CONTROL SYSTEMS OF INTERACTING OBJECTS MODELED AS A GAME AGAINST NATURE UNDER A MEAN FIELD APPROACH

CARMEN G. HIGUERA-CHAN
Departamento de Matemáticas, Universidad de Sonora
Rosales s/n, Col. Centro, 83000
Hermosillo, Sonora, México

HÉCTOR JASSO-FUENTES
Departamento de Matemáticas, CINVESTAV-IPN
Apartado Postal 14-740
México D.F., 07000, México

J. ADOLFO MINJÁREZ-SOSA*
Departamento de Matemáticas, Universidad de Sonora
Rosales s/n, Col. Centro, 83000
Hermosillo, Sonora, México

(Communicated by Onésimo Hernández-Lerma)

Abstract. This paper deals with discrete-time stochastic systems composed of a large number of $N$ interacting objects (a.k.a. agents or particles). There is a central controller whose decisions, at each stage, affect the system behavior. Each object evolves randomly among a finite set of classes, according to a transition law which depends on an unknown parameter. Such a parameter is possibly non observable and may change from stage to stage. Due to the lack of information and to the large number of agents, the control problem under study is rewritten as a game against nature according to the mean field theory; that is, we introduce a game model associated to the proportions of the objects in each class, whereas the values of the unknown parameter are now considered as "actions" selected by an opponent to the controller (the nature). Then, letting $N \rightarrow \infty$ (the mean field limit) and considering a discounted optimality criterion, the objective for the controller is to minimize the maximum cost, where the maximum is taken over all possible strategies of the nature.

1. Introduction. We consider discrete-time stochastic systems composed of a large number of $N$ interacting objects controlled by a central decision-maker (controller). The objects evolve among a finite set of classes $S$ according to a transition law $K$, homogeneous on $N$, which depends on an unknown and possibly non-observable parameter $\mu$ belonging to a suitable set $\Gamma$. Moreover, the parameter $\mu$ may change from stage to stage. Specifically, if $X_n^N(t)$ represents the class of the object $n$, $a_t$ is the...
control or action selected by some central controller, and \( \mu_t \) the unknown parameter, at time \( t \), then the conditional distribution of \( X_n^N(t + 1) \) is given by

\[
K_{ij}(a, \mu) = \Pr[X_n^N(t + 1) = j | X_n^N(t) = i, a_t = a, \mu_t = \mu], \quad i, j \in S, n = 1, \ldots, N.
\]

At each stage, a cost resulting from the movement of the objects among the classes and the selected control is generated. In this sense, the controller aims to select actions to minimize a discounted optimality criterion as (9).

Typically, systems depending on an unknown parameter, which in turn comes from a random process, are analyzed by means of statistical estimation procedures, provided, of course, that such a parameter, or the random process, is observable. However, because of the lack of observability of the parameter \( \mu \), such an approach is not applicable in our model. Instead, we assume that the only information about the parameter \( \mu \) is that, at each stage, it belongs to the set \( \Gamma \). In this scenario, the interacting objects system can be modeled as a game against nature, also known as a minimax control model, which consists in assuming that the controller has an opponent, the “nature”, who, at each stage, is selecting the parameter \( \mu \) from the set \( \Gamma \). Under this new approach, the goal of the controller is to minimize the maximum cost, where the maximum is taken over all possible strategies of the nature, as is specified in (11).

By dealing with a large number \( (N \sim \infty) \) of objects it is not possible to directly compute optimal policies (at least from a practical point of view), since for example, the corresponding dynamic programming equation contains terms depending on an \( N \)-dimensional integral. To overcome this obstacle, we will appeal to mean field theory; that is, instead of analyzing a single object, which would be almost impossible, we will focus on the proportion of objects occupying each class. This allows us to define, as a first step, a game against nature model \( M^N \) whose states are now the proportions of objects. By letting \( N \to \infty \) and applying an appropriate type of law of large numbers, a game against nature \( M \) independent of \( N \) is obtained, whose states are precisely the probability measures obtained as limit of the proportions. The model \( M \) is referred to as the mean field game against nature. Under this scheme, we can establish solutions for the game model \( M \) and obtain a “worst case” optimal control policy \( \pi^* \). Our objective is then to analyze the minimax deviation of \( \pi^* \) when it is used to control the original process defined on \( M^N \). Specifically we aim to show that \( \pi^* \) is a worst case optimal control policy in an asymptotic sense as \( N \to \infty \).

Our approach is applied to other important models. For instance, consider the case when \( S \subset \mathbb{R} \) and the process \( \{X_n^N(t)\} \) evolves according to a difference equation of the form

\[
X_n^N(t + 1) = F(X_n^N(t), a_t, \xi_n^t), \quad t = 0, 1, \ldots, \quad n = 1, 2, \ldots, N, \tag{1}
\]

where \( \{\xi_n^t\} \) is a family of possibly non observable random variables, taking values on a set \( Z \), and partially identically distributed with unknown distribution \( \mu_n^t \). The latter means that

\[
\mu_1^t = \mu_2^t = \ldots = \mu_n^t =: \mu_t, \quad t \geq 0.
\]

Considering the distribution \( \mu_t \) as the unknown parameter and \( \Gamma \) as the set of probability measures on \( Z \), the transition law takes the form

\[
K_{ij}(a, \mu) = \int_Z I_{ij}[F(i, a, z)] \mu(dz). \tag{2}
\]
In other words, at each stage $t$, the movement of objects $n = 1, 2, ..., N$ is affected by random variables $\xi^n_1, \xi^n_2, ..., \xi^n_N$, respectively, not necessarily equal, but with the same distribution $\mu_t$.

Within the context (1)-(2), the specific case where $\{\xi^n_t\}$ is a family of observable, independent and identically distributed random variables with common unknown distribution $\mu$, i.e. $\mu_t = \mu$ for all $t \geq 0$, was previously analyzed in [13]. In particular, it was assumed that the distribution $\mu$ has an unknown density $\rho$, so it was possible to implement statistical density estimation schemes for $\rho$ together with control procedures, to prove existence of nearly optimal policies.

Games against nature or minimax control problems as well as mean field theory for control systems or games, have been widely studied, separately, under several settings: For discrete- and continuous-time minimax models, see e.g., [7, 12, 17, 15, 21, 22, 24]; for discrete- and continuous-time systems of interacting objects, mean field control problems, and mean field games see e.g., [1, 2, 3, 5, 6, 9, 10, 11, 14, 16, 18, 19, 20, 23] and references therein.

To the best of our knowledge, our approach has not been previously studied in the literature and its novelty relies in mixing the aforementioned two techniques. Furthermore, this work can be considered as an extension of [13] in the sense that our results are applicable to difference-equation models as (1)-(2), but assuming the driving process $\{\xi^n_t\}$ is nonobservable with unknown distribution.

The paper is organized as follows. In Section 2 we describe the system of $N$ objects under study, and the corresponding game against nature $\mathcal{M}^N$ is introduced. Next, in Section 3, we present the game against nature $\mathcal{M}$ associated to the mean field scenario as well as the key convergence result. Furthermore we introduce our main result that establishes the existence of asymptotically minimax policies that are obtained from the model $\mathcal{M}$ but which in turn are the best approximate optimal policies for the model $\mathcal{M}^N$. Section 4 contains two examples to illustrate the elements of the proposed model. Finally, the proof of the main result is given in Section 5.

