_0 = 0.25$) and average infected agents (*i.e.*, the LHS of (13)) of the policy optimized under the threshold $\gamma_0 - \epsilon_0$ (*cf.* Remark 6.1).

We then vary the threshold γ_0 to examine how the equilibrium changes with different constraint levels. Figure 2 (Left) plots the percentage of infected agents (*i.e.*, $s_t = I$). As expected, the infection rate decreases as γ_0 becomes smaller, reflecting a stronger constraint. Figure 2 (Right) shows the proportion of agents choosing to go out (*i.e.*, $a_t = U$). Two clear patterns emerge: (1) for small γ_0 values (e.g., $\gamma_0 \leq 0.3$), agents initially prefer to stay home ($a_t = D$) to reduce infection risk, and once the infection rate drops sufficiently, they gradually resume going out to increase rewards; (2) for larger γ_0, where the constraint is relatively loose, agents go out from the start to maximize rewards. Moreover, we observe that the optimized policies coincide once γ_0 exceeds a certain threshold; for instance, the trajectories for $\gamma_0 = 0.5$ and $\gamma_0 = 1$ nearly overlap in Figure 2. This is because the constraint becomes inactive and the solution of the CMFG is the same as the the unconstrained MFG.

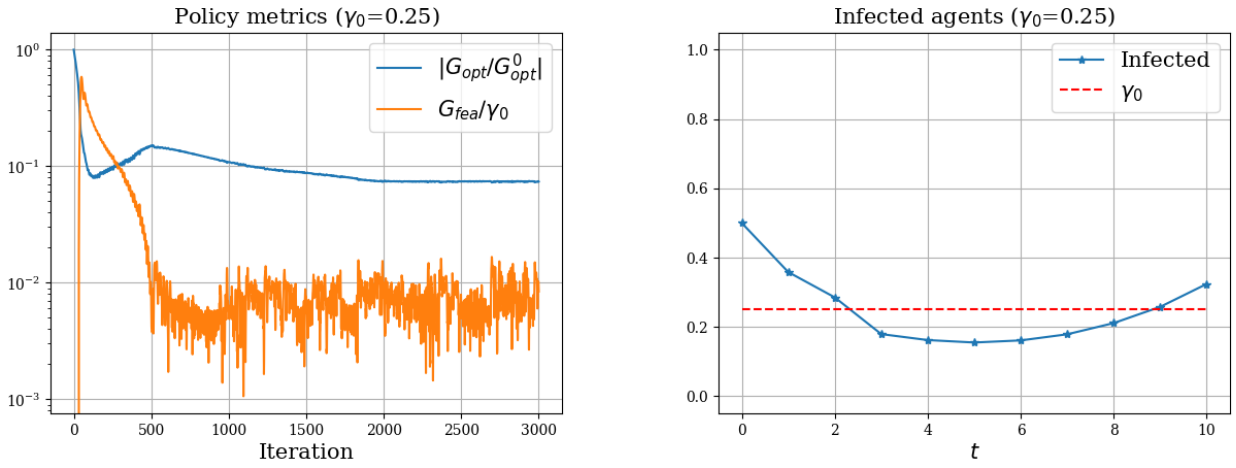

Figure 1: Optimized policy of the constrained SIS model when $\gamma_0 = 0.25$. Left: the evolution of G_{opt} and G_{fea} with respect to optimization iteration, where G^0_{opt} represents the optimality gap of the initial policy; Right: the percentage of infected agents at each time step under the optimized policy, where the red dotted line represents the threshold γ_0 of the agent-population-level constraint (13).

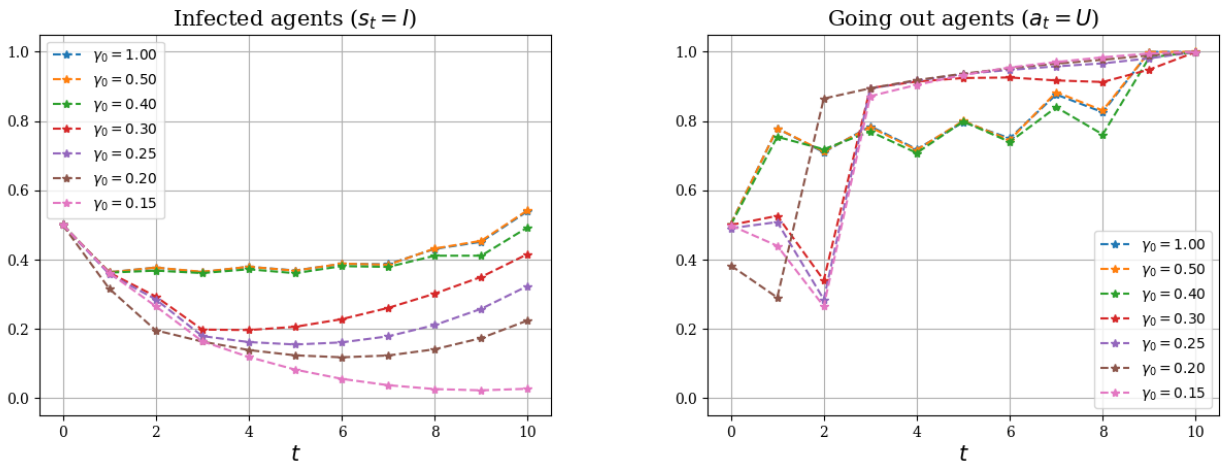

Figure 2: Sensitivity analysis under the agent-population-level constraint (13). The x-axis represents the time steps and the y-axis represents the percentage of infected agents (Left) or agents who go out (Right). The policies are optimized when γ_0 takes different values.

6.2 Agent-population-level constraint on action

In this section, we study a different agent-population-level constraint which guarantees that the probability of each agent going out is lower bounded by a certain threshold. Specifically, we impose the following agent-population-level constraint on action:

$$\frac{1}{|\mathcal{T}|}\mathbb{E}\left(\sum_{t\in\mathcal{T}}\delta_{a_t=U}\,\middle|\, s_0\sim\mu_0\right)\geq\gamma_0. \tag{14}$$

According to Theorem 3.1, a CMFG NE exist as long as $\gamma_0 < 1$. In Figure 3, we plot the percentage of infected (Left) and going out (Right) agents under different γ_0 values. The optimization procedure is exactly the same as Section 6.1. As expected, the larger the value of γ_0, the higher the infection rate will be, *i.e.*, the agents will have to go out more frequently. In addition, when γ_0 is smaller than a certain value, the optimized policies coincide (*e.g.*, the trajectories corresponding to $\gamma_0 = 0.7$ and 0.6 overlap well in Figure 3).

