Pure and Applied Functional Analysis
Volume 9, Number 3, 2024, 763–782

Yokohama Publishers
ISSN 2189-3764 ONLINE JOURNAL

AN ABELIAN THEOREM FOR A MARKOV DECISION PROCESS IN A SYSTEM OF INTERACTING OBJECTS WITH UNKNOWN RANDOM DISTURBANCE LAW

JOSÉ DANIEL LÓPEZ-BARRIENTOS, JOSÉ MANUEL MENDOZA-MADRID, AND PAOLA FRINÉ GONZÁLEZ-VEGA

Abstract. This paper studies a mean-field approach for Markov decision processes in a class of systems of a large number of objects that interact with each other according to an observable –but unknown- law for the central controller. The central controller acts under the ergodic cost criterion with Borel state and control spaces, bounded costs, and compact action space. We depart from the characterization of the discounted optimal strategies, and then, by means of an Abelian theorem, we study the existence of average cost optimal stationary policies in the original model. We also analyze the performance of the mean-field limit optimal policies in the original model.

1. Introduction

The aim of this paper is to study a discrete-time optimal control problem where the central agent tries to minimize the long-run average cost incurred by a system of a large number of interacting objects with an unknown noise law. To this end, we use the mean-field approach in the fashion of [10] and [11]. The main idea is to use [28] to extend the results in [18] to the ergodic cost criterion, and then discuss the robustness of the approximation scheme. Indeed, Higuera-Chan, Jasso-Fuentes, and Minjárez-Sosa have already provided the framework to study the class of problems we consider in this research. Our contribution lies in the use of an Abelian theorem to justify the so-called vanishing discount technique and thus establish the connection between the average and the α-discounted performance criteria. In particular, we propose two algorithms to effectively compute the dynamics of the stochastic model in a large dimension; and find the optimal control policies and the corresponding ergodic value. This fills the existing gap in the basic optimality criteria for the class of control problems of our interest and will pave the way to study (i) more advanced criteria, such as bias and overtaking optimality for both: the controlled problems, and the problem with at least two central agents, and (ii) study the problem of stable cooperation among the agents on an infinite horizon. We will use the terms "ergodic" and "average" interchangeably to avoid repetition.

2020 *Mathematics Subject Classification.* 93E20, 49J55, 46B09.
Key words and phrases. Abelian theorems, discounted and ergodic performance criteria, mean-field theory, Robustness of estimation.

In 1949, G.H. Hardy proved that if at least one of the so-called *Cesàro and Abel limits* exists for a continuous and bounded function c, then the other also exists and they coincide (see [12, Chapters 7.5 and 7.6]). There are numerous uses for this conclusion and its generalizations. In this paper, we consider an Abelian theorem for the asymptotics of optimal values in the statement of an optimal control problem. We intend to optimize the Cesàro mean so that then we move on to the limit of the optimal values corresponding to the Abel mean. It is possible to trace back the first use of this approach to [2] for a stochastic problem. The version of the Abelian theorem we use was taken from Lemma 5.6(a) in [15].

Theorem 1.1. *Let $(c_t : t = 0, 1, ...)$ be a sequence of Real numbers bounded below. Then*

$$\begin{aligned}\liminf_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T-1}c_t &\leq \liminf_{\alpha\uparrow 1}(1-\alpha)\sum_{t=1}^{\infty}\alpha^t c_t\\ &\leq \limsup_{\alpha\uparrow 1}(1-\alpha)\sum_{t=1}^{\infty}\alpha^t c_t \leq \limsup_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T-1}c_t.\end{aligned}$$

We intend to use Theorem 1.1 to compute the expected average optimal value of the so-called *N-Markov decision process* as the limit, when $\alpha \uparrow 1$, of the expected α-discounted optimal value function for the same process.

One of the difficulties we overcome in our developments is the fact that studying the deterministic optimal control problem under the ergodic cost criterion remains an open problem. We have managed to prove the existence of a solution to the average cost inequality arising from the use of standard dynamic programming techniques and our very particular set of hypotheses.

1.1. **Mathematical preliminaries and nomenclature.** A Borel space is a Borel-measurable subset of a Polish space. For a Borel space Z, we denote the corresponding metric as d_Z, and $\mathcal{B}(Z)$ stands for the Borel-σ-algebra. In this context, the term "measurable" for sets and functions, refers to "Borel-measurable".

We consider the class of Real-valued bounded functions on Z endowed with the supremum norm $\|h\| := \sup_{z\in Z}|h(z)|$, and the subspace of all Real-valued bounded and continuous functions on Z. The symbol $\mathbb{P}(Z)$ stands for the set of all probability measures on Z.

If $Z = \{z_1, z_2, ..., z_n\}$, the vector $p := (p(z_1), p(z_2), ..., p(z_n))$ such that $\sum_{i=1}^{n} p(z_i) = 1$ and $p(z_i) \geq 0$ for $i = 1, ..., n$ denotes a probability measure $p \in \mathbb{P}(Z)$. As usual, $\|\cdot\|_\infty$ stands for the norm in $\mathcal{L}_\infty(\mathbb{P}(Z))$, *i.e.*, for every $p \in \mathbb{P}(Z)$:

$$\|p\|_\infty := \max\{|p(z_1)|, |p(z_2)|, ..., |p(z_n)|\}.$$

Let Z and Y be Borel spaces. We define a stochastic kernel $Q(\cdot|\cdot)$ on Z given Y as a function $Q : \mathcal{B}(Z) \times Y \to [0, 1]$, such that $Q(\cdot|y) \in \mathbb{P}(Z)$ for each $y \in Y$, and $Q(D|\cdot)$ is a measurable function on Y for each $D \in \mathcal{B}(Z)$.

Lastly, we use $\mathbb{I}_D$ to denote the indicator function of the set D, $\mathbb{N}$ is the set of positive integers, $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$, $\mathbb{N}_N := \{1, 2, ..., N\}$ for $N \in \mathbb{N}$, and $\mathbb{R}$ denotes the set of Real numbers.

The remainder of the paper is organized as follows. In the next section, we state the features of the system that we intend to control with the ergodic criterion.

Then, in section 3, we follow [1, 7, 25, 27] to define a suitable topology for the space of Markov controllers, summarize the results we borrowed from [18, 19, 17, 20] to ensure discounted optimality in the stochastic model with N interacting objects, and use the Abelian Theorem 1.1 to ensure ergodic optimality in the fashion of [28]. Since these models are posed in a large dimension, section 4 proposes a deterministic control model that implicitly depends on the unknown disturbance law and presents an analysis of the performance of its optimal control policies in the original problem. We enlist our conclusions and further questions in section 5.

For the sake of brevity, the paper is purposefully illustrations-free. We are currently working on some examples applied to the context of the insurance industry. These will be published in our forthcoming research (see the concluding remarks in section 5).

2. A stochastic model on a set of measures

We investigate a discrete-time Markov decision process in the system containing a large number of interconnected items. There are N interacting objects, each of which can be categorized into a small number of different classes. Let $S := \{1, 2, ..., s\}$ be the set of classes, and let $X_n^N(t)$, $n \in \mathbb{N}_N$, $t \in \mathbb{N}_0$ be a random variable defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ representing the class of the n-th object at time $t \geq 0$. Hence, $X_n^N(t) \in S$ for all $n \in \mathbb{N}_N$ and $t \in \mathbb{N}_0$. The process $X_n^N(t)$ is controlled by an agent who chooses an action $u(t)$ at each time $t \in \mathbb{N}_0$, from a given Borel set U. In concrete, the evolution of $\left(X_n^N(t) : t \in \mathbb{N}_0\right)$ is given by:

$$X_n^N(t+1) = F\left(X_n^N(t), u(t), \xi(t)\right), \ t \in \mathbb{N}_0$$

where $F : S \times U \times \mathbb{R} \to S$ is a given (known) function and $(\xi(t) : t \in \mathbb{N}_0)$ is a sequence of independent and identically distributed Real random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$ with a common (but unknown) density ρ.

