Mean Field Stochastic Games with Discrete States and Mixed Players

Minyi Huang

School of Mathematics and Statistics, Carleton University,
Ottawa, ON K1S 5B6, Canada
mhuang@math.carleton.ca

Abstract. We consider mean field Markov decision processes with a major player and a large number of minor players which have their individual objectives. The players have decoupled state transition laws and are coupled by the costs via the state distribution of the minor players. We introduce a stochastic difference equation to model the update of the limiting state distribution process and solve limiting Markov decision problems for the major player and minor players using local information. Under a solvability assumption of the consistent mean field approximation, the obtained decentralized strategies are stationary and have an ε-Nash equilibrium property.

Keywords: mean field game, finite states, major player, minor player.

1 Introduction

Large population stochastic dynamic games with mean field coupling have attracted substantial interest in the recent years; see, e.g., [1,4,11,16,12,13,18,19,22,23,24,26,27]. To obtain low complexity strategies, consistent mean field approximations provide a powerful approach, and in the resulting solution, each agent only needs to know its own state information and the aggregate effect of the overall population which may be pre-computed off-line. One may further establish an ε-Nash equilibrium property for the set of control strategies [12]. The technique of consistent mean field approximations is also applicable to optimization with a social objective [5,14,23]. The survey [3] on differential games presents a timely report of recent progress in mean field game theory. This general methodology has applications in diverse areas [4,20,27]. The mean field approach has also appeared in anonymous sequential games [17] with a continuum of players individually optimally responding to the mean field. However, the modeling of a continuum of independent processes leads to measurability difficulties and the empirical frequency of the realizations of the continuum-indexed individual states cannot be meaningfully defined [2].

A recent generalization of the mean field game modeling has been introduced in [10] where a major player and a large number of minor players coexist pursuing their individual interests. Such interaction models are often seen in economic or engineering settings, simple examples being a few large corporations and many

V. Krishnamurthy et al. (Eds.): GameNets 2012, LNICST 105, pp. 138–151, 2012.

much smaller competitors, a network service provider and a large number of small users with their respective objectives. An extension of the modeling in [10] to dynamic games with Markovian switches in the dynamics is presented in [25]. The random switches model the abrupt changes of the decision environment. Traditionally, game models differentiating vastly different strengths of players have been well studied in cooperative game theory, and static models are usually considered [6,8,9]. Such players with very different strengths are called mixed players.

The linear-quadratic-Gaussian (LQG) model in [10] shows that the presence of the major player causes an interesting phenomenon called the lack of sufficient statistics. More specifically, in order to obtain asymptotic equilibrium strategies, the major player cannot simply use a strategy as a function of its current state and time; for a minor player, it cannot simply use the current states of the major player and itself. To overcome this lack of sufficient statistics for decision, the system dynamics are augmented by adding a new state, which approximates the mean field and is driven by the major player's state. This additional state enters the obtained decentralized strategy of each player and it captures the past influence of the major player. The recent work [21] considered minor players parametrized by a continuum which causes high complexity to the state space augmentation approach, and a backward stochastic differential equation based approach (see, e.g., [28]) was used to deal with the random mean field process. The resulting decentralized strategies are not Markovian.

In this paper, we consider the interaction modeling of a major player and a large number of minor players in the setting of discrete time Markov decision processes (MDPs). Although the major player modeling is conceptually very similar to [10] which considers an LQG game model, the lack of linearity in the MDP context will give rise to many challenges in analysis. Additionally, an important motivation to use the MDP framework is that our method may potentially be applicable to many practical problems. In relation to mean field games with discrete state and action spaces, related work can also be found in [15,23,7,17]; they all consider a population of comparably small decision makers which may be called peers.

A key step in our decentralized control design is to describe the evolution of the mean field, as the distribution of the minor players' states, by a stochastic difference equation driven by the major player's state. Given the above representation of the limiting mean field, we may approximate the original problems of the major player and a typical minor player by limiting MDPs with hybrid state spaces where the player in question has a finite state space and the mean field process is a continuum evolving on a simplex.

The organization of the paper is as follows. Section 2 formulates the mean field Markov decision game with a major player. Section 3 proposes a stochastic representation of the update of the mean field and analyzes two auxiliary MDPs in the mean field limit. The consistency condition for mean field approximations is introduced in Section 4, and Section 5 shows an asymptotic Nash equilibrium property. Section 6 presents concluding remarks of the paper.