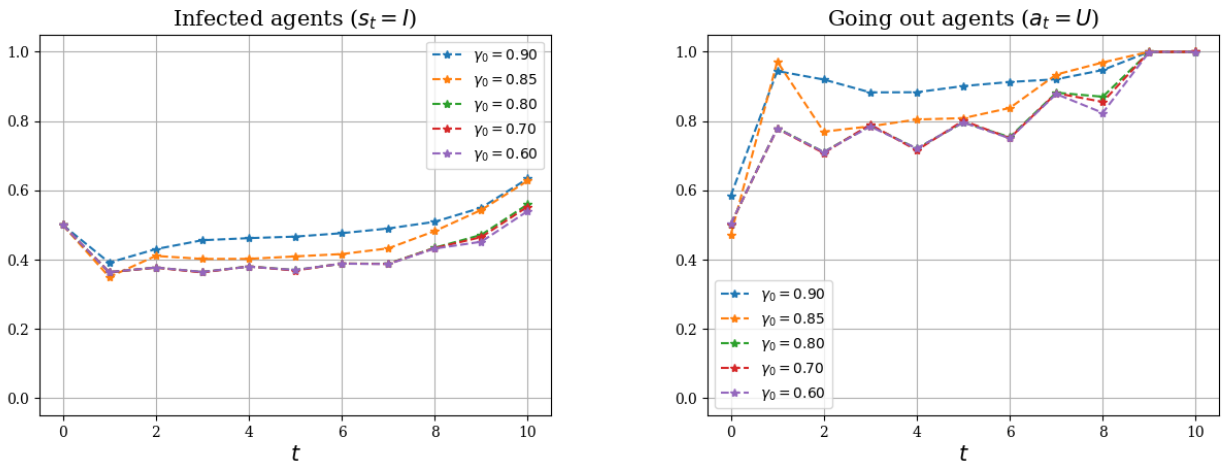

Figure 3: Sensitivity analysis under the agent-population-level constraint (14). The x-axis represents the time steps and the y-axis represents the percentage of infected agents (Left) or agents who go out (Right). The policies are optimized when γ_0 takes different values.

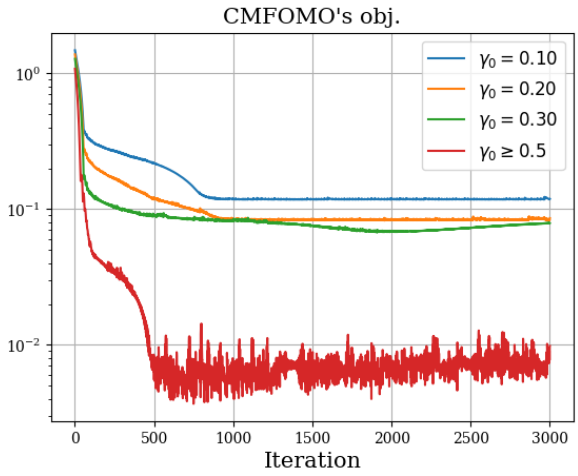

Figure 4: Evolution of the objective function (CMFOMO population) of CMFOMO defined in Theorem 4.2 under the population-level constraint (15). The x-axis represents the optimization iteration, and the y-axis represents the objective value. Different lines correspond to different values of γ_0.

6.3 Population-level constraint on state

Finally, in this section, we consider the following population-level constraint on agent state:

$$\frac{1}{|\mathcal{T}|}\sum_{t\in\mathcal{T}} L(s_t = I) \leq \gamma_0. \tag{15}$$

When optimizing the objective function (CMFOMO population) of CMFOMO defined in Theorem 4.2, we devise an algorithm similar to Algorithm 1. The core idea is that (CMFOMO population) can be viewed as plugging in $\lambda = 0$ into (CMFOMO). By Proposition 4.1 and Remark 4.1, the objective function (CMFOMO population) admits a zero point if and only if γ_0 is no smaller than a certain value.

In Figure 4, we plot the evolution of the value of the objective function (CMFOMO population) with respect to optimization iteration, with γ_0 taking different values. As shown, when γ_0 is large enough such that a CMFG NE exists (*e.g.*, $\gamma_0 \geq 0.5$), the objective decreases fast. In contrast, when γ_0 is smaller such that a CMFG NE does not exist, the objective stays at a relatively higher level.

By Proposition 4.1, under population-level constraints, if a CMFG NE exists, then it must also be an unconstrained MFG NE. In Figure 5, we plot the mean-field flow induced by the optimized policy when $\gamma_0 = 0.5$ which is large enough such that a CMFG NE exists. By observation, the trajectories in Figure 5 coincide with the trajectories corresponding to $\gamma_0 \geq 0.5$ in Figure 2. This is intuitive, since both the agent-population-level constraint (13) and the population-level constraint (15) become redundant when the value of γ_0 is large enough.

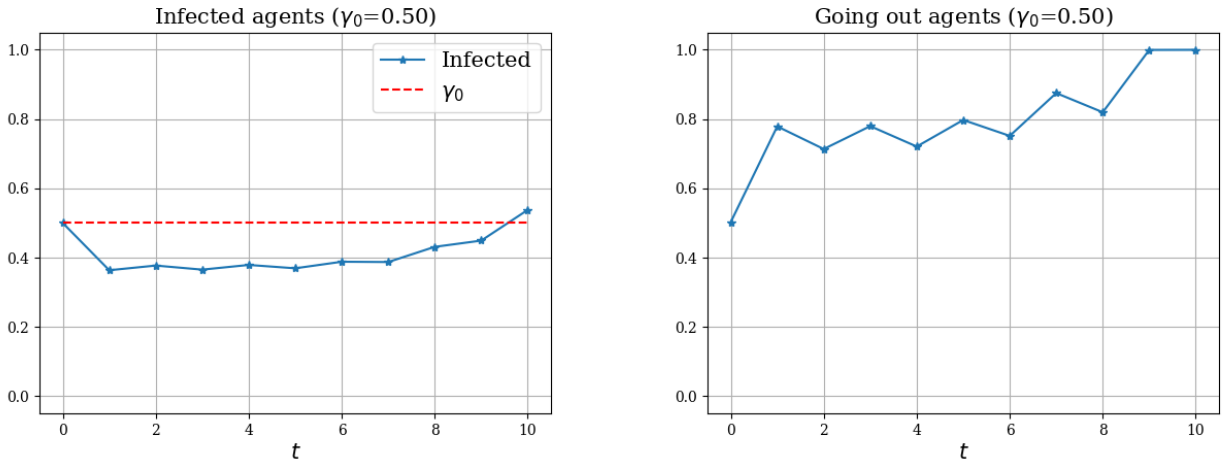

Figure 5: Mean-field flow induced by the optimized policy under the population-level constraint (15) with $\gamma_0 = 0.5$. The x-axis represents the time steps and the y-axis represents the percentage of infected agents (Left) or agents who go out (Right).

