Dynamic Games and Applications
https://doi.org/10.1007/s13235-022-00453-z

Partially Observed Discrete-Time Risk-Sensitive Mean Field Games

Naci Saldi[1] · Tamer Başar[2] · Maxim Raginsky[2]

Accepted: 4 May 2022

Abstract
In this paper, we consider discrete-time partially observed mean-field games with the risk-sensitive optimality criterion. We introduce risk-sensitivity behavior for each agent via an exponential utility function. In the game model, each agent is weakly coupled with the rest of the population through its individual cost and state dynamics via the empirical distribution of states. We establish the mean-field equilibrium in the infinite-population limit using the technique of converting the underlying original partially observed stochastic control problem to a fully observed one on the belief space and the dynamic programming principle. Then, we show that the mean-field equilibrium policy, when adopted by each agent, forms an approximate Nash equilibrium for games with sufficiently many agents. We first consider finite-horizon cost function and then discuss extension of the result to infinite-horizon cost in the next-to-last section of the paper.

Keywords Mean field games · Partial observation · Risk sensitive cost

1 Introduction

Mean-field games have been introduced in [28] and [34] to show the existence of approximate Nash equilibria for fully observed non-cooperative continuous time games, when the number of agents is large but finite. The underlying idea of the mean-field method is to transform the decentralized game problem to a centralized stochastic control problem using the so-called Nash certainty equivalence (NCE) principle [28]. The optimal solution of this control problem, calibrated appropriately using the empirical distribution of the term that (weakly) couples the players, provides an approximate Nash equilibrium for games with a sufficiently

✉ Naci Saldi
naci.saldi@bilkent.edu.tr

Tamer Başar
basar1@illinois.edu

Maxim Raginsky
maxim@illinois.edu

[1] Department of Mathematics, Bilkent University, Ankara, Turkey

[2] Coordinated Science Laboratory, University of Illinois, Urbana, IL, USA

Published online: 07 June 2022

large number of agents. To obtain the optimal solution to the associated stochastic control problem, one should simultaneously solve a Fokker–Planck equation evolving forward in time and a Hamilton–Jacobi–Bellman equation evolving backward in time. We refer the reader to [7, 13, 14, 22, 27, 29, 35, 50] for studies of fully observed continuous-time mean-field games with different models and cost functions, such as games with major-minor players, risk-sensitive games, games with Markov jump parameters, and LQG games.

In this paper, we study discrete-time partially observed mean-field games with risk-sensitive optimality criteria. Risk-sensitivity brings in an element of robustness to decision making and has been widely used in many fields, such as control, economics, financial engineering, and operations research, among others. As opposed to risk-neutral optimization where only the mean value of the cost is considered, risk-sensitive one places positive weights on also the higher moments, thus capturing the risk element (see [3, 50, 52]). In the model we study in this paper, we have a large but finite number of agents interacting with each other through their individual dynamics and cost functions via the mean-field term (i.e., the empirical distribution of their states). It is known that establishing the existence of Nash equilibria for these types of games is quite difficult due to the (almost) decentralized and noisy nature of the information structure of the problem [4, 5]. Therefore, it is of interest to find an approximate equilibrium with reduced complexity. To that end, upon letting the number of agents go to infinity, the mean-field term converges to the distribution of the state of a single generic agent. This decouples the dynamics and cost functions of the agents from each other, and because of that, in the limiting case, a generic agent is faced with a stochastic control problem with a constraint on the distribution of the state at each time (i.e., a mean-field game problem). The main goal in these problems is to show the existence of a policy and a state distribution flow such that this policy is an optimal solution of the stochastic control problem when the total population behavior is modeled by the state distribution flow and the resulting distribution of each agent's state is the same as the state distribution flow when the generic agent applies this policy. This equilibrium condition is called the Nash certainty equivalence (NCE) principle in the literature. In this paper, we first consider the existence of such an equilibrium for the limiting case and then establish that the policy in this equilibrium constitutes an approximate Nash equilibrium for finite-agent games with sufficiently many agents.

In the literature, *partially observed* mean-field games have not been studied much, especially in the discrete-time setup. Indeed, this work seems to be the first one that studies discrete-time risk-sensitive mean-field games under partial observations. Prior works have mostly considered the risk-neutral continuous-time setup. It is obvious that analyses of continuous-time and discrete-time setups are quite different, requiring different sets of tools. In [30], the authors study a partially observed continuous-time mean-field game with linear individual dynamics. In [43, 45, 46], the authors consider a continuous-time mean-field game with major-minor agents and nonlinear dynamics where the minor agents can partially observe the state of the major agent. In [44, 47], the same authors also develop a nonlinear filtering theory for McKean–Vlasov-type stochastic differential equations that arise as the infinite population limit of the partially observed differential game of the mean-field type. In [12], the authors study the linear quadratic mean-field game with major-minor agents where the minor agents can partially observe the state of the major agent. In [20, 21], the authors consider the linear quadratic mean-field game, again with major-minor agents where, in this case, both the minor agents and the major agent can partially observe the state of the major agent. In [48], the authors study a continuous-time partially observed stochastic control problem of the mean-field type and establish a maximum principle to characterize the optimal control. In [31], the authors consider a continuous-time mean-field game with linear individual dynamics where two types of partial information structure are considered: (i) agents cannot observe the white noise which is common to all agents; (ii) agents can access the additive white-noise version of their own states.

For risk-sensitive cost criteria, existing works are mostly on the continuous-time set-up, with [37], discussed further below, being one exception. Now, in continuous-time set-up, reference [50] studies a class of mean-field games with nonlinear individual dynamics and a risk-sensitive cost function. They characterize the mean-field equilibrium via coupled HJB and FP equations and explicit solutions to these equations are given when the individual state dynamics are linear. In [49], the author considers a continuous-time mean-field game with nonlinear individual dynamics, where state dynamics have L^p-norm structure. Stochastic maximum principle is used to characterize the optimal solution of the problem. In [17], the authors study a partially observed version of the continuous-time risk-sensitive mean-field game. They establish a stochastic maximum principle for the characterization of the mean-field equilibrium. Reference [36] considers continuous-time risk-sensitive mean-field games with linear individual dynamics and local state information for the players. First a generic risk-sensitive optimal control problem is solved which yields mean-field equilibrium, and then, it is shown that the policies in mean-field equilibrium lead to an approximate Nash equilibrium for games with a sufficiently large number of agents. It is also shown that this approximate Nash equilibrium is partially equivalent to the approximate Nash equilibrium of a certain robust mean-field game problem. Finally, [37] presents the counterparts of these results for the discrete-time linear-quadratic risk-sensitive mean-field game.

Here, we consider discrete-time mean-field games with Polish state, action, and observation spaces (i.e., complete and separable metric spaces) under risk-sensitive optimality criteria for the players. In the infinite population limit of such games, a generic agent should solve a partially observed stochastic control problem under the NCE principle. Due to the constraints induced by NCE principle, common techniques used to analyze partially observed stochastic control problems are not sufficient. To establish the existence of an equilibrium solution in the infinite population limit, we have to bring in the fixed-point approach that is used to obtain equilibria in classical game problems, along with the technique of converting partially observed optimal control problems to fully observed ones on the belief space. The definitions of the finite-agent game and the mean-field game problems are given in Sect. 2 and Sect. 3, respectively. In Sect. 4, we prove the existence of a mean-field equilibrium. In Sects. 5 and 6, we establish that the mean-field equilibrium policy is approximately Nash for finite-agent games with sufficiently many agents. In Sect. 7, we extend previous results to games with infinite-horizon risk-sensitive cost functions. Section 9 concludes the paper.

In an earlier paper [41], we studied the risk-neutral version of this problem under a similar set of assumptions on the system components. There are some parallels between the techniques used in this paper and those in [41] to show the existence of a mean-field equilibrium and to prove that the policies in mean-field equilibrium provide an approximate Nash equilibrium for games with large but finitely many agents. In this paper, we exploit this connection and refer the reader to [41] for proofs of certain results. We note, however, that as far as their analyses go, there are considerable technical differences between risk-sensitive and risk-neutral cost functions. The fact that, in the risk-sensitive case, the cost function is in a multiplicative form leads to complication in the analysis of the optimality condition. Therefore, to establish the existence of a mean-field equilibrium in the infinite-population limit and an approximate Nash equilibrium in the finite-agent case, we need to first transform the risk-sensitive problem to one where the cost function is risk-neutral and in an additive form. However, in this risk-neutral form, the one-stage cost function and the transition probability become non-homogeneous (i.e., time-dependent) as opposed to the risk-neutral problem in [41]. Hence, after a careful execution of this step, we can prove the existence of a mean-field equilibrium by adapting the technique developed in [41] to the non-homogeneous and finite-horizon case. We also note that in [42] we have studied the fully observed version of the same problem under a slightly different set of assumptions on the system components. Indeed, to prove the existence of an approximate Nash equilibrium, here we generalize the results established in [42] to the game models with expanding state spaces and non-homogeneous system components.

Notation. For a metric space E, we let $C_b(\mathsf{E})$ denote the set of all bounded continuous real functions on E, $\mathcal{P}(\mathsf{E})$ denote the set of all Borel probability measures on E, and $\mathcal{B}(\mathsf{E})$ denote the collection of Borel sets. For any E-valued random element x, $\mathcal{L}(x)(\,\cdot\,) \in \mathcal{P}(\mathsf{E})$ denotes the distribution of x. A sequence $\{\mu_n\}$ of measures on E is said to converge weakly to a measure μ if $\int_{\mathsf{E}} g(e)\mu_n(de) \to \int_{\mathsf{E}} g(e)\mu(de)$ for all $g \in C_b(\mathsf{E})$. For any $\nu \in \mathcal{P}(\mathsf{E})$ and measurable real function g on E, we define $\nu(g) = \int g d\nu$. For any subset B of E, we let ∂B and B^c denote the boundary and complement of B, respectively. The notation $v \sim \nu$ means that the random element v has distribution ν. Unless otherwise specified, the term “measurable" will refer to Borel measurability.

2 Finite Player Game Model

2.1 Original Game Model

Let S, A, and Y be Polish spaces. We consider a discrete-time partially observed N-agent mean-field game with a state space S, an action space A, and an observation space Y. For every $i \in \{1, 2, \ldots, N\}$, the state, the action, and the observation of Agent i at time t ($t = 0, 1, 2, \ldots$) are, respectively, denoted by $s_i^N(t) \in \mathsf{S}$, $u_i^N(t) \in \mathsf{A}$, and $g_i^N(t) \in \mathsf{Y}$. We let $d_t^{(N)}(\,\cdot\,) = \frac{1}{N}\sum_{i=1}^N \delta_{s_i^N(t)}(\,\cdot\,) \in \mathcal{P}(\mathsf{S})$ denote the empirical distribution of the states (i.e., mean-field term) at time t, where $\delta_s \in \mathcal{P}(\mathsf{S})$ is the Dirac measure at s; that is, $\delta_s(A) = 1$ if $s \in A$ and otherwise 0.

At the initial time step $t = 0$, the states $(s_1^N(0), \ldots, s_N^N(0)) \sim \kappa_0 \otimes \ldots \otimes \kappa_0$ are independent and identically distributed according to κ_0. For each $t \geq 0$, the current-observations $(g_1^N(t), \ldots, g_N^N(t))$ and the next-states $(s_1^N(t+1), \ldots, s_N^N(t+1))$ are distributed according to the probability laws:

$$\prod_{i=1}^N l\big(\mathrm{d}g_i^N(t)\big|s_i^N(t)\big) \text{ and } \prod_{i=1}^N q\big(\mathrm{d}s_i^N(t+1)\big|s_i^N(t), u_i^N(t), d_t^{(N)}\big), \tag{1}$$

where $q : \mathsf{S} \times \mathsf{A} \times \mathcal{P}(\mathsf{S}) \to \mathcal{P}(\mathsf{S})$ is the state transition kernel and $l : \mathsf{S} \to \mathcal{P}(\mathsf{Y})$ is the observation kernel. Note that the state dynamics of each agent are weakly coupled through the mean-field term $d_t^{(N)}$.

For any Agent i, define the history spaces $\mathsf{G}_0 = \mathsf{Y}$ and $\mathsf{G}_t = (\mathsf{Y} \times \mathsf{A})^t \times \mathsf{Y}$ for $t = 1, 2, \ldots$, all endowed with product Borel σ-algebras. A *policy* for Agent i is a sequence $\pi^i = \{\pi_t^i\}$ of stochastic kernels on A given G_t; that is, for any $t \geq 0$, $u_i^N(t) \sim \pi_t^i(\cdot|\gamma_i^N(t))$, where $\gamma_i^N(t) = \big(g_i^N(t), u_i^N(t-1), g_i^N(t-1) \ldots, u_i^N(0), g_i^N(0)\big)$ is the observation-action history observed by Agent i up to time t. The set of all policies for Agent i is denoted by Π_i. Let $\tilde{\Pi}_i$ be the set of policies in Π_i which only use the observations; that is, $\pi \in \tilde{\Pi}_i$ if $\pi_t : \prod_{k=0}^t \mathsf{Y} \to \mathcal{P}(\mathsf{A})$ for each $t \geq 0$. Let $\boldsymbol{\Pi}^{(N)} = \prod_{i=1}^N \Pi_i$ and $\tilde{\boldsymbol{\Pi}}^{(N)} = \prod_{i=1}^N \tilde{\Pi}_i$. We let $\boldsymbol{\pi}^{(N)} = (\pi^1, \ldots, \pi^N)$ ($\pi^i \in \Pi_i$) denote the N-tuple of joint policies of all the agents in the game. Under such an N-tuple of policies, the actions of agents at each time $t \geq 0$ are obtained with respect to the conditional probability distribution

$$\prod_{i=1}^{N} \pi_t^i \big(du_i^N(t) \big| \gamma_i^N(t)\big). \tag{2}$$

The *one-stage cost* function for a generic agent is a measurable function $m : \mathsf{S} \times \mathsf{A} \times \mathcal{P}(\mathsf{S}) \to [0, \infty)$. Then, the agent's finite-horizon *risk-sensitive* cost under a policy $\boldsymbol{\pi}^{(N)} \in \boldsymbol{\Pi}^{(N)}$ is given by

$$V_i^{(N)}(\boldsymbol{\pi}^{(N)}) = \frac{1}{\lambda} \log\left(E^{\boldsymbol{\pi}^{(N)}} \left[e^{\lambda \sum_{t=0}^{T} \beta^t m(s_i^N(t), u_i^N(t), d_t^{(N)})} \right] \right),$$

where $\beta \in (0, 1]$ is the discount factor, $\lambda > 0$ is the risk factor, and T is the finite horizon of the problem. Here, $E^{\boldsymbol{\pi}^{(N)}}\big[\,\cdot\,\big]$ denotes the expectation with respect to the probability law, which is uniquely specified by the kernels in (1) and (2) and the initial state distribution κ_0.

Since $\frac{1}{\lambda}\log(\cdot)$ is a strictly increasing function, without loss of generality, it suffices to consider only the part with expectation:

$$W_i^{(N)}(\boldsymbol{\pi}^{(N)}) = E^{\boldsymbol{\pi}^{(N)}} \left[e^{\lambda \sum_{t=0}^{T} \beta^t m(s_i^N(t), u_i^N(t), d_t^{(N)})} \right].$$

With this cost function, the equilibrium solution for the game is defined as follows:

Definition 1 A policy $\boldsymbol{\pi}^{(N*)} = (\pi^{1*}, \ldots, \pi^{N*})$ constitutes a *Nash equilibrium* for the N-player game, if

$$W_i^{(N)}(\boldsymbol{\pi}^{(N*)}) = \inf_{\pi^i \in \Pi_i} W_i^{(N)}(\boldsymbol{\pi}_{-i}^{(N*)}, \pi^i)$$

for each $i = 1, \ldots, N$, where $\boldsymbol{\pi}_{-i}^{(N*)} = (\pi^{j*})_{j \neq i}$.

As we explained in detail in [41], establishing the existence of Nash equilibria for *partially observed* mean-field games is challenging due to the (almost) decentralized and noisy nature of the information structure of the problem. To that end, we slightly change the definition of Nash equilibrium in this model and adopt the approximate Nash equilibrium concept instead of exact Nash equilibrium.

Definition 2 A policy $\boldsymbol{\pi}^{(N*)} \in \tilde{\boldsymbol{\Pi}}^{(N)}$ is a *Nash equilibrium* if

$$W_i^{(N)}(\boldsymbol{\pi}^{(N*)}) = \inf_{\pi^i \in \tilde{\Pi}_i} W_i^{(N)}(\boldsymbol{\pi}_{-i}^{(N*)}, \pi^i)$$

for each $i = 1, \ldots, N$, and an ε-*Nash equilibrium* (for a given $\varepsilon > 0$) if

$$W_i^{(N)}(\boldsymbol{\pi}^{(N*)}) \leq \inf_{\pi^i \in \tilde{\Pi}_i} W_i^{(N)}(\boldsymbol{\pi}_{-i}^{(N*)}, \pi^i) + \varepsilon$$

for each $i = 1, \ldots, N$.