**Notation and terminology.** As usual, $\mathbb{N}$ (respectively $\mathbb{N}_0$) stands for the set of positive (resp. nonnegative) integers. Similarly, $\mathbb{R}$ (resp. $\mathbb{R}_+$) denotes the set of real (resp. nonnegative real) numbers.

On the other hand, given a Borel space $Z$ (that is, a Borel subset of a complete and separable metric space) its Borel $\sigma-$algebra is denoted by $\mathcal{B}(Z)$, and the attribute “measurable” will be applied for either Borel measurable sets or Borel measurable functions.

Let $\mathbb{P}(Z)$ be the set of probability measures on $Z$. If $Z$ is finite, e.g. $Z = \{1, 2, ..., z\}$, we will identify any $p \in \mathbb{P}(Z)$ by the vector $p := (p(1), p(2), ..., p(z))$ where $p(i) \geq 0$, $i \in Z$, and $\sum_{i=1}^{z} p(i) = 1$. As usual, $|\cdot|$ will denote the norm on $\mathbb{R}$.

We shall define the norm on $\mathbb{P}(Z)$, for $Z$ finite, under the corresponding $L_\infty$ norm; that is, for each vector $p \in \mathbb{P}(Z)$:

$$||p|| := \max \{|p(1)|, ..., |p(z)|\}.$$  

Let $Z$ and $Y$ be Borel spaces. A stochastic kernel $Q(\cdot|\cdot)$ is a function $Q : \mathcal{B}(Z) \times Y \to [0, 1]$, such that $Q(\cdot|y)$ is a probability measure on $\mathcal{B}(Z)$ for each fixed $y \in Y$, and $Q(B|\cdot)$ is a measurable function on $Y$ for each fixed $B \in \mathcal{B}(Z)$.

Finally, $\mathcal{B}(Z)$ denotes the class of real-valued bounded functions on $Z$ endowed with the supremum norm $||v|| := \sup_{z \in Z} |v(z)|$, while $\mathcal{C}_b(Z)$ is the subspace of $\mathcal{B}(Z)$, consisting of all real-valued bounded continuous functions defined on $Z$.  

2. Model description. We consider a discrete-time controlled stochastic system of \( N \) (\( N \sim \infty \)) interacting objects where \( X_n^N(t) \), \( n = 1, 2, \ldots, N \), \( t \in \mathbb{N}_0 \), represents the class of the object \( n \) at time \( t \) taking values in a given finite set \( S = \{c_1, c_2, \ldots, c_s\} \) := \( \{1, 2, \ldots, s\} \), named the class set. At each stage, a central controller selects an action or control \( a_t \) from a given Borel set \( A \), which produces a random movement of the objects from a class to another. Additionally, such a movement is also influenced by a parameter which is unknown and possibly nonobservable by the controller, and it can change form stage to stage. Hence, we assume that the objects evolve according to a transition law

\[
K_{ij}(a, \mu) := \text{Pr} \left[ X_n^N(t+1) = j \mid X_n^N(t) = i, a_t = a, \mu_t = \mu \right],
\]

where \( \mu_t \) represents the parameter at time \( t \in \mathbb{N}_0 \). We assume that \( \mu_t \) belongs to a compact set \( \Gamma \), so-named parameter set.

The relation (3) defines a stochastic kernel represented in matrix notation by \( K = K(a, \mu) = [K_{ij}(a, \mu)] \).

In the present analysis, it is assumed that the objects are observable only through the class in which they are located, so that the controller can only determine the number of objects in each class \( i \in S \). In this sense, the behavior of the system can be reformulated by means of the proportions of objects at each class. Indeed, let \( M_i^N(t) \) be the proportion of objects in class \( i \in S \) at time \( t \); i.e.,

\[
M_i^N(t) := \frac{1}{N} \sum_{n=1}^{N} I \{X_n^N(t) = i\}, \quad i \in S,
\]

and we denote by \( M^N(t) \) the vector whose components are the \( M_i^N(t) \)'s:

\[
M^N(t) = (M_1^N(t), M_2^N(t), \ldots, M_s^N(t)).
\]

Note that \( M^N(t) \) belongs to the space of probability measures \( \mathbb{P}_N(S) := \{p \in \mathbb{P}(S) : Np(i) \in \mathbb{N}, \forall i \in S\} \subset \mathbb{P}(S) \).

There are several ways to represent the dynamics of the proportions \( M^N(t) \) by means of stochastic difference equations. Some of them depend on the specific control problem under study, while others are obtained according to a simulation process. That is, there exists a function \( G^N : \mathbb{P}_N(S) \times A \times \Gamma \times \mathbb{R}^N \rightarrow \mathbb{P}_N(S) \) such that

\[
M^N(t + 1) = G^N(M^N(t), a_t, \mu_t, \bar{w}_t),
\]

where \( \{\bar{w}_t\} \) is a sequence of i.i.d. random vectors on \( \mathbb{R}^N \) with common distribution \( \theta \).

In our case, we adopt the one obtained by a particular simulation process, borrowed from [9], which is in the spirit of simulations and has suitable convergence properties. Indeed, for each \( t \in \mathbb{N}_0 \), let \( \{w_n^k(t)\} \), \( n \in \{1, \cdots, N\} \) \( k \in S \), be a family of i.i.d. uniform random variables on \( [0, 1] \), and denote

\[
\bar{w}^i(t) := (w_1^i(t), \cdots, w_N^i(t)) \quad i \in S,
\]

and

\[
\bar{w}_t := (\bar{w}^1(t), \cdots, \bar{w}^s(t)).
\]

It is worth noting that \( \sum_{i=1}^s N M_i^N(t) = N \), thus \( \bar{w}_t \in [0, 1]^N \). The dynamic of the process \( \{M^N(t)\} \) is obtained as follows. We first define

\[
G_1^N(M^N(t), a_t, \mu_t, \bar{w}_t) := \frac{1}{N} \sum_{k=1}^{s} \sum_{n=1}^{N M_i^N(t)} I \{\Delta_{a_t}^n \} (w_n^k(t)), \quad i \in S,
\]
CONTROL SYSTEMS OF INTERACTING OBJECTS

where

$$\Delta_{\mu t_{ki}} := [\Psi_{\mu t_{ki}}(a), \Psi_{\mu t_{ki}+1}(a)] \subseteq [0, 1],$$

and

$$\Psi_{\mu t_{ki}}(a) := \sum_{l=1}^{i-1} K_{kl}(a, \mu_t), \ k, i \in S.$$ (7)

Then, the function $G^N$ in (5) takes the form

$$G^N(m, a, \mu, w) = (G^N_1(m, a, \mu, w), ..., G^N_s(m, a, \mu, w))$$ (8)

for $(m, a, \mu, w) \in P_N(S) \times A \times \Gamma \times [0, 1]^N$. The dynamics (5)-(6) also provide the convergence of $M^N(\cdot)$ to some other dynamics associated to the mean field model (see Section 3).

Now, let us assume that the one-stage cost depends on the configuration of the system $M^N(t)$ and on the action selected by the controller. This one-stage cost can then be represented by a measurable function $r: P(S) \times A \rightarrow \mathbb{R}$.