Next, we describe the evolution of the system. The central controller selects her/his/its action $u \in U$, and a random movement of the objects from class $i \in S$ to class $j \in S$ happens according to a transition probability

$$\begin{aligned} K_{ij}^{\rho}(u) :=& \mathbb{P}\left(X_n^N(t+1) = j | X_n^N(t) = i, u(t) = u\right) \\ &= \int_{\mathbb{R}} \mathbb{I}_j\left(F(i, u, z)\right) \rho(z) \mathrm{d}z, \end{aligned} \tag{2.1}$$

which is homogeneous in N (see (2.2) in [18]). Finally, the agent pays a cost that depends on the proportion of the objects in each state. Note that $K_{ij}^{\rho}(\cdot)$ greatly depends on ρ, which, at this point, is unknown to the controller. However, at each time, it is possible for the central agent to observe the behavior of the objects. We will deal with this in the sequel.

Assumption 2.1. (*i*) *The control space U is a compact metric Borel space. We denote its metric as d_U.*

(*ii*) *The mapping $u \mapsto K_{ij}^{\rho}(u)$ is continuous for all $i, j \in S$.*

To define the proportion we just referred to, we assume that the states of the objects can be observed at all times so that the central controller can determine

only the cardinality of the objects in each state. We define the proportion $M_i^N(t)$ of objects in state $i \in S$ at time t as

$$M_i^N(t) := \frac{1}{N} \sum_{n=1}^{N} \mathbb{I}_{\{X_n^N(t)=i\}} \text{ for } i \in S.$$

Further, $\vec{M}^N(t) := \left(M_1^N(t), M_2^N(t), ..., M_s^N(t)\right)$ is the vector whose components are these proportions.

Recall from section 1.1 that $\mathbb{P}(S)$ stands for the set of all probability measures on S. Let $\mathbb{P}_N(S) := \{p \in \mathbb{P}(S) : Np(i) \in \mathbb{N} \text{ for } i \in S\}$ and note that $\vec{M}^N(t) \in \mathbb{P}_N(S) \subset \mathbb{P}(S)$. Moreover, since for N fixed and all $i \in S$, we have $Np(i) \in \mathbb{N}$, then $p(i)$ is of the form m_i/N with $m_i \in \mathbb{N}$ for $m_i \leq N$, therefore $\mathbb{P}_N(S)$ is a finite set. Thus $\cup_{N=1}^{\infty} \mathbb{P}_N(S)$ is a denumerable set. In fact, $\cup_{N=1}^{\infty} \mathbb{P}_N(S)$ is dense subset of $\mathbb{P}(S)$ with the metric induced by $\|\cdot\|_\infty$. Indeed, for every $p \in \mathbb{P}(S)$, and every $\varepsilon > 0$, there exists $\vec{p} \in \cup_{N=1}^{\infty} \mathbb{P}_N(S)$ such that $\|p - \vec{p}\|_\infty < \varepsilon$. From this fact, we gather that $\mathbb{P}(S)$ is a Borel space with the metric induced by $\|\cdot\|_\infty$ (see [20, Remark 1]). We will use this property of $\mathbb{P}(S)$ in the sequel.

For $i \in S$, $n = 1, ..., NM_i^N(t)$ and $t \in \mathbb{N}_0$, let $w_n^i(t)$ be a uniformly distributed random variable in $[0, 1]$. Now, by the arguments in [11, 18, 19, 20], there exists a Borel-measurable function $G_\rho^N : \mathbb{P}_N(S) \times U \times \mathbb{R}^N \to \mathbb{P}_N(S)$ such that

$$\vec{M}^N(t+1) = G_\rho^N \left(\vec{M}^N(t), u(t), \vec{w}(t)\right), \tag{2.2}$$

where $\left(\vec{w}(t) \in \mathbb{R}^N : t \in \mathbb{N}_0\right)$ is a sequence of independent and identically distributed vectors with common distribution θ with $\vec{w}(t) := \left(w^1(t), ..., w^s(t)\right)$ and $w^i(t) := \left(w_1^i(t), ..., w_{NM_i^N(t)}(t)\right)$ for $i \in S$. This renders the process $\left(\vec{M}^N(t) : t \in \mathbb{N}_0\right)$ a non-homogeneous Markov chain (see [11, 22]).

It is possible to explicitly obtain the function G_ρ^N. To this end, we apply the Monte Carlo Markov chain simulation technique described in [20]. This is the purpose of Algorithm 1.

From Algorithm 1, we define

$$G_\rho^N\left(\vec{m}, u, \vec{w}\right) := \left(G_{\rho,1}^N\left(\vec{m}, u, \vec{w}\right), ..., G_{\rho,s}^N\left(\vec{m}, u, \vec{w}\right)\right) \tag{2.3}$$

for $(\vec{m}, u, \vec{w}) \in \mathbb{P}_N(S) \times U \times [0, 1]^N$, where

$$G_{\rho,j}^N\left(\vec{m}, u, \vec{w}\right) := \frac{1}{N} \sum_{i=1}^{s} \sum_{n=1}^{Nm_i} \mathbb{I}_{\Delta_{ij}(u)}\left(w_n^i\right), \tag{2.4}$$

for $j \in S$, with $\vec{m} = (m_1, ..., m_s)$.

Observe that G_ρ^N defined in (2.3)-(2.4) is the dynamic of the process $\vec{M}^N(t)$ in (2.2). At the same time, $\vec{M}^N(t) \in \mathbb{P}_N(S) \subset \mathbb{P}(S)$ stands for the layout of the system at time t, which depends on the controller's actions –see (2.2)-.

The following result is a consequence of Assumption 2.1 (see also the discussion at Remark 2.2 in [18] and Remark 2.4(b) in [17]). It is the first step to prove the lower semi-continuity of the expected α-discounted cost and the expected average cost defined below. Our proof closely follows the arguments presented in Proposition 1 in [20].

Algorithm 1: Simulation of the dynamic of a generic object

Data: Initial distribution of proportions $\vec{M}^N(0)$ and transition probability function $K^\rho(u) = \left[K^\rho_{ij}(u)\right]$

Result: General form of G^N_ρ

$\vec{w}(0) \leftarrow 0 \in \mathbb{R}^N$;
for $t \in \mathbb{N}_0$ **do**
  for $j \in S$ **do**
    $\left[G^N_\rho\left(\vec{M}^N(t), u, \vec{w}(t)\right)\right]_j \leftarrow M^N_j(t)$
  end
  for $i, j \in S$ *and* $u \in U$ **do**
    $\Psi_{ij}(u) \leftarrow \sum_{\iota=1}^{j-1} K^\rho_{i\iota}(u)$;
    $\Delta_{ij}(u) \leftarrow [\Psi_{ij}(u), \Psi_{i,j+1}(u)] \subseteq [0,1]$; ▷ The symbol $\{\Delta_{ij}(u)\}_{j\in S}$ defines a partition of $[0,1]$ for each $i \in S$. The size of $\Delta_{ij}(u)$ stands for the probability that the object moves from class i to class j when the agent selects the action $u \in U$.
  end
  for $i \in S$ **do**
    for $n = 1, ..., NM^N_i(t)$ **do**
      generate $\upsilon \sim U(0,1)$;
      $w^i_n(t) \leftarrow \upsilon$
    end
  end
  $\vec{w}(t) \leftarrow \left(w^1(t), ..., w^s(t)\right)$; ▷ since $\sum_{i=1}^{s} NM^N_i(t) = N$, then $\vec{w}(t) \in [0,1]^N$.
  for $j \in S$ **do**
    $M^N_j(t+1) \leftarrow \frac{1}{N}\sum_{i=1}^{s}\sum_{n=1}^{NM^N_i(t)} \mathbb{I}_{\Delta_{ij}(u)}\left(w^i_n(t)\right)$
  end
end

Theorem 2.2. *Let Assumption 2.1 hold. For all $\vec{w} \in \mathbb{R}^N$, the mapping $(\vec{m}, u) \mapsto G^N_\rho(\vec{m}, u, \vec{w})$ defined in* (2.3)-(2.4) *is continuous θ-a.s.*

Proof. The fact that $\vec{m} \mapsto G^N_\rho(\vec{m}, \cdot, \cdot)$ is continuous follows from (2.3) and (2.4). Now take a converging sequence $u_k \to u$ as $k \to \infty$. Since by Assumption 2.1(i), U is a compact set, we know that $u \in U$. Moreover, Assumption 2.1(ii) yields that, for all $i, j \in S$,

$$\Psi_{ij}(u_k) \to \Psi_{ij}(u) \text{ as } k \to \infty, \tag{2.5}$$

where Ψ_{ij} is the function defined in line 7 in Algorithm 1. This fact, and line 8 in Algorithm 1 enable us to assert that

$$\Delta_{ij}(u_k) = [\Psi_{ij}(u_k), \Psi_{i,j+1}(u_k)] \to [\Psi_{ij}(u), \Psi_{i,j+1}(u)] = \Delta_{ij}(u) \text{ as } k \to \infty.$$

