DISCRETE TIME MEAN FIELD GAMES: THE SHORT-STAGE LIMIT

JUAN PABLO MALDONADO LÓPEZ\textsuperscript{1,2,3}

\textsuperscript{1} Sorbonne Universités, UPMC Univ Paris 06, UMR 7586
IMJ-PRG, case 247, 4 place Jussieu, F-75005, Paris, France
\textsuperscript{2} CNRS, UMR 7586, F-75005, Paris, France
\textsuperscript{3} Univ Paris-Diderot, Sorbonne Paris Cité, UMR 7586
F-75013, Paris, France

(Communicated by Alberto A. Pinto)

Abstract. In this note we provide a model for discrete time mean field games. Our main contributions are an explicit approximation in the discounted case and an approximation result for a mean field game with short-stage duration.

1. Introduction. In this paper we study a model for a discrete time, discrete state space, finitely repeated stochastic games where the transition and the payoff of the players depend on the position in space and the actions of the adversaries, but not on their identities. We assume all the players have the same dynamics and the same payoff, thus, for each player, we can consider the influence of the adversaries only through the empirical distribution of the state-action pair.

Mean field games have been introduced independently by Huang, Caines and Malhamé [11] and by Lasry and Lions [13, 14, 15] and have received considerable attention in the literature. The aim of the mean field games paradigm is to describe situations with many interacting agents whose preferences and dynamics depend on the aggregate effect of the other agents. Mean field game models are composed by two parts: a backward component, where each agent considers the aggregate behavior as an external parameter and computes myopically his own optimal behavior and a forward component, which is the evolution of the initial distribution in the state space under a common strategy. Mean field games have found applications in many different areas, we refer to [9] and the references therein for examples.

Most of the models studied in the literature so far are in continuous time, while the discrete time case has received less attention. The discrete time case has not only independent interest, but also allows to model more general transitions, contrary to the assumption usually made in continuous time mean field games that the noise in the dynamics of the players is independent of their actions. In discrete time, we can also allow the players to choose their actions randomly, as in classical game...
theory. However, for some applications it might be relevant to consider frequent interactions between the players. This motivates the study of a limit model as the duration of each stage tends to zero, which we pursue in Section 3.

The main novelty of our work with respect to the previous work on discrete time mean field games is the short-stage version. Short-stage games have been recently introduced in [16]. The aim of this theory is to study games where players are allowed to interact more frequently. Incorporating this machinery, we obtain a limit object that provides an approximate Nash equilibrium for games with sufficiently many players and sufficiently frequent interactions.

In [8], a discrete time, finite state mean field game with a continuum of players is studied. The authors study a finite horizon game and prove the exponential convergence of the finite-horizon mean field equilibrium to a stationary solution. There are two significant differences with our work. First, we consider a fixed time horizon. Second, we are interested in constructing approximate equilibria for games with large numbers of players, while in [8] a continuum of players is considered. An anonymous referee pointed us to the recent paper [6] where a model for linear quadratic mean field games in discrete time is studied.

Our work is closer to [2], where a similar notion is studied for an infinite horizon, discounted stochastic game. While we restrict our framework to a finite state space (in [2] an unbounded state space is considered), we provide explicit approximation estimates in terms of the basic parameters of the game. Our estimate is of the same order as the one in [12] in continuous time.

Let us remark also that we study a discrete-time game by itself, and not the discretization of a continuous-time mean field game for numerical solution purposes. Numerical methods have been initially developed in [1]. A semi-Lagrangian scheme has been proposed in [4] for the deterministic, finite-horizon case. The full discretization has been studied in [5].

The paper is organized as follows, In Section 2, we describe the model and some results on the existence of mean field equilibrium, as well as the approximation results with explicit convergence rates. In Section 3 we introduce a short-stage version of the discounted stochastic mean field game. In the Appendix, we prove an approximation lemma which allows to prove the results we present in Section 2.3.

2. The discrete time model.

2.1. Mean field equilibrium. Let \( \Omega \) and \( A \) denote respectively the state and action sets. We assume both to be finite. Let \( \mathcal{Z} := \Delta(\Omega \times A) \), where, for a finite set \( \mathcal{S} \), \( \Delta(\mathcal{S}) \) denotes the set of probability distributions over \( \mathcal{S} \). Consider a bounded payoff function \( g : \Omega \times A \times \mathcal{Z} \to [0,1] \) and a transition function \( Q : \Omega \times A \times \mathcal{Z} \to \Delta(\Omega) \). Let \( n \) be a fixed positive integer. Let us define a family of auxiliary one-player games, parameterized by a vector \( z = (z_1, z_2, \ldots, z_n) \in \mathcal{Z}^n \). The one-player dynamic programming problem \( \Gamma_n \) is defined as follows: at stage \( k \), the player observes the state \( \omega_k \in \Omega \) and chooses the action \( a_k \in A \) from which he receives a payoff \( g(\omega_k, a_k, z_k) \) and the new state is chosen according to the law \( Q(\omega_k, a_k, z_k) \).

A pure behavior strategy (resp. mixed behavior strategy) is a sequence of functions \( \sigma = (\sigma_1, \ldots, \sigma_n) \) where \( \sigma_k : \mathcal{H}_k^1 \to A \) (resp. \( \sigma_k : \mathcal{H}_k^1 \to \Delta(A) \)). Here, \( \mathcal{H}_k^1 = (\Omega \times A)^{k-1} \cup \Omega \times \Delta(\Omega) \) denotes the set of histories up to time \( k \), for \( k = 1, \ldots, n \). Let \( \Sigma_n \) denote the set of pure strategies. The player knows \( z \) and observes the payoff. We introduce the value function \( V_n : \Omega \times \Delta(\Omega) \to [0,1] \) for the
game $\Gamma^n_z$:

$$V_n(\omega, z) := \max_{\sigma \in \Sigma_n} E^Q \left( \sum_{k=1}^{n} g(\omega_k, a_k, z_k) \mid \omega_1 = \omega \right)$$