With these previous ingredients, the optimal control problem for the controller reduces to find a control policy directed to minimize his/her cost flow

$$\{r(M^N(t), a_t)\}, \quad M^N(t) \in P_N(S), a_t \in A,$n

over an infinite horizon using a discounted cost criterion of the form

$$E \left[ \sum_{t=0}^{\infty} \alpha^t r(M^N(t), a_t) \big| M^N(0) = m \right], \quad 0 < \alpha < 1,$$ (9)

subject to (5). However, notice that at each stage the only information for the controller about the parameter $\mu_t$ is that it belongs to the set $\Gamma$. This means that the problem must be addressed from an alternative approach, which is introduced below.

Games against nature formulation. Roughly speaking, in games against nature formulation, it is assumed that the controller has an opponent (the nature) who selects the parameter $\mu_t$ at each stage, so the goal for the controller is to minimize the maximum cost incurred by the nature. More precisely, we define the discrete-time game against nature associated to the system of $N$ interacting objects (in short $N$-GAN) as follows:

$$\mathcal{M}^N := \left( P_N(S), A, \Gamma, G^N, \theta, r \right).$$ (10)

The evolution of the system associated to the model $\mathcal{M}^N$ can be interpreted as follows: At time $t$, the controller observes the system configuration via the state $m = M^N(t) \in P_N(S)$, representing the proportions of the objects. Then he/she selects an action $a = a_t \in A$ whereas the opponent (the nature) picks a parameter $\mu = \mu_t \in \Gamma$. As a consequence the following happens: (1) a cost $r(m, a)$ is incurred, and (2) the system moves to a new state $M^N(t+1)$ according to the transition law

$$Q(B|m, a, \mu) := \text{Pr}[M^N(t+1) \in B|M^N(t) = m, a_t = a, \mu_t = \mu]$$

$$= \int_{\mathbb{R}^N} I_B[G^N(m, a, \mu, w)] \theta(dw) \quad \forall B \in \mathcal{B}(P_N(S)),$$ with $G^N$ as in (5). Once the transition to the state $M^N(t+1)$ occurs, the process is repeated.

Observe that the changes in the system configuration governed by the transition law $Q$, rely strongly on the parameters $\{\mu_t\}$ selected by the opponent. In this sense,
the objective for the controller is to find a policy that minimizes the maximum total expected discounted cost introduced in (11) below.

In order to ensure the existence of minimizers, we impose the following conditions regarding continuity and compactness on some elements of the game model.

**Assumption 2.1.**

(a) The action space \( A \) is a compact Borel space, whose metric is denoted by \( d_A \).

(b) The set \( \Gamma \) is compact.

(c) The mapping \( a \mapsto K_{ij}(a, \mu) \), defined in (3), is continuous for all \( i, j \in S \) and \( \mu \in \Gamma \).

(d) The one-stage cost \( r \) is bounded and uniformly Lipschitz in \( m \) with constant \( L_r \); that is, for some constant \( R > 0 \),

\[
|r(a, \alpha)| \leq R \quad \forall (a, \alpha) \in \mathbb{P}(S) \times A,
\]

and for every \( a, a' \in A \), and \( m, m' \in \mathbb{P}(S) \),

\[
\sup_{(a, a') \in A \times A} |r(a, \alpha) - r(a', \alpha)| \leq L_r \|m - m'\|_\infty.
\]

A first consequence of Assumption 2.1 is the following result, whose proof is provided in Section 5.

**Lemma 2.2.** Under Assumption 2.1, the mapping \( a \mapsto G^N(m, a, \mu, w) \) defined in (8), is continuous for all \( m \in \mathbb{P}_N(S), \mu \in \Gamma \) and \( w \in \mathbb{R}^N \).

**Control policies:** Let \( \mathbb{H}_t^N := \mathbb{P}_N(S) \) and \( \mathbb{H}^N := (\mathbb{P}_N(S) \times A \times \Gamma \times \mathbb{R}^N)^t \times \mathbb{P}_N(S), t = 1 \), be the space of histories up to time \( t \). An element \( h_t^N \) of \( \mathbb{H}^N_t \) can be expressed as

\[
h_t^N = (M^N(0), a_0, \mu_0, \bar{w}_0, \ldots, M^N(t - 1), a_{t-1}, \mu_{t-1}, \bar{w}_{t-1}, M^N(t)).
\]

An admissible control policy is a sequence \( \pi^N = \{\pi_t^N\} \) of stochastic kernels \( \pi_t^N \) on \( A \) given \( \mathbb{H}^N \) such that \( \pi_t^N (A|h_t^N) = 1 \) for all \( h_t^N \in \mathbb{H}^N_t \), \( t \in \mathbb{N}_0 \). We denote by \( \Pi^N \) the set of all admissible control policies for the controller.

Let \( \mathcal{F} \) be the set consisting of all measurable functions \( f : \mathbb{P}_N(S) \to A \) and \( \mathcal{F}^N := \mathcal{F}|_{\mathbb{P}_N(S)} \) be the restriction of \( \mathcal{F} \) to \( \mathbb{P}_N(S) \). A policy \( \pi^N \in \Pi^N \) is said to be a (deterministic) Markov policy if there exists a sequence \( \{f_t^N\} \subseteq \mathcal{F}^N \) such that for all \( t \in \mathbb{N}_0 \) and \( h_t^N \in \mathbb{H}_t^N \), \( \pi_t^N (A|h_t^N) = \delta_{f_t^N(M^N(t))}(\cdot) \), where \( \delta_x(\cdot) \) is the Dirac measure concentrated at point \( a \). In this case, \( \pi^N \) takes the form \( \pi^N = \{f_t^N\} \).

In particular, if \( f_t^N = f^N \) for some \( f^N \in \mathcal{F}^N \) and for all \( t \in \mathbb{N}_0 \), we say that \( \pi^N \) is a stationary policy. We denote by \( \Pi_M^N \) and \( \mathcal{F}^N \) the set of all Markov policies and the set of stationary policies for the controller. We also denote by \( \Pi_A^N \) the set of deterministic Markov policies for the controller when we use \( \mathcal{F} \) instead of \( \mathcal{F}^N \); that is, \( \Pi_A^N \) is the family of sequences of functions \( \{f_t\} \subseteq \mathcal{F} \). Observe that any policy \( \pi = \{f_t\} \in \Pi_A^N \) whose elements \( f_t \) are restricted to \( \mathbb{P}_N(S) \) turns out to be an element of \( \Pi^N \); hence we can conclude that \( \mathcal{F}^N \subseteq \Pi_M^N \subseteq \Pi^N \).

On the other hand, we assume that the nature’s actions depend only on the time parameter; in other words, a control policy for the opponent is a sequence \( \bar{\pi} = \{\mu_t\}, \mu_t \in \Gamma, t \in \mathbb{N}_0 \). We denote by \( \Pi_{\bar{\Gamma}} \) the set of policies for the opponent.

For each pair of policies \( (\pi^N, \bar{\pi}) \in \Pi^N \times \Pi_{\bar{\Gamma}} \) and initial state \( M^N(0) = m \in \mathbb{P}_N(S) \), we define the total expected discounted cost as

\[
V^N(\pi^N, \bar{\pi}, m) := E_{m}^{\pi^N, \bar{\pi}} \left[ \sum_{t=0}^{\infty} \alpha^t r(M^N(t), a_t) \right], \tag{11}
\]
In this case, \( \pi \) as follows (see, e.g., [12, 22]). the methodology of minimax Markov control theory to characterize minimax policies.

\[ V^*_N(m) := \inf_{\pi \in \Pi^N} \sup_{\pi \in \Pi} V^N(\pi, \pi, m) = \sup_{\pi \in \Pi} V^N(\pi_m, \pi, m), \quad m \in \mathbb{P}_N(S). \] (12)

In this case, \( \pi^*_N \) is said to be a minimax policy, whereas \( V^*_N \) is the N-value function.