According to this definition, the agents can only use their local observations $(g_i^N(t), \ldots, g_i^N(0))$ to construct their policies. In real-life applications, agents typically have access only to their local observations. Hence, it suffices to establish the existence of an approximate Nash equilibrium for the game with a local information structure. In addition, in the discrete-time mean field literature, it is common to establish the existence of approximate Nash equilibria with local (decentralized) information structures (see [1] [9]). This is true for partially observed case as well (see [41]).

Here, our goal is to establish the existence of approximate Nash equilibria for games with sufficiently many agents. Indeed, if the number of agents is small, it is all but impossible to show even the existence of approximate Nash equilibria for these types of games. Therefore, it is key to assume that the number of agents is large (but finite). With this assumption, we can go to the infinite population limit, for which we can model the mean-field term as an exogenous state-measure flow, which should be consistent with the distribution of a generic agent (i.e., the NCE principle) by the law of large numbers. In this case, to establish the existence of an equilibrium, a generic agent should solve a classical partially observed stochastic control problem with a constraint on the distributions on the states (i.e., mean-field game). Then, we expect that if each agent in the finite-agent N game adopts the equilibrium policy in the infinite-population limit, the resulting policy will be an approximate Nash equilibrium for all sufficiently large N.

Our approach to prove the existence of approximate Nash equilibria can be summarized as follows: (i) Note that in the risk-sensitive criteria, the one-stage cost functions are in a multiplicative form as opposed to the risk-neutral setting. As stated earlier, this makes the analysis of the problem quite complicated. Therefore, we first construct an equivalent non-homogeneous game model, where the cost can be written in an additive form as in the risk-neutral case (see Sect. 2.2). (ii) Then, we introduce the infinite-population limit ($N \to \infty$) of the equivalent game model to approximate the finite-agent setting (see Sect. 3). (iii) By adapting the proof technique in [41] to the non-homogeneous and finite-horizon set-up, we prove the existence of an appropriately defined mean-field equilibrium for this limiting infinite-population game (see Sect. 4). (iv) Then, we return to the finite-N case for the equivalent game model and show that if each agent in the game problem adopts the mean-field equilibrium policy, then the resulting policy will be an approximate Nash equilibrium for all sufficiently large N. Since the equivalent game model is identical to the original game model in terms of cost functions, this establishes the existence of approximate Nash equilibria for the original game model (see Sects. 5 and 6).

Now, proceeding along the lines above, we first introduce the following assumptions, imposed throughout the paper.

Assumption 1 (a) The cost function m is bounded and continuous with $\|m\| = \sup_{s \in \mathsf{S}} |m(s)| \leq K$.

(b) The stochastic kernel q is weakly continuous in (s, u, κ); i.e., $q(\,\cdot\,|s(k), u(k), \kappa_k) \to q(\,\cdot\,|s, u, \kappa)$ weakly when $(s(k), u(k), \kappa_k) \to (s, u, \kappa)$.

(c) The observation kernel l is continuous in s with respect to total variation norm; i.e., for all s, $l(\,\cdot\,|s_k) \to l(\,\cdot\,|s)$ in total variation norm when $s_k \to s$.

(d) A is compact.

(e) There exist a constant $\alpha \geq 0$ and a continuous moment function $v : \mathsf{S} \to [1, \infty)$ (see [25,Definition E.7]) such that

$$\sup_{(u,\kappa) \in \mathsf{A} \times \mathcal{P}(\mathsf{S})} \int_{\mathsf{S}} v(y) q(\mathrm{d}y|s, u, \kappa) \leq \alpha v(s). \tag{3}$$

(f) The initial probability measure κ_0 satisfies $\int_{\mathsf{S}} v(s)\kappa_0(\mathrm{d}s) = M < \infty$.

2.2 Equivalent Game Model

In this section, we construct an equivalent game model whose states are the states of the original model plus the one-stage costs incurred up to that time. Namely, the state at time t for Agent i is

$$x_i^N(t) = \left(s_i^N(t), \sum_{k=0}^{t-1} \beta^k m(s_i^N(k), u_i^N(k), d_k^{(N)})\right).$$

In this new model, finite-horizon risk-sensitive cost function can be written in an additive-form like in risk-neutral case. For this new game model, we have been inspired by [6], in which the authors study the classical fully observed risk-sensitive control problem. For a generic agent, this new game model is specified by

$$\left(\mathsf{X}, \mathsf{A}, \mathsf{Y}, \{p_t\}_{t=0}^{T+1}, r, \{c_t\}_{t=0}^{T+1}, \mu_0\right),$$

where $\mathsf{X} = \mathsf{S} \times [0, L]$ is the new state space with $L = \frac{K}{1-\beta}$, where L is the maximum risk-neutral discounted-cost that can be incurred. For every t, the state transition kernel $p_t : \mathsf{X} \times \mathsf{A} \times \mathcal{P}(\mathsf{X}) \to \mathcal{P}(\mathsf{X})$ is defined as:[1]

$$p_t\big(B \times D \big| x(t), a(t), \mu_t\big) = q(B|s(t), a(t), \mu_{t,1}) \otimes \delta_{m(t)+\beta^t m(s(t),a(t),\mu_{t,1})}(D),$$

where $B \in \mathcal{B}(\mathsf{S})$, $D \in \mathcal{B}([0, L])$, $x(t) = (s(t), m(t))$, and $\mu_{t,1}$ is the marginal of μ_t on S. Here, p_t is indeed the controlled transition probability of the next state $s_i^N(t+1)$ and current risk-neutral total discounted cost

$$\sum_{k=0}^{t} \beta^k m(s_i^N(k), a_i^N(k), d_k^{(N)})$$

given the current state-action pair $(s_i^N(t), a_i^N(t))$ and past risk-neutral total discounted cost $\sum_{k=0}^{t-1} \beta^k m(s_i^N(k), a_i^N(k), d_k^{(N)})$ in the original game. The observation kernel $r : \mathsf{X} \to \mathcal{P}(\mathsf{Y})$ is equivalent to the observation kernel l in the original problem; that is, $r(\mathrm{d}y|x) = l(\mathrm{d}y|s)$ where $x = (s, m)$. For each t, the one-stage cost function $c_t : \mathsf{X} \times \mathsf{A} \times \mathcal{P}(\mathsf{X}) \to [0, \infty)$ is defined as:

$$c_t(x(t), a(t), \mu_t) = \begin{cases} 0, & \text{if } t \leq T \\ e^{\lambda m(t)}, & \text{if } t = T+1. \end{cases}$$

Finally, the initial measure μ_0 is given by $\mu_0(\mathrm{d}x(0)) = \kappa_0(\mathrm{d}s(0)) \otimes \delta_0(\mathrm{d}m(0))$, where the initial states $\{x_i^N(0)\}$ are independent and identically distributed according to μ_0. Note that, in this equivalent game model, the finite-horizon is $T+1$ instead of T and system components depend on time t. We also define the empirical distribution of the states at time t as follows:

$$e_t^{(N)}(\,\cdot\,) = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i^N(t)}(\,\cdot\,) \in \mathcal{P}(\mathsf{X}).$$

Suppose that Assumption 1 holds. Then, for each t, the following are true for the new game model:

[1] In the remainder of this paper, we use letter 'a' instead of 'u', to denote actions, to emphasize that they are generated using the new game model.

(I) The one-stage cost function c_t is bounded and continuous.

(II) The stochastic kernel p_t is weakly continuous.

(III) The observation kernel r is continuous with respect to the total variation distance.

(IV) Let $w : \mathsf{X} \to [1, \infty)$ be defined as $w(x) = w((s, m)) = v(s)$, which is a moment function. Then, we have

$$\sup_{(a,\mu)\in \mathsf{A}\times\mathcal{P}(\mathsf{X})} \int_{\mathsf{X}} w(y) p_t(\mathrm{d}y|x, a, \mu) \leq \alpha w(x). \tag{4}$$

(V) The initial probability measure μ_0 satisfies $\int_{\mathsf{X}} w(x)\mu_0(\mathrm{d}x) = M < \infty$.

Recall that $\tilde{\Pi}_i$ denotes the set of policies for Agent i that only use observations in the original game. Note that $\tilde{\Pi}_i$ is also the set of policies for Agent i that only use observations in the new game model. For Agent i, the finite-horizon risk-neutral total cost under the N-tuple of policies $\boldsymbol{\pi}^{(N)} \in \tilde{\boldsymbol{\Pi}}^{(N)}$ is denoted as $J_i^{(N)}(\boldsymbol{\pi}^{(N)})$; that is

$$J_i^{(N)}(\boldsymbol{\pi}^{(N)}) = E^{\boldsymbol{\pi}^{(N)}}\left[\sum_{t=0}^{T+1} c_t(x_i^N(t), a_i^N(t), e_t^{(N)})\right].$$

The following proposition makes the connection between this new model and the original model. The proof is straightforward, and so, we omit the details (see the proof of [42,Proposition 5.1]).

Proposition 1 *For any* $\boldsymbol{\pi}^{(N)} \in \tilde{\boldsymbol{\Pi}}^{(N)}$ *and* $i = 1, \ldots, N$, *we have* $J_i^{(N)}(\boldsymbol{\pi}^{(N)}) = W_i^{(N)}(\boldsymbol{\pi}^{(N)})$.

Proposition 1 states that the new game model is equivalent to the original game model in terms of cost functions. This is true because the new game model consists of the one-stage costs incurred up to the current time as an additional state variable. Therefore, if we take the exponent of this additional state at time $T + 1$ as in the definition of c_{T+1}, we obtain the risk-sensitive cost of the original game model. Hence, in the remainder of this paper, we replace the original game model with the new one; that is, from this point on, we have the following system components satisfying (I)-(V):

$$\left(\mathsf{X}, \mathsf{A}, \mathsf{Y}, \{p_t\}_{t=0}^{T+1}, r, \{c_t\}_{t=0}^{T+1}, \mu_0\right).$$

Remark 1 Note that in the new game model, the time horizon is $T + 1$, which means that agents should also design control policies for the time step $T + 1$. However, note that control policies at time step $T + 1$ do not affect the cost function (i.e., one-stage cost at time $T + 1$ is only a function of the state), and thus agents indeed do not need to select these policies in the new game model. Hence, we can in a sense view the time horizons of the two problems as T.

Note that the cost functions $J_i^{(N)}(\boldsymbol{\pi}^{(N)})$ of this new game model are in additive form (i.e., risk-neutral). Therefore, we can use a technique similar to the one in [41] to prove the existence of an approximate Nash equilibrium. To this end, we will first consider the infinite-population limit of the new game model and prove the existence of an equilibrium. Then, we will go back to the finite agent case and establish the existence of approximate Nash equilibrium for the new game model using the infinite population equilibrium solution. Since, by Proposition 1, the new game model has the same cost function as the original game model, the last result also implies the existence of an approximate Nash equilibrium for the original game, which was the main goal of this paper.

3 Partially Observed Mean-Field Games and Mean-Field Equilibria

In this section, we introduce the infinite population limit of the new game introduced in the preceding section. Although it is called mean-field game, it is not game in the classical sense: It is a stochastic control problem whose state distribution at each time step should satisfy a certain consistency condition. The optimal solution of this problem is referred to as mean-field equilibrium. In other words, we have a single agent and model the mean-field term by an exogenous *state-measure flow* $\boldsymbol{\mu} := (\mu_t)_{t=0}^{T+1} \subset \mathcal{P}(\mathsf{X})$ with a given initial condition μ_0, by the law of large numbers. This measure flow $\boldsymbol{\mu}$ should also be consistent with the state distributions of this single agent when the agent acts optimally. The precise mathematical description of the problem is given as follows.

The mean-field game model for a generic agent is specified by

$$\left(\mathsf{X}, \mathsf{A}, \mathsf{Y}, \{p_t\}_{t=0}^{T+1}, r, \{c_t\}_{t=0}^{T+1}, \mu_0\right),$$

where, as before, X, A, and Y are the state, action, and observation spaces, respectively. The stochastic kernel $p_t : \mathsf{X} \times \mathsf{A} \times \mathcal{P}(\mathsf{X}) \to \mathcal{P}(\mathsf{X})$ denotes the transition probability, and $r : \mathsf{X} \times \mathcal{P}(\mathsf{X}) \to \mathcal{P}(\mathsf{Y})$ denotes the observation kernel. The measurable function $c_t : \mathsf{X} \times \mathsf{A} \times \mathcal{P}(\mathsf{X}) \to [0, \infty)$ is the one-stage cost function and μ_0 is the distribution of the initial state.

Recall the history spaces $\mathsf{G}_0 = \mathsf{Y}$ and $\mathsf{G}_t = (\mathsf{Y} \times \mathsf{A})^t \times \mathsf{Y}$ for $t = 1, 2, \ldots$, all endowed with product Borel σ-algebras. A *policy* is a sequence $\pi = \{\pi_t\}$ of stochastic kernels on A given G_t. The set of all policies is denoted by Π.

We let $\mathcal{M} = \big\{\boldsymbol{\mu} \in \mathcal{P}(\mathsf{X})^{T+2} : \mu_0 \text{ is fixed}\big\}$ be the set of all state-measure flows with a given initial condition μ_0. Given any measure flow $\boldsymbol{\mu} \in \mathcal{M}$, the evolution of the states, observations, and actions is as follows:

$$\begin{aligned}
x(0) &\sim \mu_0, \\
y(t) &\sim r(\,\cdot\,|x(t)),\ t = 0, 1, \ldots \\
x(t) &\sim p_{t-1}(\,\cdot\,|x(t-1), a(t-1), \mu_{t-1}),\ t = 1, 2, \ldots \\
a(t) &\sim \pi_t(\,\cdot\,|\gamma(t)),\ t = 0, 1, \ldots,
\end{aligned}$$

where $\gamma(t) \in \mathsf{G}_t$ is the observation-action history up to time t. An initial distribution μ_0 on X, a policy π, and a state-measure flow $\boldsymbol{\mu}$ define a unique probability measure P^{π} on $(\mathsf{X} \times \mathsf{Y} \times \mathsf{A})^{T+2}$. The expectation with respect to P^{π} is denoted by $E^{\pi}[\,\cdot\,]$. A policy $\pi^* \in \Pi$ is said to be optimal for $\boldsymbol{\mu}$ if $J_{\boldsymbol{\mu}}(\pi^*) = \inf_{\pi \in \Pi} J_{\boldsymbol{\mu}}(\pi)$, where the finite-horizon cost of policy π with measure flow $\boldsymbol{\mu}$ is given by

$$J_{\boldsymbol{\mu}}(\pi) = E^{\pi}\left[\sum_{t=0}^{T+1} c_t(x(t), a(t), \mu_t)\right]$$

Using these definitions, we first define the set-valued mapping $\Psi : \mathcal{M} \to 2^{\Pi}$ as $\Psi(\boldsymbol{\mu}) = \{\pi \in \Pi : \pi \text{ is optimal for } \boldsymbol{\mu}\}$. Conversely, we define a single-valued mapping $\Lambda : \Pi \to \mathcal{M}$ as follows: given $\pi \in \Pi$, the state-measure flow $\boldsymbol{\mu} := \Lambda(\pi)$ is constructed recursively as:

$$\mu_{t+1}(\,\cdot\,) = \int_{\mathsf{X} \times \mathsf{A}} p_t(\,\cdot\,|x(t), a(t), \mu_t) P^{\pi}(\mathrm{d}a(t)|x(t)) \mu_t(\mathrm{d}x(t)),$$

where $P^{\pi}(\mathrm{d}a(t)|x(t))$ denotes the conditional distribution of $a(t)$ given $x(t)$ under π and $(\mu_\tau)_{0 \le \tau \le t}$. Using Ψ and Λ, we now introduce the mean-field equilibrium.

Definition 3 A pair $(\pi^*, \boldsymbol{\mu}^*) \in \Pi \times \mathcal{M}$ is a *mean-field equilibrium* if $\pi^* \in \Psi(\boldsymbol{\mu}^*)$ and $\boldsymbol{\mu}^* = \Lambda(\pi^*)$.

The main result of this section is the existence of a mean-field equilibrium. Later we will show that this mean-field equilibrium constitutes an approximate Nash equilibrium for games with sufficiently many agents.

Theorem 1 *The mean-field game* $\big(\mathsf{X}, \mathsf{A}, \mathsf{Y}, \{p_t\}_{t=0}^{T+1}, r, \{c_t\}_{t=0}^{T+1}, \mu_0\big)$ *admits a mean-field equilibrium* $(\pi^*, \boldsymbol{\mu}^*)$.