Let $w \in]\Psi_{ij}(u), \Psi_{i,j+1}(u)[$. Then, (2.5) yields that there exist $K_1, K_2 \in \mathbb{N}$ such that $w < \Psi_{i,j+1}(u_k)$ for all $k > K_1$ and $w > \Psi_{ij}(u_k)$ for all $k > K_2$. Hence, $\mathbb{I}_{]\Psi_{ij}(u_k),\Psi_{i,j+1}(u_k)[}(w) = 1$ for all $k > K = \max(K_1, K_2)$. Hence

$$\left|\mathbb{I}_{]\Psi_{ij}(u_k),\Psi_{i,j+1}(u_k)[}(w) - \mathbb{I}_{]\Psi_{ij}(u),\Psi_{i,j+1}(u)[}(w)\right| = 0 \text{ for all } k > K.$$

Observe that $\Delta_{ij}(u) \setminus]\Psi_{ij}(u), \Psi_{i,j+1}(u)[= \{\Psi_{ij}(u), \Psi_{i,j+1}(u)\} =: \partial\Delta_{ij}(u)$ is a finite set. Therefore, $u \mapsto \mathbb{I}_{\Delta_{ij}(u)}(w)$ is continuous for all $i, j \in S$ and for all $w \in [0,1] \setminus \cup_{j\in S}\partial\Delta_{ij}(u)$. This fact, together with (2.3) and (2.4) imply that $(m,u) \mapsto G_\rho^N(m,u,\cdot)$ is continuous θ-almost surely. This completes the proof. □

To complete the description of the model, we define a cost function $c : \mathbb{P}(S)\times U \to \mathbb{R}$ that depends on the proportion of the objects and the action selected by the agent. Once the agent has selected her/his/its action u, she/he/it incurs in the one-stage cost $c\left(\vec{M}^N, u\right)$. Now we give the hypotheses we use for this function.

Assumption 2.3. *The one-stage cost function c has the following features.*

(*i*) *For some constant L_c, and all $\vec{m}, \vec{m}' \in \mathbb{P}(S)$,*

$$\sup_{u,u'\in U} |c(\vec{m},u) - c(\vec{m}',u')| \quad \leq \quad L_c\|\vec{m} - \vec{m}'\|_\infty.$$

That is, the one-stage cost function is uniformly Lipschitz in the state argument.

(*ii*) *The one-stage cost function is lower semi-continuous on $\mathbb{P}(S)\times U$. That is, for each $\lambda \in \mathbb{R}$, the set $\{(\vec{m},u) \in \mathbb{P}(S)\times U : c(\vec{m},u) \leq \lambda\} \subseteq \mathbb{P}(S)\times U$ is closed.*

Remark 2.4. The finiteness of the state space S implies that $\mathbb{P}(S)$ is compact. Thus, Assumption 2.3(*i*) yields the existence of a constant $R > 0$ such that

$$\sup_{(\vec{m},u)\in\mathbb{P}(S)\times U} |c(\vec{m},u)| \leq R.$$

Moreover, this condition, along with Assumption 2.3(*ii*) imply that for each $\lambda > 0$, the set $\{(\vec{m},u) \in \mathbb{P}(S)\times U : c(\vec{m},u) \leq \lambda\} \subseteq \mathbb{P}(S)\times U$ is compact. That is, the one-stage cost function is inf-compact on $\mathbb{P}(S)\times U$.

3. Formulation of the controlled N-object Markov model

Now we define the discrete-time Markov decision process associated with the N-object system previously introduced (in short, N-MDP), through the following elements:

$$\mathcal{M}_N := \left(\mathbb{P}_N(S), U, G_\rho^N, \theta, c\right). \tag{3.1}$$

This model describes the evolution of the system. At the time $t \in \mathbb{N}_0$, the agent observes the state $\vec{m} = \vec{M}^N(t) \in \mathbb{P}_N(S)$, and then chooses an action $u = u(t) \in U$.

As a consequence, the agent incurs in a cost $c(\vec{m}, u)$, and the system evolves to a new state $\vec{m}' = \vec{M}^N(t+1) \in B$ according to the transition law

$$\begin{aligned} Q_\rho(B|\vec{m}, u) &:= \mathbb{P}\left(\vec{M}^N(t+1) \in B | \vec{M}^N(t) = \vec{m}, u(t) = u\right) \\ &= \int_{[0,1]^N} \mathbb{I}_B\left(G_\rho^N(\vec{m}, u, \vec{w})\right) \theta(\mathrm{d}\vec{w}), \end{aligned} \tag{3.2}$$

with G_ρ^N as in (2.3)-(2.4). Then, the process repeats itself and the one-stage costs for the agent are accumulated by means of an expected cost criterion. The ultimate goal of the central agent is to minimize the expected average cost referred to in Definition 3.6 below.

The control policies are the actions taken by the central agent. We define them by letting $\mathbb{H}_0^N := \mathbb{P}_N(S)$, now, for $t = 1, 2, ...$, we denote the *space of histories up to time* t as

$$\mathbb{H}_t^N := \left(\mathbb{P}_N(S) \times U \times \mathbb{R} \times \mathbb{R}^N\right)^t \times \mathbb{P}_N(S).$$

With this in mind, we let $h_0^N := \left(\vec{M}^N(0)\right) \in \mathbb{H}_0^N$,

$$h_1^N := \left(\vec{M}^N(0), u(0), \vec{M}^N(1)\right) \in \mathbb{H}_1^N,$$

and in general, for $t = 1, 2, ...$, we define

$$h_t^N := \left(h_{t-1}^N, u(t-1), \vec{M}^N(t)\right) \in \mathbb{H}_t^N.$$

Recall subsection 1.1. Let us denote the Borel σ-algebra of U as $\mathcal{B}(U)$. Further, let $\mathbb{P}(U)$ represent the family of all probability measures on U endowed with the topology of weak convergence. For technical reasons, we will consider the so-called *randomized control policies* defined as follows.

Definition 3.1. We will say that a *randomized control policiy* is a sequence $\pi^N := \left(\pi_t^N : t \in \mathbb{N}_0\right)$ of stochastic kernels π_t^N on U given $\mathbb{H}_t^N$. That is:

- **(a):** for each $h_t^N \in \mathbb{H}_t^N$, and $t \in \mathbb{N}_0$, $\pi_t^N(\cdot|h_t^N)$ is a probability measure on U, so that $\pi_t^N(U|h_t^N) = 1$ for all $h_t^N \in \mathbb{H}_t^N$, and $t \in \mathbb{N}_0$;
- **(b):** for each $D \in \mathcal{B}(U)$ and $t \geq 0$, $\pi_t^N(D|\cdot, \cdot)$ is a Borel function on $\mathbb{H}_t^N$; and
- **(c):** for each $B \in \mathcal{B}(U)$, $h_t^N \in \mathbb{H}_t^N$, the mapping $t \mapsto \pi_t^N(B|h_t^N)$ is Borel-measurable.

The symbol Π^N denotes the *set of admissible control policies*.

Let $\Omega' := (\mathbb{P}_N(S) \times U)^\infty$, and $\mathcal{F}'$ be a σ-algebra of events of Ω'. If $\pi^N \in \Pi^N$ is a randomized control policy and $\vec{M}(0) = \vec{m} \in \mathbb{P}_N(S)$ is an initial state of the stochastic process $\left(\vec{M}(t) : t = 0, 1, ...\right)$, the theorem of Ionescu-Tulcea (see [4, Chapter 5.4], [13, Proposition C.10 and Remark C.11] and [24, Proposition V.1.1]) yields the existence of a unique probability measure $\mathbb{P}_{\vec{m}}^{\pi^N}$ on $(\Omega', \mathcal{F}')$ such that, for all $t \in \mathbb{N}_0$,

- $\mathbb{P}_{\vec{m}}^{\pi^N}(\vec{m} \in B) = \delta_{\vec{m}}(B)$, $B \in \mathcal{B}\left(\mathbb{P}_N(S)\right)$,
- $\mathbb{P}_{\vec{m}}^{\pi^N}\left(u(t) \in C | h_t^N\right) = \pi_t^N\left(C|h_t^N\right)$ for $C \in \mathcal{B}(U)$,

- a Markov-like property holds:

$$\mathbb{P}_{\vec{m}}^{\pi^N}\left(\vec{M}(t+1)\in B|h_t^N,u(t)\right)=Q_\rho\left(B|\vec{M}(t),u(t)\right)$$
$$=\int_{[0,1]^N}\mathbb{I}_B\left[G_\rho^N\left(\vec{M}(t),u(t),\vec{w}\right)\right]\theta(\mathrm{d}\vec{w})\text{ for }B\in\mathcal{B}\left(\mathbb{P}_N(S)\right),$$

  where Q_ρ is as in (3.2).