According to the description previously presented, and taking into account the continuity of the mapping \( a \rightarrow G^N(m, a, \mu, w) \) stated in Lemma 2.2, we can apply the methodology of minimax Markov control theory to characterize minimax policies as follows (see, e.g., [12, 22]).

For \( u \in \mathbb{B}(\mathbb{P}(S)) \) and \( (m, a, \mu) \in \mathbb{P}(S) \times A \times \Gamma \), we define

\[ T_{(a, \mu)}u(m) := r(m, a) + \alpha \int_{\mathbb{R}^N} u[G^N(m, a, \mu, w)] \theta(dw), \] (13)

and

\[ Tu(m) := \min_{a \in A} \sup_{\mu \in \Gamma} T_{(a, \mu)}u(m) \] (14)

**Remark 1.** Taking into account Lemma 2.2, under Assumption 2.1, we can apply similar arguments as in [12] to prove that the operator \( T \) satisfies the following properties:

(a) \( T \) is a contraction operator with modulus \( \alpha \),

(b) \( T \) maps \( C_0(\mathbb{P}_N(S)) \) into itself, and

(c) For each \( u \in C_0(\mathbb{P}_N(S)) \), there exists \( f^*_N \in \mathbb{F} \) such that

\[ Tu(m) = \sup_{\mu \in \Gamma} T_{(f^*_N, \mu)}u(m), \quad m \in \mathbb{P}_N(S) \]

Then, under these properties, Banach’s Fixed Point Theorem yields the following minimax theorem.

**Theorem 2.3.** If Assumption 2.1 holds, then:

(a) The optimal value function (12) is the unique solution in \( C_0(\mathbb{P}_N(S)) \) satisfying

\[ V^*_N(m) = TV^*_N(m), \quad m \in \mathbb{P}_N(S), \] (15)

where \( T \) is the operator in (14).

(b) For each \( n \in \mathbb{N} \), there exists \( v_n \in C_0(\mathbb{P}_N(S)) \), with \( v_0 := 0 \) satisfying

\[ \|v_n - V^*_N\| \leq R \frac{\alpha^n}{1 - \alpha} \]

(c) There exists \( f^*_N \in \mathbb{F} \) such that, for all \( m \in \mathbb{P}_N(S) \),

\[ V^*_N(m) = \sup_{\mu \in \Gamma} \left\{ r(m, f^*_N) + \alpha \int_{A} V^*_N[G^N(m, f^*_N, \mu, w)] \theta(dw) \right\} \] (16)

or, equivalently,

\[ \Phi^N(m, f^*_N) = 0, \] (17)

where, for each \( m \in \mathbb{P}_N(S) \) and \( a \in A \),

\[ \Phi^N(m, a) = \sup_{\mu \in \Gamma} \left\{ r(m, a) + \alpha \int_{A} V^*_N[G^N(m, a, \mu, w)] \theta(dw) \right\} - V^*_N(m). \] (18)
Moreover \( \{f^N\} \) is a minimax policy for the controller; that is,

\[
V^*_N(m) = \sup_{\pi \in \Pi} V^N(m, f^N, \pi), \quad m \in \mathcal{P}(S).
\]

Although Theorem 2.3 provides a way of finding minimax policies, from a practical point of view it is not quite useful because \( N \) is too large (\( N \sim \infty \)). Just to mention one of these difficulties, observe that the equations (15)-(16) highly depend on the \( N \)-dimensional integral whose analysis is prohibitive for large \( N \). To overcome this difficulty, we introduce a new game against nature, denoted by \( \mathcal{M} \), so-called *mean field game against nature*, that becomes the limit of the GAN model \( \mathcal{M}^N \) as \( N \to \infty \), which, of course, is independent of \( N \). Thus, our challenge is to show that the resulting minimax policy on \( \mathcal{M} \) is minimax in the original model \( \mathcal{M}^N \), in an asymptotic sense as \( N \to \infty \); in other words, \( \mathcal{M} \) will be used as an approximating model for \( \mathcal{M}^N \).

3. Mean field game against nature. Consider the function \( G : \mathcal{P}(S) \times A \times \Gamma \to \mathcal{P}(S) \) defined as

\[
G(m, a, \mu) := mK(a, \mu),
\]

where \( K \) is the transition kernel. Let \( \{m(t)\} \subset \mathcal{P}(S) \) be a controlled process defined by the deterministic difference equation

\[
m(t + 1) = G(m(t), a_t, \mu_t);
\]

with initial condition \( m(0) = m \in \mathcal{P}(S) \). As before, \( a_t \in A \) represents the control selected and \( \mu_t \) stands for the unknown parameter at time \( t \).

As is established in (23) below, \( \{m(t)\} \) is the limit process of \( \{M^N(t)\} \), as \( N \to \infty \), which plays the role of the mean field process. Hence, using the same one-stage cost \( r \) defined for the \( N \)-GAN (10) we can then define the mean field game against nature (MFGAN–\( \mathcal{M} \) for short) as

\[
\mathcal{M} = (\mathcal{P}(S), A, \Gamma, G, r).
\]

Considering the deterministic nature of the process (19), it is clear that the dynamics is completely determined by the sequences \( \{a_t\} \subset A \) and \( \{\mu_t\} \), and by the initial condition \( m \in \mathcal{P}(S) \). Furthermore, it is also well recognized that a control policy \( \pi \) for deterministic control systems is a sequence of decision rules \( \pi = \{f_t\} \subseteq \mathcal{F} \) (see [4]). Therefore, in terms of the notation given in the definition of control policies, we can consider the set \( \Pi_M \) as the set of all control policies for the controller on the model \( \mathcal{M} \). In this way, for each \( (\pi, \bar{\pi}) \in \Pi_M \times \Pi_{\Gamma} \) initial condition \( m(0) = m \in \mathcal{P}(S) \), we define the total discounted cost for the mean field model as

\[
v(\pi, \bar{\pi}, m) = \sum_{t=0}^{\infty} \alpha^t r(m(t), a_t).
\]

Thus, the mean field game against nature consists in finding a policy \( \pi_* \in \Pi_M \) such that

\[
v_*(m) := \inf_{\pi \in \Pi_M} \sup_{\bar{\pi} \in \Pi_{\Gamma}} v(\pi, \bar{\pi}, m) = \sup_{\pi \in \Pi_M} v(\pi, \bar{\pi}, m), \quad m \in \mathcal{P}(S),
\]

where \( v_* \) is the mean field value function and \( \pi_* \) is said to be a mean field-minimax policy for the MFGAN–\( \mathcal{M} \).