The proof of Theorem 1 is given in Sect. 4. Our approach to prove Theorem 1 can be summarized as follows: (i) First, we lift the partially observed stochastic control problem a generic agent is faced with for a given measure flow to a fully observed stochastic control problem; (ii) we then transform the fixed point equation $\pi \in \Psi(\Lambda(\pi))$ characterizing the mean-field equilibrium into a fixed point equation of a set-valued mapping from the set of state-action measure flows into itself using the Bellman optimality operator; (iii) then, we prove that this set-valued mapping has a closed graph; and (iv) finally, we deduce the existence of a mean-field equilibrium using Kakutani's fixed point theorem.

4 Proof of Theorem 1

Note that any measure flow $\boldsymbol{\mu} \in \mathcal{M}$ leads to a non-homogenous partially observed Markov decision process (POMDP). Hence, before starting the proof of Theorem 1, we first review a few relevant results on POMDPs. To this end, fix any $\boldsymbol{\mu} \in \mathcal{M}$ and consider the corresponding optimal control problem.

Let $\mathcal{P}_w(\mathsf{X}) = \big\{\mu \in \mathcal{P}(\mathsf{X}) : \int_{\mathsf{X}} w(x)\mu(\mathrm{d}x) < \infty\big\}$. It is known that any POMDP can be reduced to a (completely observable) MDP (see [53], [39]), whose states are the posterior state distributions or beliefs of the observer; that is, the state at time t is

$$z(t) = \mathsf{Pr}\{x(t) \in \cdot\,|y(0), \ldots, y(t), a(0), \ldots, a(t-1)\} \in \mathcal{P}(\mathsf{X}).$$

We call this equivalent MDP the belief-state MDP. Note that since $\mathcal{L}(x(t)) \in \mathcal{P}_w(\mathsf{X})$ under any policy by (IV)-(V), we have

$$\mathsf{Pr}\{x(t) \in \cdot\,|y(0), \ldots, y(t), a(0), \ldots, a(t-1)\} \in \mathcal{P}_w(\mathsf{X})$$

almost everywhere. Therefore, the belief-state MDP has state space $\mathsf{Z} = \mathcal{P}_w(\mathsf{X})$ and action space A. Here, Z is endowed with the Borel σ-algebra generated by the topology of weak convergence. Next, we construct the transition probabilities $\{\eta_t\}_{t=0}^{T+1}$ of the belief-state MDP (see also [24]). Let z denote the generic state variable for the belief-state MDP. Fix any t. First consider the transition probability on $\mathsf{X} \times \mathsf{Y}$ given $\mathsf{Z} \times \mathsf{A}$

$$R_t(x \in A, y \in B|z, a) = \int_{\mathsf{X}} \kappa_t(A, B|x', a) z(\mathrm{d}x'),$$

where $\kappa_t(\mathrm{d}x, \mathrm{d}y|x', a) = r(\mathrm{d}y|x) \otimes p_t(\mathrm{d}x|x', a, \mu_t)$. Let us disintegrate R_t as follows $R_t(\mathrm{d}x, \mathrm{d}y|z, a) = H_t(\mathrm{d}y|z, a) \otimes F_t(\mathrm{d}x|z, a, y)$. Then, we define the mapping $F_t : \mathsf{Z} \times \mathsf{A} \times \mathsf{Y} \to \mathsf{Z}$ as:

$$F_t(z, a, y)(\,\cdot\,) = F_t(\,\cdot\,|z, a, y). \tag{5}$$

Then, $\eta_t : \mathsf{Z} \times \mathsf{A} \to \mathcal{P}(\mathsf{Z})$ is defined as:

$$\eta_t(\,\cdot\,|z(t), a(t)) = \int_{\mathsf{Y}} \delta_{F_t(z(t),a(t),y(t+1))}(\,\cdot\,)\; H_t(\mathrm{d}y(t+1)|z(t), a(t)).$$

The initial point for the belief-state MDP is μ_0; that is, $\mathcal{L}(z(0)) \sim \delta_{\mu_0}$. Finally, for each t, the one-stage cost function C_t of the belief-state MDP is given by

$$C_t(z, a) = \int_{\mathsf{X}} c_t(x, a, \mu_t) z(\mathrm{d}x). \tag{6}$$

Hence, the belief-state MDP is a Markov decision process with the components $\big(\mathsf{Z}, \mathsf{A}, \{\eta_t\}_{t=0}^{T+1}, \{C_t\}_{t=0}^{T+1}, \delta_{\mu_0}\big)$.

For the belief-state MDP define the history spaces $\mathsf{K}_0 = \mathsf{Z}$ and $\mathsf{K}_t = (\mathsf{Z} \times \mathsf{A})^t \times \mathsf{Z}$, $t = 1, 2, \ldots$. A *policy* is a sequence $\varphi = \{\varphi_t\}$ of stochastic kernels on A given K_t. The set of all policies is denoted by Φ. A *Markov* policy is a sequence $\varphi = \{\varphi_t\}$ of stochastic kernels on A given Z. The set of Markov policies is denoted by M. Let $\tilde{J}(\varphi, \mu_0)$ denote the finite-horizon cost function of policy $\varphi \in \Phi$ for initial point μ_0 of the belief-state MDP. Notice that any history vector $s(t) = (z(0), \ldots, z(t), a(0), \ldots, a(t-1))$ of the belief-state MDP is a function of the history vector $\gamma(t) = (y(0), \ldots, y(t), a(0), \ldots, a(t-1))$ of the POMDP. Let us write this relation as $i(\gamma(t)) = s(t)$. Hence, for a policy $\varphi = \{\varphi_t\} \in \Phi$, we can define a policy $\pi^{\varphi} = \{\pi_t^{\varphi}\} \in \Pi$ as $\pi_t^{\varphi}(\,\cdot\,|\gamma(t)) = \varphi_t(\,\cdot\,|i(\gamma(t)))$. Let us write this as a mapping from Φ to Π: $\Phi \ni \varphi \mapsto i(\varphi) = \pi^{\varphi} \in \Pi$. It is straightforward to show that the cost functions $\tilde{J}(\varphi, \mu_0)$ and $J_{\boldsymbol{\mu}}(\pi^{\varphi})$ are the same. One can also prove that (see [53], [39])

$$\inf_{\varphi \in \Phi} \tilde{J}(\varphi, \mu_0) = \inf_{\pi \in \Pi} J_{\boldsymbol{\mu}}(\pi) \tag{7}$$

and furthermore, that if φ is an optimal policy for belief-state MDP, then π^{φ} is optimal for the POMDP as well. Therefore, the optimal control problem for the mean-field game is equivalent to the optimal control of belief-state MDP.

We now derive the conditions that are satisfied by belief-state MDP. To that end, define $W : \mathsf{Z} \to \mathbb{R}$ as:

$$W(z) = \int_{\mathsf{X}} w(x) z(\mathrm{d}x).$$

Note that W is a lower semi-continuous moment function on Z. One can prove that (see [41,Section 4]) the belief-state MDP satisfies the following conditions under Assumption 1:

(i) The cost functions $\{C_t\}$ are bounded and continuous.
(ii) The stochastic kernels $\{\eta_t\}$ are weakly continuous.
(iii) A is compact and Z is σ-compact.
(iv) There exists a constant $\alpha \geq 0$ such that

$$\sup_{a \in \mathsf{A}} \int_{\mathsf{Z}} W(y) \eta_t(\mathrm{d}y|z, a) \leq \alpha W(z), \text{ for all } t.$$

(v) The initial probability measure δ_{μ_0} satisfies $W(\delta_{\mu_0}) = M < \infty$.

With these conditions, we are now ready to prove Theorem 1 by adapting techniques in [41] to the non-homogeneous and finite-horizon set-up.

We first define the mapping $\mathsf{B} : \mathcal{P}(\mathsf{Z}) \to \mathcal{P}(\mathsf{X})$, which will define the relation between state-measure flows in the mean-field game and state-measure flows in the belief-state MDP, as follows:

$$\mathsf{B}(\nu)(\,\cdot\,) = \int_{\mathsf{Z}} z(\,\cdot\,)\;\nu(\mathrm{d}z).$$

Using this definition, for any $\boldsymbol{\nu} \in \mathcal{P}(\mathsf{Z} \times \mathsf{A})^{T+2}$, we define the measure flow $\boldsymbol{\mu}^{\boldsymbol{\nu}} \in \mathcal{P}(\mathsf{X})^{T+2}$ as follows:

$$\boldsymbol{\mu}^{\boldsymbol{\nu}} = \big(\mathsf{B}(\nu_{t,1})\big)_{t=0}^{T+1},$$

where for any $\nu \in \mathcal{P}(\mathsf{Z} \times \mathsf{A})$, we let ν_1 denote the marginal of ν on Z. Let $\{\eta_t^{\boldsymbol{\nu}}\}_{t=0}^{T+1}$ and $\{C_t^{\boldsymbol{\nu}}\}_{t=0}^{T+1}$ be, respectively, the transition probabilities and one-stage cost functions of belief-state MDP induced by the measure flow $\boldsymbol{\mu}^{\boldsymbol{\nu}}$. We let $J_{*,t}^{\boldsymbol{\nu}} : \mathsf{Z} \to [0, \infty)$ denote the optimal value function at time t of this belief-state MDP; that is,

$$J_{*,t}^{\boldsymbol{\nu}}(z) = \inf_{\varphi \in \Phi} E^{\varphi}\left[\sum_{k=t}^{T+1} C_k^{\boldsymbol{\nu}}(z(k), a(k)) \middle| z(t) = z\right].$$

Let $J_*^{\boldsymbol{\nu}} = \big(J_{*,t}^{\boldsymbol{\nu}}\big)_{t=0}^{T+1}$.

To prove the existence of a mean-field equilibrium, we use the technique in [32]. To that end, we first transform the fixed point equation $\pi \in \Psi(\Lambda(\pi))$ characterizing the mean-field equilibrium into a fixed-point equation of a set-valued mapping from the set of state-action measure flows $\mathcal{P}(\mathsf{Z} \times \mathsf{A})^{T+2}$ into itself. Then, using Kakutani's fixed point theorem ([2,Corollary 17.55]), we deduce the existence of a mean-field equilibrium.

For any t, the *Bellman optimality operator* $T_t^{\boldsymbol{\nu}} : C_b(\mathsf{Z}) \to C_b(\mathsf{Z})$ is given by

$$T_t^{\boldsymbol{\nu}} u(z) = \min_{a \in \mathsf{A}}\left[C_t^{\boldsymbol{\nu}}(z, a) + \int_{\mathsf{Z}} u(y)\eta_t^{\boldsymbol{\nu}}(\mathrm{d}y|z, a)\right].$$

Note that $T_t^{\boldsymbol{\nu}} J_{*,t+1}^{\boldsymbol{\nu}} = J_{*,t}^{\boldsymbol{\nu}}$ for every t. The following theorem is a known result in the theory of nonhomogeneous Markov decision processes (see [26,Theorems 14.4 and 17.1]). For any given $\boldsymbol{\nu}$, it characterizes the optimal policy of the belief-state MDP.

Theorem 2 *For any $\boldsymbol{\nu}$, a policy $\varphi \in M$ is optimal if and only if, for all t,*

$$\nu_t^{\varphi}\left(\left\{(z, a) : C_t^{\boldsymbol{\nu}}(z, a) + \int_Z J_{*,t+1}^{\boldsymbol{\nu}}(y)\eta_t^{\boldsymbol{\nu}}(dy|z, a) = T_t^{\boldsymbol{\nu}} J_{*,t+1}^{\boldsymbol{\nu}}(z)\right\}\right) = 1, \tag{8}$$

where $\nu_t^{\varphi} = \mathcal{L}\big(z(t), a(t)\big)$ under φ and $\boldsymbol{\nu}$.

Using Theorem 2, we now define the set-valued map from $\mathcal{P}(\mathsf{Z} \times \mathsf{A})^{T+2}$ into itself. To that end, for any $\boldsymbol{\nu} \in \mathcal{P}(\mathsf{Z} \times \mathsf{A})^{T+2}$, let us define the following sets:

$$C(\boldsymbol{\nu}) = \left\{\boldsymbol{\nu}' \in \mathcal{P}(\mathsf{Z} \times \mathsf{A})^{T+2} : \nu'_{0,1} = \delta_{\mu_0},\; \nu'_{t+1,1}(\,\cdot\,) = \int_{\mathsf{Z} \times \mathsf{A}} \eta_t^{\boldsymbol{\nu}}(\,\cdot\,|z, a)\nu_t(\mathrm{d}z, \mathrm{d}a)\right\}$$

and

$$B(\boldsymbol{\nu}) = \Bigg\{\boldsymbol{\nu}' \in \mathcal{P}(\mathsf{Z} \times \mathsf{A})^{T+2} : \forall 0 \le t \le T+1,$$
$$\nu'_t\left(\left\{(z, a) : C_t^{\boldsymbol{\nu}}(z, a) + \int_{\mathsf{Z}} J_{*,t+1}^{\boldsymbol{\nu}}(y)\eta_t^{\boldsymbol{\nu}}(\mathrm{d}y|z, a) = T_t^{\boldsymbol{\nu}} J_{*,t+1}^{\boldsymbol{\nu}}(z)\right\}\right) = 1\Bigg\}.$$

Here, the set $C(\boldsymbol{\nu})$ characterizes the consistency of the mean-field term with the state distribution of a generic agent, and the set $B(\boldsymbol{\nu})$ characterizes optimality of the policy for the mean-field term. The set-valued mapping $\Gamma : \mathcal{P}(\mathsf{Z} \times \mathsf{A})^{T+2} \rightarrow 2^{\mathcal{P}(\mathsf{Z}\times\mathsf{A})^{T+2}}$ is given as follows:

$$\Gamma(\boldsymbol{\nu}) = C(\boldsymbol{\nu}) \cap B(\boldsymbol{\nu}).$$

Note that the fixed-point equation $\pi \in \Psi(\Lambda(\pi))$ characterizes the behavior of the state distribution and the control law in mean-field equilibrium separately. To establish the existence of mean-field equilibrium via Kakutani's Fixed Point Theorem or Banach Fixed Point Theorem using this equation, one needs to put some topology on the policy space. However, by combining the state distribution with the control law, which gives the joint distribution of the state and the action, we can characterize via the set-valued mapping Γ the behavior of the state and the control law together in mean-field equilibrium. This will enable us to deduce the existence of a mean-field equilibrium without introducing a topology for the control laws, which is in general the solution technique in continuous time setup (see [28]).

An element $\boldsymbol{\nu}$ is a fixed point of Γ if $\boldsymbol{\nu} \in \Gamma(\boldsymbol{\nu})$. The following proposition makes the connection between mean-field equilibria and fixed points of Γ.

Proposition 2 *Suppose that Γ has a fixed point $\boldsymbol{\nu} = (\nu_t)_{t=0}^{T+1}$. Construct a Markov policy $\varphi = \{\varphi_t\}$ for belief-state MDP by disintegrating each ν_t as $\nu_t(dz, da) = \nu_{t,1}(dz)\varphi_t(da|z)$. Let $\pi^* = \pi^{\varphi}$ and $\boldsymbol{\mu}^* = (\mathsf{B}(\nu_{t,1}))_{t=0}^{T+1}$. Then, the pair $(\pi^*, \boldsymbol{\mu}^*)$ is a mean-field equilibrium.*

Proof Note that, since $\boldsymbol{\nu} \in C(\boldsymbol{\nu})$, we have $\nu_t = \mathcal{L}\big(z(t), a(t)\big)$ for belief-state MDP under the policy φ and the measure flow $\boldsymbol{\mu}^*$. Then, for any $f \in C_b(\mathsf{X})$, we have

$$\begin{aligned}
\mu^*_{t+1}(f) &= \mathsf{B}(\nu_{t+1,1})(f) \\
&= \int_{\mathsf{Z}\times\mathsf{A}} \int_{\mathsf{Z}} z'(f)\eta_t^{\boldsymbol{\nu}}(\mathrm{d}z'|z, a)\nu_t(\mathrm{d}z, \mathrm{d}a) \\
&= \int_{\mathsf{Z}\times\mathsf{A}} \left\{ \int_{\mathsf{X}} \int_{\mathsf{X}} f(y)p_t(\mathrm{d}y|x, a, \mu_t^*)z(\mathrm{d}x) \right\} \nu_t(\mathrm{d}z, \mathrm{d}a) \\
&= E^{\varphi}\big[l_t(z(t), a(t))\big] \left(\text{here} l_t(z, a) = \int_{\mathsf{X}} \int_{\mathsf{X}} f(y)p_t(\mathrm{d}y|x, a, \mu_t^*)z(\mathrm{d}x) \right) \\
&= E^{\pi^*}\left[\int_{\mathsf{X}} f(y)p_t(\mathrm{d}y|x(t), a(t), \mu_t^*) \right].
\end{aligned} \tag{9}$$

Since (9) is true for all $f \in C_b(\mathsf{X})$, we have

$$\mu^*_{t+1}(\,\cdot\,) = \int_{\mathsf{X}\times\mathsf{A}} p_t(\,\cdot\,|x(t), a(t), \mu_t^*)P^{\pi^*}(\mathrm{d}a(t)|x(t))\mu_t^*(\mathrm{d}x(t)),$$

where $P^{\pi^*}(\mathrm{d}a(t)|x(t))$ denotes the conditional distribution of $a(t)$ given $x(t)$ under π^* and $(\mu^*_\tau)_{0\leq\tau\leq t}$. Hence, $\Lambda(\pi^*) = \boldsymbol{\mu}^*$.