The stochastic process $\left(\Omega',\mathcal{F}',\mathbb{P}_{\vec{m}}^{\pi^N},\left(\vec{M}(t)\right)\right)$ is a discrete-time Markov control process. We will refer to the probability measure $\mathbb{P}_{\vec{m}}^{\pi^N}$ as the *strategic probability measure*, in the fashion of [7]. See the paragraph about the *canonical construction* in [13, p.15-16], Remark 2.3 in [18], the first three lines on [19, p.65], Remark 2.4 in [17] and Remark 2 in [20].

Definition 3.2. (*i*) Let $\mathbb{F}$ be the class of all measurable functions $f:\mathbb{P}(S)\to U$, and $\mathbb{F}^N:=\mathbb{F}|_{\mathbb{P}_N(S)}$ be the restriction of $\mathbb{F}$ over $\mathbb{P}_N(S)$. A policy $\pi^N\in\Pi^N$ is said to be a *(deterministic) Markov policy for the N-MDP* $\mathcal{M}_N$ if there exists a sequence $\left(f_t^N:t\in\mathbb{N}_0\right)\subseteq\mathbb{F}^N$ such that, for all $t\in\mathbb{N}_0$ and $h_t^N\in\mathbb{H}_t^N$, we have that $\pi_t^N(\cdot|h_t^N)=\delta_{f_t^N(\vec{M}^N(t))}(\cdot)$. In this case, $\pi^N=\left(f_t^N:t\in\mathbb{N}_0\right)$. The symbol Π_M^N stands for the *class of all Markov policies in the model* $\mathcal{M}_N$.

(*ii*) If there exists a sequence $(f_t:t\in\mathbb{N}_0)\subseteq\mathbb{F}$ such that, for all $t\in\mathbb{N}_0$ and $h_t^N\in\mathbb{H}_t^N$, we have that $\pi_t(\cdot|h_t^N)=\delta_{f_t(\vec{M}^N(t))}(\cdot)$, then the set of (deterministic) Markov policies is denoted as Π_M. In other words, Π_M is the class of sequences in $\mathbb{F}$.

(*iii*) If in (*ii*), $f_t^N\equiv f^N$ for some $f^N\in\mathbb{F}$ and all $t\in\mathbb{N}_0$, then π^N is a *stationary policy*. By abusing the notation, we denote the *set of stationary policies* as $\mathbb{F}$ (and $\mathbb{F}^N$).

The next result is a consequence of Theorem 2.2 and the dominated convergence theorem. It matches Assumption 3.18(c) in [16] and Assumption 3.1(c) in [28].

Proposition 3.3. *Under Assumption 2.1, the transition law $Q_\rho(\cdot|\cdot,\cdot)$ is weakly continuous on $\mathbb{P}_N(S)\times U$. That is, the mapping*

$$(\vec{m},u)\mapsto\int_{\mathbb{P}(S)}v(y)Q_\rho(\mathrm{d}y|\vec{m},u)$$

is continuous for each continuous and bounded function v.

Proof. Note that, by Algorithm 1 and (3.2), for each bounded and continuous function $v:\mathbb{P}(S)\to\mathbb{R}$,

$$\int_{\mathbb{P}(S)}v(y)Q_\rho(\mathrm{d}y|\vec{m},u)=\int_{[0,1]^N}v\left[G_\rho^N\left(\vec{m},u,\vec{w}\right)\right]\theta\left(\mathrm{d}\vec{w}\right).$$

Let $(\vec{m}_k)\subset\mathbb{P}_N(S)$ and $(u_k)\subset U$ be a couple of sequences such that $\|\vec{m}_k-\vec{m}\|_\infty\to0$ and $d_U(u_k,u)\to0$ as $k\to\infty$, for $(\vec{m},u)\in\mathbb{P}_N(S)\times U$. Since v is a continuous and bounded function, the dominated convergence theorem and Theorem 2.2 yield that

$$\lim_{k\to\infty}\int_{[0,1]^N}v\left[G_\rho^N\left(\vec{m}_k,u_k,\vec{w}\right)\right]\theta\left(\mathrm{d}\vec{w}\right)=\int_{[0,1]^N}v\left[G_\rho^N\left(\vec{m},u,\vec{w}\right)\right]\theta\left(\mathrm{d}\vec{w}\right).$$

This completes the proof. □

Assumption 2.1(i) and Proposition 3.3 correspond to assumptions (W1)-(W2) in [7]. This fact enables us to use [7, Lemma 1] and thus assert that the set of strategic probability measures $\mathcal{P} := \left\{\mathbb{P}_{\vec{m}}^{\pi^N} : \pi^N \in \Pi^N\right\}$ is a compact subset of $\mathbb{P}(\Omega')$ endowed with the topology of weak convergence. Moreover, the functional $P \mapsto \int_{\Omega'} v \mathrm{d}P$ is lower semi-continuous on $\mathcal{P}$ for every bounded from below and lower semi-continuous function v on Ω'.

3.1. **Discounted and ergodic optimalities in the N-MDP.** In the former subsection we presented the system of our interest. Now we present the criteria we will study.

For each control policy $\pi^N \in \Pi^N$ and initial state $\vec{M}^N(0) = \vec{m} \in \mathbb{P}_N(S)$, define the expected α-discounted cost as

$$V^N\left(\pi^N, \vec{m}; \alpha\right) := \mathbb{E}_{\vec{m}}^{\pi^N}\left[\sum_{t=0}^{\infty} \alpha^t c\left(\vec{M}^N(t), u(t)\right)\right], \tag{3.3}$$

where the so-called *discount factor* α is in the interval $]0,1[$, and $\mathbb{E}_{\vec{m}}^{\pi^N}[\cdot]$ stands for the conditional expectation operator with respect to the probability measure $\mathbb{P}_{\vec{m}}^{\pi^N}$ when the agent chooses the control policy π^N given $\vec{M}^N(0) = \vec{m}$. Analogously, for each control policy $\pi^N \in \Pi^N$ and initial state $\vec{M}^N(0) = \vec{m} \in \mathbb{P}_N(S)$, we define the expected average cost as

$$J^N\left(\pi^N, \vec{m}\right) := \limsup_{T\to\infty} \frac{1}{T} \mathbb{E}_{\vec{m}}^{\pi^N}\left[\sum_{t=0}^{T-1} c\left(\vec{M}^N(t), u(t)\right)\right]. \tag{3.4}$$

In view of Assumption 2.1(i), Remark 2.4, Proposition 3.3, we can quote [7, Lemma 2] to claim that both the expected α-discounted and average costs from (3.3) and (3.4), respectively, are lower semi-continuous in $\left(\pi^N, \vec{m}\right) \in \Pi^N \times \mathbb{P}(S)$. (In fact, Lemma 2 in [7] proves only that (3.3) is lower semi-continuous with respect to the strategic probability measure $\mathbb{P}_{\vec{m}}^{\pi^N} \in \mathcal{P}$. However, the proof that also (3.4) is lower semi-continuous with respect to $\mathbb{P}_{\vec{m}}^{\pi^N} \in \mathcal{P}$ can be obtained from the same result, and Abelian Theorem 1.1.)