Now, similarly as in Theorem 2.3, we can apply arguments on minimax theory (see, e.g., [12, 22]) to get the following results.
Theorem 3.1. If Assumption 2.1 holds true, then:

(a) The optimal value function (21) is the unique solution in \( C_b(\mathcal{P}(S)) \) satisfying
\[
v_\ast (m) = T v_\ast (m), \quad m \in \mathcal{P}(S),
\]
where, for each \( u \in C_b(\mathcal{P}(S)) \) and \( m \in \mathcal{P}(S) \),
\[
Tu(m) := \min_{a \in \Lambda} \sup_{\mu \in \Gamma} \{ r(m, a) + \alpha u[G(m, a, \mu)] \}.
\]
(b) For each \( n \in \mathbb{N} \), there is a function \( v_n \in C_b(\mathcal{P}(S)) \), with \( v_0 := 0 \) satisfying
\[
\| v_n - v_* \| \leq R \frac{\alpha^n}{1 - \alpha}.
\]
(c) There exists a control policy \( f^* \in \mathcal{F} \) minimizing the right-hand side of (22); i.e.,
\[
v_\ast (m) = \sup_{\mu \in \Gamma} \{ r(m, f^*) + \alpha v_\ast [G(m, f^*, \mu)] \};
\]
and furthermore \( \pi^* = \{ f^* \} \in \Pi_M \) becomes a mean field-minimax policy; that is,
\[
v_\ast (m) = \sup_{\tilde{\pi} \in \Pi_T} v(m, f^*, \tilde{\pi}).
\]

Observe that the existence of the minimax policy \( \pi^* \) does not depend on \( N \).

**Mean field convergence.** Under the assumption of \( M^N(0) = m(0) = m \in \mathcal{P}_N(S) \), for any \( N \in \mathbb{N}, m \in \mathcal{P}_N(S) \) and \( T \in \mathbb{N} \), we can use the same arguments in [13, pp. 217-219], to obtain the estimate
\[
sup_{(\pi, \tilde{\pi}) \in \Pi_M \times \Pi_T} P_{m, \tilde{\pi}} \left( \sup_{0 \leq t \leq T} \| M^N(t) - m(t) \|_{\infty} > \gamma_T(\varepsilon) \right) \leq K T e^{-\lambda N \varepsilon^2}, \quad (23)
\]
where \( \varepsilon > 0, \gamma_T(\varepsilon) \to 0 \) as \( \varepsilon \to 0 \) and \( T \to \infty \).

As a consequence, for each \( m \in \mathcal{P}_N(S) \), \( (\pi, \tilde{\pi}) \in \Pi_M \times \Pi_T, T \in \mathbb{N} \), and \( \varepsilon > 0 \), we can deduce
\[
sup_{(\pi, \tilde{\pi}) \in \Pi_M \times \Pi_T} E_{m, \tilde{\pi}}^n \left( \sup_{0 \leq t \leq T} \| M^N(t) - m(t) \|_{\infty} \right) \leq K T e^{-\lambda N \varepsilon^2} + \gamma_T(\varepsilon). \quad (24)
\]
Indeed, if we call \( Y_T := \sup_{0 \leq t \leq T} \| M^N(t) - m(t) \|_{\infty} \), then for any \( (\pi, \tilde{\pi}) \in \Pi_M \times \pi_T \),
\[
E_{m, \tilde{\pi}}^n [Y_T] = E_{m, \tilde{\pi}}^n [Y_T 1_{\{Y_T > \gamma_T(\varepsilon)\}} + Y_T 1_{\{Y_T \leq \gamma_T(\varepsilon)\}}]\n\leq P_{m, \tilde{\pi}}(Y_T > \gamma_T(\varepsilon)) + \gamma_T(\varepsilon) P_{m, \tilde{\pi}}(Y_T \leq \gamma_T(\varepsilon))\n\leq K T e^{-\lambda N \varepsilon^2} + \gamma_T(\varepsilon)
\]

Under the above estimate it is clear that the MFGAN-\( \mathcal{M} \) can be regarded as the limit of the GAN-\( \mathcal{M}_N \) when \( N \to \infty \).

We have arrived to our main result concerning the existence of asymptotically minimax policies.

**Theorem 3.2.** Under Assumption 2.1, if \( M^N(0) = m(0) = m \in \mathcal{P}_N(S) \), the mean field-minimax control policy \( \pi^* = \{ f^* \} \) associated to the MFGAN model \( \mathcal{M} \) is asymptotically minimax for the N-GAN model \( \mathcal{M}_N \) as \( N \to \infty \); that is
\[
\lim_{N \to \infty} \Phi^N(m, f^*) = 0, \quad m \in \mathcal{P}_N(S),
\]
where \( \Phi^N \) is the function defined in (18).
Theorem 3.2 establishes that when the controller uses the policy $\pi^*$ to control the original process (5), corresponding to the model $\mathcal{M}_N$, the "minimax deviation" vanishes to zero, as $N \to \infty$; that is, for $N$ sufficiently large, $\pi^*$ is a minimax policy for the $N$–GAN $\mathcal{M}_N$.

4. Examples. In this section we describe some examples in order to identify the elements of our model.

4.1. An investment system. We consider the following investment system which is composed by $N$ "small" investors whose actions have no economic impact on the market prices. Each investor is ranked in $s$ different wealthiness classes, according to their capital. We denote by $S := \{1, 2, \ldots, s\}$ the set of such classes. Assume that there is a central controller (government or public body) who provides a subsidy or imposes a fee $a \in A$ to the investors. In addition, at time $t \in \mathbb{N}_0$, each investor is able to invest his/her money in different risky and risk-free assets, which, depending on the controller’s decision, could cause an investor’s class change at the next stage. We shall suppose that $a \in [-a^*, a^*] = A$ for some $a^* \geq 0$, whose interpretation is as follows:

- $a_t = -a_t$ (if $a_t < 0$) or subsidy of size $a_t$ (if $a_t > 0$), at time $t$.

Let $\xi_n^t$ be a random variable with values on a set $Z$, representing the return rate related with the risky asset for investor $n$ at time $t$, and assume that $\xi_n^t$ has distribution $\gamma_{\mu_t}$, independent of $n$, that is,

$$\gamma_{\mu_t}(B) = P[\xi_n^t \in B], \quad \forall n = 1, 2, \ldots, N, \quad B \in \mathcal{B}(Z),$$

where $\{\mu_t\} \subset \Gamma$ is a sequence of unknown parameters. Clearly the investors evolve among the classes according to a transition law determined by distributions $\gamma_{\mu_t}$. In this case, we suppose that there exists a function $\Psi : S^2 \times A \times \mathcal{P}(Z) \to [0, 1]$ such that

$$K_{ij}(a, \mu) = \Psi(i, j, a, \gamma_{\mu_t}).$$

and $a \mapsto \Psi(i, j, a, \gamma_{\mu_t})$ is continuous. Finally, the one-stage cost can be taken as an arbitrary measurable function satisfying Assumption 2.1(d).

4.2. Financial strategy of a company. A company that sells a certain product of brand $\mathcal{L}$ is interested in applying marketing techniques to a group of $N$ families to improve its income. Each family has their own perception about the quality of the product, which is determined by several factors such as price, relative quality, appearance, presentation, among others. Of course, this perception is difficult to measure by the company, which makes it an uncertain entity. Instead, the company can only observe and accurately determine when each family is in one of the following 3 classes: $S = \{c_1, c_2, c_3\}$, where

- $c_1$ := the family does not buy the product;
- $c_2$ := the family buys the product of brand $\mathcal{L}$;
- $c_3$ := the family buys the product of a brand different to $\mathcal{L}$.

Let $\xi_n^t$ be the random variable representing the perception of the family $n$ about brand $\mathcal{L}$ at time $t$. Thus, it is reasonable to assume that $\xi_n^t$ has a distribution depending on an unknown and nonobservable parameter by the company, denoted by $\mu$.