One can also consider an infinitely repeated game $\Gamma^n_\lambda$ with parameters $z \in Z$ and $\lambda \in (0, 1]$, played as before but where the payoff is evaluated by

$$V_\lambda(\omega, z) := \max_{\sigma \in \Sigma} E^Q \left( \sum_{k=1}^{\infty} (1 - \lambda)^{k-1} g(\omega_k, a_k, z) \mid \omega_1 = \omega \right).$$

From the familiar arguments\(^1\), one can prove that the value functions satisfy the following recursive formulae (dynamic programming principle):

$$V_n(\omega, z) = \max_{a \in A} \left\{ g(\omega, a, z_1) + (n - 1) \sum_{\omega' \in \Omega} V_{n-1}(\omega', z_1)Q(\omega, a, z_1)(\omega') \right\} \quad (1)$$

and

$$V_\lambda(\omega, z) = \max_{a \in A} \left\{ g(\omega, a, z) + (1 - \lambda) \sum_{\omega' \in \Omega} V_\lambda(\omega', z)Q(\omega, a, z)(\omega') \right\}. \quad (2)$$

In (1), if $z = (z_1, z_2, \ldots, z_n)$, then $z^+$ denotes the vector $(z_2, z_3, \ldots, z_n)$.

The dynamic programming principle (1) also tells us that the player can restrict his attention to the set of Markovian strategies $\Sigma^M_n \subset \Sigma_n$, which consists of all the functions $\sigma = (\sigma_1, \ldots, \sigma_n)$ such that $\sigma_k : \Omega \to \Delta(A)$.

Let $m_1 \in \Delta(\Omega)$ given and let $Z^n_1 := \{z \in Z^n : z_1|_{\Omega} = m_1 \}$. For the rest of the paper, $z_k|_{\Omega}$ denotes the marginal distribution of $z_k \in Z$ on the set $\Omega$. Define $\Psi_n : Z^n_1 \to \Sigma^M_n$ as the set valued map that associates to every $z \in Z^n_1$ the set of optimal Markovian strategies in $\Gamma^n_z$.

Let $\Phi_n : \Sigma^M_n \to Z^n_1$ defined by $\sigma \mapsto z^\sigma$ where the sequence $z^\sigma$ is recursively defined by setting $z^\sigma_1(\omega, a) := m_1(\omega) \cdot \sigma_1[\omega](a)$ and

$$z^\sigma_{k+1}(\omega, a) := \sum_{(\omega', a') \in \Omega \times A} Q(\omega', a', z^\sigma_k)(\omega) \cdot z^\sigma_k(\omega', a') \cdot \sigma_{k+1}[\omega](a). \quad (3)$$

We are interested in the fixed points of $\Phi_n \circ \Psi_n$. In order to apply fixed point theorems, one needs to ensure certain continuity and convexity properties, which will hold under the following assumptions.

**Assumption 1.** (Lipschitz continuity) There exists positive real numbers $L_Q, L_g$ such that for all $(\omega, a, y, z) \in \Omega \times A \times Z \times Z$,

$$\|Q(\omega, a, y) - Q(\omega, a, z)\|_1 \leq L_Q \|y - z\|_1$$

and

$$\|g(\omega, a, y) - g(\omega, a, z)\|_\infty \leq L_g \|y - z\|_\infty.$$

For the existence results in these sections, this continuity assumption can be relaxed, however, for our main approximation results we need Lipschitz continuity. One way to ensure convexity properties is the following:

**Assumption 2.** (Independent transitions) For all $(\omega, a, y, z) \in \Omega \times A \times Z \times Z$,

$$Q(\omega, a, y) = Q(\omega, a, z) =: Q(\omega, a).$$

\(^1\)This is a one-player Markov decision process, which are well understood. See for instance [10].
In order to avoid this assumption, one needs to impose a different assumption so that a convexity property can still be preserved.

**Assumption 3. (Uniqueness of the maximizer)** The right hand side of equations (1), (2) admits a unique maximizer, i.e. there exists a unique pure Markovian optimal strategy.

It is possible to provide conditions on the basic model data that ensure that Assumption 3 holds, see for example Assumption 2 and 3 in [2] or Assumptions 1-3 in [8]. As uniqueness of the maximizer might hold under other circumstances, we prefer not to write down explicit conditions.

A straightforward application of Brouwer’s fixed point theorem yields the following result.

**Proposition 1.** If $Q$ satisfies Assumptions 1 and 3, then $\Phi_n \circ \Psi_n$ has a fixed point.

**Definition 2.1.** A mean field equilibrium with initial mass $m_1$ for the $n$–stage game is a pair $(\sigma, z) \in \Sigma^M_n \times \mathcal{Z}^n_1$ such that $z$ is a fixed point of $\Phi_n \circ \Psi_n$ and $\sigma = \Psi_n(z)$.

For the discounted case, from (2) one obtains that there exist optimal stationary strategies, i.e. functions of the form $\sigma : \Omega \to \Delta(A)$. Let $\Sigma$ denote the set of stationary strategies. Define $\Psi_\lambda : \mathcal{Z} \to \mathcal{Z} \times \Sigma$ as the set valued map that associates to every $z \in \mathcal{Z}$ the pair $(z, S^1_\lambda(z))$, where $S^1_\lambda$ is the set of optimal stationary strategies in $\Gamma^1_\lambda$.

We will make the following ergodicity assumption throughout the paper.