Definition 3.4. We say that $\pi_*^N \in \Pi^N$ is an *optimal control policy for the N-MDP $\mathcal{M}_N$ under the expected α-discounted cost criterion* (3.3) if

$$\begin{aligned} V_*^N(\vec{m}; \alpha) &:= \inf_{\pi^N \in \Pi^N} V^N\left(\pi^N, \vec{m}; \alpha\right) \\ &= V^N\left(\pi_*^N, \vec{m}; \alpha\right) \text{ for } \vec{m} \in \mathbb{P}_N(S). \end{aligned} \tag{3.5}$$

We call $V_*^N(\cdot; \alpha)$ the *expected α-discounted optimal value function for the N-MDP $\mathcal{M}_N$*.

The following is a very important result, which we borrow from Theorem 3.21(d) in [16] (see also Theorem 4.2.3 in [13]).

Proposition 3.5. *Let Assumptions 2.1 and 2.3 hold, and let $\pi_* \in \mathbb{F}^N$ be an optimal control policy with respect to the Markov strategies,* i.e.*,*

$$V^N(\pi_*, \vec{m}; \alpha) \leq V^N(\pi, \vec{m}; \alpha) \text{ for all } \pi \in \Pi_M^N, \vec{m} \in \mathbb{P}_N(S). \tag{3.6}$$

Then π_* is an optimal control policy with respect to all control policies. That is, (3.6) *holds.*

By virtue of Proposition 3.5, we analyze only stationary control policies.

Now we introduce the definition of the average value function. It is analogous to Definition 3.4.

Definition 3.6. We say that $\pi_*^N \in \Pi^N$ is an *optimal control policy for the N-MDP $\mathcal{M}_N$ under the expected average cost criterion* (3.4) if

$$J_*^N(\vec{m}) := \inf_{\pi^N \in \Pi^N} J^N\left(\pi^N, \vec{m}\right) = J^N\left(\pi_*^N, \vec{m}\right) \text{ for } \vec{m} \in \mathbb{P}_N(S). \tag{3.7}$$

We call $J_*^N(\cdot)$ the *expected average optimal value function for the N-MDP $\mathcal{M}_N$*.

The expected α-discounted and the expected average cost criteria from Definitions 3.4 and 3.6 are related by the Abelian Theorem 1.1 (see for instance, the discussion immediately after Lemma 5.3.1 in [13] and Remark 2.1 in [28]). Indeed, take $c_t := \mathbb{E}_{\vec{m}}^{\pi^N} c\left(\vec{M}(t), u(t)\right)$ and observe that, by linearity, the third inequality in the Abelian Theorem 1.1, (3.3) and (3.4) give

$$\limsup_{\alpha \uparrow 1}(1-\alpha) V^N\left(\pi^N, \vec{m}; \alpha\right) \leq J^N\left(\pi^N, \vec{m}\right) \text{ for all } \pi^N \in \Pi^N, \vec{m} \in \mathbb{P}_N(S).$$

Thus, from (3.5),

$$\limsup_{\alpha \uparrow 1}(1-\alpha) V_*^N\left(\vec{m}; \alpha\right) \leq J^N\left(\pi^N, \vec{m}\right) \text{ for all } \pi^N \in \Pi^N, \vec{m} \in \mathbb{P}_N(S).$$

Which in turn, since $\pi^N \in \Pi^N$ was arbitrary, by Definition 3.6 implies that

$$\limsup_{\alpha \uparrow 1}(1-\alpha) V_*^N\left(\vec{m}; \alpha\right) \leq J_*^N\left(\vec{m}\right) \text{ for all } \vec{m} \in \mathbb{P}_N(S). \tag{3.8}$$

(See also Lemma 5.7(b) in [15].) This means that the product of $(1-\alpha)$ with the expected α-discounted optimal value function for the N-MDP is a lower bound of the expected average optimal value function for the N-MDP, when α is *close* to the unit.

We begin by characterizing the optimal control policies for the expected α-discounted cost criterion and the corresponding value function for the N-MDP $\mathcal{M}_N$ by means of the following result (see Proposition 2.4 in [18], Theorem 3.3 in [28]).

Proposition 3.7. *Let $\vec{m} \in \mathbb{P}_N(S)$, α be the discount factor referred to in* (3.3)*, and Assumptions 2.1 and 2.3 hold. Then:*

(*i*) *The expected α-discounted optimal value function for the N-MDP $\mathcal{M}_N$* (3.5) *satisfies the dynamic programming equation:*

$$V_*^N(\vec{m}; \alpha) = \min_{u \in U}\left[c(\vec{m}, u) + \alpha \int_{[0,1]^N} V_*^N\left(G_\rho^N\left(\vec{m}, u, \vec{w}\right); \alpha\right) \theta(\mathrm{d}\vec{w})\right]. \tag{3.9}$$

Moreover $\left|V_*^N(\vec{m}; \alpha)\right| \leq \frac{R}{1-\alpha}$.

(*ii*) *There exists* $f_*^N \in \mathbb{F}^N$ *such that* $f_*^N(\vec{m}) \in U$ *attains the minimum in* (3.9). *That is,*

$$V_*^N(\vec{m};\alpha) = c\left(\vec{m}, f_*^N\right) + \alpha \int_{[0,1]^N} V_*^N\left(G_\rho^N\left(\vec{m}, f_*^N, \vec{w}\right);\alpha\right)\theta(\mathrm{d}\vec{w}).$$

In fact, the stationary policy $\pi_*^N = \left(f_*^N\right) \in \Pi_M^N$ *is* α*-discounted optimal for the control model* $\mathcal{M}_N$.

Definition 3.8. Let $\mu_\alpha^N := \inf_{\vec{m}\in\mathbb{P}(S)} V_*^N(\vec{m};\alpha)$, and $J_*^N := \limsup_{\alpha\uparrow 1}(1-\alpha)\mu_\alpha^N$; and define the *relative discounted value function* as:

$$r_\alpha^N(\vec{m}) := V_*^N(\vec{m};\alpha) - \mu_\alpha^N, \text{ for each } \alpha \in]0,1[.$$

It is straightforward that we can re-state (3.9) as

$$(1-\alpha)\mu_\alpha^N + r_\alpha^N(m) = \min_{u\in U}\left[c(\vec{m},u) + \alpha\int_{[0,1]^N} r_\alpha\left(G_\rho^N\left(\vec{m},u,\vec{w}\right)\right)\theta(\mathrm{d}\vec{w})\right].$$

One would hope that, letting $\alpha\uparrow 1$ in the former equation, we could obtain

$$J_*^N + r^N(\vec{m}) = \min_{u\in U}\left[c(\vec{m},u) + \int_{[0,1]^N} r^N\left(G_\rho^N\left(\vec{m},u,\vec{w}\right)\right)\theta(\mathrm{d}\vec{w})\right]$$

for some $r^N(\cdot)$. Asserting this is very difficult in general. However, the following hypothesis will enable us to state Proposition 3.10 (which is the next best thing).

Assumption 3.9. *Let* $\inf_{\vec{m}\in\mathbb{P}_N(S)} J_*^N(\vec{m})$ *be a finite-valued constant, and* $r^N(\cdot) := \liminf_{\alpha\uparrow 1} r_\alpha^N(\cdot)$ *be a finite-valued function.*

The first part of Assumption 3.9 and (3.8) enable us to avoid the trivial case where the expected optimal value functions for the N-MDP from (3.5) and (3.7) fail to exist. We use the second part of Assumption 3.9 to define the *lower semi-continuous envelope of* $r^N(\cdot)$ as

$$r_*^N(\vec{m}) := \sup_{s>0}\inf_{\vec{\ell}\in B_s(\vec{m})} r^N(\vec{\ell}) \text{ for all } \vec{m}\in\mathbb{P}_N(S), \tag{3.10}$$

where $B_s(\vec{m})$ stands for the open ball with center $\vec{m}\in\mathbb{P}_N(S)$ and radius $s>0$.

The following result is a consequence of the Abelian Theorem 1.1 and of Proposition 3.3. The details can be consulted in Theorem 4.5 in [28] (see also Theorem 5.9 in [15], Theorem 3.31 in [16] and Theorem 4 in [8]). One of the hypotheses of [28, Theorem 4.5] is the weak continuity of the transition probability of the model $\mathcal{M}_N$, which holds by virtue of Proposition 3.3. Another assumption made in [28, Theorem 4.5] is the inf-compactness of the one-stage cost function c, which holds by virtue of Remark 2.4.