Since $\boldsymbol{\nu} \in B(\boldsymbol{\nu})$, the corresponding Markov policy φ satisfies (8) for $\boldsymbol{\nu}$. Therefore, by Theorem 2 and the fact that $\nu_t = \mathcal{L}\big(z(t), a(t)\big)$ for belief-state MDP under the policy φ and the measure flow $\boldsymbol{\mu}^*$, φ is optimal for belief-state MDP induced by the measure flow $\boldsymbol{\mu}^*$ (or, equivalently, $\boldsymbol{\nu}$). Therefore, $\pi^* \in \Psi(\boldsymbol{\mu}^*)$. □

By Proposition 2, it suffices to prove that Γ has a fixed point in order to establish the existence of a mean-field equilibrium. To prove this, we use Kakutani's fixed point theorem, which is stated below:

Theorem 3 [2,Corollary 17.55] *Let K be a non-empty compact convex subset of a locally convex Hausdorff space, and let the set-valued mapping $\phi : K \rightarrow 2^K$ have closed graph and non-empty convex values. Then, the set of fixed points of ϕ is compact and non-empty.*

Hence, in order to use Kakutani's fixed point theorem, the set-valued mapping Γ should be defined on a convex and compact set. However, the set $\mathcal{P}(\mathsf{Z} \times \mathsf{A})^{T+2}$ in the definition of Γ is not compact. To get around that, we will prove that the image of $\mathcal{P}(\mathsf{Z} \times \mathsf{A})^{T+2}$ under Γ is in fact a subset of some convex and compact set, and it is sufficient to consider this convex and compact set in the definition of Γ. To that end, for each t, define the set

$$\mathcal{P}^t(\mathsf{Z}) = \left\{\mu \in \mathcal{P}(\mathsf{Z}) : \int_{\mathsf{Z}} W(z)\mu(\mathrm{d}z) \leq \alpha^t M\right\}.$$

Since W is a lower semi-continuous moment function, the set $\mathcal{P}^t(\mathsf{Z})$ is compact with respect to the weak topology [25,Proposition E.8, p. 187]. Let us define

$$\mathcal{P}^t(\mathsf{Z} \times \mathsf{A}) = \left\{\nu \in \mathcal{P}(\mathsf{Z} \times \mathsf{A}) : \nu_1 \in \mathcal{P}^t(\mathsf{Z})\right\}.$$

Since A is compact, $\mathcal{P}^t(\mathsf{Z} \times \mathsf{A})$ is tight. Furthermore, $\mathcal{P}^t(\mathsf{Z} \times \mathsf{A})$ is closed with respect to the weak topology since W is lower semi-continuous. Hence, $\mathcal{P}^t(\mathsf{Z} \times \mathsf{A})$ is compact. Let $\Xi = \prod_{t=0}^{T+1} \mathcal{P}^t(\mathsf{Z} \times \mathsf{A})$, which is convex and compact with respect to the product topology.

Proposition 3 *We have* $\Gamma\big(\mathcal{P}(Z \times A)^{T+2}\big) = \left\{\boldsymbol{\nu}' : \boldsymbol{\nu}' \in \Gamma(\boldsymbol{\nu}),\ \boldsymbol{\nu} \in \mathcal{P}(Z \times A)^{T+2}\right\} \subset \Xi.$

Proof Fix any $\boldsymbol{\nu} \in \mathcal{P}(\mathsf{Z} \times \mathsf{A})^{T+2}$. It is sufficient to prove that $C(\boldsymbol{\nu}) \subset \Xi$ as $\Gamma(\boldsymbol{\nu}) = C(\boldsymbol{\nu}) \cap B(\boldsymbol{\nu})$. Let $\boldsymbol{\nu}' \in C(\boldsymbol{\nu})$. We prove by induction that $\nu'_{t,1} \in \mathcal{P}^t_v(\mathsf{Z})$ for all t. The claim trivially holds for $t = 0$ as $\nu'_{0,1} = \delta_{\mu_0}$. Assume that the claim holds for t and consider $t+1$. We have

$$\begin{aligned}\int_{\mathsf{Z}} W(y)\nu'_{t+1,1}(\mathrm{d}y) &= \int_{\mathsf{Z}\times\mathsf{A}} \int_{\mathsf{Z}} W(y)\eta^{\boldsymbol{\nu}}_t(\mathrm{d}y|z,a)\nu_t(\mathrm{d}z,\mathrm{d}a) \\ &\leq \int_{\mathsf{Z}} \alpha W(z)\nu_{t,1}(\mathrm{d}z) \text{ (by (iv))} \\ &\leq \alpha^{t+1}M \text{ (as } \nu_{t,1} \in \mathcal{P}^t_v(\mathsf{Z})).\end{aligned}$$

Hence, $\nu'_{t+1,1} \in \mathcal{P}^{t+1}_v(\mathsf{Z})$. □

By Proposition 3, we can now consider Γ as a multi-valued mapping from Ξ into itself. It can be proved that $C(\boldsymbol{\nu}) \cap B(\boldsymbol{\nu}) \neq \emptyset$ for any $\boldsymbol{\nu} \in \Xi$. Indeed, for any $t \geq 0$, we define

$$\mu_{t+1}(\,\cdot\,) = \int_{\mathsf{Z}\times\mathsf{A}} \eta^{\boldsymbol{\nu}}_t(\,\cdot\,|z,a)\,\nu_t(\mathrm{d}x,\mathrm{d}a).$$

Moreover, for any $t \geq 0$, let $f_t : \mathsf{Z} \to \mathsf{A}$ be the minimizer of the following optimality equation:

$$C^{\boldsymbol{\nu}}_t(z, f_t(z)) + \int_{\mathsf{Z}} J^{\boldsymbol{\nu}}_{*,t+1}(y)\eta^{\boldsymbol{\nu}}_t(\mathrm{d}y|z, f_t(z)) = T^{\boldsymbol{\nu}}_t J^{\boldsymbol{\nu}}_{*,t+1}(z).$$

Existence of such an f_t follows from the Measurable Selection Theorem [25,Section D] since $C^{\boldsymbol{\nu}}_t$ is continuous in a, $\eta^{\boldsymbol{\nu}}_t$ is weakly continuous in a, and A is compact. If we define $\nu'_t(\mathrm{d}z,\mathrm{d}a) = \mu_t(\mathrm{d}z)\,\delta_{f_t(z)}(\mathrm{d}a)$, then it is straightforward to prove that $\boldsymbol{\nu}' \in C(\boldsymbol{\nu}) \cap B(\boldsymbol{\nu})$, and thus $C(\boldsymbol{\nu}) \cap B(\boldsymbol{\nu}) \neq \emptyset$. Moreover, both $C(\boldsymbol{\nu})$ and $B(\boldsymbol{\nu})$ are convex, and so, their intersection is also convex. Ξ is a convex compact subset of a locally convex topological space $\mathcal{M}(\mathsf{Z} \times \mathsf{A})^{T+2}$, where $\mathcal{M}(\mathsf{Z} \times \mathsf{A})$ denotes the set of all finite signed measures on $\mathsf{Z} \times \mathsf{A}$. Hence, in order to deduce the existence of a fixed point of Γ, we only need to prove that it has a closed graph. Before stating this result, we state the following proposition which is a key element of the proof.

Proposition 4 ([41,Proposition 4.3]) *Let $\boldsymbol{\nu}^{(n)} \to \boldsymbol{\nu}$ in product topology. Then, for all t, $\eta_t^{\boldsymbol{\nu}^{(n)}}(\,\cdot\,|z_n, a_n)$ weakly converges to $\eta_t^{\boldsymbol{\nu}}(\,\cdot\,|z, a)$ for all $(z_n, a_n) \to (z, a) \in Z \times A$.*

Using Proposition 4, we can now prove the following result.

Proposition 5 *The graph of Γ, i.e., the set*

$$\mathrm{Gr}(\Gamma) := \{(\boldsymbol{\nu}, \boldsymbol{\xi}) \in \Xi \times \Xi : \boldsymbol{\xi} \in \Gamma(\boldsymbol{\nu})\},$$

is closed.

Proof The graph $\mathrm{Gr}(\Gamma)$ of Γ is closed if and only if when $(\boldsymbol{\nu}^{(n)}, \boldsymbol{\xi}^{(n)}) \to (\boldsymbol{\nu}, \boldsymbol{\xi})$ as $n \to \infty$ for some $\left\{(\boldsymbol{\nu}^{(n)}, \boldsymbol{\xi}^{(n)})\right\} \subset \Xi$, then we must have $\boldsymbol{\xi} \in \Gamma(\boldsymbol{\nu})$. To that end, let $\left\{(\boldsymbol{\nu}^{(n)}, \boldsymbol{\xi}^{(n)})\right\} \subset \mathrm{Gr}(\Gamma)$ be such that $(\boldsymbol{\nu}^{(n)}, \boldsymbol{\xi}^{(n)}) \to (\boldsymbol{\nu}, \boldsymbol{\xi})$ as $n \to \infty$ for some $(\boldsymbol{\nu}, \boldsymbol{\xi}) \in \Xi \times \Xi$. We prove that $\boldsymbol{\xi} \in \Gamma(\boldsymbol{\nu})$.

Using Proposition 4, we first prove that $\boldsymbol{\xi} \in C(\boldsymbol{\nu})$; that is, for all t, we have

$$\xi_{t+1,1}(\,\cdot\,) = \int_{\mathsf{Z}\times\mathsf{A}} \eta_t^{\boldsymbol{\nu}}(\,\cdot\,|z, a)\nu_t(\mathrm{d}z, \mathrm{d}a).$$

For all n and t, we have

$$\xi_{t+1,1}^{(n)}(\,\cdot\,) = \int_{\mathsf{Z}\times\mathsf{A}} \eta_t^{\boldsymbol{\nu}^{(n)}}(\,\cdot\,|z, a)\nu_t^{(n)}(\mathrm{d}z, \mathrm{d}a). \tag{10}$$

Since $\boldsymbol{\xi}^{(n)} \to \boldsymbol{\xi}$ in Ξ, $\xi_{t+1}^{(n+1)} \to \xi_{t+1}$ weakly. Let $g \in C_b(\mathsf{Z})$. Then, by [33,Theorem 3.5], we have

$$\lim_{n\to\infty} \int_{\mathsf{Z}\times\mathsf{A}} \int_{\mathsf{Z}} g(z')\eta_t^{\boldsymbol{\nu}^{(n)}}(\mathrm{d}z'|z, a)\nu_t^{(n)}(\mathrm{d}z, \mathrm{d}a) = \int_{\mathsf{Z}\times\mathsf{A}} \int_{\mathsf{Z}} g(z')\eta_t^{\boldsymbol{\nu}}(\mathrm{d}z'|z, a)\nu_t(\mathrm{d}x, \mathrm{d}a)$$

since $\boldsymbol{\nu}_t^{(n)} \to \boldsymbol{\nu}_t$ weakly and $\int_{\mathsf{Z}} g(y)\eta_t^{\boldsymbol{\nu}^{(n)}}(\,\cdot\,|z, a)$ converges to $\int_{\mathsf{Z}} g(y)\eta_t^{\boldsymbol{\nu}}(\,\cdot\,|z, a)$ continuously[2] (see [33,Theorem 3.5]). This implies that the measure on the right-hand side of (10) converges weakly to $\int_{\mathsf{Z}\times\mathsf{A}} \eta_t^{\boldsymbol{\nu}}(\,\cdot\,|z, a)\nu_t(\mathrm{d}z, \mathrm{d}a)$. Therefore, we have

$$\xi_{t+1,1}(\,\cdot\,) = \int_{\mathsf{Z}\times\mathsf{A}} \eta_t^{\boldsymbol{\nu}}(\,\cdot\,|z, a)\nu_t(\mathrm{d}z, \mathrm{d}a),$$

from which we conclude that $\boldsymbol{\xi} \in C(\boldsymbol{\nu})$.

To complete the proof, it suffices to prove that $\boldsymbol{\xi} \in B(\boldsymbol{\nu})$. To that end, for each n and t, let us define the following functions:

$$F_t^{(n)}(z, a) = C_t^{\boldsymbol{\nu}^{(n)}}(z, a) + \int_{\mathsf{Z}} J_{*,t+1}^{\boldsymbol{\nu}^{(n)}}(y)\eta_t^{\boldsymbol{\nu}^{(n)}}(\mathrm{d}y|z, a)$$

and

$$F_t(z, a) = C_t^{\boldsymbol{\nu}}(z, a) + \int_{\mathsf{Z}} J_{*,t+1}^{\boldsymbol{\nu}}(y)\eta_t^{\boldsymbol{\nu}}(\mathrm{d}y|z, a).$$

By definition, $J_{*,t}^{\boldsymbol{\nu}^{(n)}}(z) = \min_{a\in\mathsf{A}} F_t^{(n)}(z, a)$ and $J_{*,t}^{\boldsymbol{\nu}}(z) = \min_{a\in\mathsf{A}} F_t(z, a)$. Define also the following sets:

$$A_t^{(n)} = \left\{(z, a) : F_t^{(n)}(z, a) = J_{*,t}^{\boldsymbol{\nu}^{(n)}}(z)\right\} \text{ and } A_t = \left\{(z, a) : F_t(z, a) = J_{*,t}^{\boldsymbol{\nu}}(z)\right\}.$$

[2] Suppose g, g_n $(n \geq 1)$ are measurable functions on metric space E. The sequence g_n is said to converge to g continuously if $\lim_{n\to\infty} g_n(e_n) = g(e)$ for any $e_n \to e$ where $e \in \mathsf{E}$.

Since $\boldsymbol{\xi}^{(n)} \in B(\boldsymbol{\nu}^{(n)})$, we have $1 = \xi_t^{(n)}\big(A_t^{(n)}\big),$ for all n and$t$. To prove to $\boldsymbol{\xi} \in B(\boldsymbol{\nu})$, we need to show that $1 = \xi_t\big(A_t\big),$ for all t.

First note that since both $F_t^{(n)}$ and $J_{*,t}^{\boldsymbol{\nu}^{(n)}}$ are continuous, $A_t^{(n)}$ is closed. Moreover, A_t is also closed as both F_t and $J_{*,t}^{\boldsymbol{\nu}}$ are continuous. Using Proposition 4, one can also prove as in [40,Proposition 3.10], [42,Proposition 4.4] that $F_t^{(n)}$ converges to F_t continuously and $J_{*,t}^{\boldsymbol{\nu}^{(n)}}$ converges to $J_{*,t}^{\boldsymbol{\nu}}$ continuously, as $n \to \infty$.

For each $M \geq 1$, define the closed set $B_t^M = \big\{(z,a) : F_t(z,a) \geq J_{*,t}^{\boldsymbol{\nu}}(z) + \epsilon(M)\big\}$, where the sequence $\{\epsilon(M)\}$ is decreasing and $\epsilon(M) \to 0$ as $M \to \infty$. Since both F_t and $J_{*,t}^{\boldsymbol{\nu}}$ are continuous, we can choose $\{\epsilon(M)\}_{M \geq 1}$ so that $\xi_t(\partial B_t^M) = 0$ for each M. Note that by the monotone convergence theorem, we have

$$\xi_t^{(n)}\big(A_t^c \cap A_t^{(n)}\big) = \liminf_{M \to \infty} \xi_t^{(n)}\big(B_t^M \cap A_t^{(n)}\big).$$

This implies that

$$\begin{aligned} 1 &= \limsup_{n\to\infty} \liminf_{M\to\infty} \Big\{\xi_t^{(n)}\big(A_t \cap A_t^{(n)}\big) + \xi_t^{(n)}\big(B_t^M \cap A_t^{(n)}\big)\Big\} \\ &\leq \liminf_{M\to\infty} \limsup_{n\to\infty} \Big\{\xi_t^{(n)}\big(A_t \cap A_t^{(n)}\big) + \xi_t^{(n)}\big(B_t^M \cap A_t^{(n)}\big)\Big\}. \end{aligned}$$

For any fixed M, we prove that the limit of the second term in the last expression converges to zero. To that end, we first note that $\xi_t^{(n)}$ converges weakly to ξ_t as $n \to \infty$ when both measures are restricted to B_t^M, as B_t^M is closed and $\xi_t(\partial B_t^M) = 0$ [10,Theorem 8.2.3]. Furthermore, since $F_t^{(n)}$ converges to F_t continuously and $J_{*,t}^{\boldsymbol{\nu}^{(n)}}$ converges to $J_{*,t}^{\boldsymbol{\nu}}$ continuously, $1_{A_t^{(n)} \cap B_t^M}$ converges continuously to 0, which implies by [33,Theorem 3.5] that

$$\limsup_{n\to\infty} \xi_t^{(n)}\big(B_t^M \cap A_t^{(n)}\big) = 0.$$

Therefore, we obtain

$$1 \leq \limsup_{n\to\infty} \xi_t^{(n)}\big(A_t \cap A_t^{(n)}\big) \leq \limsup_{n\to\infty} \xi_t^{(n)}(A_t) \leq \xi_t(A_t),$$

where the last inequality follows from the Portmanteau theorem [8,Theorem 2.1] and the fact that A_t is closed. Hence, $\xi_t(A_t) = 1$. Since t is arbitrary, this is true for all t. This means that $\boldsymbol{\xi} \in B(\boldsymbol{\nu})$. Therefore, $\boldsymbol{\xi} \in \Gamma(\boldsymbol{\nu})$. □

As a result of Proposition 5, we now conclude via Kakutani's fixed point theorem ([2,Corollary 17.55]) that Γ has a fixed point. Therefore, the pair $(\pi^*, \boldsymbol{\mu}^*)$ in Proposition 2 is a mean field equilibrium. This completes the proof of Theorem 1.