2 The Mean Field Game Model

We adopt the framework of Markov decision processes to formulate the mean field game which involves a major player $\mathcal{A}_0$ and a large population of minor players $\{\mathcal{A}_i, 1 \le i \le N\}$. The state and action spaces of all players are finite, and denoted by $S_0 = \{1, \dots, K_0\}$ and $A_0 = \{1, \dots, L_0\}$, respectively, for the major player. For simplicity, we consider uniform minor players which share common state and action spaces denoted by $S = \{1, \dots, K\}$ and $A = \{1, \dots, L\}$, respectively. At time $t \in \mathbb{Z}_+ = \{0, 1, 2, \dots\}$, the state and action of $\mathcal{A}_j$ are, respectively, denoted by $x_j(t), u_j(t)$, $0 \le j \le N$. To model the mean field interaction of the players, we denote the random measure process as follows

$$I^{(N)}(t) = (I_1^{(N)}(t), \dots, I_K^{(N)}(t)), \quad t \ge 0,$$

where $I_k^{(N)}(t) = (1/N) \sum_{i=1}^N 1_{(x_i(t)=k)}$. The process $I^{(N)}(t)$ describes the frequency of occurrence of the states in S at time t.

For the major player, the state transition law is determined by the stochastic kernel

$$Q_0(z|y, a_0) = P(x_0(t+1) = z | x_0(t) = y, u_0(t) = a_0), \tag{1}$$

where y, $z \in S_0$ and $a_0 \in A_0$. Following the usual convention in Markov decision processes, the transition probability of the process x_0 from t to $t+1$ is solely determined by $x_0(t) = y$ and $u_0(t) = a_0$ observed at t even if additional state and action information before t is known.

The one-stage cost of the decision problem of the major player is given by $c_0(x_0, \theta, a_0)$, where θ is the state distribution of the minor players. The infinite horizon discounted cost is

$$J_0 = E \sum_{t=0}^{\infty} \rho^t c_0(x_0(t), I^{(N)}(t), u_0(t)),$$

where $\rho \in (0,1)$ is the discount factor.

The state transition of minor player $\mathcal{A}_i$ is specified by

$$Q(z|y, a) = P(x_i(t+1) = z | x_i(t) = y, u_i(t) = a), \tag{2}$$

where y, $z \in S$ and $a \in A$. The one-stage cost is $c(x, x_0, \theta, a)$ and the infinite horizon discounted cost is

$$J_i = E \sum_{t=0}^{\infty} \rho^t c(x_i(t), x_0(t), I^{(N)}(t), u_i(t)).$$

Due to the structure of the costs J_0 and J_i, the major player has a significant impact on each minor player. By contrast, each minor player has a negligible impact on another minor player or the major player. Also, from the point of view of the major player or a fixed minor player, it does not distinguish other specific individual minor players. Instead, only the aggregate state information $I^{(N)}(t)$ matters at each step, which is an important feature of mean field decision problems.

For the $N+1$ decision processes, we specify the joint distribution as follows. Given the states and actions of all players at time t, the transition probability to a value of $(x_0(t+1), x_1(t+1), \dots, x_N(t+1))$ is simply given by the product of the individual transition probabilities under their respective actions.

For integer $k \geq 2$, denote the simplex

$$\mathcal{D}_k = \left\{ (\lambda_1, \dots, \lambda_k) \in \mathbb{R}_+^k \Big| \sum_{j=1}^{k} \lambda_j = 1 \right\}.$$

To ensure that the individual costs are finite, we introduce the assumption.

(**A1**) The one-stage costs c_0 and c are functions on $S_0 \times \mathcal{D}_K \times A_0$ and $S \times S_0 \times \mathcal{D}_K \times A$, respectively, and they are both continuous in θ. $\diamondsuit$

Remark 1. By the continuity condition in (**A1**), there exists a fixed constant C such that $|c_0| + |c| \leq C$ for all $x_0 \in S_0$, $x \in S$, $a_0 \in A_0$, $a \in A$ and $\theta \in \mathcal{D}_K$.

We further assume the following condition on the initial state distribution of the minor players.

(**A2**) The initial states $x_1(0), \dots, x_N(0)$ are independent and there exists a deterministic $\theta_0 \in \mathcal{D}_K$ such that

$$\lim_{N \to \infty} I^{(N)}(0) = \theta_0$$

with probability one. $\diamondsuit$

2.1 The Traditional Approach and Complexity

Denote the so-called t-history

$$h_t = (x_j(s), u_j(s-1), s \leq t, j = 0, \dots, N), \qquad t \geq 1, \tag{3}$$

and $h_0 = (x_0)$. We may further specify mixed strategies (or policies; we shall use the two names strategy and policy interchangeably), as a probability measure on the action space, of each player depending on h_t, and use the method of dynamic programming to identify Nash strategies for the mean field game. However, for a large population of minor players, this traditional approach is impractical. First, each player must use centralized information which causes high complexity in implementation; second, numerically solving the dynamic programming equation is a prohibitive or even impossible task when the number of minor players exceeds a few dozen.