**Assumption 4. (Ergodicity)** For all $y \in \mathcal{Z}$, $\sigma \in \Sigma$, and $(\omega', a'), (\omega, a) \in \Omega \times A$, the Markov chain $(Z_k)_{k \in \mathbb{N}}$ over $\Omega \times A$ with transition probability

$$Q(z_{k+1} = (\omega, a) \mid Z_k = (\omega', a')) = Q(\omega', a') \cdot \sigma(\omega) \cdot (y, \sigma(a)) =: Q[y, \sigma](\omega', a'),$$

(4) has a stationary distribution $z = z(y, \sigma)$. In other words, for $y, \sigma$ given, there exists $z$ such that the following holds

$$Q[y, \sigma] z = z.$$  

(5)

Let $\Phi_\lambda : \mathcal{Z} \times \Sigma \to \mathcal{Z}$ be defined by $(y, \sigma) \mapsto z(y, \sigma)$ where $z(y, \sigma) \in \mathcal{Z}$ is defined by

$$z(y, \sigma)(\omega, a) := \sum_{(\omega', a') \in \Omega \times A} Q(\omega', a') \cdot \sigma(\omega) \cdot z(y, \sigma)(\omega', a').$$

Under Assumption 4, $\Phi_\lambda$ is well defined. One obtains analogous results to Proposition 1.

**Proposition 2.** Under Assumptions 1 and 4, $\Phi_\lambda \circ \Psi_n$ has a fixed point.

**Proof.** The upper semicontinuity follows easily from the assumptions. For the convexity, let $z \in \mathcal{Z}$ and $z_1, z_2 \in \Phi_\lambda \circ \Psi_n(z)$. Let $z_\theta := \theta z_1 + (1 - \theta) z_2$. Consider two stationary strategies $\sigma_1, \sigma_2 \in \Psi_n(z)$ such that $\Phi_\lambda(\sigma_1) = z_1, \Phi_\lambda(\sigma_2) = z_2$. Let $\sigma_\theta$ be the following strategy:

$$\sigma_\theta[\omega](a) := \frac{\theta \cdot z_1(\omega, a) \cdot \sigma_1[\omega](a) + (1 - \theta) \cdot z_2(\omega, a) \cdot \sigma_2[\omega](a)}{z_\theta(\omega, a)}.$$

Observe that $\sigma_\theta$ is optimal for $\Gamma^1_\lambda$ from the optimality of $\sigma_1, \sigma_2$. We make the convention that when $z_\theta(\omega, a) = 0$, $\sigma_\theta[\omega](a) = 0$. Note also that

$$z_\theta(\omega, a) = 0 \iff z_1(\omega, a), z_2(\omega, a) = 0.$$
We have, up to excluding the above trivial cases,
\[
\begin{align*}
    z_0(\omega, a) &= \theta z_1(\omega, a) + (1 - \theta) z_2(\omega, a) \\
    &= \theta \sum_{\omega', a'} Q(\omega', a', z)(\omega) \cdot \sigma_1[\omega](a) \cdot z_1(\omega, a) + \\
    &+ (1 - \theta) \sum_{\omega', a'} Q(\omega', a', z)(\omega) \cdot \sigma_2[\omega](a) \cdot z_2(\omega, a) \\
    &= \sum_{\omega', a'} Q(\omega', a', z)(\omega) \cdot \sigma_\theta[\omega](a) \cdot z_0(\omega, a)
\end{align*}
\]

Hence, \( z_0 \in \Phi_\lambda \circ \Psi_\lambda(z) \).

In the case of pure strategies, one has the following:

**Proposition 3.** Under Assumptions 1, 3 and 4, \( \Phi_\lambda \circ \Psi_\lambda \) has a fixed point.

The proof of this proposition is a straightforward application of Brouwer’s fixed point theorem.

**Definition 2.2.** A stationary mean field equilibrium is a pair \((\sigma, z) \in \Sigma \times Z\) such that \( z \) is a fixed point of \( \Phi_\lambda \circ \Psi_\lambda \) and \( \sigma \) is the strategy associated to \( \Psi_\lambda(z) \).

### 2.2. The \( N \)-player game.

We consider a \( n \)-stage stochastic game \( \Gamma_{n,N}[m_1] \) with \( N + 1 \) identical players, i.e. with common state space \( \Omega \), action set \( A \), stage payoff \( g : \Omega \times A \times Z \rightarrow [0, 1] \) and transition function \( Q : \Omega \times A \times Z \rightarrow \Delta(\Omega) \) played as follows: at stage \( k \), for \( k = 1, \ldots, n \), each player \( i \) observes his own state \( \omega_k^i \) and the state of each of the adversaries and chooses his action \( a_k^i \). The initial state of each player is sampled i.i.d using the lottery \( m_1 \). The actions of the players are chosen simultaneously and independently. After the actions were chosen, each player has a state-action pair \( z_{k,N} := (\omega_{k,N}^i, a_{k,N}^i) \). The payoff for player \( i \) is \( g(\omega_{k,N}^i, a_{k,N}^i, z_{k,N}) \) where

\[
    z_{k,N}(\omega, a) := \frac{1}{N} \sum_{j=1}^N 1_{\{z_{k,N}^j = (\omega, a)\}}
\]

denotes the empirical distribution of the state-action pairs of the players after the play at stage \( k \). The new state for player \( i \), \( \omega_{k+1,N}^i \), is chosen according to the law \( Q(\omega_{k,N}^i, a_{k,N}^i, z_{k,N}) \) and the situation is repeated.

A **behavioral strategy** for player \( i \) is a vector \( \pi^i = (\pi^i_k)_{k=1}^n \) where \( \pi^i_k : \mathcal{H}_k \rightarrow \Delta(A) \) and \( \mathcal{H}_k = (\Omega \times A \times Z)^{k-1} \cup (\Omega \times \Omega^{N-1}) \) is the set of all possible histories up to stage \( k \). Denote by \( \Sigma_{n,N} \) the set of behavioral strategies for each player and note that \( \Sigma_M \subset \Sigma_{n,N} \). A **strategy profile** is a vector \( \pi = (\pi^i)_{i=1,\ldots,N} \), where \( \pi^i \) is a behavioral strategy of player \( i \). The average payoff of player \( i \), when using the strategy \( \pi^i \) and when his adversaries use the strategy profile \( \pi^{-i} \in (\Sigma_{n,N})^{N-1} \) is

\[
    J_N(\omega, m_1, \pi^i, \pi^{-i}) := E\left[ \sum_{k=1}^n g(\omega_{k,N}^i, a_{k,N}^i, z_{k,N}) \mid \omega_{1,N}^i = \omega \right].
\]