Proposition 3.10. *If Assumptions 2.1, 2.3 and 3.9 hold then*

(*i*) *The expected ergodic optimal value function for the N-MDP* $\mathcal{M}_N$ (3.7) *is such that*

$$J_*^N + r_*^N(\vec{m}) \geq \min_{u\in U}\left[c(\vec{m},u) + \int_{[0,1]^N} r_*^N\left(G_\rho^N\left(\vec{m},u,\vec{w}\right)\right)\theta(\mathrm{d}\vec{w})\right], \tag{3.11}$$

where $r^N(\cdot)$ is as in (3.10).

(*ii*) *There exists $f_*^N \in \mathbb{F}^N$ such that $f_*^N(\vec{m}) \in U$ attains the minimum in* (3.11). *That is,*

$$\begin{aligned} &\min_{u\in U}\left[c(\vec{m},u)+\int_{[0,1]^N} r_*^N\left(G_\rho^N\left(\vec{m},u,\vec{w}\right)\right)\theta(\mathrm{d}\vec{w})\right] \\ &\quad = c\left(\vec{m},f_*^N\right)+\int_{[0,1]^N} r^N\left(G_\rho^N\left(m,f_*^N,\vec{w}\right)\right)\theta(\mathrm{d}\vec{w}) \end{aligned} \tag{3.12}$$

for $\vec{m} \in \mathbb{P}_N(S)$. In fact, the stationary policy $\pi_*^N = \left(f_*^N\right) \in \Pi_M^N$ is an optimal control policy for the N-MDP $\mathcal{M}_N$ under the expected average cost criterion (3.4) *and $J_*^N = J_*^N(\vec{m}) = J^N\left(\pi_*^N, \vec{m}\right)$ for all $\vec{m} \in \mathbb{P}(S)$.*

Proposition 3.10 represents a possibility to study ergodic optimality in systems with interacting objects. However, as is the case of the problem studied in [18], from a practical standpoint, its applicability is severely constrained because N is too large, and there is no information so as to the functional form of the density ρ. In fact, in order to examine (3.12), it is necessary to solve a multiple integral of dimension N, and the dynamics of the system depends on the unidentified density ρ.

4. A deterministic control model

Recall the model from (3.1) and consider instead the *deterministic* model $\mathcal{M} = (\mathbb{P}(S), U, G_\rho, c)$, where $G_\rho : \mathbb{P}(S) \times U \to \mathbb{P}(S)$ is a Lipschitz-continuous function with Lipschitz constant L_G (that is implicitly dependent on ρ). That is

$$\left\|G_\rho(\vec{m},u) - G_\rho(\vec{m}',u')\right\|_\infty \leq L_G \max\left(\left\|\vec{m}-\vec{m}'\right\|_\infty, d_U(u,u')\right), \tag{4.1}$$

for $\vec{m}, \vec{m}' \in \mathbb{P}(S)$ and $u, u' \in U$. We will refer to $\mathcal{M}$ as the *mean-field control model* and use Proposition 3.5 to consider that the set of control policies for the model $\mathcal{M}$ is Π_M. With this in mind, recall (2.2) and let

$$\vec{m}(t+1) := G_\rho\left(\vec{m}(t), u(t)\right),$$

with the initial condition $\vec{m} = \vec{m}(0) \in \mathbb{P}(S)$.

Now define the *total discounted cost for the mean-field model* as

$$v(\pi, \vec{m}; \alpha) = \sum_{t=0}^{\infty} \alpha^t c(\vec{m}(t), u(t)).$$

Proposition 3.2 in [18] asserts the existence of a control policy $\pi_* \in \Pi_M$ such that the so-called *mean-field value function under the total discounted criterion* $v_*(\vec{m}; \alpha) := \inf_{\pi\in\Pi_M} v(\pi, \vec{m}; \alpha)$ for $\vec{m} \in \mathbb{P}(S)$ satisfies

$$v_*(\vec{m};\alpha) = \inf_{u\in U}\left[c(\vec{m},u) + \alpha v_*\left(G_\rho(\vec{m},u);\alpha\right)\right] \text{ for } \vec{m} \in \mathbb{P}(S). \tag{4.2}$$

Next define the *average cost for the mean-field model* $\mathcal{M}$ as

$$j(\pi, \vec{m}) \;:=\; \limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} c\left(\vec{m}(t), u(t)\right).$$

Section 5.4 in [15] proves that it is possible to use the Abelian Theorem 1.1 to prove the existence of a stationary policy $\pi_* = (f_*) \in \Pi_M$ such that $f_* \in \mathbb{F}$ such that the *average optimal value function for the mean-field model* $\mathcal{M}$ $j(\vec{m}) := \inf_{\pi\in\Pi} j(\pi, \vec{m})$ is attained for $\vec{m} \in \mathbb{P}(S)$ (provided that $j(\vec{m}) < \infty$ for all $\vec{m} \in \mathbb{P}(S)$). We summarize this procedure for we will need it in the sequel.

For each $\alpha \in]0, 1[$, we define

$$\mu_\alpha := \inf_{\vec{m}\in\mathbb{P}(S)} v_*(\vec{m}; \alpha),$$

the *relative discounted value function*

$$r_\alpha(\vec{m}) := v_*(\vec{m}; \alpha) - \mu_\alpha,$$

and $j_\alpha := (1 - \alpha)\mu_\alpha$. Rewrite (4.2) as

$$j_\alpha + r_\alpha(\vec{m}) = \inf_{u\in U} [c(\vec{m}, u) + \alpha r_\alpha(G_\rho(\vec{m}, u))]$$

Observe that the third inequality in the Abelian Theorem 1.1 yields that

$$j_* := \limsup_{\alpha\uparrow 1} j_\alpha \leq \inf_{\vec{m}\in\mathbb{P}(S)} j(\vec{m}). \tag{4.3}$$

Now we quote Theorem 5.9 in [15], which is somewhat of a deterministic analog of our Proposition 3.10 (see also Remark 2.60 in [16]).

Proposition 4.1. *Let Assumptions 2.1 and 2.3 hold, and the function G_ρ be as in* (4.1)*. Furthermore, let $\inf_{\vec{m}\in\mathbb{P}_N(S)} j(\vec{m})$ be a finite-valued constant, and $r(\cdot) := \liminf_{\alpha\uparrow 1} r_\alpha(\cdot)$ be a finite-valued function.*

Then, there exists a finite-valued lower semi-continuous and bounded below function r_* with $r_*(\cdot) \leq r(\cdot)$ such that the pair (j_*, r_*) satisfies the average control mean-field optimality inequality*:*

$$j_* + r_*(\vec{m}) \geq \inf_{u\in U} [c(\vec{m}, u) + r_* (G_\rho (\vec{m}, u))] . \tag{4.4}$$

In fact, there exists $f_* \in \mathbb{F}$ such that $f_*(\vec{m}) \in U$ attains the minimum in (4.4)*. That is,*

$$\inf_{u\in U} [c(\vec{m}, u) + r_* (G_\rho (\vec{m}, u))] = c (\vec{m}, f_*) + r_* (G_\rho (\vec{m}, f_*)) \tag{4.5}$$

for $\vec{m} \in \mathbb{P}(S)$. Hence, the policy $\pi_* = (f_*) \in \Pi_M$, such that $f \in \mathbb{F}$, is optimal.

Remark 4.2. The inequalities (4.4)-(4.5) give that $j_* \geq j(\pi_*, \vec{m})$ for all $\vec{m} \in \mathbb{P}(S)$. Indeed, let $\vec{m}^{f_*}(t + 1) := G_\rho (\vec{m}(t), f_*)$ be the dynamic of $\mathcal{M}$ under the policy $\pi_* = (f_*) \in \Pi_M$, for $f_* \in \mathbb{F}$. By virtue of (4.4)-(4.5), we can assert that

$$\begin{aligned} j_* &\geq c (\vec{m}(t), f_*) + r_* (G_\rho (\vec{m}(t), f_*)) - r_* (\vec{m}(t)) \\ &= c (\vec{m}(t), f_*) + r_* \left(\vec{m}^{f_*}(t + 1)\right) - r_* (\vec{m}(t)) . \end{aligned}$$

Therefore

$$\begin{aligned} Tj_* &\geq \sum_{t=0}^{T-1} \left[c (\vec{m}(t), f_*) + r_* \left(\vec{m}^{f_*}(t + 1)\right) - r_* (\vec{m}(t))\right] \\ &= \sum_{t=0}^{T-1} c (\vec{m}(t), f_*) + r_* \left(\vec{m}^{f_*}(T)\right) - r_* (\vec{m}(0)) . \end{aligned}$$

Multiplying both sides by $1/T$ and letting $T \to \infty$ give that $j_* \geq j(\pi_*, \vec{m})$. The reverse inequality is given by (4.3), which implies that $j_* = \inf_{\vec{m}\in\mathbb{P}(S)} j(\vec{m})$.