3 The Mean Field Approximation

To overcome the fundamental complexity difficulty, we use the mean field approximation approach. The basic idea is to introduce a limiting process to approximate the random measure process $I^{(N)}(t)$ and solve localized optimization problems for both the major player and a representative minor player.

Regarding the informational requirement in our decentralized strategy design, we assume (i) the limiting distribution θ_0 and the state $x_0(t)$ of the major player are known to all players, (ii) each minor player knows its own state but not the state of any other particular minor player.

We use a process $\theta(t)$ with state space $\mathcal{D}_K$ to approximate $I^{(N)}(t)$ when $N \to \infty$. Before specifying the rule governing the evolution of $\theta(t)$, we give some intuitive explanation. Due to the presence of the major player, the action of each minor player should be affected by $x_0(t)$ and its own state $x_i(t)$, and this causes the correlation of the individual state processes $\{x_i(t), 1 \le i \le N\}$ in the closed-loop system. The resulting process $\theta(t)$ should be a random process. We propose the updating rule

$$\theta(t+1) = \psi(x_0(t), \theta(t)), \tag{4}$$

where $\theta(0) = \theta_0$. The specific form of ψ will be determined by a procedure of consistent mean field approximations. We consider ψ from the following function class

$$\Psi = \{\phi(i,\theta) = (\phi_1, \ldots, \phi_K) | \phi_k \ge 0, \textstyle\sum_{k\in S} \phi_k = 1\},$$

where $\phi(i,\cdot)$ is continuous on $\mathcal{D}_K$ for all $i \in S_0$. The structure of (4) is analogous to the stochastic ordinary differential equation (ODE) modeling of the random mean field in the mean field LQG game model in [10], where the the evolution of the ODE is driving by the state of the major player.

It is possible to consider a function of the form $\psi(t, x_0, \theta)$, which is more general than in (4). For computational efficiency, we will not seek this generality. And on the other hand, the consideration of a time-invariant function will be sufficient for developing our mean field approximation scheme. More specifically, by introducing (4), we may develop stationary feedback strategies for all the players, and furthermore, the mean field limit of the closed-loop will regenerate a stationary transition law of $\theta(t)$ which is in agreement with the initial assumption of time-invariant dynamics.

3.1 The Limiting Problem of the Major Player

Suppose the function ψ in (4) has been given. The original problem of the major player is now approximated by a new Markov decision process. We will often use x_0, x_i, θ to denote a value of the corresponding processes.

Problem (P0): Minimize

$$\bar{J}_0 = E \sum_{t=0}^{\infty} \rho^t c_0(x_0(t), \theta(t), u_0(t)),$$

where $x_0(t)$ has the transition law (1) and $\theta(t)$ satisfies (4).

Problem (P0) gives a standard Markov decision process. To solve this problem, we use the dynamic programming approach by considering a family of optimization problems associated with different initial conditions. Given the initial state $(x_0, \theta) \in S_0 \times \mathcal{D}_K$ at $t = 0$, define the cost function

$$\bar{J}_0(x_0, \theta, u(\cdot)) = E\left[\sum_{t=0}^{\infty} \rho^t c_0(x_0(t), \theta(t), u_0(t)) | x_0, \theta\right].$$

Denote the value function $v(x_0, \theta) = \inf \bar{J}_0(x_0, \theta, u(\cdot))$, where the infimum is with respect to all mixed policies/strategies of the form $\pi = (\pi(0), \pi(1), \ldots,)$ such that each $\pi(s)$ is a probability measure on A_0, indicating the probability to take a particular action, and depends on all past history $(\ldots, x_0(s-1), \theta(s-1), u_0(s-1), x_0(s), \theta(s))$. By taking two different initial conditions (x_0, θ) and (x_0, θ') and comparing the associated optimal costs, we may easily obtain the following continuity property.

Proposition 1. *For each x_0, the value function $v(x_0, \cdot)$ is continuous on $\mathcal{D}_K$.* □

We write the dynamic programming equation

$$\begin{aligned} & v(x_0, \theta) \\ &= \min_{a_0 \in A_0} \{c_0(x_0, \theta, a_0) + \rho E v(x_0(t+1), \theta(t+1))\} \\ &= \min_{a_0 \in A_0} \left\{ c_0(x_0, \theta, a_0) + \rho \sum_{k \in S_0} Q_0(k | x_0, a_0) v(k, \psi(x_0, \theta)) \right\}. \end{aligned}$$

Since the action space is finite, an optimal policy $\hat{\pi}_0$ solving the dynamic programming equation exists and is determined as a stationary Markov policy of the form $\hat{\pi}_0(x_0, \theta)$, i.e., $\hat{\pi}_0$ is a function of the current state. Let the set of optimal policies be denoted by Π_0. It is possible that Π_0 consists of more than one element.