The company wants to implement new policies to make a greater number of families to buy its brand $\mathcal{L}$; i.e., families that do not buy its product start do it,
but also making sure that families who regularly buy the product keep it buying. For this, the company has a set of marketing strategies $A = \{a_1, \cdots, a_k\}$. Let $X_n^N(t) \in S$ be the class of family $n$ at time $t$. Then the transition among classes is modeled by a transition $K_{ij}(a, \mu), i, j \in S, a \in A, \mu \in \Gamma$ which depends on the parameter $\mu$ related with the perception of families about the quality of product.

Finally, we can assume a one-stage reward of the form

$$r(m, a) = m_2 u(a), \quad m = (m_1, m_2, m_3) \in \mathbb{P}(S), a \in A,$$

where $m_2$ is the proportion of products of brand $L$ sold by the company, and $u : A \to \mathbb{R}$ is a suitable utility function satisfying Assumption 2.1(d).

5. Proofs.

5.1. Proof of Lemma 2.2. From the compactness of set $A$, let $\{a_k\} \in A$ be a sequence such that $a_k \to a^* \in A$. Note that Assumption 2.1(c) yields that the mapping $a \mapsto \Psi_{ij}^\mu(a)$ (with $\Psi_{ij}^\mu(\cdot)$ defined as in (7)) is continuous for all $i, j \in S$, $\mu \in \Gamma$. Hence, $\Delta_{ij}^\mu(a_k) \to \Delta_{ij}^\mu(a^*), a_k \to \infty$, in the set sense.

Now, for $i, j \in S$, $w \in [0, 1]$, $\mu \in \Gamma$, and $a \in A$, let $\delta_w(\Delta_{ij}^\mu(a))$ be the Dirac measure corresponding to the indicator function $I(\Delta_{ij}^\mu(a))$. Then, since $\delta_w(\cdot)$ is a probability measure (therefore it is continuous), we conclude that

$$\delta_w(\Delta_{ij}^\mu(a_k)) \to \delta_w(\Delta_{ij}^\mu(a^*)), \quad a_k \to \infty.$$

Hence, from (6), we obtain the continuity of the function $G^N$ defined in (8).

5.2. Proof of Theorem 3.2. The proof of Theorem 3.2 will be consequence of the following facts.

For each $(\pi, \pi) \in \Pi_M \times \Pi_R$, $m \in \mathbb{P}_N(S) \subset \mathbb{P}(S)$, and $T \in \mathbb{N}$, we define the discounted criteria with finite horizon $T$ as:

$$V_T^N(\pi, \pi, m) := E_m^\pi,\pi \sum_{k=0}^{T-1} \alpha^k r(M^N(k), a_k), \quad v_T(\pi, \pi, m) := \sum_{k=0}^{T-1} \alpha^k r(m(k), a_k).$$

We will use the following notation. For any fixed policy $\pi = \{f_t\} \in \Pi_M$, we denote

$$a^{\pi, N}_t := f_t(M^N(t)) \quad \text{and} \quad a^{\pi}_t := f_t(m(t))$$

the actions at time $t$ corresponding to the application of the policy $\pi$ under the process $\{M^N(t)\}$ and $\{m(t)\}$, respectively.

Lemma 5.1. Suppose that $M^N(0) = m(0) = m \in \mathbb{P}_N(S)$, and let $L_\varepsilon$ and $R$ be the constants in Assumption 2.1(d). Then, under Assumption 2.1, for each $m \in \mathbb{P}_N(S)$, $\varepsilon > 0$, $T \in \mathbb{N}$, $0 \leq t \leq T$, and for all $(\varphi, \tilde{\varphi}) \in \Pi_M \times \Pi_R$, the following statements hold true:

\begin{enumerate}
\item[(a)]

\begin{align}
E_m^{\varphi, \tilde{\varphi}} \left[ \sup_{(\pi, \tilde{\pi}) \in \Pi_M \times \Pi_R} \left| V_T^N(\pi, \tilde{\pi}, M^N(t)) - v_T(\pi, \tilde{\pi}, m(t)) \right| \right] & \leq L_\varepsilon \frac{1 - \alpha^T}{1 - \alpha} \\
\times K T e^{-\lambda N \varepsilon^2} + \gamma_T(\varepsilon) \right) ;
\end{align}

\item[(b)]

\begin{align}
E_m^{\varphi, \tilde{\varphi}} \left[ \sup_{(\pi, \tilde{\pi}) \in \Pi_M \times \Pi_R} \left| V_T^N(\pi, \tilde{\pi}, M^N(t)) - V_T^N(\pi, \tilde{\pi}, M^N(0)) \right| \right] & \leq \frac{R \alpha^T}{1 - \alpha} ;
\end{align}
\end{enumerate}
Hence, (27) follows by taking expectation

\[ E_m^{\varphi, \tilde{\varphi}} \left[ \sup_{(\pi, \tilde{\pi}) \in \Pi_M \times \Pi_N} \left| v(\pi, \tilde{\pi}, m(t)) - v_T(\pi, \tilde{\pi}, m(t)) \right| \right] \leq \frac{R\alpha^T}{1 - \alpha} \quad (28) \]

Proof. (a) First observe that by Assumption 2.1(d) and (24)

\[ E_m^{\varphi, \tilde{\varphi}} \left| V_N^T (M^N (t)) - v_\ast (m(t)) \right| \leq \frac{2R\alpha^T}{1 - \alpha} + L_r \frac{1 - \alpha^T}{1 - \alpha} \left[ KT e^{-\lambda T \varepsilon^2} + \gamma_T (\varepsilon) \right] \quad (29) \]

Now, for each pair of policies \((\varphi, \tilde{\varphi}) \in \Pi_M \times \Pi_N\), let us denote the corresponding trajectories as \(M^N (t) := M_{\varphi, \tilde{\varphi}}^N (t)\) and \(m(t) := m_{\varphi, \tilde{\varphi}} (t)\). Then we have

\begin{align*}
&\sup_{(\pi, \tilde{\pi}) \in \Pi_M \times \Pi_N} \left| V_T (\pi, \tilde{\pi}, M^N (t)) - v_T (\pi, \tilde{\pi}, m(t)) \right| \\
&= \sup_{(\pi, \tilde{\pi}) \in \Pi_M \times \Pi_N} \left| E_{M^N (t)}^{\pi, \tilde{\pi}} \left[ \sum_{k=0}^{T-1} \alpha^k r(M^N (k), a_k^\pi) - \sum_{k=0}^{T-1} \alpha^k r(M^N (k), a_k^\tilde{\pi}) \right] \right| \\
&\leq \sum_{k=0}^{T-1} \alpha^k \sup_{(\pi, \tilde{\pi}) \in \Pi_M \times \Pi_N} \left| E_{M^N (t)}^{\pi, \tilde{\pi}} \left[ r(M^N (k), a_k^\pi) - r(M^N (k), a_k^\tilde{\pi}) \right] \right| \\
&\leq L_r \frac{1 - \alpha^T}{1 - \alpha} \left[ KT e^{-\lambda T \varepsilon^2} + \gamma_T (\varepsilon) \right] \quad P_m^{\varphi, \tilde{\varphi}} - \text{almost surely,} \quad (31)
\end{align*}

where the last inequality follows from (30). Hence (26) follows by taking expectation \(E_m^{\varphi, \tilde{\varphi}} (\cdot)\) in both sides of (31).