One can also consider a game \( \Gamma_{\lambda,N}[m_1] \) with infinite horizon and payoff

\[
    J_\lambda(\omega, m_1, \pi^i, \pi^{-i}) := E\left[ \sum_{k=1}^\infty (1 - \lambda)^k g(\omega_{k,N}^i, a_{k,N}^i, z_{k,N}) \mid \omega_{1,N}^i = \omega \right]. \quad (6)
\]

where \( \lambda \in (0, 1) \).
Definition 2.3. An $\epsilon$-Nash equilibrium for the average payoff, where $\epsilon > 0$, is a strategy profile $(\pi^i)_{i=1,\ldots,N}$ such that, for all player $i$ and all behavioral strategy $\tau^i$,

\[ J_N^i(\omega, m_1, \tau, \pi^{-i}) - \epsilon \leq J_N^i(\omega, m_1, \pi^i, \pi^{-i}). \]

Analogously, an $\epsilon$-Nash equilibrium for the $\lambda$-discounted payoff is a strategy profile $(\pi^i)_{i=1,\ldots,N}$ such that, for all player $i$ and all behavioral strategy $\tau^i$,

\[ J^i_\lambda(\omega, m_1, \tau, \pi^{-i}) - \epsilon \leq J^i_\lambda(\omega, m_1, \pi^i, \pi^{-i}). \]

2.3. Approximation results. We are ready to state our first main result in the finite horizon case, which is an easy consequence of the approximation lemma in Section 4. This result is an estimate of the maximal deviation between the trajectories followed by a player if the observed aggregate state-action of his adversaries affects his own transition functions and an independent game in which he plays the same action, but the transition function takes as argument the corresponding aggregate state-action of the mean field equilibrium.

Throughout this Section Assumption 1 holds.

Proposition 4. Let $(\sigma, z)$ be a mean field equilibrium. Assume player $i$’s adversaries play the mean field equilibrium strategy, whereas player $i$’s action at time $s$, $a^i_s$, is arbitrary.

- Let $Z^i_{s+1}$ denote the state-action pair of player $i$ at time $s+1$ when his transitions are influenced by the mean field term $z_s$, i.e. $Z^i_{s+1} \sim Q(\omega^i_s, a^i_s, z_s)$.

- Let $Z^i_{s+1,N}$ denote the state-action pair of player $i$ at time $s+1$ when the empirical state-action pair of the adversaries influences his transitions, i.e. $Z^i_{s+1,N} \sim Q(\omega^i_s, a^i_s, z_{s,N})$.

Then we have:

\[ \mathbb{E} \left( \max_{s \leq n} \| Z^i_s - Z^i_{s,N} \|_{\infty} \right) \leq \frac{L_Q n |\Omega \times A| \exp(n(\|Q\|_{\infty} + L_Q))}{\sqrt{N}}. \]

Proof. Let $S := \Omega \times A$ in Lemma 4.1. In this lemma, we take $P(\cdot) := Q(a^i_s, \cdot)$. \qed

Once the difference of the trajectories of player $i$ in the $N$ player game and in the one-player game where the state-action term enters as a parameter is bounded, we obtain the following result:

Theorem 2.4. Let $\epsilon > 0$ be given. In a finite horizon, $N$ player game, there exists $N_0$ such that for $N > N_0$ the mean field equilibrium is an $\epsilon$-equilibrium, where the approximation error is given by:

\[ \frac{L_Q(L_g + \|g\|_{\infty}) n |\Omega \times A| \exp(n(\|Q\|_{\infty} + L_Q))}{\sqrt{N}}. \]

Proof. Consider a game with $N$ players and let us focus on the payoff function of player $i$. Let $(\sigma, z)$ denote a mean field equilibrium for a given initial distribution
and let $\omega_1 = \omega$. We have that, for all behavior strategy $\tau$:

$$D := J_N^i(\omega, m_1, \tau, \sigma^{-i}) - J_N^i(\omega, m_1, \sigma, \sigma^{-i})$$

$$= J_N^i(\omega, m_1, \tau, \sigma^{-i}) - \mathbb{E}^Q \sum_{k=1}^n g(\omega_k, \sigma(\omega_k), z_k) +$$

$$+ \mathbb{E}^Q \sum_{k=1}^n g(\omega_k, \sigma(\omega_k), z_k) - J_N^i(\omega, m_1, \sigma, \sigma^{-i})$$

$$\leq \left[ J_N^i(\omega, m_1, \tau, \sigma^{-i}) - \mathbb{E}^Q \sum_{k=1}^n g(\omega_k, \tau(\omega_k), z_k) \right] +$$

$$+ \left[ \mathbb{E}^Q \sum_{k=1}^n g(\omega_k, \sigma(\omega_k), z_k) - J_N^i(\omega, m_1, \sigma, \sigma^{-i}) \right]$$

The above inequality comes from the optimality of $\sigma$. The result now follows applying Lemma 4.2 to each of the terms in brackets.

**Remark 1.** Note that the discounted case can be reduced to the finite horizon case: indeed, it suffices to find $K \in \mathbb{N}$ large enough so that

$$\lambda(1 - \lambda)^K \|g\|_\infty < \epsilon/2$$

and consider the $N_0$ for $\epsilon/2$ in Theorem 2.4. However, this may not be appropriate when $\lambda$ is small because the number of stages will be too large. For small $\lambda$, it makes more sense to consider the construction we proposed. This will be the case for instance in Section 3.

**Remark 2.** Our bound suggests that the number of players should be much larger than the length of the game. This seems intuitive, since one would expect that if there are not enough players and they play for many stages, it could happen that the empirical distribution at early stages of the game is too far from the predicted distribution and this error would be propagated.