3.2 The Limiting Problem of the Minor Player

Suppose a particular optimal strategy $\hat{\pi}_0 \in \Pi_0$ has been fixed for the major player. The resulting state process is $x_0(t)$. The decision problem of the minor player is approximated by the following limiting problem.

Problem (P1): Minimize

$$\bar{J}_i = E \sum_{t=0}^{\infty} \rho^t c(x_i(t), x_0(t), \theta(t), u_i(t)),$$

where $x_i(t)$ has the state transition law (2); $\theta(t)$ satisfies (4); and $x_0(t)$ is subject to the control policy $\hat{\pi}_0 \in \Pi_0$. This leads to a Markov decision problem with the state $(x_i(t), x_0(t), \theta(t))$ and control action $u_i(t)$. Following the steps in Section 3.1, we define the value function $w(x_i, x_0, \theta)$.

Before analyzing the value function w, we specify the state transition law of the major player under any mixed strategy π_0. Suppose

$$\pi_0 = (\alpha_1, \ldots, \alpha_{L_0}), \tag{5}$$

which is a probability vector. By the standard convention in Markov decision processes, the strategy π_0 selects action k with probability α_k. We further define

$$Q_0(z|y, \pi_0) = \sum_{l \in A_0} \alpha_l Q_0(z|y, l),$$

where π_0 is given by (5).

The dynamic programming equation is now given by

$$\begin{aligned} & w(x_i, x_0, \theta) \\ = & \min_{a \in A} \{c(x_i, x_0, \theta, a) + \rho E w(x_i(t+1), x_0(t+1), \theta(t+1))\} \\ = & \min_{a \in A} \Big\{ c(x_i, x_0, \theta, a_0) + \rho \sum_{j \in S, k \in S_0} Q(j|x_i, a) Q_0(k|x_0, \hat{\pi}_0) w(j, k, \psi(x_0, \theta)) \Big\}. \end{aligned}$$

The following continuity property parallels Proposition 1.

Proposition 2. *For each pair (x_i, x_0), the value function $w(x_i, x_0, \cdot)$ is continuous on $\mathcal{D}_K$.* □

Again, since the action space in Problem (P1) is finite, the value function is attained by at least one optimal strategy. Let the optimal strategy set be denoted by Π. Note that Π is determined after $\hat{\pi}_0$ is selected first.

Let π be a mixed strategy of the minor player and represented in the form

$$\pi = (\beta_1, \ldots, \beta_L).$$

We determine the state transition law of the minor player as follows

$$Q(z|y, \pi) = \sum_{l \in A} \beta_l Q(z|y, l). \tag{6}$$

We have the following theorem on the closed-loop system.

Theorem 1. *Suppose $\hat{\pi}_0 \in \Pi_0$ and $\hat{\pi} \in \Pi$ is determined after $\hat{\pi}_0$. Under the policy pair $(\hat{\pi}_0, \hat{\pi})$, $(x_i(t), x_0(t), \theta(t))$ is a Markov chain with stationary transition probabilities.*

Proof. It is clear that $\hat{\pi}_0$ and $\hat{\pi}$ are stationary feedback policies as a function of the current state of the corresponding system. They may be represented as two probability vectors

$$\begin{aligned} \hat{\pi}_0 &= (\hat{\pi}_0^1(x_0, \theta), \ldots, \hat{\pi}_0^{L_0}(x_0, \theta)), \\ \hat{\pi} &= (\hat{\pi}^1(x_i, x_0, \theta), \ldots, \hat{\pi}^L(x_i, x_0, \theta)). \end{aligned}$$

The process $(x_i(t), x_0(t), \theta(t))$ is a Markov chain since the transition probability from time t to $t+1$ depends only on the value of $(x_i(t), x_0(t), \theta(t))$ and not on the past history. Suppose at time t, $(x_i(t), x_0(t), \theta(t)) = (j, k, \theta)$. Then at $t+1$, we have the transition probability

$$\begin{aligned} &P\Big(x_i(t+1) = j', x_0(t+1) = k', \theta(t+1) = \theta' \Big| x_i(t), x_0(t), \theta(t)) = (j, k, \theta)\Big) \\ &= Q(j'|j, \hat{\pi}(j, k, \theta)) Q_0(k'|k, \hat{\pi}_0(k, \theta)) \delta_{\psi(k,\theta)}(\theta'). \end{aligned}$$

We use $\delta_a(x)$to denote the dirac function, i.e., $\delta_a(x) = 1$ if $x = a$, and $\delta_a(x) = 0$ elsewhere. It is seen that the transition probability is determined by (j, k, θ) and does not depend on time. □

3.3 Discussions on Mixed Strategies

If Problems (P0) and (P1) are considered alone, one may always select an optimal policy which is a pure policy, i.e., given the current state, the action can be selected in a deterministic manner. However, in the mean field game setting we need to eventually determine the function ψ by a fixed point argument. For this reason, it is generally necessary to consider the optimal policies from the larger class of mixed policies. The restriction to deterministic policies may potentially lead to a nonexistence situation when the consistency requirement is imposed later on the mean field approximation.