References

[1] Yaodong Yang and Jun Wang. An overview of multi-agent reinforcement learning from game theoretical perspective. arXiv preprint arXiv:2011.00583, 2020.

[2] Kaiqing Zhang, Zhuoran Yang, and Tamer Başar. Multi-agent reinforcement learning: A selective overview of theories and algorithms. Handbook of reinforcement learning and control, pages 321–384, 2021.

[3] Minyi Huang, Roland P Malhamé, and Peter E Caines. Large population stochastic dynamic games: closed-loop mckean-vlasov systems and the nash certainty equivalence principle. 2006.

[4] Jean-Michel Lasry and Pierre-Louis Lions. Mean field games. Japanese journal of mathematics, 2(1):229–260, 2007.

[5] François Delarue, Daniel Lacker, and Kavita Ramanan. From the master equation to mean field game limit theory: a central limit theorem. 2019.

[6] Naci Saldi, Tamer Basar, and Maxim Raginsky. Markov–Nash equilibria in mean-field games with discounted cost. SIAM Journal on Control and Optimization, 56(6):4256–4287, 2018.

[7] Eitan Altman. Constrained Markov Decision Processes. Routledge, 2021.

[8] Harry Markowitz. Portfolio selection. The Journal of Finance, 7(1):77–91, 1952.

[9] Andrew J Schaefer, Matthew D Bailey, Steven M Shechter, and Mark S Roberts. Modeling medical treatment using Markov decision processes. Operations research and health care: A handbook of methods and applications, pages 593–612, 2004.

[10] Akifumi Wachi and Yanan Sui. Safe reinforcement learning in constrained markov decision processes. In International Conference on Machine Learning, pages 9797–9806. PMLR, 2020.

[11] Eitan Altman and Adam Shwartz. Markov decision problems and state-action frequencies. SIAM journal on control and optimization, 29(4):786–809, 1991.

[12] Eitan Altman. Denumerable constrained markov decision processes and finite approximations. Mathematics of operations research, 19(1):169–191, 1994.

[13] Eitan Altman. Constrained Markov decision processes with total cost criteria: Occupation measures and primal LP. Mathematical methods of operations research, 43(1):45–72, 1996.

[14] Thomas G Kurtz and Richard H Stockbridge. Linear programming formulations of singular stochastic control problems: time-homogeneous problems. arXiv preprint arXiv:1707.09209, 2017.

[15] Géraldine Bouveret, Roxana Dumitrescu, and Peter Tankov. Mean-field games of optimal stopping: a relaxed solution approach. SIAM Journal on Control and Optimization, 58(4):1795–1821, 2020.

[16] Roxana Dumitrescu, Marcos Leutscher, and Peter Tankov. Control and optimal stopping mean field games: a linear programming approach. Electronic Journal of Probability, 26:1–49, 2021.

[17] Xin Guo, Anran Hu, and Junzi Zhang. MF-OMO: An optimization formulation of mean-field games. SIAM Journal on Control and Optimization, 62(1):243–270, 2024.

[18] Xin Guo, Anran Hu, Jiacheng Zhang, and Yufei Zhang. Continuous-time mean field games: a primal-dual characterization. arXiv preprint arXiv:2503.01042, 2025.

[19] Anran Hu and Junzi Zhang. Mf-oml: Online mean-field reinforcement learning with occupation measures for large population games. arXiv preprint arXiv:2405.00282, 2024.

[20] Piermarco Cannarsa and Rossana Capuani. Existence and uniqueness for mean field games with state constraints. PDE models for multi-agent phenomena, pages 49–71, 2018.

[21] Piermarco Cannarsa, Rossana Capuani, and Pierre Cardaliaguet. Mean field games with state constraints: from mild to pointwise solutions of the pde system. Calculus of Variations and Partial Differential Equations, 60(3):108, 2021.

[22] Piermarco Cannarsa, Wei Cheng, Cristian Mendico, and Kaizhi Wang. Weak kam approach to first-order mean field games with state constraints. Journal of Dynamics and Differential Equations, pages 1–32, 2021.

[23] Saeed Sadeghi Arjmand and Guilherme Mazanti. Nonsmooth mean field games with state constraints. ESAIM: Control, Optimisation and Calculus of Variations, 28:74, 2022.

[24] Jameson Graber and Sergio Mayorga. A note on mean field games of controls with state constraints: existence of mild solutions. arXiv preprint arXiv:2109.11655, 2021.

[25] J Frédéric Bonnans, Justina Gianatti, and Laurent Pfeiffer. A lagrangian approach for aggregative mean field games of controls with mixed and final constraints. SIAM Journal on Control and Optimization, 61(1):105–134, 2023.

[26] Guanxing Fu, Paulwin Graewe, Ulrich Horst, and Alexandre Popier. A mean field game of optimal portfolio liquidation. Mathematics of Operations Research, 46(4):1250–1281, 2021.

[27] Guanxing Fu, Paul Hager, and Ulrich Horst. A mean-field game of market entry. Technical report, CRC TRR 190 Rationality and Competition, 2024.

[28] David Evangelista and Yuri Thamsten. On finite population games of optimal trading. arXiv preprint arXiv:2004.00790, 2020.

[29] Kai Cui and Heinz 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.

[30] David G Luenberger, Yinyu Ye, et al. Linear and nonlinear programming, volume 2. Springer, 1984.

[31] Xin Guo, Anran Hu, Matteo Santamaria, Mahan Tajrobehkar, and Junzi Zhang. MFGLib: A library for mean-field games. arXiv preprint arXiv:2304.08630, 2023.

A Proofs of results in §3.2

Proof of Theorem 3.7

Proof. We shall first prove the upper bound of the optimality gap. This part will be divided into three steps.

$G_{\mathbf{opt}}$ step 1: bound the gap between L and $\Psi(\pi)$.

By the same argument as the proof of [17, Theorem 8], we have:

$$||\Psi_t(\pi) - L_t||_1 \leq \frac{\epsilon}{c_2}\frac{(C_p+1)^{t+1}-1}{C_p}\sqrt{|\mathcal{S}|}, \quad \forall t \in \mathcal{T},$$

which implies:

$$||\Psi(\pi) - L||_1 \leq \frac{\epsilon}{c_2}\frac{(C_p+1)^{|\mathcal{T}|+1}-2C_p-1}{C_p^2}\sqrt{|\mathcal{S}|} =: \epsilon\alpha(c_2).$$

$G_{\mathbf{opt}}$ step 2: near-optimality of $(\Psi(\pi), y, z, \lambda, w)$.