5 Approximation of Nash Equilibria

We are now ready to prove that the policy in the mean-field equilibrium, when applied by every agent, is approximately Nash equilibrium for mean-field games with a sufficiently large number of agents. Let $(\pi^{'*}, \boldsymbol{\mu}^*)$ denote the pair in the mean-field equilibrium. In order to prove the existence of an approximate Nash equilibrium, we need Assumption 2 in addition to Assumption 1.

Our approach can be summarized as follows: (i) First, Assumption 2 enables us to define another mean-field equilibrium, in which the policy deterministically and continuously depends on only the observations; (ii) we then construct an equivalent game model whose states are the states of the game model in Sect. 2.2 plus the current and past observations; (iii) in this equivalent model, the new mean-field equilibrium policy becomes Markov; (iv) using this Markov structure, we prove that the cost function of a generic agent under any policy in the finite-agent regime, where the rest of the agents adopt mean-field equilibrium policy, converges to the cost function in the infinite-population limit as the number of agents goes to infinity; (v) since the mean-field equilibrium policy is optimal in the infinite-population limit, we establish the existence of an approximate Nash equilibrium via the result in step (iv).

Let d_{BL} denote the bounded Lipschitz metric on $\mathcal{P}(\mathsf{S})$, which metrizes the weak topology [18,Proposition 11.3.2].

Assumption 2 (a) $\omega_q(r) \to 0$ and $\omega_m(r) \to 0$ as $r \to 0$, where

$$\omega_q(r) = \sup_{(s,u)\in\mathsf{S}\times\mathsf{A}} \sup_{\substack{\mu,\nu:\\ d_{BL}(\mu,\nu)\leq r}} \|q(\,\cdot\,|s,u,\mu) - q(\,\cdot\,|s,u,\nu)\|_{TV}$$

$$\omega_m(r) = \sup_{(s,u)\in\mathsf{S}\times\mathsf{A}} \sup_{\substack{\mu,\nu:\\ d_{BL}(\mu,\nu)\leq r}} |m(s,u,\mu) - m(s,u,\nu)|.$$

(b) For each $t \geq 0$, $\pi_t^{'*} : \mathsf{G}_t \to \mathcal{P}(\mathsf{A})$ is deterministic; that is, $\pi_t^{'*}(\,\cdot\,|g(t)) = \delta_{f_t(g(t))}(\,\cdot\,)$ for some measurable function $f_t : \mathsf{G}_t \to \mathsf{A}$, and weakly continuous.

In Appendix 1, we give sufficient conditions for Assumption 2-(b) in terms of the system components.

We now construct another mean-field equilibrium in which the policy deterministically depends on only the observations. For t, let $\mathsf{Y}^{t+1} = \prod_{k=0}^{t} \mathsf{Y}$. Then, for each $t \geq 1$, define $\tilde{f}_t : \mathsf{Y}^{t+1} \to \mathsf{A}$ as:

$$\tilde{f}_t(y(t),\ldots,y(0)) = f_t\big(y(t),\ldots,y(0),\tilde{f}_{t-1}(y(t-1),\ldots,y(0)),\ldots,\tilde{f}_0(y(0))\big),$$

where $\tilde{f}_0 = f_0$. Let $\pi_t^*(\,\cdot\,|y(t),\ldots,y(0)) = \delta_{\tilde{f}_t(y(t),\ldots,y(0))}(\,\cdot\,)$. Note that π_t^* is a weakly continuous stochastic kernel on A given Y^{t+1} under Assumption 2-(b). Moreover, π^* and $\pi^{'*}$ are equivalent because, for all t, we have

$$\begin{aligned} P^{\pi^{'*}}\big(a(t) \in \,\cdot\,|g(t)\big) &= P^{\pi^{'*}}\big(a(t) \in \,\cdot\,|y(t),\ldots,y(0)\big) \\ &= P^{\pi^{*}}\big(a(t) \in \,\cdot\,|y(t),\ldots,y(0)\big). \end{aligned}$$

Hence, $(\pi^*, \boldsymbol{\mu}^*)$ is also a mean-field equilibrium. In the sequel, we use $(\pi^*, \boldsymbol{\mu}^*)$ to prove the approximation result. The reason for passing from f_t to $\tilde{f}_t$ is that the latter policy becomes Markov in the equivalent game model that will be introduced in the proof of Theorem 4. Then, we can prove the existence of an approximate Nash equilibrium by adapting the proof techniques and results in [40, 42] to the game models with expanding state spaces and non-homogeneous system components.

The following theorem is the main result of this section, which states that the policy $\boldsymbol{\pi}^{(N,*)} = (\pi^*,\ldots,\pi^*)$, where π^* is repeated N times, is an ε-Nash equilibrium for sufficiently large N. Its proof appears in the next section.

Theorem 4 *For any $\varepsilon > 0$, there exists $N(\varepsilon)$ such that for $N \geq N(\varepsilon)$, the policy $\boldsymbol{\pi}^{(N,*)}$ is an ε-Nash equilibrium for the game with N agents that is introduced in Sect. 2.2. Since the original N-agent game model is equivalent to the one in Sect. 2.2 by Proposition 1, the policy $\boldsymbol{\pi}^{(N,*)}$ is also an ε-Nash equilibrium for the original game with N agents.*

Remark 2 Note that to obtain an explicit relation between ε and $N(\varepsilon)$, one needs to establish that the optimal policy π^* in mean-field equilibrium is Lipschitz continuous. In the *fully observed continuous-time* setup, this is in general established easily due to very restrictive structural assumptions on the system components. In a recent monograph [16], Lipschitz continuity of the optimal policy in mean-field equilibrium was established in Lemma 3.3 using regularity properties of system components. However, in our setup, in order to establish this, we need Lipschitz continuity, strong convexity, and differentiability conditions on one-stage cost functions $\{C_t\}$ and transition probabilities $\{\eta_t\}$ of the fully observed reduction. However, establishing Lipschitzness of the transition probabilities $\{\eta_t\}$ is in general prohibitive. Indeed, even weak continuity of the transition probabilities $\{\eta_t\}$, which is a much weaker condition than Lipschitz continuity, has been established relatively recently in [19]. Moreover, it was discussed in that paper that even if very restrictive conditions are imposed on the system components, it is not possible to extend weak continuity of the transition probability to setwise continuity, which is also a very weak condition that is used in the stochastic control literature. Therefore, establishing Lipschitz continuity of the transition probabilities $\{\eta_t\}$ is in general prohibitive. This would also be the case for the partially observed continuous-time setup, since the above-mentioned result pertains to the fully observed case.

Remark 3 In the mean-field games literature, uniqueness of the mean-field equilibrium can be established using a monotonicity condition as introduced by Lasry and Lions in [34] (see also [15]). However, in addition to the monotonicity condition, we should also have the following conditions in order to have uniqueness (see, e.g., [15,Assumption U]):

a) The cost function should be in additive form.
b) The one-stage cost function can be additively decomposed into two functions, where the first function is a function of the state and the mean-field term, and the second function is a function of the state and the action.
c) The dynamics of a generic agent should be independent of the mean-field term.
d) For any state-measure flow, there exists a unique optimal policy.

Under these conditions, one can prove that if $(\pi^{\boldsymbol{\mu}}, \boldsymbol{\mu})$ and $(\pi^{\boldsymbol{\nu}}, \boldsymbol{\nu})$ are two mean-field equilibria, then

$$J_{\boldsymbol{\mu}}(\pi^{\boldsymbol{\mu}}) + J_{\boldsymbol{\nu}}(\pi^{\boldsymbol{\nu}}) \geq J_{\boldsymbol{\mu}}(\pi^{\boldsymbol{\nu}}) + J_{\boldsymbol{\nu}}(\pi^{\boldsymbol{\mu}}) \tag{11}$$

in the equivalent game model. This implies that $J_{\boldsymbol{\mu}}(\pi^{\boldsymbol{\mu}}) = J_{\boldsymbol{\mu}}(\pi^{\boldsymbol{\nu}})$ and $J_{\boldsymbol{\nu}}(\pi^{\boldsymbol{\mu}}) = J_{\boldsymbol{\nu}}(\pi^{\boldsymbol{\mu}})$. Then, conditions c) and d) ensure that these mean-field equilibria must be the same, which implies uniqueness. However, note that to have inequality (11), conditions a), b), and c) must hold. Indeed, to state the monotonicity condition, we should have condition b).

In our case, the cost function in the equivalent game model is in additive form, and thus we do have condition a). Moreover, we can assume the decomposition in condition b). However, if we assume that transition probabilities $\{p_t\}$ are independent of the mean-field term, then it implies that the transition probability q and the one-stage cost function m of the original game model are independent of the mean-field term since

$$p_t\big(B \times D \big| x(t), a(t), \mu_t\big) = q(B|s(t), a(t), \mu_{t,1}) \otimes \delta_{m(t)+\beta^t m(s(t),a(t),\mu_{t,1})}(D).$$

But this is merely a risk-sensitive stochastic control setup.

Conversely, if we consider the original game model instead of the equivalent one, then, in this case, the cost function is not in additive form and thus, we cannot achieve inequality (11) because we cannot have conditions a) and b), which are needed along with the monotonicity condition to have unique mean-field equilibrium.

6 Proof of Theorem 4

For the game model introduced in Sect. 2.2, the policy π^* in the mean-field equilibrium is not necessarily Markov, and so, the joint process of the state, observation, and mean-field term does not have the Markov property as well. To prove Theorem 4, we will first introduce another equivalent game model whose states are the state of the original game model[3] plus the current and past observations. In this new model, the mean-field equilibrium policy automatically becomes Markov.

In the infinite-population limit, this new mean-field game model is specified by

$$\left(\{\mathsf{S}_t\}_{t=0}^{T+1}, \mathsf{A}, \{P_t\}_{t=0}^{T+1}, \{\mathcal{C}_t\}_{t=0}^{T+1}, \lambda_0\right),$$

where, for each t, $\mathsf{S}_t = \mathsf{X} \times \underbrace{\mathsf{Y} \times \ldots \times \mathsf{Y}}_{t+1\text{-times}}$ and A are the Polish state and action spaces at time t, respectively. The stochastic kernel $P_t : \mathsf{S}_t \times \mathsf{A} \times \mathcal{P}(\mathsf{S}_t) \to \mathcal{P}(\mathsf{S}_{t+1})$ is defined as:

$$\begin{aligned} &P_t\big(B_{t+1} \times D_{t+1} \times \ldots \times D_0 \big| b(t), a(t), \Delta_t\big) \\ &= \int_{B_{t+1}} r(D_{t+1}|x(t+1)) \prod_{k=0}^{t} 1_{D_k}(y(k)) p_t(\mathrm{d}x(t+1)|x(t), a(t), \Delta_{t,1}), \end{aligned}$$

where $B_{t+1} \in \mathcal{B}(\mathsf{X})$, $D_k \in \mathcal{B}(\mathsf{Y})$ $(k = 0, \ldots, t+1)$, $b(t) = (x(t), y(t), y(t-1), \ldots, y(0))$, and $\Delta_{t,1}$ is the marginal of Δ_t on X. Indeed, P_t is the controlled transition probability of next state-observation pair, current observation, and past observations, i.e., $\big(x(t+1), y(t+1), y(t), \ldots, y(0)\big)$, given the current state-observation pair and past observations, i.e., $\big(x(t), y(t), y(t-1), \ldots, y(0)\big)$, in the original mean-field game. For each t, the one-stage cost function $\mathcal{C}_t : \mathsf{S}_t \times \mathsf{A} \times \mathcal{P}(\mathsf{S}_t) \to [0, \infty)$ (do not confuse this with C_t in Sect. 4) is defined as:

$$\mathcal{C}_t(b(t), a(t), \Delta_t) = c_t(x(t), a(t), \Delta_{t,1}).$$

Finally, the initial measure λ_0 is given by $\lambda_0(\mathrm{d}b) = r(\mathrm{d}y|x)\mu_0(\mathrm{d}x)$, where $b = (x, y)$. Suppose that Assumption 1 and Assumption 2 hold. Then, for each t, the following are satisfied:

(I) The one-stage cost function $\mathcal{C}_t$ is bounded and continuous.
(II) The stochastic kernel P_t is weakly continuous.

It is straightforward to prove that (I) and (II) hold since c_t is continuous, p_t is weakly continuous, and r is continuous in total variation norm. Recall the set of policies $\tilde{\Pi}$ in the original mean-field game which only use the observations; that is, $\pi \in \tilde{\Pi}$ if $\pi_t : \mathsf{Y}^{t+1} \to \mathcal{P}(\mathsf{A})$ for each $t \geq 0$. Note that $\tilde{\Pi}$ is a subset of the set of Markov policies in the new model. For

[3] When we say original game model in this section, it means the game model introduced in Sect. 2.2 in place of the risk-sensitive game model.

any measure flow $\boldsymbol{\Delta} = (\Delta_t)_{t\geq 0}$, where $\Delta_t \in \mathcal{P}(\mathsf{S}_t)$, we denote by $\hat{J}_{\boldsymbol{\Delta}}(\pi)$ the finite-horizon risk-neutral total cost of the policy $\pi \in \tilde{\Pi}$ in this new mean-field game model.

We also define the corresponding N agent game as follows. We have the Polish state spaces $\{\mathsf{S}_t\}_{t=0}^{T+1}$ and action space A. For every t and every $i \in \{1, 2, \dots, N\}$, let $b_i^N(t) \in \mathsf{S}_t$ and $a_i^N(t) \in \mathsf{A}$ denote the state and the action of Agent i at time t, and let

$$\Delta_t^{(N)}(\,\cdot\,) = \frac{1}{N}\sum_{i=1}^{N} \delta_{b_i^N(t)}(\,\cdot\,) \in \mathcal{P}(\mathsf{S}_t)$$

denote the empirical distribution of the state configuration at time t. The initial states $b_i^N(0)$ are independent and identically distributed according to λ_0, and, for each t, the next-state configuration $(b_1^N(t+1), \dots, b_N^N(t+1))$ is generated according to the probability laws

$$\prod_{i=1}^{N} P_t\big(\mathrm{d}b_i^N(t+1)\big|b_i^N(t), a_i^N(t), \Delta_t^{(N)}\big).$$

Recall that $\tilde{\Pi}_i$ denotes the set of policies that only use local observations for Agent i in the original game. Note that policies in $\tilde{\Pi}_i$ are Markov for the new model since they partly use the state information. We let $\tilde{\Pi}_i^c$ denote the set of all policies in $\tilde{\Pi}_i$ for Agent i that are weakly continuous; that is, $\pi = \{\pi_t\} \in \tilde{\Pi}_i^c$ if for all $t \geq 0$, $\pi_t : \mathsf{Y}^{t+1} \to \mathcal{P}(\mathsf{A})$ is continuous when $\mathcal{P}(\mathsf{A})$ is endowed with the weak topology. For Agent i, the finite-horizon risk-neutral total cost under the initial distribution λ_0 and N-tuple of policies $\boldsymbol{\pi}^{(N)} \in \tilde{\boldsymbol{\Pi}}^{(N)}$ is denoted by $\hat{J}_i^{(N)}(\boldsymbol{\pi}^{(N)})$.

The following proposition makes the connection between this new model and the original model.