4 Replication of the Frequency Process

This section develops the procedure to replicate the dynamics of $\theta(t)$ from the closed-loop system when the minor players apply the control strategies obtained from the limiting Markov decision problems.

We start with a system of N minor players. Suppose the major player has selected its optimal policy $\hat{\pi}_0(x_0, \theta)$ from Π_0. Note that for the general case of Problem (P1), there may be more than one optimal policy. We make the convention that the same optimal policy $\hat{\pi}(x_i, x_0, \theta)$ is used by all the minor players while each minor player substitutes its own state into the feedback policy $\hat{\pi}$. It is necessary to make this convention since otherwise the mean field limit cannot be properly defined if there are multiple optimal policies and if each minor player can take an arbitrary one.

We have the following key theorem on the asymptotic property of the update of $I^{(N)}(t)$ when $N \to \infty$. Note that the range of $I^{(N)}(t)$ is a discrete set. For any $\theta \in \mathcal{D}_K$, we take an approximation procedure. We suppose the vector θ has been used by the minor players (of the finite population) at time t in solving their limiting control problems and used in their optimal policy.

Theorem 2. *Fix any $\theta = (\theta_1, \ldots, \theta_K) \in \mathcal{D}_K$. Suppose the major player applies $\hat{\pi}_0$ and the N minor players apply $\hat{\pi}$, and at time t the state of the major player* *is x_0 and $I^{(N)}(t) = (s_1, \dots, s_K)$, where $(s_1, \dots, s_K) \to \theta$ as $N \to \infty$. Then given $(x_0, I^{(N)}(t), \hat{\pi})$, as $N \to \infty$,*

$$I^{(N)}(t+1) \to \Big(\sum_{l=1}^{K} \theta_l Q(1|l, \hat{\pi}(l, x_0, \theta)), \dots, \sum_{l=1}^{K} \theta_l Q(K|l, \hat{\pi}(l, x_0, \theta)) \Big) \tag{7}$$

with probability one.

Proof. By the assumption on $I^{(N)}(t)$, there are $s_k N$ minor players in state $k \in S$ at time t. In determining the distribution of $I^{(N)}(t+1)$, by symmetry of the minor players, we may assume without loss of generality that at time t minor players $\mathcal{A}_1, \dots, \mathcal{A}_{s_1 N}$ are in state 1, $\mathcal{A}_{s_1 N+1}, \dots, \mathcal{A}_{(s_1+s_2)N}$ are in state 2, etc. We check the contribution of $\mathcal{A}_1$ alone in generating different states in S. Due to the transition of $\mathcal{A}_1$, state $k \in S$ will appear with probability

$$Q(k|1, \hat{\pi}(1, x_0, \theta)).$$

We further obtain a probability vector $Q_1 := (Q(k|1, \hat{\pi}(1, x_0, \theta)))_{k=1}^{K}$ with its entries assigned on the set S indicating the probability that each state appears resulting from the transition of $\mathcal{A}_1$.

An important fact is that in the closed-loop system with $x_0(t) = x_0$, conditional independence holds for the transition from $x_i(t)$ to $x_i(t+1)$ for the N processes.

Thus, the distribution of $NI^{(N)}(t+1)$ given $(x_0, I^{(N)}(t), \hat{\pi})$ is obtained as the convolution of N independent distributions corresponding to all N minor players. And Q_1 is one of these N distributions. We have

$$E_{x_0, I^{(N)}(t), \hat{\pi}} I^{(N)}(t+1) = \Big(\sum_{l=1}^{K} s_l Q(1|l, \hat{\pi}(l, x_0, \theta)), \dots, \sum_{l=1}^{K} s_l Q(K|l, \hat{\pi}(l, x_0, \theta)) \Big), \tag{8}$$

where $E_{x_0, I^{(N)}(t), \hat{\pi}}$ denotes the conditional mean given $(x_0, I^{(N)}(t), \hat{\pi})$.