(b) Again, for each pair of policies \((\varphi, \tilde{\varphi}) \in \Pi_M \times \Pi_N\), let us denote the corresponding trajectories as \(M^N (t) := M_{\varphi, \tilde{\varphi}}^N (t)\) and \(m(t) := m_{\varphi, \tilde{\varphi}} (t)\). Hence, we obtain

\begin{align*}
&\sup_{(\pi, \tilde{\pi}) \in \Pi_M \times \Pi_N} \left| V_N^T (\pi, \tilde{\pi}, M^N (t)) - V_N^T (\pi, \tilde{\pi}, M^N (t)) \right| \\
&\leq \sup_{(\pi, \tilde{\pi}) \in \Pi_M \times \Pi_N} \left| E_{M^N (t)}^{\pi, \tilde{\pi}} \left[ \sum_{k=0}^{\infty} \alpha^k r(M^N (k), a_k^\pi) - \sum_{k=0}^{T-1} \alpha^k r(M^N (k), a_k^\pi) \right] \right| \\
&\leq \sum_{k=T}^{\infty} \alpha^k \sup_{(\pi, \tilde{\pi}) \in \Pi_M \times \Pi_N} \left| E_{M^N (t)}^{\pi, \tilde{\pi}} \left[ r(M^N (k), a_k^\pi) \right] \right| \\
&\leq R \sum_{k=T}^{\infty} \alpha^k \leq \frac{R\alpha^T}{1 - \alpha} \quad P_m^{\varphi, \tilde{\varphi}} - \text{almost surely,} \quad (32)
\end{align*}

Hence, (27) follows by taking expectation \(E_m^{\varphi, \tilde{\varphi}} (\cdot)\) in both sides of (32). Similar arguments prove part (c).
(d) We define the stationary policy \( \tilde{\pi} = \{ \tilde{f} \} \in \Pi_M \), where \( \tilde{f} : \mathbb{P}(S) \rightarrow A \) is given by
\[
\tilde{f}(m) = f^N_* (m) I_{\mathbb{P}_N(S)}(m) + \tilde{f}(m) I_{\mathbb{P}_N(S)}^\dagger (m),
\]
with \( \pi_*^N = \{ f_*^N \} \in \Pi_M^N \) being a minimax stationary policy for the \( N \)-GAN \( M_N \) (see Theorem 2.3), and \( \tilde{f} \in \mathbb{F} \), an arbitrary selector. Let \( M^N(t) := M^N_{\tilde{\varphi}, \tilde{\xi}}(t) \) and \( m(t) := m_{\tilde{\varphi}, \tilde{\xi}}(t) \) as before. Observe that for each \( t \in \mathbb{N}_0 \) and \( \tilde{\pi} \in \Pi_G \),
\[
V^N_*(M^N(t)) := \sup_{\tilde{\pi} \in \Pi_G} V^N(\pi^*, \tilde{\pi}, M^N(t)) = \sup_{\tilde{\pi} \in \Pi_G} V^N(\tilde{\pi}, \tilde{\pi}, M^N(t))
\]
Hence,
\[
V^N_*(M^N(t)) - v_*(m(t)) \leq \sup_{(\pi, \tilde{\pi}) \in \Pi_M \times \Pi_G} \left| V^N(\pi, \tilde{\pi}, M^N(t)) - v(\pi, \tilde{\pi}, m(t)) \right|
\]
which gives
\[
\left| V^N_*(M^N(t)) - v_*(m(t)) \right| \leq \sup_{(\pi, \tilde{\pi}) \in \Pi_M \times \Pi_G} \left| V^N(\pi, \tilde{\pi}, M^N(t)) - v(\pi, \tilde{\pi}, m(t)) \right|, \quad t \in \mathbb{N}_0.
\]
Therefore, for each \( m \in \mathbb{P}_N(S), 0 \leq t < T, \)
\[
\left| V^N_*(M^N(t)) - v_*(m(t)) \right| \leq \sup_{(\pi, \tilde{\pi}) \in \Pi_M \times \Pi_G} \left\{ \left| V^N(\pi, \tilde{\pi}, M^N(t)) - V^N_*(\pi, \tilde{\pi}, M^N(t)) \right| + \left| V^N_*(\pi, \tilde{\pi}, M^N(t)) - v_*(\pi, \tilde{\pi}, m(t)) \right| \right. \right.
\]
\[
+ \left| v_*(\pi, \tilde{\pi}, m(t)) - v(\pi, \tilde{\pi}, m(t)) \right| \right\} \leq \sup_{(\pi, \tilde{\pi}) \in \Pi_M \times \Pi_G} \left| V^N(\pi, \tilde{\pi}, M^N(t)) - V^N_*(\pi, \tilde{\pi}, M^N(t)) \right| + \sup_{(\pi, \tilde{\pi}) \in \Pi_M \times \Pi_G} \left| V^N_*(\pi, \tilde{\pi}, M^N(t)) - v_*(\pi, \tilde{\pi}, m(t)) \right| \right.
\]
\[
+ \sup_{(\pi, \tilde{\pi}) \in \Pi_M \times \Pi_G} \left| v_*(\pi, \tilde{\pi}, m(t)) - v(\pi, \tilde{\pi}, m(t)) \right| \leq 2\alpha T \frac{1}{1-\alpha} + L_\gamma \frac{1}{1-\alpha} \left[ \lambda N e^{-\lambda N \epsilon^2} + \gamma T (\epsilon) \right] P_{\tilde{\varphi}, \tilde{\xi}}^\dagger -\text{almost surely}. \tag{33}
\]
where the last inequality is due to parts (a)-(c). Finally, taking expectation \( E_m^\varphi \tilde{\varphi} (\cdot) \) in both sides of (33), we can easily get (29).

\[\square\]

Remark 2. Observe that from (5) and (19), we have:
\[
M^N(1) = G^N(M^N(0), a_0, \mu_0, \bar{w}_0) := M^N_{\mu_0}(1)
\]
and
\[
m(1) = G(m(0), a_0, \mu_0) := m_{\mu_0}(1).
\]
Hence, for each \( m \in \mathbb{P}_N(S) \subset \mathbb{P}(S), \epsilon > 0, T \in \mathbb{N} \), and for all \( (\varphi, \tilde{\varphi}) \in \Pi_M \times \Pi_G \), the relation in (29) implies
\[
\sup_{\mu \in \Pi_G} E_m^\varphi \tilde{\varphi} \left| V^N_*(M^N_{\mu}(1)) - v_*(m_{\mu}(1)) \right| \leq 2\alpha T \frac{1}{1-\alpha}
\]
\[
+ L_\gamma \frac{1}{1-\alpha} \left[ \lambda N e^{-\lambda N \epsilon^2} + \gamma T (\epsilon) \right]. \tag{34}
\]
Proof of Theorem 3.2

Proof. Let $\Phi : \mathbb{P}(S) \times A \rightarrow \mathbb{R}_+$ be the function defined as

$$
\Phi (m, a) := \sup_{\mu \in \Gamma} \{ r(m, a) + \alpha v_* [G(m, a, \mu)] \} - v_* (m)
$$

(compare with (18)). Note that, from the definition of the mean field-minimax policy $\pi^* = \{ f^* \}$ (see Theorem 3.1), we have $\Phi (m, f^*) = 0$.