Proposition 6 *For any $N \geq 1$, $\boldsymbol{\pi}^{(N)} \in \tilde{\boldsymbol{\Pi}}^{(N)}$, and $i = 1, \dots, N$, we have $\hat{J}_i(\boldsymbol{\pi}^{(N)}) = J_i(\boldsymbol{\pi}^{(N)})$. Similarly, for any $\pi \in \tilde{\Pi}$ and measure flow $\boldsymbol{\Delta}$, we have $\hat{J}_{\boldsymbol{\Delta}}(\pi) = J_{\boldsymbol{\mu}}(\pi)$ where $\boldsymbol{\mu} = (\Delta_{t,1})_{t\geq 0}$.*

Proof The result can easily be proved as in [41,Proposition 5.1], and thus, we do not include the details. □

By Proposition 6, in the remainder of this section we consider the new game model in place of the one introduced in Sect. 2.2. Define the measure flow $\boldsymbol{\Delta} = (\Delta_t)_{t\geq 0}$ as follows:

$$\Delta_t = \mathcal{L}(x(t), y(t), \dots, y(0)),$$

where $\mathcal{L}(x(t), y(t), \dots, y(0))$ denotes the probability law of $(x(t), y(t), \dots, y(0))$ in the original mean-field game under the policy π^* in the mean-field equilibrium. For each $t \geq 0$, define the stochastic kernel $P_t^{\pi^*}(\,\cdot\,|b, \Delta)$ on S_{t+1} given $\mathsf{S}_t \times \mathcal{P}(\mathsf{S}_t)$ as

$$P_t^{\pi^*}(\,\cdot\,|b, \Delta) = \int_{\mathsf{A}} P_t(\,\cdot\,|b, a, \Delta)\pi_t^*(\mathrm{d}a|b).$$

Since π_t^* is weakly continuous, $P_t^{\pi^*}(\,\cdot\,|b, \Delta)$ is also weakly continuous in (b, Δ). In the sequel, to ease the notation, we will also write $P_t^{\pi^*}(\,\cdot\,|b, \Delta)$ as $P_{t,\Delta}^{\pi^*}(\,\cdot\,|b)$.

Lemma 1 *Measure flow $\boldsymbol{\Delta}$ satisfies*

$$\begin{aligned}\Delta_{t+1}(\,\cdot\,) &= \int_{\mathsf{S}_t} P_t^{\pi^*}(\,\cdot\,|b, \Delta_t)\Delta_t(db)\\ &= \Delta_t P_{t,\Delta_t}^{\pi^*}(\,\cdot\,).\end{aligned}$$

Proof The result can easily be proved as in [41,Lemma 5.1], and thus, we do not include the details. □

For each $N \geq 1$, let $\{b_i^N(t)\}_{1 \leq i \leq N}$ denote the states of agents at time t in the N-agent new game model under the policy $\boldsymbol{\pi}^{(N,*)} = \{\pi^*, \pi^*, \ldots, \pi^*\}$. Define the empirical distribution

$$\Delta_t^{(N)}(\,\cdot\,) = \frac{1}{N} \sum_{i=1}^{N} \delta_{b_i^N(t)}(\,\cdot\,).$$

Proposition 7 *For all $t \geq 0$, we have $\mathcal{L}(\Delta_t^{(N)}) \to \delta_{\Delta_t}$ weakly in $\mathcal{P}(\mathcal{P}(\mathsf{S}_t))$, as $N \to \infty$.*

Proof Weak topology on $\mathcal{P}(\mathsf{S}_t)$ can be metrized using the following metric:

$$\rho(\mu, \nu) = \sum_{m=1}^{\infty} 2^{-(m+1)} |\mu(f_m) - \nu(f_m)|,$$

where $\{f_m\}_{m \geq 1}$ is a sequence of real continuous and bounded functions on S_t such that $\|f_m\| \leq 1$ for all $m \geq 1$ (see [38,Theorem 6.6, p. 47]). Define the Wasserstein distance of order 1 on the set of probability measures $\mathcal{P}(\mathcal{P}(\mathsf{S}_t))$ as follows (see [51,Definition 6.1]):

$$W_1(\Phi, \Psi) = \inf\big\{E[\rho(X, Y)] : \mathcal{L}(X) = \Phi \text{ and } \mathcal{L}(Y) = \Psi\big\}.$$

Note that since δ_{Δ_t} is a Dirac measure, we have

$$\begin{aligned} W_1(\mathcal{L}(\Delta_t^{(N)}), \delta_{\Delta_t}) &= \big\{E[\rho(X, Y)] : \mathcal{L}(X) = \mathcal{L}(\Delta_t^{(N)}) \text{ and } \mathcal{L}(Y) = \delta_{\Delta_t}\big\} \\ &= E\bigg[\sum_{m=1}^{\infty} 2^{-(m+1)} |\Delta_t^{(N)}(f_m) - \Delta_t(f_m)|\bigg]. \end{aligned}$$

Since convergence in W_1 distance implies weak convergence (see [51,Theorem 6.9]), it suffices to prove that

$$\lim_{N \to \infty} E\big[|\Delta_t^{(N)}(f) - \Delta_t(f)|\big] = 0$$

for any $f \in C_b(\mathsf{S}_t)$ and for all t. We prove this by induction on t.

As $\{b_i^N(0)\}_{1 \leq i \leq N}$ are i.i.d. with common distribution Δ_0, the claim is true for $t = 0$. We suppose that the claim holds for t and consider $t + 1$. Fix any $g \in C_b(\mathsf{S}_{t+1})$. Then, we have

$$\begin{aligned} &|\Delta_{t+1}^{(N)}(g) - \Delta_{t+1}(g)| \\ &\leq |\Delta_{t+1}^{(N)}(g) - \Delta_t^{(N)} P_{t,\Delta_t^{(N)}}^{\pi^*}(g)| + |\Delta_t^{(N)} P_{t,\Delta_t^{(N)}}^{\pi^*}(g) - \Delta_t P_{t,\Delta_t}^{\pi^*}(g)|. \end{aligned} \tag{12}$$

We first prove that the expectation of the second term on the right-hand side (RHS) of (12) converges to 0 as $N \to \infty$. To that end, define $F : \mathcal{P}(\mathsf{S}_t) \to \mathbb{R}$ as:

$$F(\Delta) = \Delta P_{t,\Delta}^{\pi^*}(g) = \int_{\mathsf{S}_t} \int_{\mathsf{S}_{t+1}} g(b') P_t^{\pi^*}(\mathrm{d}b' | b, \Delta) \Delta(\mathrm{d}b).$$

One can prove that $F \in C_b(\mathcal{P}(\mathsf{S}_t))$. Indeed, suppose that Δ_n converges to Δ. Let us define

$$l_n(b) = \int_{\mathsf{S}_{t+1}} g(b') P_t^{\pi^*}(\mathrm{d}b' | b, \Delta_n) \text{ and } l(b) = \int_{\mathsf{S}_{t+1}} g(b') P_t^{\pi^*}(\mathrm{d}b' | b, \Delta).$$

Since $P_t^{\pi^*}$ is weakly continuous, one can prove that l_n converges to l continuously. By [33,Theorem 3.5], we have $F(\Delta_n) \to F(\Delta)$, and so, $F \in C_b(\mathcal{P}(\mathsf{S}_t))$. This implies that the expectation of the second term on the RHS of (12) converges to zero as $\mathcal{L}(\Delta_t^{(N)}) \to \delta_{\Delta_t}$ weakly, by the induction hypothesis.

Now, let us write the expectation of the first term on the RHS of (12) as:

$$E\left[E\left[|\Delta_{t+1}^{(N)}(g) - \Delta_t^{(N)} P_{t,\Delta_t^{(N)}}^{\pi^*}(g)| \middle| b_1^N(t), \ldots, b_N^N(t)\right]\right].$$

Then, by [11,Lemma A.2], we have

$$E\left[|\Delta_{t+1}^{(N)}(g) - \Delta_t^{(N)} P_{t,\Delta_t^{(N)}}^{\pi^*}(g)| \middle| b_1^N(t), \ldots, b_N^N(t)\right] \leq 2\frac{\|g\|}{\sqrt{N}}.$$

Therefore, the expectation of the first term on the RHS of (12) also converges to zero as $N \to \infty$. Since g was arbitrary, this completes the proof. □

The implication of Proposition 7 is the key to prove the main theorem. It basically says that in the infinite-population limit, the empirical distribution of the states under the mean-field policy converges to the deterministic measure flow $\boldsymbol{\Delta}$ (i.e., the principle of law of large numbers). This result leads to the following important proposition.

Proposition 8 *We have*

$$\lim_{N\to\infty} \hat{J}_1^{(N)}(\boldsymbol{\pi}^{(N,*)}) = \hat{J}_{\boldsymbol{\Delta}}(\pi^*) = \inf_{\pi' \in \Pi} \hat{J}_{\boldsymbol{\Delta}}(\pi').$$

Proof As the transition probabilities $P_t(\cdot\,|d, a, \Delta)$ are continuous in Δ, the dynamics of the state of a generic agent in the finite-agent game with sufficiently many agents and the dynamics of the state in the mean-field game under policies $\boldsymbol{\pi}^{(N,*)} = (\pi^*, \ldots, \pi^*)$ and π^*, respectively, should therefore be close. Hence, the distributions of the states in these games should also be close, from which we obtain the proposition. The precise mathematical proof is given below.

For each $t \geq 0$, let us define

$$\mathcal{C}_{\pi_t^*}(b, \Delta) = \int_{\mathsf{A}} \mathcal{C}_t(b, a, \Delta)\pi_t^*(\mathrm{d}a|b).$$

Note that random elements $\big(b_1^N(t), \ldots, b_N^N(t), \Delta_t^{(N)}\big)$ are exchangeable; that is, for any permutation σ of $\{1, \ldots, N\}$, we have

$$\mathcal{L}\big(b_1^N(t), \ldots, b_N^N(t), \Delta_t^{(N)}\big) = \mathcal{L}\big(b_{\sigma(1)}^N(t), \ldots, b_{\sigma(N)}^N(t), \Delta_t^{(N)}\big).$$

Hence, the cost function at time t can be written as:

$$\begin{aligned} E\big[\mathcal{C}_t(b_1^N(t), a_1^N(t), \Delta_t^{(N)})\big] &= \frac{1}{N}\sum_{i=1}^{N} E\big[\mathcal{C}_t(b_i^N(t), a_i^N(t), \Delta_t^{(N)})\big] \\ &= E\big[\Delta_t^{(N)}\big(\mathcal{C}_{\pi_t^*}(b, \Delta_t^{(N)})\big)\big]. \end{aligned}$$

Define $F : \mathcal{P}(\mathsf{S}_t) \to \mathbb{R}$ as

$$F(\Delta) = \int_{\mathsf{S}_t} \mathcal{C}_{\pi_t^*}(b, \Delta)\Delta(\mathrm{d}b).$$

One can show that $F \in C_b(\mathcal{P}(\mathsf{S}_t))$ as π_t^* is weakly continuous. Hence, by Proposition 7, we obtain

$$\begin{aligned} \lim_{N\to\infty} E\big[\mathcal{C}_t(b_1^N(t), a_1^N(t), \Delta_t^{(N)})\big] &= \lim_{N\to\infty} E\big[\Delta_t^{(N)}\big(\mathcal{C}_{\pi_t^*}(b, \Delta_t^{(N)})\big)\big] \\ &= \lim_{N\to\infty} E[F(\Delta_t^{(N)})] \\ &= F(\Delta_t) \\ &= \Delta_t(\mathcal{C}_{\pi_t^*}(\,\cdot\,, \Delta_t)). \end{aligned} \tag{13}$$

Note that by Lemma 1, the cost in the mean-field game can be written as:

$$\hat{J}_{\boldsymbol{\Delta}}(\pi^*) = \sum_{t=0}^{T+1} \Delta_t(\mathcal{C}_{\pi_t^*}(\,\cdot\,, \Delta_t)).$$

Therefore, by (13) and the dominated convergence theorem, we obtain

$$\lim_{N\to\infty} \hat{J}_1^{(N)}(\boldsymbol{\pi}^{(N,*)}) = \hat{J}_{\boldsymbol{\Delta}}(\pi^*),$$

which completes the proof. □

To obtain the approximation result, we should show that if the policy of some agent deviates from the mean-field equilibrium policy, then the corresponding cost of this agent should be close to the cost in the mean-field limit as in Proposition 8, for N sufficiently large. Since the transition probabilities and the one-stage cost functions are identical for all agents in the game model, it is sufficient to change the policy of Agent 1 for each N. To that end, let $\{\tilde{\pi}^{(N)}\}_{N\geq 1} \subset \tilde{\Pi}_1^c$ be an arbitrary sequence of policies for Agent 1; that is, for each $N \geq 1$ and $t \geq 0$, $\tilde{\pi}_t^{(N)} : \mathsf{Y}^{t+1} \to \mathcal{P}(\mathsf{A})$ is weakly continuous. For each $N \geq 1$, let $\big\{\tilde{b}_i^N(t)\big\}_{1\leq i\leq N}$ be the collection of states in the N-person game under the policy $\tilde{\boldsymbol{\pi}}^{(N)} = \{\tilde{\pi}^{(N)}, \pi^*, \ldots, \pi^*\}$. Define

$$\tilde{\Delta}_t^{(N)}(\,\cdot\,) = \frac{1}{N}\sum_{i=1}^N \delta_{\tilde{b}_i^{(N)}(t)}(\,\cdot\,).$$

The following result says that the asymptotic behavior of the empirical distribution of the states at each time t is insensitive to local deviations from the mean-field equilibrium policy.

Proposition 9 *For all $t \geq 0$, we have $\mathcal{L}(\tilde{\Delta}_t^{(N)}) \to \delta_{\Delta_t}$ weakly $\mathcal{P}(\mathcal{P}(\mathsf{S}_t))$, as $N \to \infty$.*

Proof The proof can be done by slightly modifying the proof of Proposition 7, and therefore will not be included here. □

For each $N \geq 1$, let $\{\hat{b}^N(t)\}_{t\geq 0}$ denote the state trajectory of the generic agent in the mean-field game (i.e., infinite-population limit) under policy $\tilde{\pi}^{(N)}$; that is, $\hat{b}^N(t)$ evolves as follows:

$$\hat{b}^N(0) \sim \lambda_0 \text{ and } \hat{b}^N(t+1) \sim P_{t,\Delta_t}^{\tilde{\pi}^{(N)}}(\,\cdot\,|\hat{b}^N(t)).$$

The cost function of this mean-field game is given by

$$\hat{J}_{\boldsymbol{\Delta}}(\tilde{\pi}^{(N)}) = \sum_{t=0}^{T+1} E\big[C_t(\hat{b}^N(t), \hat{a}^N(t), \Delta_t)\big], \tag{14}$$

where the actions at each time $t \geq 0$ is generated according to the probability law

$$\tilde{\pi}_t^{(N)}(d\hat{a}^N(t)|\hat{b}^N(t)) = \tilde{\pi}_t^{(N)}(d\hat{a}^N(t)|\hat{y}^N(t), \ldots, \hat{y}^N(0)).$$

The following result is a bit technical but very important for proving the main result. Its proof is quite long and complicated, and thus can be found in Appendix 2.

Proposition 10 *For any $t \geq 0$, we have*

$$\lim_{N\to\infty} \left|\mathcal{L}(\tilde{b}_1^N(t))(g_N) - \mathcal{L}(\hat{b}^N(t))(g_N)\right| = 0$$

for any sequence $\{g_N\} \subset C_b(\mathsf{S}_t)$ such that $\sup_{N\geq 1} \|g_N\| < \infty$ and $\omega_g(r) \to 0$ as $r \to 0$, where

$$\omega_g(r) = \sup_{\substack{s\in S \\ y^t\in Y^t}} \sup_{N\geq 1} \sup_{\substack{m,m' \\ |m-m'|\leq r}} |g_N(s,m,y^t) - g_N(s,m',y^t)|.$$

Using Proposition 10, we now prove the following result.

Theorem 5 *Let $\{\tilde{\pi}^{(N)}\}_{N\geq 1} \subset \tilde{\Pi}_1^c$ be an arbitrary sequence of policies for Agent 1. Then, we have*

$$\lim_{N\to\infty} \left|\hat{J}_1^{(N)}(\tilde{\pi}^{(N)}, \pi^*, \ldots, \pi^*) - \hat{J}_{\boldsymbol{\Delta}}(\tilde{\pi}^{(N)})\right| = 0,$$

where $\hat{J}_{\boldsymbol{\Delta}}(\tilde{\pi}^{(N)})$ is given in (14).

Proof Since $\mathcal{C}_t = 0$ for $t \leq T$, we set $t = T + 1$. We have

$$\left|\hat{J}_1^{(N)}(\tilde{\pi}^{(N)}, \pi^*, \ldots, \pi^*) - \hat{J}_{\boldsymbol{\Delta}}(\tilde{\pi}^{(N)})\right| = \left|E\big[\mathcal{C}_t(\tilde{b}_1^N(t))\big] - E\big[\mathcal{C}_t(\hat{b}_1^N(t))\big]\right|.$$

Note that $\mathcal{C}_t(b) = \mathcal{C}_t((s, m, y_0, \ldots, y_t)) = e^{\lambda m}$, where $m \in [0, L]$, is Lipschitz. Therefore, the term in the above equation converges to zero by Proposition 10. □

As a corollary of Proposition 8 and Theorem 5, we obtain the following result.