After replacing L with $\Psi(\pi)$ in CMFOMO's objective (CMFOMO), we have the following estimates. For notational simplicity, we write Ψ in the place of $\Psi(\pi)$.

$$\begin{aligned}
\left|\left|A_\Psi^\top y + z + r_\Psi - c_\Psi^\top\lambda\right|\right|_2 &\leq \left|\left|A_\Psi^\top y + z + r_\Psi - c_\Psi^\top\lambda\right|\right|_1 \\
&\leq \left|\left|(A_\Psi - A_L)^\top y\right|\right|_1 + ||r_\Psi - r_L||_1 + \left|\left|(c_\Psi - c_L)^\top\lambda\right|\right|_1 + \frac{\epsilon}{c_1} \\
&\leq ||\Psi - L||_1\big(C_p||y||_1 + C_r + C_c||\lambda||_1\big) + \frac{\epsilon}{c_1} \\
&\leq \epsilon\left[\frac{1}{c_1} + \alpha(c_2)\left(C_p\beta_y + C_r + kC_c\frac{|\mathcal{T}|r_{\max}}{\delta}\right)\right] =: \epsilon\zeta_1(c_1, c_2), \\
A_\Psi\Psi - b &= 0, \\
0 \leq z^\top\Psi &\leq \epsilon\left[\frac{1}{c_3} + \beta_z\alpha(c_2)\right] =: \epsilon\zeta_2(c_2, c_3), \\
||\gamma_0 - c_\Psi\Psi - w||_2 &\leq ||c_\Psi\Psi - c_L L||_2 + \frac{\epsilon}{c_4} \\
&\leq ||c_\Psi(\Psi - L)||_2 + ||(c_\Psi - c_L)L||_2 + \frac{\epsilon}{c_4}
\end{aligned}$$

$$\leq \epsilon \left[\frac{1}{c_4} + \alpha(c_2)\Big(kc_{\max} + C_c|\mathcal{S}||\mathcal{A}||\mathcal{T}|\Big)\right] =: \epsilon\zeta_3(c_2, c_4),$$

$$\begin{aligned}\Big|\lambda^\top(\gamma_0 - c_\Psi\Psi)\Big| &\leq \Big|\lambda^\top(c_\Psi\Psi - c_L L)\Big| + \frac{\epsilon}{c_5} \\ &\leq \epsilon\left[\frac{1}{c_5} + k\alpha(c_2)\frac{|\mathcal{T}|r_{\max}}{\delta}\Big(kc_{\max} + C_c|\mathcal{S}||\mathcal{A}||\mathcal{T}|\Big)\right] \\ &=: \epsilon\zeta_4(c_2, c_5).\end{aligned}$$

G_{opt} step 3: estimate the upper bound.

Throughout this step of the proof, we shall always take the mean-field flow $\Psi(\pi)$. Therefore, we omit the subscript for notational simplicity.

To start with, define $\Delta := A^\top y + z + r - c^\top\lambda$. Consider the following maximization problem:

$$\begin{aligned}\text{maximize}_{\hat{y},\hat{z},\hat{\lambda}} \quad & b^\top\hat{y} - \gamma_0^\top\hat{\lambda} \\ \text{subject to} \quad & A^\top\hat{y} + \hat{z} = c^\top\hat{\lambda} - r + \Delta, \quad \hat{\lambda} \geq 0, \quad \hat{z} \geq 0,\end{aligned}$$

whose optimal solution is denoted by $\Big(\hat{y}^*, \hat{z}^*, \hat{\lambda}^*\Big)$. The dual problem is:

$$\begin{aligned}\text{minimize}_{\hat{d}} \quad & (-r + \Delta)^\top\hat{d} \\ \text{subject to} \quad & A\hat{d} = b, \quad c\hat{d} \leq \gamma_0, \quad \hat{d} \geq 0,\end{aligned}$$

whose optimal solution is denoted by $\hat{d}^*$. In addition, denote by d^* the optimal solution to the LP problem defined in Lemma 3.2 driven by $\Psi(\pi)$. Then, we have the estimates:

$$\begin{aligned}G_{\text{opt}}(\pi) := r^\top d^* - r^\top\Psi(\pi) &\leq r^\top d^* + \Big(b^\top\hat{y}^* - \gamma_0^\top\hat{\lambda}^*\Big) - \Big(b^\top y - \gamma_0^\top\lambda\Big) - r^\top\Psi(\pi) \\ &\leq r^\top d^* - (r - \Delta)^\top\hat{d}^* - \Big[\Big(b^\top y - \gamma_0^\top\lambda\Big) + r^\top\Psi(\pi)\Big] \\ &\leq ||\Delta||_1 + \epsilon\Big(|\mathcal{T}|\zeta_1 + \zeta_2 + \zeta_4\Big) \qquad (16) \\ &\leq \epsilon\Big[(|\mathcal{T}| + 1)\zeta_1 + \zeta_2 + \zeta_4\Big].\end{aligned}$$

where the inequality (16) comes from the estimate:

$$\begin{aligned}\Big|\Big(b^\top y - \gamma_0^\top\lambda\Big) + r^\top\Psi(\pi)\Big| &= \Big|\Psi(\pi)^\top A^\top y - \gamma_0^\top\lambda + r^\top\Psi(\pi)\Big| \\ &\leq \Big|\Psi(\pi)^\top(A^\top y + z + r - c^\top\lambda)\Big| + \Psi(\pi)^\top z + \Big|\lambda^\top(\gamma_0 - c\Psi(\pi))\Big| \\ &\leq \epsilon\Big(|\mathcal{T}|\zeta_1 + \zeta_2 + \zeta_4\Big).\end{aligned}$$

This finishes our proof for the upper bound of the optimality gap.

Next, we shall prove the upper bound of the feasibility gap.

G_{fea}: estimate the upper bound.

Again, for notational simplicity, we write Ψ in the place of $\Psi(\pi)$. It is enough to notice that:

$$\begin{aligned}G_{\text{fea}}(\pi) &= ||\min(0, \gamma_0 - c_\Psi\Psi)||_2 \\ &\leq ||\min(0, \gamma_0 - c_\Psi\Psi) - \min(0, \gamma_0 - c_L L)||_2 + ||\min(0, \gamma_0 - c_L L)||_2 \\ &\leq ||c_\Psi\Psi - c_L L||_2 + ||\min(0, \gamma_0 - c_L L)||_2 \\ &\leq ||c_\Psi\Psi - c_L L||_2 + ||\gamma_0 - c_L L - w||_2\end{aligned}$$

$$\leq \epsilon\zeta_3.$$

This finishes our proof for the upper bound of the feasibility gap.