Let $\tilde{\pi} \in \Pi_\Gamma$ be an arbitrary policy for the opponent (the nature) with $\mu_0 = \mu \in \Gamma$. Then, for each $M_N(0) = m(0) = m \in \mathbb{P}_N(S) \subset \mathbb{P}(S)$, from the Markov property we have

$$
\int_{\mathbb{R}^N} V_*^N \left[ G^N (m, f^*, \mu, w) \right] \theta (dw) - v_* [G (m, f^*, \mu)]
$$

$$
= E_{m, \tilde{\pi}}^* \left[ V_*^N (M_N (1)) - v_* (m (1)) | h_0^N \right],
$$

which in turns implies

$$
\left| \int_{\mathbb{R}^N} V_*^N \left[ G^N (m, f^*, \mu, w) \right] \theta (dw) - v_* [G (m, f^*, \mu)] \right|
$$

$$
\leq E_{m, \tilde{\pi}}^* \left| V_*^N (M_N (1)) - v_* (m (1)) \right|. \quad (36)
$$

This yields

$$
\Phi^N (m, f^*) = |\Phi^N (m, f^*) - \Phi (m, f^*)|
$$

$$
= \left| \sup_{\mu \in \Gamma} \left\{ r(m, f^*) + \alpha \int_{\mathbb{R}^N} V_*^N \left[ G^N (m, f^*, \mu, w) \right] \theta (dw) \right\} - V_*^N (m) \right|
$$

$$
- \sup_{\mu \in \Gamma} \left\{ r(m, f^*) + \alpha v_* [G (m, f^*, \mu)] \right\} - v_* (m) \right|
$$

$$
\leq |V_*^N (m) - v_* (m)|
$$

$$
+ \alpha \sup_{\mu \in \Gamma} \left| \int_{\mathbb{R}^N} V_*^N \left[ G^N (m, f^*, \mu, w) \right] \theta (dw) - v_* [G (m, f^*, \mu)] \right|
$$

$$
\leq |V_*^N (m) - v_* (m)|
$$

$$
+ \alpha \sup_{\mu \in \Gamma} E_{m, \tilde{\pi}}^* \left| V_*^N (M_N (1)) - v_* (m (1)) \right|,
$$

where the last inequality is due to (36). Using this last fact, together with relation (34) and Lemma 5.1, we can deduce

$$
0 \leq \Phi^N (m, f^*) \leq \frac{4R \alpha T}{1 - \alpha} + 2\epsilon \frac{1 - \alpha T}{1 - \alpha} \left[ K Te^{-\lambda N \epsilon^2} + \gamma_T (\epsilon) \right] \quad \forall \epsilon > 0, \ T \in \mathbb{N}.
$$

Therefore, letting $N \rightarrow \infty$ and hence $T \rightarrow \infty$ (recall that $T$ was taken arbitrary), we prove Theorem 3.2; that is

$$
\lim_{N \rightarrow \infty} \Phi^N (m, f^*) = 0.
$$

\[\square\]
REFERENCES

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

[2] M. Aoki, New macroeconomic modeling approaches: Hierarchical dynamics and mean field approximation, *J. Econ. Dyn. Control*, 18 (1994), 865–877.

[3] A. Bensoussan, J. Frehse and P. Yam, *Mean Field Games and Mean Field Control Theory*, Springer Briefs in Mathematics, New York, 2013.

[4] D. P. Bertsekas, *Dynamic Programming: Deterministic and Stochastic Models*, Prentice-Hall, Englewood Cliffs, N.J., 1987.

[5] A. Budhiraja, P. Dupuis and M. Fischer, Large deviation properties of weakly interacting processes via weak convergence methods, *Annals Probab.*, 40 (2012), 74–102.

[6] R. Carmona, J. P. Fouque and D. Vestal, Interacting particle systems for the computation of rare credit portfolio losses, *Finance Stoch.*, 13 (2009), 613–633.

[7] S. P. Coraluppi and S. I. Marcus, Mixed risk-neutral/minimax control of discrete-time finite state Markov decision processes, *IEEE Trans. Autom. Control*, 45 (2000), 528–532.

[8] E. B. Dynkin and A. A. Yushkevich, *Controlled Markov Processes*, Springer-Verlag, New York, 1979.

[9] N. Gast and B. Gaujal, A mean field approach for optimization in discrete time, *Discrete Event Dyn. Syst.*, 21 (2011), 63–101.

[10] N. Gast, N. B. Gaujal and J. Y. Le Boudec, Mean field for Markov decision processes: From discrete to continuous optimization, *IEEE Trans. Autom. Control*, 57 (2012), 2266–2280.

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

[12] T. J. González-Trejo, O. Hernández-Lerma and L. F. Hoyos-Reyes, Minimax control of discrete-time stochastic systems, *SIAM J. Control Optim.*, 41 (2002), 1626–1659.

[13] C. G. Higuera-Chan, H. Jasso-Fuentes and J. A. Minjárez-Sosa, Discrete-time control for systems of interacting objects with unknown random disturbance distributions: a mean field approach, *Appl. Math. Optim.*, 74 (2016), 197–227.

[14] M. Huang, Large-population LQG games involving a major player: The Nash certainty equivalence principle, *SIAM J. Control Optim.*, 48 (2010), 3318–3353.

[15] H. Jasso-Fuentes and J. D. López-Barrientos, On the use of stochastic differential games against nature to ergodic control problems with unknown parameters, *Internat. J. Control*, 88 (2015), 897–909.

[16] V. N. Kolokoltsov, M. Troeva and W. Yang, On the rate of convergence for the mean-field approximation of controlled diffusions with large number of players, *Dyn. Games Appl.* 4 (2014), 208–230.

[17] M. Kurano, Minimax strategies for average cost stochastic games with an application to inventory models, *J. Oper. Res. Soc. Jpn.*, 30 (1987), 232–247.

[18] A. Lachapelle, J. Salomon and G. Turinici, Computation of mean field equilibria in economics, *Math. Models Meth. Appl. Sci.*, 20 (2010), 567–588.

[19] J. M. Lasry and P. L. Lions, Mean field games, *Jap. J.Math.*, 2 (2007), 229–260.

[20] J. Y. Le Boudec, D. McDonald and J. Mundinger, A generic mean field convergence result for systems of interacting objects, *4th Int. Conf. Quantitative Evaluation of Systems*, (2007).

[21] J. D. López-Barrientos, H. Jasso-Fuentes and B. A. Escobedo Trujillo, Discounted robust control for Markov diffusion processes, *TOP.*, 23 (2015), 53–76.

[22] F. Luque-Vásquez, J. A. Minjárez-Sosa and L. C. Rosas-Rosas, Semi-Markov control models with partially known holding times distribution: Discounted and Average criteria, *Appl. Math.*, 114 (2011), 135–156.

[23] N. Peyrard and R. Sabbadin, Mean field approximation of the policy iteration algorithm for graph-based Markov decision processes, Biewka, G., Coradeschi, S., Perini, A., Traverso,
[24] A. Schield, Robust optimal control for a consumption-investment problem, Math. Methods Oper. Res., 67 (2008), 1–20.

Received December 2015; revised December 2016.

E-mail address: aminjare@gauss.mat.uson.mx
E-mail address: hjasso@math.cinvestav.mx
E-mail address: carhiguera@mat.uson.mx