**Remark 3.** Our result is on the spirit of [12]: construct a limit object that induces $\epsilon$-Nash equilibria in games with large players. Our limit object corresponds heuristically to a game with a continuum of players. The complementary approach of studying the limits of a sequence of Nash equilibria of games with finitely many players has been explored in some cases, see for instance [3, 13] but the general case remains open. We will illustrate this remark in Example 2.

We conclude this Section with two illustrative examples:

**Example 1.** As an application let us consider the following example, adapted from *learning by doing* [7]. Consider the industry of online hotel booking, where many firms offer accommodation. In this case, the state space $\Omega$ is the reputation of the firm, the action set $A$ is the capital to be invested. Each firm aims to improve their reputation by making investments and/or adjusting their offers. The reputation changes according to the transition $Q$. Note that in this context makes sense to consider independent transitions, as in Assumption 2. The firms interact with each other through the payoff function $g$, which represents their market share. For instance, if all the firms have similar reputation, customers might be indifferent and the utilities will be shared evenly, whereas if there are few firms with outstanding reputation, they may have higher revenues.
Example 2. Consider a game with \( N \) players, where each player chooses whether to drive on the left or right side of the street. Assume that the payoff for driving on the same side as everyone else is 1 and zero otherwise. Observe that 'everyone drives left' and 'everyone drives right' are Nash equilibria. However, this game has more equilibria which are sensible to the number of players present in the game, for instance 'everyone drives right if \( N \) is even'. In this case it does not make sense to consider limits of Nash equilibria, it is rather desirable to have equilibria that are independent of the number of players.

3. A mean field game with frequent actions. The aim of this section is to study a model for mean field games where players are allowed to play more frequently.

3.1. The one-player game. Let \( \delta > 0 \) and \( z \in \mathbb{Z} \). In the spirit of [16], we consider a family of discrete time repeated games parametrized by \( \delta \) as follows: let \( \mu : \Omega \times \Omega \times A \times \mathbb{Z} \to \mathbb{R}^+ \) bounded and such that, for all \( (\omega, a) \in \Omega \times A, \)

\[
\mu(\omega, \omega, a, z) = - \sum_{\omega' \neq \omega} \mu(\omega, \omega', a, z)
\]

That is, \( \mu(\omega, \omega', a, z) \) defines the escape velocity from \( \omega \) to \( \omega' \).

For \( \delta \) small enough, the function

\[
Q_\delta(\omega, a, z)(\omega') := \delta \mu(\omega', \omega, a, z), \quad \text{for } \omega' \neq \omega, \quad Q_\delta(\omega, a, z)(\omega) = 1 + \delta \mu(\omega, \omega, a, z)
\]

(7)

defines transition probabilities. Introduce the notation \( g_\delta := \delta g \) and let \( \Gamma^\lambda_\delta \) denote the one player game with stage payoff \( g_\delta \) and transition function \( Q_\delta \). For a fixed \( \delta \), this is exactly a discounted one-player game as introduced in Section 2 to define a stationary mean field equilibrium. The stationary mean field equilibrium defined through these games enjoys, for a fixed \( \delta \), identical approximation properties as in Section 2.3 in terms of the number of players. Our goal in this section is to provide a limit object that provides simultaneously good approximations for a large enough population of players and for a short enough time between plays.

Let \( 0 < \rho < 1 \). Informally, our aim is to approximate a mean field equilibrium for the stochastic game in continuous time with payoff \( \int_{0}^{\infty} e^{-\rho t} g(\omega_t, a_t, z)dt \) via mean field equilibria of the discrete time games \( \Gamma^\lambda_\delta \). The discount factor needs to be adjusted so that the accumulated payoff at the fraction \( t \) of the continuous time game is indeed the limit of the accumulated payoffs during the first \( \lfloor \frac{t}{\delta} \rfloor \) stages of the discrete time game. This is achieved by taking the discount factor \( \lambda = \rho \delta \). Denote by \( V_\delta^\lambda \) the value function of the game \( \Gamma^\lambda_\delta \). Taking \( \lambda = \rho \delta \) in (2) and dividing by \( \delta \) yields

\[
\rho V_{\rho \delta}^\delta(\omega, z) = \max_{a \in A} \left\{ g(\omega, a, z) + (1 - \rho \delta) \sum_{\omega' \in \Omega} \mu(\omega, \omega', a, z) V_{\rho \delta}^\delta(\omega', z) \right\}
\]

(8)

which suggests that if \( f \) is an accumulation point of \( \left\{ V_{\rho \delta}^\delta \right\}_{\delta > 0} \), then it should satisfy

\[
\rho f(\omega, z) = \max_{a \in A} \left\{ g(\omega, a, z) + \sum_{\omega' \in \Omega} \mu(\omega, \omega', a, z) f(\omega', z) \right\}.
\]

(9)

Let us provide a proof of this result. The proof is inspired from the proof of Theorem 1 in [16].
Proposition 5. The equation (9) has a unique solution, denoted $V_\rho$. Moreover, $V^\delta_{\rho \delta} \to V_\rho$ uniformly as $\delta \to 0$.

Proof. Let $C(\Omega \times \mathcal{Z})$ denote the set of continuous real-valued functions over $\Omega \times \mathcal{Z}$. Let $T : C(\Omega \times \mathcal{Z}) \to C(\Omega \times \mathcal{Z})$ be the following operator:

$$T f(\omega, z) = \frac{1}{\left\| \mu \right\|_\infty + \rho} \max_{a \in A} \left( g(\omega, a, z) + \sum_{\omega' \in \Omega} \mu(\omega, \omega', a, z) f(\omega', z) + \left\| \mu \right\|_\infty f(\omega, z) \right).$$