So by the law of large numbers $I^{(N)}(t+1) - E_{x_0, I^{(N)}(t), \hat{\pi}} I^{(N)}(t+1)$ converges to zero with probability one, as $N \to \infty$. We obtain (7). □

Based on the right hand side of (7), we introduce the $N \times N$ matrix

$$Q^*(x_0, \theta) = \begin{bmatrix} Q(1|1, \hat{\pi}(1, x_0, \theta)) & \dots & Q(N|1, \hat{\pi}(1, x_0, \theta)) \\ Q(1|2, \hat{\pi}(2, x_0, \theta)) & \dots & Q(N|2, \hat{\pi}(2, x_0, \theta)) \\ \vdots & \ddots & \vdots \\ Q(1|N, \hat{\pi}(N, x_0, \theta)) & \dots & Q(N|N, \hat{\pi}(N, x_0, \theta)) \end{bmatrix}. \tag{9}$$

Theorem 2 implies that within the infinite population limit if the random measure of the states of the minor players is $\theta(t)$ at time t, then $\theta(t+1)$ should be generated as

$$\theta(t+1) = \theta(t) Q^*(x_0(t), \theta(t)). \tag{10}$$

4.1 The Consistent Condition

The fundamental requirement of consistent mean field approximations is that the mean field initially assumed should be the same as what is replicated by the closed-loop system when the number of minor players tends to infinity. By comparing (4) with (10), this consistency requirement reduces to the following condition

$$\psi(x_0, \theta) = \theta Q^*(x_0, \theta), \tag{11}$$

where Q^* is given by (9). Recall that when we introduce the class Ψ for ψ, we have a continuity requirement. By imposing (11), we implicitly require a continuity property of Q^* with respect to the variable θ.

Combining the solutions to Problems (P0) and (P1) and the consistent requirement, we write the so-called mean field equation system

$$\theta(t+1) = \psi(x_0(t), \theta(t)), \tag{12}$$

$$v(x_0, \theta) = \min_{a_0 \in A_0} \Big\{ c_0(x_0, \theta, a_0) + \rho \sum_{k \in S_0} Q_0(k|x_0, a_0) v(k, \psi(x_0, \theta)) \Big\}, \tag{13}$$

$$\begin{aligned} w(x_i, x_0, \theta) = \min_{a \in A} \Big\{ & c(x_i, x_0, \theta, a_0) + \\ & \rho \sum_{j \in S, k \in S_0} Q(j|x_i, a) Q_0(k|x_0, \hat{\pi}_0) w(j, k, \psi(x_0, \theta)) \Big\}, \end{aligned} \tag{14}$$

$$\psi(x_0, \theta) = \theta Q^*(x_0, \theta). \tag{15}$$

In the above, we use x_i to denote the state of the generic minor player. Note that only a single generic minor player appears in this mean field equation system.

Definition 1. *We call* $(\hat{\pi}_0, \hat{\pi}, \psi(x_0, \theta))$ *a consistent solution to the mean field equation system* (12)-(15) *if* $\hat{\pi}_0$ *solves* (13) *and* $\hat{\pi}$ *solves* (14) *and if the constraint* (15) *is satisfied.* ♢

5 Decentralized Strategies and Performance

We consider a system of $N+1$ players. We specify randomized strategies with centralized information and decentralized information, respectively.

Centralized Information. Define the t-history h_t by (3). For any $j = 0, ..., N$, the admissible control set $\mathcal{U}_j$ of player $\mathcal{A}_j$ consists of control $(u_j(0), u_j(1), \ldots)$, where each $u_j(t)$ is a mixed strategy as a mapping from h_t to $\mathcal{D}_{L_0}$ if $j = 0$, and to $\mathcal{D}_L$ if $1 \le j \le N$.

Decentralized Information. For the major player, denote

$$h_t^{0,\text{dec}} = \Big(x_0(0), \theta(0), u_0(0), \ldots, x_0(t-1), \theta(t-1), u_0(t-1), x_0(t), \theta(t) \Big).$$

A decentralized strategy at time t is such that $u_0(t)$ is a randomized strategy depending on $h_t^{0,\text{dec}}$. For minor player $\mathcal{A}_i$, denote

$$h_t^{i,\text{dec}} = \Big(x_i(0), x_0(0), \theta(0), u_i(0), \ldots, \\ x_i(t-1), x_0(t-1), \theta(t-1), u_0(t-1), x_i(t), x_0(t), \theta(t)\Big).$$

A decentralized strategy at time t is such that $u_i(t)$ depends on $h_t^{i,\text{dec}}$.

For the mean field equation system, if a solution triple $(\hat{\pi}_0, \hat{\pi}, \psi)$ exists, we will obtain $\hat{\pi}_0$ and $\hat{\pi}$ as decentralized Markov strategies as a function of the current state $(x_0(t), \theta(t))$ and $(x_i(t), x_0(t), \theta(t))$, respectively.