Finally, we shall prove the lower bound of the optimality gap.

$G_{\mathbf{opt}}$: estimate the lower bound.

In the rest, we shall always take the mean-field flow $\Psi(\pi)$. Therefore, we omit the subscript for notational simplicity.

To start with, define $\tilde{\Delta} := \max\left(0, c^\top \Psi(\pi) - \gamma_0\right)$. Consider the following maximization problem:

$$\begin{aligned} \text{maximize}_{\tilde{y},\tilde{z},\tilde{\lambda}} \quad & b^\top \tilde{y} - \left(\gamma_0 + \tilde{\Delta}\right)^\top \tilde{\lambda} \\ \text{subject to} \quad & A^\top \tilde{y} + \tilde{z} = c^\top \tilde{\lambda} - r, \quad \tilde{\lambda} \geq 0, \quad \tilde{z} \geq 0, \end{aligned}$$

whose optimal solution is denoted by $\left(\tilde{y}^*, \tilde{z}^*, \tilde{\lambda}^*\right)$. The dual problem is:

$$\begin{aligned} \text{minimize}_{\tilde{d}} \quad & -r^\top \tilde{d} \\ \text{subject to} \quad & A\tilde{d} = b, \quad c\tilde{d} \leq \gamma_0 + \tilde{\Delta}, \quad \tilde{d} \geq 0, \end{aligned}$$

whose optimal solution is denoted by $\tilde{d}^*$. In addition, denote by $(d^*, y^*, z^*, \lambda^*)$ a solution of $\mathcal{K}(\Psi(\pi))$ (*i.e.*, the KKT conditions induced by $\Psi(\pi)$). Then, we have the estimates:

$$\begin{aligned} G_{\text{opt}}(\pi) := r^\top d^* - r^\top \Psi(\pi) &\geq r^\top d^* - r^\top \tilde{d}^* \\ &= b^\top \tilde{y}^* - \left(\gamma_0 + \tilde{\Delta}\right)^\top \tilde{\lambda}^* - \left(b^\top y^* - \gamma_0 \lambda^*\right) \\ &\geq b^\top y^* - \left(\gamma_0 + \tilde{\Delta}\right)^\top \lambda^* - \left(b^\top y^* - \gamma_0 \lambda^*\right) \\ &= -\tilde{\Delta}^\top \lambda^* \\ &\geq -\epsilon\sqrt{k}\frac{|\mathcal{T}| r_{\max}}{\delta}\zeta_3, \end{aligned}$$

where the last inequality comes from the facts that $G_{\text{fea}}(\pi) \leq \epsilon\zeta_3$ and $||\lambda^*||_\infty \leq \frac{|\mathcal{T}| r_{\max}}{\delta}$ (see Lemma 3.4). This finalizes our proof. □

Proof of Theorem 3.8

Proof. The feasibility gap is a direct corollary of Theorem 3.7, which further implies that the lower bound of the optimality gap is 0. Next, we prove the upper bound of the optimality gap. The proof utilizes similar techniques as the proof of Theorem 3.7.

First, the inequalities in G_{opt} step 2 of the proof of Theorem 3.7 become:

$$\begin{aligned} \left|\left|A_\Psi^\top y + z + r_\Psi - c_\Psi^\top \lambda\right|\right|_1 &\leq \frac{\epsilon_0}{\zeta_3}\left[\frac{1}{c_1} + \alpha(c_2)\left(C_p\tilde{\beta}_y + C_r + kC_c\frac{|\mathcal{T}| r_{\max}}{\delta - \epsilon_0}\right)\right] =: \frac{\epsilon_0}{\zeta_3}\tilde{\zeta}_1(c_1, c_2), \\ A_\Psi \Psi - b &= 0, \\ 0 \leq z^\top \Psi &\leq \frac{\epsilon_0}{\zeta_3}\left[\frac{1}{c_3} + \tilde{\beta}_z\alpha(c_2)\right] =: \frac{\epsilon_0}{\zeta_3}\tilde{\zeta}_2(c_2, c_3), \\ ||\gamma_0 - c_\Psi \Psi - w||_2 &\leq \frac{\epsilon_0}{\zeta_3}\left[\frac{1}{c_4} + \alpha(c_2)\Big(kc_{\max} + C_c|\mathcal{S}||\mathcal{A}||\mathcal{T}|\Big)\right] + \epsilon_0 = \frac{\epsilon_0}{\zeta_3}\zeta_3(c_2, c_4) + \epsilon_0, \\ \left|\lambda^\top(\gamma_0 - c_\Psi \Psi)\right| &\leq \frac{\epsilon_0}{\zeta_3}\left[\frac{1}{c_5} + k\alpha(c_2)\frac{|\mathcal{T}| r_{\max}}{\delta - \epsilon_0}\Big(kc_{\max} + C_c|\mathcal{S}||\mathcal{A}||\mathcal{T}|\Big)\right] + \epsilon_0\frac{k|\mathcal{T}| r_{\max}}{\delta - \epsilon_0} \end{aligned}$$

$$=: \frac{\epsilon_0}{\zeta_3}\tilde{\zeta}_4(c_2, c_5) + \epsilon_0 \frac{k|\mathcal{T}|r_{\max}}{\delta - \epsilon_0}.$$

Then, the upper bound of the optimality gap is obtained by repeating exactly the same calculation as G_{opt} step 3 (plugging the upper bounds of the above inequalities). The whole proof completes. □

Proof of Theorem 3.10

Proof. We shall always take the mean-field flow $L = \Psi(\pi)$. Therefore, we omit the subscript for notational simplicity.