Corollary 1 *We have*

$$\begin{aligned}\lim_{N\to\infty} \hat{J}_1^{(N)}(\tilde{\pi}^{(N)}, \pi^*, \ldots, \pi^*) &\geq \inf_{\pi'\in\tilde{\Pi}} \hat{J}_{\boldsymbol{\Delta}}(\pi') = \hat{J}_{\boldsymbol{\Delta}}(\pi^*) \\ &= \lim_{N\to\infty} \hat{J}_1^{(N)}(\pi^*, \pi^*, \ldots, \pi^*),\end{aligned}$$

where $\{\tilde{\pi}^{(N)}\}_{N\geq 1} \subset \tilde{\Pi}_1^c$ is an arbitrary sequence of policies for Agent 1.

Now, we are ready to prove the main result of this section.

Proof of Theorem 4 One can prove that for any policy $\boldsymbol{\pi}^{(N)} \in \tilde{\boldsymbol{\Pi}}^{(N)}$, we have

$$\inf_{\pi^i\in\tilde{\Pi}_i} \hat{J}_i^{(N)}(\boldsymbol{\pi}_{-i}^{(N)}, \pi^i) = \inf_{\pi^i\in\tilde{\Pi}_i^c} \hat{J}_i^{(N)}(\boldsymbol{\pi}_{-i}^{(N)}, \pi^i)$$

for each $i = 1, \ldots, N$ (see the proof of [40,Theorem 2.3]). Hence, it is sufficient to consider weakly continuous policies in ${}^{\cdot(N)}$ to establish the existence of ε-Nash equilibrium in the new model.

We prove that, for sufficiently large N, we have

$$\hat{J}_i^{(N)}(\boldsymbol{\pi}^{(N,*)}) \leq \inf_{\pi^i \in \tilde{\Pi}_i^c} \hat{J}_i^{(N)}(\boldsymbol{\pi}_{-i}^{(N,*)}, \pi^i) + \varepsilon \tag{15}$$

for each $i = 1, \ldots, N$. As indicated earlier, since the transition probabilities and the one-stage cost functions are the same for all agents in the new game, it is sufficient to prove (15) for Agent 1 only. Given $\epsilon > 0$, for each $N \geq 1$, let $\tilde{\pi}^{(N)} \in \tilde{\Pi}_1^c$ be such that

$$\hat{J}_1^{(N)}(\tilde{\pi}^{(N)}, \pi^*, \ldots, \pi^*) < \inf_{\pi' \in \tilde{\Pi}_1^c} \hat{J}_1^{(N)}(\pi', \pi^*, \ldots, \pi^*) + \frac{\varepsilon}{3}.$$

Then, by Corollary 1, we have

$$\begin{aligned}
\lim_{N\to\infty} \hat{J}_1^{(N)}(\tilde{\pi}^{(N)}, \pi^*, \ldots, \pi^*) &= \lim_{N\to\infty} \hat{J}_{\boldsymbol{\Delta}}(\tilde{\pi}^{(N)}) \\
&\geq \inf_{\pi'} \hat{J}_{\boldsymbol{\Delta}}(\pi') \\
&= \hat{J}_{\boldsymbol{\Delta}}(\pi^*) \\
&= \lim_{N\to\infty} \hat{J}_1^{(N)}(\pi^*, \pi^*, \ldots, \pi^*).
\end{aligned}$$

Therefore, there exists $N(\varepsilon)$ such that for $N \geq N(\varepsilon)$, we have

$$\begin{aligned}
\inf_{\pi' \in \tilde{\Pi}_1^c} \hat{J}_1^{(N)}(\pi', \pi^*, \ldots, \pi^*) + \varepsilon &> \hat{J}_1^{(N)}(\tilde{\pi}^{(N)}, \pi^*, \ldots, \pi^*) + \frac{2\varepsilon}{3} \\
&\geq \hat{J}_{\boldsymbol{\Delta}}(\pi^*) + \frac{\varepsilon}{3} \\
&\geq \hat{J}_1^{(N)}(\pi^*, \pi^*, \ldots, \pi^*).
\end{aligned}$$

The result then follows from Proposition 6. □

7 Infinite Horizon Cost Function

In this section, we extend Theorem 4 to games with infinite-horizon risk-sensitive cost functions; that is, a generic agent's infinite-horizon risk-sensitive cost under the initial distribution κ_0 and the N-tuple of infinite-horizon policies $\boldsymbol{\pi}^{(N,\infty)} = (\pi^{(1,\infty)}, \ldots, \pi^{(N,\infty)}) \in {}^{\cdot(N)}$ is given by

$$W_i^{(N,\infty)}(\boldsymbol{\pi}^{(N,\infty)}) = E^{\boldsymbol{\pi}^{(N,\infty)}}\left[e^{\lambda \sum_{t=0}^{\infty} \beta^t m(s_i^N(t), u_i^N(t), d_t^{(N)})}\right],$$

where, for each Agent j, $\pi^{(j,\infty)} = \{\pi_0^{(j,\infty)}, \pi_1^{(j,\infty)}, \ldots\}$ (i.e., infinitely many stochastic kernels). Note that, by [42,Lemma 4.3], any infinite-horizon risk sensitive cost can be approximated by finite T-horizon one with the error bound $\theta\beta^{T+1}$ for some constant $\theta > 0$, which is independent of the policy $\boldsymbol{\pi}^{(N,\infty)}$; i.e.,

$$\left|W_i^{(N,\infty)}(\boldsymbol{\pi}^{(N,\infty)}) - W_i^{(N)}(\boldsymbol{\pi}^{(N,\infty)})\right| \leq \theta\beta^{T+1}. \tag{16}$$

Then, the following theorem is a consequence of (16) and Theorem 4.

Theorem 6 *For any $\varepsilon > 0$, choose T such that $\theta\beta^{T+1} < \frac{\varepsilon}{3}$ and let $N(\frac{\varepsilon}{3})$ be the constant in Theorem 4 for the finite horizon T. Then, for $N \geq N(\frac{\varepsilon}{3})$, the policy $\boldsymbol{\pi}^{(N,\infty)}$ is an ε-Nash* *equilibrium for the infinite-horizon risk-sensitive game with N agents, where $\boldsymbol{\pi}^{(N,\infty)} = (\pi^{\infty}, \ldots, \pi^{\infty})$,*

$$\pi^{\infty} = \{\underbrace{\pi_0^*, \ldots, \pi_T^*}_{T+1\text{-times}}, \pi_{T+1}, \pi_{T+2}, \ldots\},$$

$\pi^* = \{\pi_t^*\}_{t=0}^T$ is the policy in the mean-field equilibrium of the T-horizon game, and $\{\pi_t\}_{t=T+1}^{\infty}$ is some arbitrary policy.

8 An Example

In this section, we consider an additive noise model to illustrate our results. In this model, the state and observation dynamics of a generic agent for the infinite-population game are given, respectively, by

$$\begin{aligned} s(t+1) &= \int_{\mathsf{S}} f(s(t), u(t), s) d_t(\mathrm{d}s) + g(s(t), u(t)) w(t) \\ &=: F(s(t), u(t), d_t) + g(s(t), u(t)) w(t) \end{aligned}$$

and

$$g(t) = h(s(t)) + v(t),$$

where $s(t) \in \mathsf{S}$, $g(t) \in \mathsf{Y}$, $u(t) \in \mathsf{A}$, $w(t) \in \mathsf{W}$, and $v(t) \in \mathsf{V}$. Here, we assume that $\mathsf{S} = \mathsf{Y} = \mathsf{W} = \mathsf{V} = \mathbb{R}$, $\mathsf{A} \subset \mathbb{R}$, and $\{w(t)\}$ and $\{v(t)\}$ are sequences of i.i.d. standard normal random variables independent of each other. The one-stage cost function of a generic agent is given by

$$m(s(t), u(t), d_t) = \int_{\mathsf{S}} b(s(t), u(t), s) d_t(\mathrm{d}s),$$

for some measurable function $b : \mathsf{S} \times \mathsf{A} \times \mathsf{S} \to [0, \infty)$.

This model is the infinite-population limit of the N-agent game model with state and observation dynamics

$$\begin{aligned} s_i^N(t+1) &= \frac{1}{N} \sum_{j=1}^{N} f(s_i^N(t), u_i^N(t), s_j^N(t)) + g(s_i^N(t), u_i^N(t)) w_i^N(t) \\ y_i^N(t) &= h(s_i^N(t)) + v_i^N(t) \end{aligned}$$

and the one-stage cost function

$$m(s_i^N(t), u_i^N(t), d_t^{(N)}) = \frac{1}{N} \sum_{j=1}^{N} b(s_i^N(t), u_i^N(t), s_j^N(t)).$$

For this model, Assumption 1 holds with $v(s) = 1 + s^2$ and $\alpha = \max\{1 + \|f\|^2, L\}$ under the following conditions: (i) A is compact, (ii) b is continuous and bounded, (iii) g is continuous, and f is bounded and continuous, (iv) $\sup_{u \in \mathsf{A}} g^2(s, u) \leq L s^2$ for some $L > 0$, (v) h is continuous and bounded. Note that $\|f\|$ is defined as:

$$\|f\| := \sup_{(s,u,s') \in \mathsf{S} \times \mathsf{A} \times \mathsf{S}} |f(s, u, s')|.$$

Moreover, Assumption 2-(a) holds under the following conditions: (vi) $b(s, u, s')$ is (uniformly) Lipschitz in s', (vii) $f(s, u, s')$ is (uniformly) Lipschitz in s', and (viii) g is bounded and $\inf_{(s,u)\in \mathsf{S}\times\mathsf{A}} |g(s, u)| > 0$. For the proofs of these facts, we refer the reader to [41,Section 7].

In order to have Assumption 2-(b), we need to assume that A is convex. In addition, suppose that $q(ds'|s, a, \mu) = \varrho(s'|s, a, \mu)\nu(ds')$ and $l(dy|s) = \zeta(y|s)\nu(dy)$, where ν denotes the Lebesgue measure. Assume that both ϱ and ζ are continuous and bounded, and ϱ and m are strictly convex in a. For the justification of Assumption 2-(b) in this case, we refer the reader to Sect. 1.

Remark 4 Note that Assumption 1 also holds for finite models (i.e., S, A, and Y are finite) without any structure on the dynamics of the state and observation if the transition probability and the one-stage cost function are continuous with respect to the mean-field term. Moreover, Assumption 2-(a) holds if the transition probability and the one-stage cost function are Lipschitz continuous with respect to the mean-field term. In finite models, the only missing condition is the existence of deterministic policy in mean-field equilibrium. This can be established if we have the uniqueness condition in (17).

9 Conclusion

This paper has considered discrete-time finite-horizon partially observed risk-sensitive mean-field games. We have first constructed an equivalent game model whose states are the state of the original model plus the one-stage costs incurred up to that time. In this new model, the finite-horizon risk-sensitive cost function can be written in an additive-form as in the risk-neutral case. Then, letting the number of agents go to infinity, we have first established the existence of a mean-field equilibrium in the limiting mean-field game problem. We have then shown that the policy in the mean-field equilibrium constitutes an approximate Nash equilibrium for similarly structured games with a sufficiently large number of agents. Finally, we have extended our results to the case of infinite-horizon cost functions.

Acknowledgements This work was partly supported by The Scientific and Technological Research Council of Turkey (TÜBİTAK) BİDEB 2232 Research Grant.

Appendix

Continuous and Deterministic Equilibrium Policy

A common way to establish Assumption 2-(b) is as follows. Suppose that, for the measure-flow $\boldsymbol{\mu}$ in mean-field equilibrium, there exists a unique minimizer $a_z \in \mathsf{A}$ of

$$C_t^{\boldsymbol{\mu}}(z, \cdot) + \int_{\mathsf{Z}} J_{*,t+1}^{\boldsymbol{\mu}}(z')\eta_t^{\boldsymbol{\mu}}(dz'|z, \cdot) = R_t(z, \cdot), \tag{17}$$

for each $z \in \mathsf{Z}$ and for all t. In addition, suppose that $F_t : \mathsf{Z} \times \mathsf{A} \times \mathsf{Y} \to \mathsf{Z}$ in (5) is continuous. Note that uniqueness conditions analogous to (17) are quite common in the mean field literature (see, e.g., [23,Assumption 4], [47,Assumption A5], [28,Assumption H5], [43,Assumption A9]).

Under the condition of a unique minimizer to (17), one can prove that the policy φ in Proposition 2 is deterministic and weakly continuous (see [41,Remark 5.2]). Indeed, fix any

$t \geq 0$ and consider the policy φ_t at time t in φ. By the unique minimizer condition (17), we must have $\varphi_t(\,\cdot\,|z) = \delta_{f_t(z)}(\,\cdot\,)$ for some deterministic function $f_t : \mathsf{Z} \to \mathsf{A}$ which minimizes $R_t(z,\,\cdot\,)$; that is, $\min_{a \in \mathsf{A}} R_t(z,a) = R_t(z, f_t(z))$ for all $z \in \mathsf{Z}$. If f_t is continuous, then φ_t is also weakly continuous. Hence, in order to prove the assertion, it is sufficient to prove that f_t is continuous. Suppose that $z_n \to z$ in Z. Note that $l_t(\,\cdot\,) = \min_{a \in \mathsf{A}} R_t(\,\cdot\,, a)$ is continuous. Therefore, every accumulation point of the sequence $\{f_t(z_n)\}_{n \geq 1}$ must be a minimizer for $R_t(z,\,\cdot\,)$. Since there exists a unique minimizer $f_t(z)$ of $R_t(z,\,\cdot\,)$, the set of all accumulation points of $\{f_t(z_n)\}_{n \geq 1}$ must be the singleton $\{f_t(z)\}$. This implies that $f_t(z_n)$ converges to $f_t(z)$ since A is compact. Hence, f_t is continuous.

Recall that the mean-field equilibrium policy is given by

$$\pi_t(\,\cdot\,|g(t)) = \varphi_t(\,\cdot\,|i(g(t))).$$

Hence, π is also a deterministic policy as i is a deterministic function. The function i can be generated recursively using $F_t : \mathsf{Z} \times \mathsf{A} \times \mathsf{Y} \to \mathsf{Z}$ ($t \geq 0$) in (5) and the policy φ. Since F_t is continuous for all t and φ is also weakly continuous, we can conclude that the mean-field policy π is deterministic and weakly continuous. Hence, Assumption 2-(b) holds.

For instance, we can prove the existence of a unique minimizer to (17) and the continuity of F_t for all t under the following conditions on the system components. Suppose that $\mathsf{S} = \mathbb{R}^d$, $\mathsf{Y} = \mathbb{R}^p$, and $\mathsf{A} \subset \mathbb{R}^m$ is convex. In addition, suppose that $q(\mathrm{d}s'|s,a,\mu) = \varrho(s'|s,a,\mu)\nu(\mathrm{d}s')$ and $l(\mathrm{d}y|s) = \zeta(y|s)\nu(\mathrm{d}y)$, where ν denotes the Lebesgue measure. Assume that both ϱ and ζ are continuous and bounded, and ϱ and m are strictly convex in a, where m is the one-stage cost function of the original problem. Then, we have $H_t(\mathrm{d}y|z,a) = h_t(y|z,a)\nu(\mathrm{d}y)$, where $h_t(y|z,a)$ is given by

$$h_t(y|z,a) = \int_{\mathsf{S}} \int_{\mathsf{S}} \zeta(y|s)\varrho(s|s',a,\mu_t)\nu(\mathrm{d}s)z_1(\mathrm{d}s'),$$

where $z_1(\mathrm{d}s') = z(\mathrm{d}s' \times [0,L])$. Similarly, we have

$$F_t(\mathrm{d}x|z,a,y) = \frac{\int_{\mathsf{X}} f_t(s|s',a,y)\nu(\mathrm{d}s) \otimes \delta_{m'+\beta^t m(s',a,\mu_t)}(\mathrm{d}m)z(\mathrm{d}s',\mathrm{d}m')}{h_t(y|z,a)},$$

where $f_t(s|s',a,y)$ is given by $f_t(s|s',a,y) = \zeta(y|s)\varrho(s|s',a,\mu_t)$. Then, one can prove that F_t is continuous. To show uniqueness of the minimizer to (17), note that

$$\begin{aligned} J^{\boldsymbol{\mu}}_{*,t+1}(z) &= \inf_{\varphi \in \Phi} E^{\varphi}\left[\sum_{k=t+1}^{T+1} C^{\boldsymbol{\mu}}_k(z(k),a(k)) \middle| z(t+1) = z\right] \\ &= \inf_{\pi \in \Pi} E^{\pi}\left[\sum_{k=t+1}^{T+1} c_k(x(k),a(k),\mu_k) \middle| x(t+1) \sim z\right] \\ &= \int_{\mathsf{X}} V_{*,t+1}(x)z(\mathrm{d}x), \end{aligned}$$

where

$$V_{*,t+1}(x) = \inf_{\pi \in \Pi} E^{\pi}\left[\sum_{k=t+1}^{T+1} c_k(x(k),a(k),\mu_k) \middle| x(t+1) = x\right].$$

Hence, for any $a \in \mathsf{A}$, (17) can be written as:

$$\begin{aligned}&\int_{\mathsf{X}} c_t(x,a,\mu_t)z(\mathrm{d}x) + \int_{\mathsf{Y}}\int_{\mathsf{X}} V_{*,t+1}(x)F_t(z,a,y)(\mathrm{d}x)H_t(\mathrm{d}y|z,a)\\ &= \int_{\mathsf{X}} c_t(x,a,\mu_t)z(\mathrm{d}x)\\ &+ \int_{\mathsf{X}}\int_{\mathsf{S}} V_{*,t+1}(s,m'+\beta^t m(s',a,\mu_t))\varrho(s|s',a,\mu_t)\nu(\mathrm{d}s)z(\mathrm{d}s',\mathrm{d}m').\end{aligned}$$

Note that

$$V_{*,t+1}(s,m) = e^{\lambda m}\inf_{\pi\in\Pi} e^{\lambda E[\sum_{k=t+1}^{T}\beta^k m(s(k),u(k),\mu_{1,k})|s(k)=s]},$$

and thus $V_{*,t+1}(s,m)$ is strictly convex in m. Since m and ϱ are strictly convex in a, the last expression is also strictly convex in a. Hence, there exists a unique minimizer $a_z \in \mathsf{A}$ for (17).