Note that $T(f + c 1) = Tf + c \frac{\left\| \mu \right\|_\infty}{\left\| \mu \right\|_\infty + \rho}$, and that $T$ is monotone, i.e. $Tf \geq Tg$ whenever $f \geq g$. Here, 1 denotes the constant function 1. Consequently, $T$ is a $\frac{\left\| \mu \right\|_\infty}{\left\| \mu \right\|_\infty + \rho}$-contraction and has a unique fixed point $V_\rho$. Besides, note that $Tv = v$ if and only if $v$ is a solution to the following implicit equation:

$$\rho v(\omega, z) = \max_{a \in A} \left( g(\omega, a, z) + \sum_{\omega' \in \Omega} \mu(\omega, \omega', a, z) v(\omega', z) \right)$$

(10)

Denote with $V_\rho$ the unique solution of (10) and let $\sigma^\delta_\rho$ be an optimal stationary strategy in (10). Consider the stochastic one-player game $\Gamma^\delta_{\rho, \omega}$ with initial state $\omega$ and let $Y_m := E \left[ \delta g(\omega_m, \sigma^\delta_\rho(\omega_m), z) + (1 - \rho \delta) V_\rho(\omega_{m+1}, z) \right]$. Let us set $\lambda_\delta := 1 - \rho \delta$ temporarily to alleviate the notation. We have that:

$$Y_m = \delta g(\omega_m, \sigma^\delta_\rho(\omega_m), z) + \lambda_\delta \sum_{\omega' \in \Omega} Q_\delta(\omega_m, a, z)(\omega') V_\rho(\omega', z)$$

$$= \delta g(\omega_m, \sigma^\delta_\rho(\omega_m), z) + \lambda_\delta \sum_{\omega' \in \Omega} \delta \mu(\omega_m, \omega', a, z) V_\rho(\omega', m_1) + \lambda_\delta V_\rho(\omega_m, z)$$

$$\geq V_\rho(\omega_m, z) - \rho \delta^2 \sum_{\omega' \in \Omega} \mu(\omega, \omega', a, z) V_\rho(\omega', m_1)$$

$$\geq V_\rho(\omega_m, z) - 2\delta^2 \left\| \mu \right\|_\infty \left\| g \right\|_\infty.$$

The last inequality follows directly from (10) and using the fact that $V_\rho < \left\| g \right\|_\infty$ and $\rho < 1$.

Hence,

$$E^{Q_\delta} \lambda^m_\delta \delta g(\omega_m, \sigma^\rho(\omega_m), z) \geq E^{Q_\delta} (1 - \rho \delta)^m V_\rho(\omega_m, z) - E^{Q_\delta} \lambda^{m+1}_\delta V_\rho(\omega_{m+1}, z) - \lambda^m_\delta 2\delta^2 \left\| \mu \right\|_\infty \left\| g \right\|_\infty.$$

Summing over $m$ (since we consider the payoff as in (6)) we obtain that $V^\delta_{\rho \delta}(\omega, m_1) \geq V_\rho(\omega, z) - \frac{2\left\| \mu \right\|_\infty \left\| g \right\|_\infty}{\rho} \delta$, and $\sigma^\delta_\rho$ is asymptotically optimal in $\Gamma^\delta_{\rho, \omega}$, as $\delta \to 0$. \hfill \Box

3.2. The evolution of the state-action pair. Let $y \in \mathcal{Z}$ and $\sigma_\delta$ an optimal stationary strategy in the game $\Gamma^\delta_{y, \omega}$. Let $Q_\delta$ be a transition as defined before and $Q_\delta$ its associated transition matrix, as defined in (4). Denote by $z_\delta$ the invariant state-action pair under the transition $Q_\delta$, as introduced in (7), and the strategy $\sigma_\delta$. Hence, by (5) in Assumption 4 it should satisfy:

$$Q_\delta z_\delta = z_\delta.$$

Since $Q_\delta = I + \delta L[\sigma_\delta, y]$, where $L[\sigma_\delta, y](\omega, a), (\omega', a') := \mu(\omega, \omega', \sigma_\delta(\omega), y) \cdot \sigma_\delta(\omega') \cdot (a')$, $L[\sigma_\delta, y]z_\delta = 0$. 

Heuristically, the limit state-action pair as the stage duration goes to zero, corresponding to the limit strategy $\sigma^y_\rho$ should solve:

$$L[\sigma^y_\rho, y]z = 0.$$ 

**Assumption 5.** For every $y \in \mathcal{Z}$ and $\sigma \in \Sigma$, optimal stationary strategy given by (9) with $z = y$, the equation

$$L[\sigma, y]z = 0. \quad (11)$$

has a solution, denoted $z[\sigma, y]$. 

Let $\Psi^\rho : \mathcal{Z} \rightarrow \mathcal{Z} \times \Sigma$ be defined by $y \mapsto (y, \sigma^y_\rho)$ and $\Phi^\rho : \mathcal{Z} \times \Sigma \rightarrow \mathcal{Z}$ be defined by $(y, \sigma) \mapsto z[\sigma, y]$. Under this assumption, let us introduce the following definition:

**Definition 3.1.** A limit stationary mean field equilibrium is a pair $(\sigma, z) \in \Sigma \times \mathcal{Z}$ such that $z$ is a fixed point of $\Phi^\rho \circ \Psi^\rho$ and $\sigma$ is the strategy associated to $\Psi^\rho(z)$.

One can prove existence of the limit stationary mean field equilibrium under the appropriate uniqueness assumption of the optimal action, analogous to Assumption 3.

**Assumption 6.** The right hand side of (9) has a unique maximizer.

By a straightforward application of Brouwer’s fixed point theorem, one has:

**Proposition 6.** Under Assumptions 5 and 6, the operator $T^\rho := \Phi^\rho \circ \Psi^\rho$ has a fixed point.

### 3.3. The approximation of the $N$ player game with frequent actions.

Let us state now our second approximation result, namely the approximation of the game with sufficiently short stage and sufficiently many players.

**Theorem 3.2.** For every $\epsilon > 0$ there exist $\delta_0 > 0$ and $N_0 \in \mathbb{N}$ such that for all $\delta < \delta_0$ and $N > N_0$ the strategy provided by the limit stationary mean field equilibrium is a $2\epsilon$-Nash equilibrium of the discounted mean field game with discount factor $\lambda = \rho \delta$ and $N$ players.