To start with, notice that $AL = b$, which comes from the fact that $L = \Psi(\pi)$. Then, taking advantage of Lemma 3.4, direct calculations give the following estimates:

$$\begin{aligned}
0 \le z^\top L &= (c^\top \lambda - A^\top y - r)^\top L \\
&= \lambda^\top \gamma_0 + \lambda^\top (cL - \gamma_0) - y^\top b - V^\pi(L) \\
&= \lambda^\top \gamma_0 + \lambda^\top (cL - \gamma_0) + \left(r_L - c^\top \lambda\right)^\top d - V^\pi(L) \\
&= \lambda^\top (cL - \gamma_0) + V_c^*(L) - V_{\mu_0}^\pi(L) \\
&\le \lambda^\top \max(0, cL - \gamma_0) + |G_{\text{opt}}(\pi)| \\
&\le \sqrt{k}\frac{|\mathcal{T}|r_{\max}}{\delta}\epsilon_2 + \epsilon_1, \\
||\gamma_0 - cL - w||_2 &= ||\min(0, \gamma_0 - cL)||_2 \le \epsilon_2, \\
\left|\lambda^\top(\gamma_0 - cL)\right| &= \left|\lambda^\top c(d - L)\right| \\
&= \left|y^\top A(d - L) + z^\top(d - L) + r^\top(d - L)\right| \\
&= \left|-z^\top L + r^\top(d - L)\right| \\
&\le z^\top L + |G_{\text{opt}}(\pi)| \\
&\le \sqrt{k}\frac{|\mathcal{T}|r_{\max}}{\delta}\epsilon_2 + 2\epsilon_1.
\end{aligned}$$

Putting everything together yields the desired result. □

B Proofs of results in §5

With an abuse of notation, denote by $\mathcal{M}_c(L, \gamma_0)$ the constrained MDP problem defined in §2 with the driving mean-field flow L and threshold γ_0 (*i.e.*, we explicitly write out the threshold γ_0 in $\mathcal{M}_c(L)$), and denote by $V_c^*(L, \gamma_0)$ its maximum cumulative expected reward, if exists.

Lemma B.1. *Fix the mean-field flow L. Let $\epsilon > 0$ and $\gamma_1, \gamma_2 \in \mathbb{R}^k$ be such that $||\gamma_1 - \gamma_2||_1 \le \epsilon$. If the constrained MDP problems $\mathcal{M}_c(L, \gamma_1)$ and $\mathcal{M}_c(L, \gamma_2)$ satisfy Assumption 1, then their optimal values will satisfy:*

$$|V_c^*(L, \gamma_1) - V_c^*(L, \gamma_2)| \le \epsilon \frac{|\mathcal{T}|r_{\max}}{\delta}.$$

Proof. By Lemma 3.2, the MDP problems are equivalent to the following LP problems ($i = 1, 2$):

$$\begin{aligned}
&\text{minimize}_d \quad -r_L^\top d \\
&\text{subject to} \quad A_L d = b, \quad c_L d \le \gamma_i, \quad d \ge 0,
\end{aligned}$$

whose dual problems are:

$$\begin{aligned} \text{maximize}_{y,z,\lambda} \quad & b^\top y - \gamma_i^\top \lambda \\ \text{subject to} \quad & A_L^\top y + z - c_L^\top \lambda = -r_L, \quad z \geq 0, \quad \lambda \geq 0. \end{aligned}$$

Denote by $(d_i^*, y_i^*, z_i^*, \lambda_i^*)$ the primal and dual optimal solutions to the above LP problems for $i = 1, 2$. By strong duality, we have:

$$\begin{aligned} -r_L^\top d_1^* &= b^\top y_1^* - \gamma_1^\top \lambda_1^* \\ &= b^\top y_1^* - \gamma_2^\top \lambda_1^* + (\gamma_2^\top \lambda_1^* - \gamma_1^\top \lambda_1^*) \\ &\leq b^\top y_2^* - \gamma_2^\top \lambda_2^* + (\gamma_2^\top \lambda_1^* - \gamma_1^\top \lambda_1^*) \\ &= -r_L^\top d_2^* + (\gamma_2^\top \lambda_1^* - \gamma_1^\top \lambda_1^*) \\ &\leq -r_L^\top d_2^* + \epsilon \frac{|\mathcal{T}| r_{\max}}{\delta}, \end{aligned}$$

where the last inequality comes from Lemma 3.4. By exactly the same steps, we have:

$$-r_L^\top d_2^* \leq -r_L^\top d_1^* + \epsilon \frac{|\mathcal{T}| r_{\max}}{\delta}.$$

This completes the proof. □

In the following, with an abuse of notation, denote by $L^{(N)}\left(\pi_1, \pi^{\otimes(N-1)}\right)$ the empirical distribution of the N players when the first player takes policy π_1 and the rest $N-1$ players take policy π (*i.e.*, $\pi^{\otimes(N-1)}$ denotes $N-1$ copies of π). Note that $L^{(N)}\left(\pi_1, \pi^{\otimes(N-1)}\right)$ is a random variable taking values in $[\Delta(\mathcal{S}\times\mathcal{A})]^{|\mathcal{T}|}$. Recall that we embed $L^{(N)}$ in $\mathbb{R}_{\geq 0}^{|\mathcal{S}||\mathcal{A}||\mathcal{T}|}$, and for each $t \in \mathcal{T}$, we embed $L_t^{(N)}$ in $\mathbb{R}_{\geq 0}^{|\mathcal{S}||\mathcal{A}|}$.

Lemma B.2. *Under Assumption 2, let π and π_1 be two policies. We have:*

$$\max_{t\in\mathcal{T}} \mathbb{E}\left|\left|L_t^{(N)}\left(\pi_1, \pi^{\otimes(N-1)}\right) - \Psi_t(\pi)\right|\right|_1 \leq \left(\frac{1}{2\sqrt{N}} + \frac{2}{N}\right)|\mathcal{S}||\mathcal{A}|\frac{C_{p,\mathcal{S},\mathcal{A}}^{|\mathcal{T}|} - 1}{C_{p,\mathcal{S},\mathcal{A}} - 1},$$

where the constant

$$C_{p,\mathcal{S},\mathcal{A}} := (C_p + 1)\,|\mathcal{S}||\mathcal{A}|.$$

Proof. For notational simplicity, we use $L^{(N)}$ to represent $L_t^{(N)}\left(\pi_1, \pi^{\otimes(N-1)}\right)$ in this proof. Observe that for each $1 \leq t \leq T$ and $(s,a) \in \mathcal{S}\times\mathcal{A}$, we have:

$$\begin{aligned} \mathbb{E}\left|L_t^{(N)}(s,a) - \Psi_t(\pi)(s,a)\right| &= \mathbb{E}\left[\mathbb{E}\left(\left|L_t^{(N)}(s,a) - \Psi_t(\pi)(s,a)\right| \,\middle|\, L_{t-1}^{(N)}\right)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left(\left|L_t^{(N)}(s,a) - \left[\Psi_{t-1}(\pi)\cdot p_{t-1}\left(s \mid \Psi_{t-1}(\pi)\right)\right]\pi_t(s)(a)\right| \,\middle|\, L_{t-1}^{(N)}\right)\right] \\ &\leq \mathbb{E}(J_1 + J_2), \end{aligned}$$