Proof of Proposition 10

We prove the result by induction on t. The claim trivially holds for $t=0$ as $\mathcal{L}(\tilde{b}_1^N(0)) = \mathcal{L}(\hat{b}^N(0)) = \lambda_0$ for all $N \geq 1$. Suppose that the claim holds for t and consider $t+1$. Set $\sup_{N\geq 1}\|g_N\| =: L < \infty$ and define

$$T_N(b,\Delta) := \int_{\mathsf{A}\times\mathsf{B}_{t+1}} g_N(b')P_t(\mathrm{d}b'|b,a,\Delta)\tilde{\pi}_t^{(N)}(\mathrm{d}a|b).$$

We can write

$$\begin{aligned}&\left|\mathcal{L}(\tilde{b}_1^N(t+1))(g_N) - \mathcal{L}(\hat{b}^N(t+1))(g_N)\right|\\ &= \left|\int_{\mathsf{B}_t\times\mathcal{P}(\mathsf{B}_t)} T_N(b,\Delta)\mathcal{L}(\tilde{b}_1^N(t),\tilde{\Delta}_t^{(N)})(\mathrm{d}b,d\Delta)\right.\\ &\left.- \int_{\mathsf{B}_t\times\mathcal{P}(\mathsf{B}_t)} T_N(b,\Delta)\mathcal{L}(\hat{b}^N(t),\delta_{\Delta_t})(\mathrm{d}b,d\Delta)\right|\\ &= \left|\mathcal{L}(\tilde{b}_1^N(t),\tilde{\Delta}_t^{(N)})(T_N) - \mathcal{L}(\hat{b}^N(t),\delta_{\Delta_t})(T_N)\right|.\end{aligned}$$

Note that, for any $b \in \mathsf{B}_t$ and $(\Delta,\Delta') \in \mathcal{P}(\mathsf{B}_t)^2$, we have

$$|T_N(b,\Delta) - T_N(b,\Delta')| \leq \omega_g\big(\omega_m(d_{BL}(\Delta_1,\Delta'_1))\big) + L\omega_q(d_{BL}(\Delta_1,\Delta'_1)),$$

where Δ_1 is the marginal distribution of s under Δ (recall that $b=(s,m,y^t)$). Hence, the family $\{T_N(b,\cdot) : b \in \mathsf{B}_t, N \geq 1\}$ is uniformly bounded and equi-continuous. Moreover, for any $\Delta \in \mathcal{P}(\mathsf{B}_t)$, we have

$$\begin{aligned}\omega_{T,\Delta}(r) &:= \sup_{s,y^t}\sup_{N\geq 1}\sup_{\substack{m,m'\\|m-m'|\leq r}} |T_N(s,m,y^t,\Delta) - T_N(s,m',y^t,\Delta)|\\ &\leq \sup_{s,y^t}\sup_{N\geq 1}\sup_{\substack{m,m'\\|m-m'|\leq r}} \omega_g(|m-m'|) = \omega_g(r).\end{aligned}$$

Hence, $\omega_{T,\Delta}(r) \to 0$ as $r \to 0$. Therefore, $\{T_N\} \subset C_b(\mathsf{B}_t\times\mathcal{P}(\mathsf{B}_t))$ is a sequence of functions such that the family $\big\{T_N(b,\cdot) : b \in \mathsf{B}_t, N \geq 1)\big\}$ is equi-continuous, $\sup_{N\geq 1}\|T_N\| < \infty$, and $\omega_{T,\varDelta}(r) \to 0$ as $r \to 0$ for any $\varDelta \in \mathcal{P}(\mathsf{B}_t)$. We now prove that

$$\lim_{N\to\infty} \left|\mathcal{L}(\tilde{b}_1^N(t), \tilde{\varDelta}_t^{(N)})(T_N) - \mathcal{L}(\hat{b}^N(t), \delta_{\varDelta_t})(T_N)\right| = 0, \tag{18}$$

which would then complete the proof. Indeed, we have

$$\begin{aligned}
&\left|\mathcal{L}(\tilde{b}_1^N(t), \tilde{\varDelta}_t^{(N)})(T_N) - \mathcal{L}(\hat{b}^N(t), \delta_{\varDelta_t})(T_N)\right| \\
&\le \left|\int_{\mathsf{B}_t\times\mathcal{P}(\mathsf{B}_t)} T_N(b,\varDelta)\mathcal{L}(\tilde{b}_1^N(t), \tilde{\varDelta}_t^{(N)})(\mathrm{d}b, d\varDelta)\right. \\
&\quad \left. - \int_{\mathsf{B}_t\times\mathcal{P}(\mathsf{B}_t)} T_N(b,\varDelta)\mathcal{L}(\tilde{b}_1^N(t), \delta_{\varDelta_t})(\mathrm{d}b, d\varDelta)\right| \\
&\quad + \left|\int_{\mathsf{B}_t\times\mathcal{P}(\mathsf{B}_t)} T_N(b,\varDelta)\mathcal{L}(\tilde{b}_1^N(t), \delta_{\varDelta_t})(\mathrm{d}b, d\varDelta)\right. \\
&\quad \left. - \int_{\mathsf{B}_t\times\mathcal{P}(\mathsf{B}_t)} T_N(b,\varDelta)\mathcal{L}(\hat{b}^N(t), \delta_{\varDelta_t})(\mathrm{d}b, d\varDelta)\right|.
\end{aligned} \tag{19}$$

First, note that since the family $\{T_N(\,\cdot\,, \varDelta_t)\}_{N\ge1} \subset C_b(\mathsf{B}_t)$ satisfies the hypothesis of the proposition and the proposition is true for t, by induction hypothesis, we have

$$\lim_{N\to\infty}\left|\int_{\mathsf{B}_t} T_N(b,\varDelta_t)\mathcal{L}(\tilde{b}_1^N(t))(\mathrm{d}b) - \int_{\mathsf{B}_t} T_N(b,\varDelta_t)\mathcal{L}(\hat{b}^N(t))(\mathrm{d}b)\right| = 0.$$

Hence, the second term in (19) converges to zero as $N \to \infty$.

Now, let us consider the first term in (19). To that end, define $\mathcal{F} := \big\{T_N(b,\,\cdot\,) : b \in \mathsf{B}_t, N \ge 1)\big\}$. Note that $\mathcal{F}$ is a uniformly bounded and equi-continuous family of functions on $\mathcal{P}(\mathsf{B}_t)$, and therefore

$$\lim_{N\to\infty} E\left[\sup_{F\in\mathcal{F}}\left|F(\tilde{\varDelta}_t^{(N)}) - F(\varDelta_t)\right|\right] = 0$$

as $\mathcal{L}(\tilde{\varDelta}_t^{(N)}) \to \mathcal{L}(\varDelta_t)$ weakly. Then, we have

$$\begin{aligned}
&\lim_{N\to\infty}\left|\int_{\mathsf{B}_t\times\mathcal{P}(\mathsf{B}_t)} T_N(b,\varDelta)\mathcal{L}(\tilde{b}_1^N(t), \tilde{\varDelta}_t^{(N)})(\mathrm{d}b, d\varDelta)\right. \\
&\left. - \int_{\mathsf{B}_t\times\mathcal{P}(\mathsf{B}_t)} T_N(b,\varDelta)\mathcal{L}(\tilde{b}_1^N(t), \delta_{\varDelta_t})(\mathrm{d}b, d\varDelta)\right| \\
&\le \lim_{N\to\infty}\int_{\mathsf{B}_t}\left|\int_{\mathcal{P}(\mathsf{B}_t)} T_N(b,\varDelta)\mathcal{L}(\tilde{\varDelta}_t^{(N)}|\tilde{b}_1^N(t))(d\varDelta|b)\right. \\
&\quad \left. - \int_{\mathcal{P}(\mathsf{B}_t)} T_N(b,\varDelta)\mathcal{L}(\delta_{\varDelta_t})(d\varDelta)\right|\mathcal{L}(\tilde{b}_1^N(t))(\mathrm{d}b) \\
&\le \lim_{N\to\infty} E\left[E\left[\left|T_N(\tilde{b}_1^N(t), \tilde{\varDelta}_t^{(N)}) - T_N(\tilde{b}_1^N(t), \varDelta_t)\right|\middle|\tilde{b}_1^N(t)\right]\right] \\
&\le \lim_{N\to\infty} E\left[\sup_{F\in\mathcal{F}}\left|F(\tilde{\varDelta}_t^{(N)}) - F(\varDelta_t)\right|\right] \\
&= 0.
\end{aligned}$$

This completes the proof.

References

1. Adlakha S, Johari R, Weintraub G (2015) Equilibria of dynamic games with many players: Existence, approximation, and market structure. J Econ Theory 156:269–316
2. Aliprantis C, Border K (2006) Infinite dimensional analysis, 3rd edn. Springer, Berlin
3. Başar T (2000) Risk-averse designs: From exponential cost to stochastic games. In: Djaferis T, Schick I (eds) System theory: modeling, analysis and control. Kluwer, pp 131–144
4. Başar T (1978) Decentralized multicriteria optimization of linear stochastic systems. IEEE Trans Autom Control 23(2):233–243
5. Başar T (1978) Two-criteria LQG decision problems with one-step delay observation sharing pattern. Inf Control 38(1):21–50
6. Bauerle N, Rieder U (2014) More risk-sensitive Markov decision processes. Math Oper Res 39(1):105–120
7. Bensoussan A, Frehse J, Yam P (2013) Mean field games and mean field type control theory. Springer, New York
8. Billingsley P (1999) Convergence of probability measures, 2nd edn. Wiley, New York
9. Biswas A (2015) Mean field games with ergodic cost for discrete time Markov processes. arXiv:1510.08968
10. Bogachev V (2007) Measure theory, vol II. Springer, Berlin
11. Budhiraja A, Majumder A (2015) Long time results for a weakly interacting particle system in discrete time. Stoch Anal Appl 33(3):429–463
12. Caines PE, Kizilkale AC (2017) ϵ-Nash equilibria for partially observed LQG mean field games with a major player. IEEE Trans Autom Control 62(7):3225–3234
13. Cardaliaguet P (2011) Notes on mean-field games
14. Carmona R, Delarue F (2013) Probabilistic analysis of mean-field games. SIAM J Control Optim 51(4):2705–2734
15. Carmona R, Lacker D (2015) A probabilistic weak formulation of mean field games and applications. Ann Appl Prob 25(3):1189–1231
16. Carmona R, Delarue F (2018) Probabilistic theory of mean field games with applications I: mean field FBSDEs, control, and games. Springer
17. Djehiche B, Tembine H (2016) Risk-sensitive mean-field type control under partial observation. In: Benth F, Nunno GD (eds) Stochastics of Environmental and Financial Economics. Springer International Publishing, Cham, pp 243–263
18. Dudley RM (2004) Real analysis and probability. Cambridge University Press, Cambridge
19. Feinberg E, Kasyanov P, Zgurovsky M (2016) Partially observable total-cost markov decision processes with weakly continuous transition probabilities. Math Oper Res 41(2):656–681
20. Firoozi D, Caines P (2015) ε-Nash equilibria for partially observed lqg mean field games with major agent: Partial observations by all agents. In: CDC 2015. Japan
21. Firoozi D, Caines PE (2019) ϵ-Nash equilibria for major minor LQG mean field games with partial observations of all agents. arXiv:1810.04369
22. Gomes D, Saúde J (2014) Mean field games models—a brief survey. Dyn. Games Appl. 4(2):110–154
23. Gomes D, Mohr J, Souza R (2010) Discrete time, finite state space mean field games. J. Math. Pures Appl. 93:308–328
24. Hernández-Lerma O (1989) Adaptive Markov control processes. Springer-Verlag, Berlin
25. Hernández-Lerma O, Lasserre J (1996) Discrete-time Markov control processes: basic optimality criteria. Springer
26. Hinderer K (1970) Foundations of non-stationary dynamic programming with discrete time parameter. Springer-Verlag, Berlin
27. Huang M (2010) Large-population LQG games involving major player: the Nash certainity equivalence principle. SIAM J Control Optim 48(5):3318–3353
28. Huang M, Malhamé R, Caines P (2006) Large population stochastic dynamic games: Closed loop McKean-Vlasov sysyems and the Nash certainity equivalence principle. Commun Inf Syst 6:221–252
29. Huang M, Caines P, Malhamé R (2007) Large-population cost coupled LQG problems with nonuniform agents: Individual-mass behavior and decentralized ϵ-Nash equilibria. IEEE Trans Autom Control 52(9):1560–1571
30. Huang M, Caines P, Malhame R (2006) Distributed multi-agent decision-making with partial observations: asymptotic Nash equilibria. In: Theory of networks and systems. Japan
31. Huang J, Wang S (2014) A class of mean-field LQG games with partial information. arXiv:1403.5859v1
32. Jovanovic B, Rosenthal R (1988) Anonymous sequential games. J Math Econ 17:77–87
33. Langen H (1981) Convergence of dynamic programming models. Math Oper Res 6(4):493–512

34. Lasry J, Lions P (2007) Mean field games. Japan J Math 2:229–260
35. Moon J, Başar T (2016) Robust mean field games for coupled Markov jump linear systems. Int J Control 89(7):1367–1381
36. Moon J, Başar T (2017) Linear quadratic risk-sensitive and robust mean field games. IEEE Trans Autom Control 62(3):1062–1077
37. Moon J, Başar T (2015) Discrete-time decentralized control using the risk-sensitive performance criterion in the large population regime: a mean field approach. In: ACC 2015. Chicago, pp 4779–4784
38. Parthasarathy K (1967) Probability measures on metric spaces. AMS Bookstore
39. Rhenius D (1974) Incomplete information in Markovian decision models. Ann Statist 2:1327–1334
40. Saldi N, Başar T, Raginsky M (2018) Markov-Nash equilibria in mean-field games with discounted cost. SIAM J Control Optim 56(6):4256–4287
41. Saldi N, Başar T, Raginsky M (2019) Approximate Nash equilibria in partially observed stochastic games with mean-field interactions. Math Oper Res 44(3):1006–1033
42. Saldi N, Başar T, Raginsky M (2020) Approximate Markov-Nash equilibria for discrete-time risk-sensitive mean-field games. Math Oper Res 45(4):1596–1620
43. Şen N, Caines P (2016) Mean field game theory with a partially observed major agent. SIAM J Control Optim 54(6):3174–3224
44. Şen N, Caines P (2016) Nonlinear filtering theory for McKean-Vlasov type stochastic differential equations. SIAM J Control Optim 54(1):153–174
45. Şen N, Caines P (2014) Mean field games with partially observed major player and stochastic mean field. In: CDC 2014. Los Angeles
46. Şen N, Caines P (2015) ϵ-Nash equilibria for a partially observed mean field game with major player. In: ACC 2015. Chicago
47. Şen N, Caines P (2016) On mean field games and nonlinear filtering for agents with individual-state partial observations. In: ACC 2016. Boston
48. Tang M, Meng Q (2016) Partially observed optimal control for mean-field SDEs. arXiv:1610.02587v1
49. Tembine H (2015) Risk-sensitive mean-field-type games with Lp-norm drifts. Automatica 59:224–237
50. Tembine H, Zhu Q, Başar T (2014) Risk-sensitive mean field games. IEEE Trans Autom Control 59(4):835–850
51. Villani C (2009) Optimal transport: old and new. Springer, Berlin
52. Whittle P (1990) Risk-sensitive optimal control. Wiley Interscience Series in Systems and Optimization. Wiley
53. Yushkevich A (1976) Reduction of a controlled Markov model with incomplete data to a problem with complete information in the case of Borel state and control spaces. Theory Prob Appl 21:153–158

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.