Since the unknown density function ρ is very important to characterize the optimal policies, we need to estimate it to actually obtain average optimal policies and values. We aim for this next.

4.1. **Estimation procedure in the Mean-Field model.** Our approach is a modified version of the one presented in [18]. To this end, consider the unknown density ρ referred to in (2.1) and let $\rho_k(\cdot) := \rho_k(\cdot; \tilde{\xi}_0, ..., \tilde{\xi}_{k-1})$ be an estimator of ρ such that, as $k \to \infty$,

$$\int_{\mathbb{R}} |\rho_k(z) - \rho(z)| \mathrm{d}z \quad \to \quad 0 \text{ a.s.}, \tag{4.6}$$

$$\sup_{(\vec{m},u)\in\mathbb{P}(S)\times U} \|G_{\rho_k}(\vec{m}, u) - G_\rho(\vec{m}, u)\|_\infty \quad \to \quad 0 \text{ a.s.} \tag{4.7}$$

where $\tilde{\xi}_0, ..., \tilde{\xi}_{k-1}$ are independent observations of the random variable with density ρ. By virtue of (4.7) and the dominated convergence theorem, it is straightforward that, for all $\pi \in \Pi_M$,

$$\mathbb{E}^\pi_{\vec{m}} \left[\sup_{(\vec{x},u)\in\mathbb{P}(S)\times U} \|G_{\rho_k}(\vec{x}, u) - G_\rho(\vec{x}, u)\|_\infty \right] \to 0 \text{ as } k \to \infty.$$

We use this certainty as a termination criterion in Algorithm 2.

Recall the notation used in Remark 4.2 for the dynamic of $\mathcal{M}$ under the policy $\pi_k = (f_k \in \mathbb{F}) \in \Pi_M$, and replace G_ρ by G_{ρ_k} to obtain $\vec{m}^{f_k}(t+1) := G_{\rho_k}(\vec{m}(t), f_k)$. Provided that the pair (j_*, r_*) satisfies the average control mean-field optimality inequality (4.4), Algorithm 2 is our implementation of the classic Howard's *policy iteration procedure* that obtains an optimal policy $(f_*) \subset \mathbb{F}$ for the mean-field model $\mathcal{M}$ and a pair (j_*, r_*) that satisfies (4.4). Section 2.4 in [16] presents it for discrete-time control problems under the discounted cost criterion, while Remark 2.4 in [14] states the stochastic version of the algorithm for MDPs under the long-run cost criterion.

Remark 4.3. (1) Line 1 in Algorithm 2 is the so-called *initialization* phase of the policy iteration algorithm. Then, the algorithm goes into the loop from lines 4-7. Lines 4 and 5 stand for the phase of policy evaluation, while line 6 improves (reduces) the value of j_k. Indeed, from lines 4-6, it is clear that

$$j_k + r_k \geq j_{k+1} + r_{k+1}.$$

(2) By the feature we imposed on j_k at line 4, it is straightforward that the function sought by line 5 is such that

$$\begin{aligned} r_k\left(\vec{m}^{f_k}(t)\right) &\leq r_k\left(\vec{m}^{f_k}(0)\right) + \sum_{s=0}^{t-1} \left[j_k - c\left(\vec{m}^{f_k}(s), f_k\right)\right] \\ &\leq r_k\left(\vec{m}^{f_k}(0)\right) + \frac{1-\varepsilon^t}{1-\varepsilon} \\ &\leq r_k\left(\vec{m}^{f_k}(0)\right) + \frac{1}{1-\varepsilon}. \end{aligned}$$

Algorithm 2: The policy iteration algorithm

Data: Estimated densities $(\rho_k(\cdot))$ that meet (4.7) and tolerance level $0 < \varepsilon < 1$.

Result: Triplet $(j_*, r_*, (f_*))$ that meets (4.4).

1 $k \leftarrow 0$;

2 Select $(f_k) \subset \mathbb{F}$;

3 **do**

4 Find a constant j_k such that $\left|j_k - c\left(\vec{m}^{f_k}(t), f_k\right)\right| < \varepsilon^t$ for all $t \geq 0$;

5 Find a lower semi-continuous function $r_k : \mathbb{P}(S) \to \mathbb{R}$ such that $j_k + r_k(\vec{m}) \geq c(\vec{m}, f_k) + r_k\left(G_{\rho_k}(\vec{m}, f_k)\right)$ for all $\vec{m} \in \mathbb{P}(S)$; ▷ `Note that, in particular,` $j_k + r_k(\vec{m}) \geq \inf_{u \in U}\left[c(\vec{m}, u) + r_k\left(G_{\rho_k}(\vec{m}, u)\right)\right]$.

6 Find $(f_{k+1}) \subset \mathbb{F}$ such that $c(\vec{m}, f_{k+1}) + r_k\left(G_{\rho_k}(\vec{m}, f_{k+1})\right) = \inf_{u \in U}\left[c(\vec{m}, u) + r_k\left(G_{\rho_k}(\vec{m}, u)\right)\right]$;

7 $k \leftarrow k + 1$;

8 **while** $|j_{k-1} - j_k| \geq \varepsilon$ *or* $\sup_{(\vec{m},u) \in \mathbb{P}(S) \times U} \left\|G_{\rho_k}(\vec{m}, u) - G_{\rho_{k-1}}(\vec{m}, u)\right\|_\infty \geq \varepsilon$;

9 **return** $(j_k, r_k, (f_k))$;

This implies that the sequence (r_k) generated by Algorithm 2 is bounded. This fact will be used in the proof of Lemma 4.5 below.

(3) The procedure leaves the loop only if the termination criteria from line 8 are eventually met. That is, if the approximations j_k to j_* in (4.4); and G_{ρ_k} to G_ρ in (4.7) are *good enough*.

Our next result is a consequence of Proposition 4.1. It ensures that Algorithm 2 converges, that is $j_k \downarrow j_*$ as $k \to \infty$.

Proposition 4.4. *Let Assumptions 2.1 and 2.3 hold, the function G_ρ be as in (4.1), $j_* = \inf_{\vec{m} \in \mathbb{P}_N(S)} j(\vec{m}) < \infty$, and $r_*(\vec{m}) := \sup_{s>0} \inf_{\vec{\ell} \in B_s(\vec{m})} r(\vec{\ell})$ for all $\vec{m} \in \mathbb{P}_N(S) < \infty$ (that is, $r_* < \infty$ is the lower semi-continuous envelope of $r(\cdot) = \liminf_{\alpha \uparrow 1} r_\alpha(\cdot)$). Let (j_k, r_k) be as in Algorithm 2. Then:*

(*i*) *If there exists $k \in \mathbb{N}$ for which*

$$|j_{k+1} - j_k| < \varepsilon, \text{ and} \tag{4.8}$$

$$\sup_{(\vec{m},u) \in \mathbb{P}(S) \times U} \left\|G_{\rho_{k+1}}(\vec{m}, u) - G_{\rho_k}(\vec{m}, u)\right\|_\infty < \varepsilon \text{ for all } \varepsilon > 0, \tag{4.9}$$

then $j_* \equiv j_k$ and $r_*(\cdot) = r_k(\cdot)$ satisfy (4.4) and f_k is an optimal control.

(*ii*) *As $k \to \infty$, $j_k \downarrow j_*$.*

(*iii*) *The sequence of functions (r_k) generated by Algorithm 2 is such that, for all $\vec{m} \in \mathbb{P}(S)$, $t = 0, 1, \ldots$ and $f \in \mathbb{F}$,*

$$\mathbb{E}^f_{\vec{m}} \left|r_k\left(\vec{m}(t)\right) - r_*\left(\vec{m}(t)\right)\right| \to 0 \text{ as } k \to \infty.$$

To prove Proposition 4.4, we need the following ancillary result.