where

$$\begin{aligned} J_1 &= \mathbb{E}\left(\left|\frac{1}{N}\sum_{i=1}^N \delta_{\left(s_t^{(i)}, a_t^{(i)}\right)=(s,a)} - \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^N \delta_{\left(s_t^{(i)}, a_t^{(i)}\right)=(s,a)}\right)\right| \,\middle|\, L_{t-1}^{(N)}\right) \leq \frac{1}{2\sqrt{N}}, \\ J_2 &= \mathbb{E}\left(\left|\mathbb{E}\left(\frac{1}{N}\sum_{i=1}^N \delta_{\left(s_t^{(i)}, a_t^{(i)}\right)=(s,a)}\right) - \left[\Psi_{t-1}(\pi)\cdot p_{t-1}\left(s \mid \Psi_{t-1}(\pi)\right)\right]\pi_t(s)(a)\right| \,\middle|\, L_{t-1}^{(N)}\right) \\ &\leq \frac{1}{N}\left|p_{t-1}\left(s_{t-1}^{(1)}, a_{t-1}^{(1)}, L_{t-1}^{(N)}\right)(s)\pi_{1,t}(s)(a) - \left[\Psi_{t-1}(\pi)\cdot p_{t-1}\left(s \mid \Psi_{t-1}(\pi)\right)\right]\pi_t(s)(a)\right| \end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{N}\left|\sum_{i=2}^{N}\left[p_{t-1}\left(s_{t-1}^{(i)},a_{t-1}^{(i)},L_{t-1}^{(N)}\right)(s)\pi_t(s)(a)-\left[\Psi_{t-1}(\pi)\cdot p_{t-1}\left(s\,|\,\Psi_{t-1}(\pi)\right)\right]\pi_t(s)(a)\right]\right|\\
&\leq \frac{1}{N}+\frac{1}{N}\left|\sum_{i=2}^{N}\left[p_{t-1}\left(s_{t-1}^{(i)},a_{t-1}^{(i)},L_{t-1}^{(N)}\right)(s)-p_{t-1}\left(s_{t-1}^{(i)},a_{t-1}^{(i)},\Psi_{t-1}(\pi)\right)(s)\right]\right|\\
&\quad+\frac{1}{N}\left|\sum_{i=2}^{N}\left[p_{t-1}\left(s_{t-1}^{(i)},a_{t-1}^{(i)},\Psi_{t-1}(\pi)\right)(s)-\Psi_{t-1}(\pi)\cdot p_{t-1}\left(s\,|\,\Psi_{t-1}(\pi)\right)\right]\right|\\
&\leq \frac{1}{N}+\frac{N-1}{N}C_p\left|\left|L_{t-1}^{(N)}-\Psi_{t-1}(\pi)\right|\right|_1+\left(\left|\left|L_{t-1}^{(N)}-\Psi_{t-1}(\pi)\right|\right|_1+\frac{1}{N}\right)\\
&\leq \frac{2}{N}+(C_p+1)\left|\left|L_{t-1}^{(N)}-\Psi_{t-1}(\pi)\right|\right|_1.
\end{aligned}
$$

When $t=0$, by a similar argument as above, we get:

$$
\begin{aligned}
\mathbb{E}\left|L_0^{(N)}(s,a)-\Psi_0(\pi)(s,a)\right| &= \mathbb{E}\left|L_0^{(N)}(s,a)-\mu_0(s)\pi(s)(a)\right|\\
&\leq \frac{1}{2\sqrt{N}}+\frac{1}{N}.
\end{aligned}
$$

The desired result is obtained by an induction argument on time stamps. □

Similarly, we denote by $L^{(N)}\left(\pi_1\,\middle|\,\pi^{\otimes(N-1)}\right)$ the empirical distribution of the first player when the first player takes policy π_1 and the rest players take policy π. Note that $L^{(N)}\left(\pi_1\,\middle|\,\pi^{\otimes(N-1)}\right)$ is a random variable taking values in $[\Delta(\mathcal{S}\times\mathcal{A})]^{|\mathcal{T}|}$. Different from our previously defined $L^{(N)}\left(\pi_1,\pi^{\otimes(N-1)}\right)$, here $L^{(N)}\left(\pi_1\,\middle|\,\pi^{\otimes(N-1)}\right)$ only characterizes the distribution of the first player. In addition, in the CMFG, we denote by $L(\pi_1\,|\,\pi)$ the distribution of a specific player who takes policy π_1, provided that that population distribution is $\Psi(\pi)$ (*i.e.*, all the rest players take policy π).

Lemma B.3. *Under Assumption 2, let π and π_1 be two policies. Then, we have:*

$$
\max_{t\in\mathcal{T}}\left|\left|\mathbb{E}\left[L_t^{(N)}\left(\pi_1\,\middle|\,\pi^{\otimes(N-1)}\right)\right]-L_t(\pi_1\,|\,\pi)\right|\right|_1\leq\left(\frac{1}{2\sqrt{N}}+\frac{2}{N}\right)C_p|\mathcal{S}||\mathcal{A}|(|\mathcal{T}|-1)\frac{C_{p,\mathcal{S},\mathcal{A}}^{|\mathcal{T}|}-1}{C_{p,\mathcal{S},\mathcal{A}}-1},
$$

where $C_{p,\mathcal{S},\mathcal{A}}$ is defined in Lemma B.2.

Proof. For notational simplicity, in the rest of the proof, we use $L^{(N)}$ to represent $L^{(N)}\left(\pi_1\,\middle|\,\pi^{\otimes(N-1)}\right)$, and use L to represent $L(\pi_1\,|\,\pi)$.

For each $1\leq t\leq T$ and $(s,a)\in\mathcal{S}\times\mathcal{A}$, we have:

$$
\begin{aligned}
&\left|\mathbb{E}\left[L_t^{(N)}\right](s,a)-L_t(s,a)\right|\\
&=\left|\mathbb{E}\left[L_{t-1}^{(N)}\cdot p_{t-1}\left(s\,\middle|\,L_{t-1}^{(N)}(\pi_1,\pi)\right)\right]\pi_1(s)(a)-\left[L_{t-1}\cdot p_{t-1}(s\,|\,\Psi_{t-1}(\pi))\right]\pi_1(s)(a)\right|\\
&\leq\left|\mathbb{E}\left[L_{t-1}^{(N)}\cdot p_{t-1}\left(s\,\middle|\,L_{t-1}^{(N)}(\pi_1,\pi)\right)\right]-\left[L_{t-1}\cdot p_{t-1}(s\,|\,\Psi_{t-1}(\pi))\right]\right|\\
&\leq\left|\mathbb{E}\left[L_{t-1}^{(N)}\cdot p_{t-1}\left(s\,\middle|\,L_{t-1}^{(N)}(\pi_1,\pi)\right)-L_{t-1}^{(N)}\cdot p_{t-1}(s\,|\,\Psi_{t-1}(\pi))\right]\right|\\
&\quad+\left|\mathbb{E}\left[L_{t-1}^{(N)}\cdot p_{t-1}(s\,|\,\Psi_{t-1}(\pi))\right]-L_{t-1}\cdot p_{t-1}(s\,|\,\Psi_{t-1}(\pi))\right|\\
&\leq\left(\frac{1}{2\sqrt{N}}+\frac{2}{N}\right)C_p|\mathcal{S}||\mathcal{A}|\frac{C_{p,\mathcal{S},\mathcal{A}}^{|\mathcal{T}|}-1}{C_{p,\mathcal{S},\mathcal{A}}-1}+\left|\left|\mathbb{E}\left[L_{t-1}^{(N)}\right]-L_{t-1}\right|\right|_1.
\end{aligned}
$$