**Proof.** Let $\epsilon > 0$ fixed and consider a limit stationary mean field equilibrium $(\sigma, z)$. Observe from the proof of Proposition 5 that one can choose $\delta_0$ small enough so that $\sigma$ is $\epsilon$–optimal for the one-player discounted game with discount $\rho \delta$, for all $\delta < \delta_0$. Let $K_0$ such that

$$\rho \delta_0 (1 - \rho \delta_0)^K \|g\|_\infty < \epsilon/2.$$ 

Finally, let us take the $N_0$ given by Theorem 2.4 for the game of $n = K_0$ stages and error $\epsilon/2$. 

To conclude this Section, let us provide an example of a possible application of our model.

**Example 3.** As an example of application, let us revisit the example of the online booking industry (Example 1). We consider again the state space as the reputation of the firm but restrict the action set to the offers the firm can post online. By monitoring each other actions, firms can frequently update their offers and promotions (with the help perhaps of automated software) to change their reputation levels.
4. Concluding remarks. An interesting feature of the mean field game models from the point of view of applications is the simplification it entails: on the equilibrium, each player has at his disposal an extremely simple strategy that depends only on his current state and he does not need to keep track of the other players, provided the number of players is large enough. This is because the aggregate state-action of the other players is regarded as a parameter, which deviates from the actual realization of the aggregated state-action with very small probability.

However, this nice feature is also its curse. One problem is that the mean field equilibrium need not be unique. If there is a coordinator of the game that informs the players which mean field equilibrium should be played, there are no problems. In applications, this will typically not be the case. One way around would be to provide the players with an adaptation mechanism. To explain this point, let us revisit the example of the driving game:

Example 4. Consider the driving game of Example 2 with $N$ players. The only equilibria that do not depend on $N$ are everyone on the left and everyone on the right. Consider the following adaptation mechanism: each player chooses left or right with probability $\frac{1}{2}$ on the first stage. On the second stage, observing the realizations of the first stage, each player looks at everyone’s choice (and recalls its own) and imitates the choice of the majority. Thus, from stage three, the players will be on an equilibrium path if $N$ is odd. If $N$ is even, there is positive probability that none of the equilibria is reached.

A proper study of adaptation mechanisms for mean field games in the general case is clearly an interesting direction of future research.

Acknowledgments. I am grateful to Martino Bardi, who motivated this research during a very pleasant stay at the University of Padova and to an anonymous referee for useful comments.

Appendix: An approximation Lemma. Let $S$ denote a finite set. We identify the set $S$ with the canonical basis of $\mathbb{R}^{|S|}$. Denote by $\mathcal{M}$ the subset of $\mathbb{R}^{S \times S}$ consisting of transition matrices for Markov chains over $S$.

Let $P : \Delta(S) \to \mathcal{M}$ denote a Lipschitz continuous function with respect to the $L_1$ norm with Lipschitz constant $L_P$. Since $S$ is finite, we have that the total variation distance in $\Delta(S)$, defined as $\|\mu - \nu\|_\infty := \max_{A \subseteq S} |\mu(A) - \nu(A)|$ is related to the $L_1$ distance by $\|\mu - \nu\|_\infty = \frac{1}{2}\|\mu - \nu\|_1$, so that a $L_P$-Lipschitz function in the $L_1$ norm is $2L_P$ in the total variation norm.

Let $T > 1$ be an integer, representing the number of stages. For $i = 1, \ldots, N$ and $k = 0, 1 \ldots, T - 1$ define the following

$$X_{k+1}^i = P(m_k)X_k^i$$

$$X_{k+1,N}^i = P(m_{k,N})X_{k,N}^i$$

where $X_0^i = X_{0,N}^i$ is a random variable with law $m_0$ and $X_{0,N}^i$ are sampled i.i.d with probability $m_0$. Here, $m_k$ denotes the law of $X_k^i$ and $m_{k,N} := \frac{1}{N} \sum_{i=1}^N X_{k,N}^i$.

Observe that

$$X_{k+1}^i = X_0^i + \sum_{\ell=0}^{k} X_{\ell+1}^i - X_\ell^i$$
so that

\[ X_{k+1,N}^i = X_{0,N}^i + \sum_{\ell=0}^{k} X_{\ell+1,N}^i - X_{\ell,N}^i \]

Before we proceed to the approximation lemma, let us introduce \( \xi_{k,N} := \frac{1}{N} \sum_{i=1}^{N} X_k^i \). Observe that \( m_k = E \xi_{k,N} \). We have the following

**Proposition 7.** The following estimate holds:

\[ E \| \xi_{k,N} - m_k \|_1 \leq |S|/2\sqrt{N}. \]

**Proof.** For every \( s \in S \), the random variable \( \xi_{k,N}(s) \) is the average of \( N \) independent Bernoulli variables. Hence, by definition of the variance and Jensen’s inequality,

\[ E|\xi_{k,N}(s) - m_k(s)| \leq \sqrt{\text{var}(\xi_{k,N}(s))} \leq \frac{1}{2\sqrt{N}} \]

Summing over \( s \) gives the result. \( \square \)

Let \( F_k \) denote the filtration generated by the observed history up to stage \( k \). We are ready to prove the following approximation lemma.