Suppose all the players use their decentralized strategies $\hat{\pi}_0(x_0, \theta)$, $\hat{\pi}(x_i, x_0, \theta)$, $1 \leq i \leq N$, respectively. In the setup of mean field decision problems, a central issue is to examine the performance change for player $\mathcal{A}_j$ if it unilaterally changes to a policy in $\mathcal{U}_j$ by utilizing extra information.

For examining the performance, we have the following error estimate on the mean field approximation.

Theorem 3. *Suppose (i) $\theta(t)$ is generated by* (4)*, where θ_0 is given by (***A2***); (ii) $(\hat{\pi}_0, \hat{\pi}, \psi(x_0, \theta))$ is a consistent solution to the mean field equation system* (12)-(15)*. Then we have*

$$\lim_{N\to\infty} E|I^{(N)}(t) - \theta(t)| = 0$$

for each given t.

Proof. We use the technique introduced in the proof of Theorem 2. Fix any $\epsilon > 0$. We have

$$P(|I^{(N)}(0) - \theta_0| \geq \epsilon) \leq E|I^{(N)}(0) - \theta(0)|/\epsilon.$$

We take a sufficiently large N_0 such that for all $N \geq N_0$, we have

$$P(|I^{(N)}(0) - \theta_0| < \epsilon) > 1 - \epsilon. \tag{16}$$

Then following the method for (8), we may estimate $I^{(N)}(1)$. By the consistency condition (11), we further obtain

$$\lim_{N\to\infty} E|I^{(N)}(1) - \theta(1)| = 0.$$

Carrying out the estimates recursively, we obtain the desired result for each fixed t. □

For $j = 0, ..., N$, denote $u_{-j} = (u_0, u_1, ..., u_{j-1}, u_{j+1}, ..., u_N)$.

Definition 2. *A set of strategies $u_j \in \mathcal{U}_j$, $0 \leq j \leq N$, for the $N+1$ players is called an ϵ-Nash equilibrium with respect to the costs J_j, $0 \leq j \leq N$, where $\epsilon \geq 0$, if for any j, $0 \leq j \leq N$, we have $J_i(u_j, u_{-j}) \leq J_j(u_j', u_{-j}) + \epsilon$, when any alternative u_j' is applied by player $\mathcal{A}_j$.* ♢

Theorem 4. *Assume the conditions in Theorem 3 hold. Then the set of strategies $\hat{u}_j$, $0 \le j \le N$, for the $N+1$ players is an ϵ_N-Nash equilibrium, i.e., for $0 \le j \le N$,*

$$J_j(\hat{u}_j, \hat{u}_{-j}) - \epsilon_N \le \inf_{u_j} J_j(u_j, \hat{u}_{-j}) \le J_j(\hat{u}_j, \hat{u}_{-j}),$$

where $0 \le \epsilon_N \to 0$ as $N \to \infty$ and u_j is a centralized information based strategy.

Proof. The theorem may be proven by following the usual argument in our previous work [12,10]. First, by using Theorem 3, we may approximate $I^{(N)}(t)$ in the original game by $\theta(t)$. Then the optimization problems of the major player and any minor player are approximated by Problems (P0) and (P1), respectively. Finally, it is seen that each player can gain little if it deviates from the decentralized strategy determined from the mean field equation system. □

6 Conclusion Remarks and Future Work

This paper considers a class of Markov decision processes involving a major player and a large population of minor players. The players have independent dynamics for fixed actions and have mean field coupling in their costs according to the state distribution process of the minor players. We introduce a stochastic difference equation depending on the state of the major player to characterize the evolution of the minor players' state distribution process in the infinite population limit and solve local Markov decision problems. This approach provides decentralized stationary strategies and offers a low complexity solution.

This paper presents the main conceptual framework for decentralized decision making in the setting of Markov decision processes. The existence analysis and the associated computation of a solution to the mean field equation system is more challenging than in linear models. It is of interest to develop fixed point analysis to study the existence of solutions. Also, the development of iterative computation procedures for solutions is of practical interest.