We mention that the first term of the last inequality comes from Lemma B.2, where $C_{p,\mathcal{S},\mathcal{A}}$ is defined.

Notice that when $t = 0$, we have:

$$\left|\left|\mathbb{E}\left[L_0^{(N)}\right] - L_0\right|\right|_1 = 0.$$

The desired result is obtained by an induction argument on time stamps. □

With an abuse of notation, denote by $V^{\pi_1}\left(\pi^{\otimes(N-1)}\right)$ the expected cumulative reward of the first player who takes policy π_1, provided that the rest players take policy π, *i.e.*,

$$V^{\pi_1}\left(\pi^{\otimes(N-1)}\right) := \mathbb{E}_{a^{(1)}\sim\pi_1,\,\forall 2\le j\le N,\,a^{(j)}\sim\pi}\left[\sum_{t\in\mathcal{T}} r_t\left(s_t^{(1)}, a_t^{(1)}, L_t^{(N)}\right) \,\middle|\, s_0^{(1)}\sim\mu_0\right].$$

On top of Lemma B.2 and Lemma B.3, a direct calculation produces the following estimate.

Lemma B.4. *Under Assumption 2, let π and π_1 be two policies. Then, we have:*

$$\left|V^{\pi_1}\left(\pi^{\otimes(N-1)}\right) - V^{\pi_1}(\Psi(\pi))\right| \le \left(\frac{1}{2\sqrt{N}} + \frac{2}{N}\right) C_p(C_r + r_{\max})|\mathcal{S}||\mathcal{A}|(|\mathcal{T}|-1)^2 \frac{C_{p,\mathcal{S},\mathcal{A}}^{|\mathcal{T}|} - 1}{C_{p,\mathcal{S},\mathcal{A}} - 1},$$

where $C_{p,\mathcal{S},\mathcal{A}}$ is defined in Lemma B.2.

Similarly, in the N-player game, we denote:

$$\mathcal{C}^{\pi_1}\left(\pi^{\otimes(N-1)}\right) := \mathbb{E}_{a^{(1)}\sim\pi_1,\,\forall 2\le j\le N,\,a^{(j)}\sim\pi}\left[\sum_{t\in\mathcal{T}} c_t\left(s_t^{(1)}, a_t^{(1)}, L_t^{(N)}\right) \,\middle|\, s_0^{(1)}\sim\mu_0\right],$$

and in the CMFG, we denote (with an abuse of notation):

$$\mathcal{C}^{\pi_1}(L) := \mathbb{E}_{a\sim\pi_1}\left[\sum_{t\in\mathcal{T}} c_t(s_t, a_t, L_t) \,\middle|\, s_0\sim\mu_0\right].$$

By replacing r with c in Lemma B.4, we get the following estimate on the value of the cost functions.

Lemma B.5. *Under Assumption 2, let π and π_1 be two policies. Then, we have:*

$$\left|\left|\mathcal{C}^{\pi_1}\left(\pi^{\otimes(N-1)}\right) - \mathcal{C}^{\pi_1}(\Psi(\pi))\right|\right|_\infty \le \left(\frac{1}{2\sqrt{N}} + \frac{2}{N}\right) C_p(C_c + c_{\max})|\mathcal{S}||\mathcal{A}|(|\mathcal{T}|-1)^2 \frac{C_{p,\mathcal{S},\mathcal{A}}^{|\mathcal{T}|} - 1}{C_{p,\mathcal{S},\mathcal{A}} - 1},$$

where $C_{p,\mathcal{S},\mathcal{A}}$ is defined in Lemma B.2.

Finally, we are prepared to prove the main result Theorem 5.1.

Proof of Theorem 5.1. Fix ϵ and N that satisfy the conditions. In the rest of the proof, we let all players in the N-player game take the policy π^*. Due to the symmetry of the players, it is enough to analyze the first player.

First, by Lemma B.5, it is not hard to see that $\mathcal{F}^{(1)}$ is non-empty (*e.g.*, the policy in Assumption 5 belongs to $\mathcal{F}^{(1)}$). Then, we have:

$$\sup_{\pi_1\in\mathcal{F}^{(1)}} V^{\pi_1}\left((\pi^*)^{\otimes(N-1)}\right) = V^{\hat{\pi}_1}\left((\pi^*)^{\otimes(N-1)}\right) \le V^{\hat{\pi}_1}(\Psi(\pi^*)) + \epsilon,$$

where we denote that the supremum is achieved at $\hat{\pi}_1$. Since $\hat{\pi}_1 \in \mathcal{F}^{(1)}$, by Lemma B.5, we have:

$$G_{\text{fea}}(\hat{\pi}_1, \Psi(\pi^*)) \le \sqrt{k}\epsilon,$$

where G_{fea} is defined in Definition 3. Therefore, Lemma B.1 gives the following estimate:

$$V^{\hat{\pi}_1}(\Psi(\pi^*)) \leq V^{\pi^*}(\Psi(\pi^*)) + \epsilon \frac{k|\mathcal{T}|r_{\max}}{\delta}.$$

Putting everything together, we get:

$$\begin{aligned}\sup_{\pi_1 \in \mathcal{F}^{(1)}} V^{\pi_1}\left((\pi^*)^{\otimes(N-1)}\right) &\leq V^{\pi^*}(\Psi(\pi^*)) + \epsilon\left(1 + \frac{k|\mathcal{T}|r_{\max}}{\delta}\right) \\ &\leq V^{\pi^*}\left((\pi^*)^{\otimes(N-1)}\right) + \epsilon\left(2 + \frac{k|\mathcal{T}|r_{\max}}{\delta}\right).\end{aligned}$$

This completes the proof of the ϵ_1 part. The ϵ_2 part is a direct corollary of Lemma B.5. This completes the whole proof. □