**Lemma 4.1.** The following estimate holds: for all \( i = 1, \ldots, N \),

\[ E\left( \max_{t \leq T} \| X_{s+1}^i - X_{s+1,N}^i \|_\infty \right) \leq \frac{L_p T |S| \exp(T(\|P\|_\infty + L_P))}{\sqrt{N}} \]

**Proof.** Let \( D_k^i := E \left( \max_{s \leq k} \| X_{s+1}^i - X_{s+1,N}^i \|_\infty \mid F_k \right) \) and \( D_k := \max_{1 \leq i \leq N} D_k^i \). Now observe that, for any \( i \):

\begin{align*}
D_{k+1}^i &\leq \sum_{\ell=0}^{k} E \| P(m_\ell) X_\ell^i - P(\xi_{\ell,N}) X_\ell^i \|_\infty + \beta_i \left( \| P \|_\infty + L_P \right) D_k^i \quad (12) \\
&+ E \| P(\xi_{\ell,N}) X_\ell^i - P(m_\ell) X_\ell^i \|_\infty + E \| X_\ell^i - X_{\ell,N}^i \|_\infty \quad (13) \\
&\leq \sum_{\ell=0}^{k} L_P E \| m_\ell - \xi_{\ell,N} \|_1 + (\| P \|_\infty + L_P) D_k \quad (14)
\end{align*}

From here it follows that

\[ D_{k+1} \leq \frac{L_p T |S|}{\sqrt{N}} + (\| P \|_\infty + L_P) \sum_{\ell=0}^{k} D_k. \]

The first inequality ((12) and (13)) follows from the triangle inequality. For (14), we use the fact that

\[ E \left( \| P(m_\ell) X_\ell^i - P(\xi_{\ell,N}) X_\ell^i \|_\infty \right) \leq E \left( \| P(m_\ell) - P(\xi_{\ell,N}) \|_\infty \| X_\ell^i \|_1 \right) \leq L_P E \| m_\ell - \xi_{\ell,N} \|_1. \]

and that

\[ E \| P(\xi_{\ell,N}) X_\ell^i - P(m_\ell) X_\ell^i \|_\infty \leq E \| P(\xi_{\ell,N}) - P(m_\ell) \|_\infty + E \| P(\xi_{\ell,N}) X_\ell^i - P(m_\ell) X_{\ell,N}^i \|_\infty \]

\[ \leq \| P \|_\infty E \| X_\ell^i - X_{\ell,N}^i \|_\infty + L_P D_k. \]

The conclusion follows from induction. \( \square \)
Let us prove a useful lemma.

**Lemma 4.2.** Let $f : S \times \Delta(S) \to \mathbb{R}$ be a bounded and $L_f$-Lipschitz continuous function with the second variable respect to the $L_1$ norm. Then we have

$$
E \left( |f(X_{k+1,N}^i, m_{k+1}) - f(X_{k+1,N}^i, m_k)| \mid F_k \right) \leq (L_f + \|f\|_\infty)D_{k+1}
$$

**Proof.** It suffices to observe that

$$
E \left( |f(X_{k+1,N}^i, m_{k+1}) - f(X_{k+1,N}^i, m_k)| \mid F_k \right)
$$

is less or equal than

$$
L_f E \left( \|m_{k+1,N} - m_k\|_1 \mid F_k \right)
$$

which is in turn bounded by $L_f D_{k+1}$.

We also have that

$$
E \left( |f(X_{k+1,N}^i, m_{k+1}) - f(X_{k+1,N}^i, m_k)| \mid F_k \right) \leq \|f\|_\infty D_{k+1}
$$

Combining these two inequalities yields the result. \hfill \Box

**REFERENCES**

[1] Y. Achdou and I. Capuzzo-Dolcetta, Mean field games: Numerical methods, *SIAM J. Numer. Anal.*, 48 (2010), 1136–1162.

[2] S. Adlakha, R. Johari and G. Weintrub, Equilibria of dynamic games with many players: Existence, approximation, and market structure, *J. Econom. Theory*, 156 (2015), 269–316.

[3] M. Bardi, Explicit solutions of some linear-quadratic mean field games, *Networks and Heterogeneous Media*, 7 (2012), 243–261.

[4] F. Camilli and F. Silva, A semi-discrete approximation for a first order mean field game problem, *Networks and Heterogeneous Media*, 7 (2012), 263–277.

[5] E. Carlini and F. Silva, A fully-discrete semi-Lagrangian approximation for a first order mean field game problem, *SIAM J. Numer. Anal.*, 49 (2013), 3222–3233.

[6] R. Elliot, X. Li and Y.-H. Ni, Discrete time mean-field stochastic linear-quadratic optimal control problems, *Automatica*, 49 (2013), 3222–3233.

[7] D. Fudenberg and J. Tirole, Learning-by-doing and market performance, *The Bell J. of Economics*, 14 (1983), 522–530.

[8] D. Gomes, J. Mohr and R. Souza, Discrete time, finite state space mean field games, *J. Math. Pures et Ap.*, 93 (2010), 308–328.

[9] O. Guéant, J. Lasry and P. Lions, Mean Field Games and Applications, in *Paris Princeton Lectures on Mathematical Finance 2010*, Lecture Notes in Math., 2003, Springer-Verlag, 2011, 205–266.

[10] O. Hernández-Lerma and J. Lasserre, *Discrete-Time Markov Control Problems*, Springer-Verlag, 1996.

[11] M. Huang, P. Caines and R. Malhamé, Individual and mass behavior in large population stochastic wireless power control problems: Centralized and nash equilibrium solutions, in *Proceedings of the 42nd IEEE Conference on Decision and Control*, 2003.

[12] M. Huang, P. Caines and R. Malhamé, Large population stochastic dynamic games: Closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle, *Communications in Information and Systems*, 6 (2006), 221–251.

[13] J. Lasry and P. Lions, Jeux à champ moyen. I. Le cas stationnaire, *C. R. Math. Acad. Sci. Paris*, 343 (2006), 619–625.

[14] J. Lasry and P. Lions, Jeux à champ moyen. II. Horizon fini et controle optimal, *C. R. Math. Acad. Sci. Paris*, 343 (2006), 679–684.

[15] J. Lasry and P. Lions, Mean field games, *Japanese Journal of Mathematics*, 2 (2007), 229–260.

[16] A. Neyman, Stochastic games with short-stage duration, *Dyn. Games Appl.*, 3 (2013), 236–278.

Received March 2014; revised February 2015.

*E-mail address:* pablo.maldonado@imj-prg.fr