References

1. Adlakha, S., Johari, R., Weintraub, G., Goldsmith, A.: Oblivious equilibrium for large-scale stochastic games with unbounded costs. In: Proc. IEEE CDC 2008, Cancun, Mexico, pp. 5531–5538 (December 2008)
2. Al-Najjar, N.I.: Aggregation and the law of large numbers in large economies. Games and Economic Behavior 47(1), 1–35 (2004)
3. Buckdahn, R., Cardaliaguet, P., Quincampoix, M.: Some recent aspects of differential game theory. Dynamic Games and Appl. 1(1), 74–114 (2011)
4. Dogbé, C.: Modeling crowd dynamics by the mean field limit approach. Math. Computer Modelling 52, 1506–1520 (2010)
5. Gast, N., Gaujal, B., Le Boudec, J.-Y.: Mean field for Markov decision processes: from discrete to continuous optimization (2010) (Preprint)

6. Galil, Z.: The nucleolus in games with major and minor players. Internat. J. Game Theory 3, 129–140 (1974)
7. Gomes, D.A., Mohr, J., Souza, R.R.: Discrete time, finite state space mean field games. J. Math. Pures Appl. 93, 308–328 (2010)
8. Haimanko, O.: Nonsymmetric values of nonatomic and mixed games. Math. Oper. Res. 25, 591–605 (2000)
9. Hart, S.: Values of mixed games. Internat. J. Game Theory 2, 69–86 (1973)
10. Huang, M.: Large-population LQG games involving a major player: the Nash certainty equivalence principle. SIAM J. Control Optim. 48(5), 3318–3353 (2010)
11. Huang, M., Caines, P.E., Malhamé, R.P.: Individual and mass behaviour in large population stochastic wireless power control problems: centralized and Nash equilibrium solutions. In: Proc. 42nd IEEE CDC, Maui, HI, pp. 98–103 (December 2003)
12. Huang, M., Caines, P.E., Malhamé, R.P.: 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 (2007)
13. Huang, M., Caines, P.E., Malhamé, R.P.: The NCE (mean field) principle with locality dependent cost interactions. IEEE Trans. Autom. Control 55(12), 2799–2805 (2010)
14. Huang, M., Caines, P.E., Malhamé, R.P.: Social optima in mean field LQG control: centralized and decentralized strategies. IEEE Trans. Autom. Control (in press, 2012)
15. Huang, M., Malhamé, R.P., Caines, P.E.: On a class of large-scale cost-coupled Markov games with applications to decentralized power control. In: Proc. 43rd IEEE CDC, Paradise Island, Bahamas, pp. 2830–2835 (December 2004)
16. Huang, M., Malhamé, R.P., Caines, P.E.: Nash equilibria for large-population linear stochastic systems of weakly coupled agents. In: Boukas, E.K., Malhamé, R.P. (eds.) Analysis, Control and Optimization of Complex Dynamic Systems, pp. 215–252. Springer, New York (2005)
17. Jovanovic, B., Rosenthal, R.W.: Anonymous sequential games. Journal of Mathematical Economics 17, 77–87 (1988)
18. Lasry, J.-M., Lions, P.-L.: Mean field games. Japan. J. Math. 2(1), 229–260 (2007)
19. Li, T., Zhang, J.-F.: Asymptotically optimal decentralized control for large population stochastic multiagent systems. IEEE Trans. Automat. Control 53(7), 1643–1660 (2008)
20. Ma, Z., Callaway, D., Hiskens, I.: Decentralized charging control for large populations of plug-in electric vehicles. IEEE Trans. Control Systems Technol. (to appear, 2012)
21. Nguyen, S.L., Huang, M.: Mean field LQG games with a major player: continuum-parameters for minor players. In: Proc. 50th IEEE CDC, Orlando, FL, pp. 1012–1017 (December 2011)
22. Nourian, M., Malhamé, R.P., Huang, M., Caines, P.E.: Mean field (NCE) formulation of estimation based leader-follower collective dyanmics. Internat. J. Robotics Automat. 26(1), 120–129 (2011)
23. Tembine, H., Le Boudec, J.-Y., El-Azouzi, R., Altman, E.: Mean field asymptotics of Markov decision evolutionary games and teams. In: Proc. International Conference on Game Theory for Networks, Istanbul, Turkey, pp. 140–150 (May 2009)
24. Tembine, H., Zhu, Q., Basar, T.: Risk-sensitive mean-field stochastic differential games. In: Proc. 18th IFAC World Congress, Milan, Italy (August 2011)

25. Wang, B.-C., Zhang, J.-F.: Distributed control of multi-agent systems with random parameters and a major agent (2012) (Preprint)
26. Weintraub, G.Y., Benkard, C.L., Van Roy, B.: Markov perfect industry dynamics with many firms. Econometrica 76(6), 1375–1411 (2008)
27. Yin, H., Mehta, P.G., Meyn, S.P., Shanbhag, U.V.: Synchronization of coupled oscillators is a game. IEEE Trans. Autom. Control 57(4), 920–935 (2012)
28. Yong, J., Zhou, X.Y.: Stochastic Controls: Hamiltonian Systems and HJB Equations. Springer, New York (1999)