Lemma 4.5. *Let (r_k) be the sequence of functions generated by Algorithm 2; and Assumptions 2.1 and 2.3 hold. Then, there is a measurable function $r_* : \mathbb{P}(S) \to \mathbb{R}$ and a subsequence $(k_\iota) \equiv (\iota)$ of (k) such that $r_\iota \to r_*$ as $\iota \to \infty$.*

Proof. By Remark 4.3(2), it is straightforward that every member of the sequence (r_k) generated by Algorithm 2 is bounded in $\mathcal{L}_\infty\left(\mathbb{P}(S), \mathcal{B}\left(\mathbb{P}(S)\right), \nu\right)$, where ν is any σ-finite measure (for instance, Lebesgue's measure). Now, Banach-Alaoglu theorem for separable spaces (see, for instance, Theorem 5.1 in [3]) yields the result. □

Now we are ready to prove Proposition 4.4.

Proof of Proposition 4.4. (*i*) Let (r_ι) be the subsequence of functions generated by Algorithm 2 referred to by Lemma 4.5, and (f_ι) be the corresponding subsequence. By [26, Proposition 12.2] (or [13, Proposition D.7]), we can assert the existence of an accumulation point f_* of the latter. That is, for each $\vec{m} \in \mathbb{P}(S)$, there exists a subsequence (ι_κ) of (ι) such that

$$\lim_{\kappa \to \infty} f_{\iota_\kappa}(\vec{m}) = f_*(\vec{m}). \tag{4.10}$$

Now fix an arbitrary $\vec{m} \in \mathbb{P}(S)$ and let ι_κ as in (4.10). Replace k by ι_κ in lines 4-5 of Algorithm 2, use (4.8) and (4.9); and let $\kappa \to \infty$ to see that

$$j_* + r_*(\vec{m}) \geq c\left(\vec{m}, f_*\right) + r\left(G_\rho(\vec{m}, f_*)\right).$$

(*ii*) This is a consequence of part (*i*) and Remark 4.2.

(*iii*) Lemma 4.5 and the dominated convergence theorem yield the desired outcome.

This completes the proof. □

4.2. **Performance analysis.** We complete this paper by analyzing how well the policies found by Algorithm 2 perform in the original model $\mathcal{M}_N$. We adapt [18, Assumption 5.1] to our context (sufficient conditions to ensure that it holds are given in Theorem 1 in [20]).

Assumption 4.6. *Let $\vec{m} := \vec{M}^N(0) = \vec{m}(0)$ for all $N \in \mathbb{N}$. For each $\vec{m}(t) \in \mathbb{P}(S)$, $\vec{M}^N(t) \in \mathbb{P}_N(S)$; $T \in \mathbb{N}$ and $\varepsilon > 0$ there exist positive constants K and λ such that*

$$\sup_{\pi \in \Pi_M} \mathbb{P}^\pi_{\vec{m}}\left(\sup_{0 \leq t \leq T} \left\| \vec{M}^N(t) - \vec{m}(t) \right\|_\infty \geq \gamma_T(\varepsilon)\right) \leq KT\mathrm{e}^{-\lambda N \varepsilon^2},$$

where $\gamma_T(\varepsilon)$ is a finite-valued $o(\varepsilon)$ function.

Now we establish our final result.

Theorem 4.7. *If Assumptions 2.1, 2.3, 3.9 and 4.6 hold, then*

$$\sup_{\varphi \in \Pi_M} \mathbb{E}^\varphi_{\vec{m}} \left| J^N_* - j_* \right| \to 0 \text{ as } N \to \infty.$$

Proof. Note that, for all $\alpha \in]0, 1[$,

$$\begin{aligned} \sup_{\varphi \in \Pi_M} \mathbb{E}^\varphi_{\vec{m}} \left| J^N_* - j_* \right| \quad \leq \quad & \sup_{\varphi \in \Pi_M} \mathbb{E}^\varphi_{\vec{m}} \left| (1-\alpha) v_*\left(\vec{m}; \alpha\right) - j_* \right| \\ & + \sup_{\varphi \in \Pi_M} \mathbb{E}^\varphi_{\vec{m}} \left| J^N_* - (1-\alpha) V^N_*\left(\vec{M}^N; \alpha\right) \right| \end{aligned}$$

$$+ \sup_{\varphi\in\Pi_M} \mathbb{E}^{\varphi}_{\vec{m}} \left|(1-\alpha)V^N_*\left(\vec{M}^N;\alpha\right) - (1-\alpha)v_*\left(\vec{m};\alpha\right)\right|.$$

Since the choice of α is abritrary, we can use Proposition 3.10 to see that

$$\limsup_{\alpha\uparrow 1}(1-\alpha)v_*\left(\vec{m};\alpha\right) = j_*. \tag{4.11}$$

So that the first term in the right-hand side of the inequality above nullifies. Analogously, Proposition 4.1 gives us that

$$\limsup_{\alpha\uparrow 1}(1-\alpha)V^N_*\left(\vec{M};\alpha\right) = J^N_*. \tag{4.12}$$

So, the second term in the right-hand side of the inequality tends to zero as $\alpha\uparrow 1$. Now let $\mathcal{K}(T) := L_g^T \max\left(L_g, \sup_{(u,u')\in U\times U} d(u,u')\right)$ for all T. Theorem 5.3(a) in [18] ensures that

$$\begin{aligned}
&\sup_{\varphi\in\Pi_M} \mathbb{E}^{\varphi}_{\vec{m}} \left|(1-\alpha)V^N_*(m;\alpha) - (1-\alpha)v_*(m;\alpha)\right| \\
\leq\ & 2R\alpha^T + L_c\left(1-\alpha^T\right)\left(KT\mathrm{e}^{-\lambda N\varepsilon^2}(1+\mathcal{K}(T)) + \gamma_T(\varepsilon)\right) \\
\to\ & 2R\alpha^T + L_c\left(1-\alpha^T\right)\gamma_T(\varepsilon) \text{ as } N\to\infty.
\end{aligned}$$

The finiteness of the function $\gamma_T(\cdot)$ quoted in Assumption 4.6 enables us to assert that

$$\sup_{\varphi\in\Pi_M} \mathbb{E}^{\varphi}_{\vec{m}} \left|(1-\alpha)V^N_*(\vec{m};\alpha) - (1-\alpha)v_*(\vec{m};\alpha)\right| \to 0 \text{ as } T\to\infty.$$

for all $0<\alpha<1$. Finally, take the suprema in Π_M and use (4.11) and (4.12) to complete the proof. □

5. Concluding remarks

This paper represents an extension of the results presented in [18] to analyze the ergodic cost criterion by means of the vanishing discount technique spawned by the Abelian Theorem 1.1, which can be traced back to Hardy's work (see [12]). One of the main difficulties that we tackled was that the discrete-time deterministic version of the optimal control problem under the ergodic crtierion remains an open problem, so we took advantage of the particularities of the problem at hand to provide a framework where we could find optimal policies in the space of stationary Markovian policies. To this end, we based our developments on the valuable survey [15] and the results presented in Chapter 2 in [16].

The potential applications of the theory presented here include the study of market shares in different industries, such as insurance companies, financial markets, and commercial models, where the central controller is uncertain about the exact form of the density function of the noise affecting the behavior of the interacting objects (people, market-makers, consumers). The main strength of the extension studied here is that it overcomes the fact that the discounted cost criterion emphasizes the weight of early stages and puts little attention on the later phases of the horizon.

Open problems for further studies are:

- How do we compute the estimation of ρ that meets (4.6)? A possibility is the use of a discrete-time version of the principle of estimation control presented in [6], or the games against nature approach used in [19] and [21].
- Is it possible to define zero-sum (as those studied in [20]) and nonzero-sum games between an oligopoly under the ergodic cost criterion? Moreover, can we establish stable cooperations among the agents and along the horizon, in the fashion of [9]?
- Finally, can we obtain control policies that attain optimality under the average cost criterion for multiple players in finite time? We believe a possibility in this direction is to work on a discrete-time version of [5]and [23].

Acknowledgement

The authors are truly indebted to the anonymous reviewer appointed by Professor Alexander Zaslavski. His/her thorough reading of the versions of our manuscript, and his/her valuable questions, remarks and suggestions made it possible to produce this version of our work. We also sincerely thank professor Onésimo Hernández-Lerma for his kind invitation to publish our work in the *Pure and Applied Functional Analysis*.

References

[1] E. J. Balder, *On compactness of the space of policies in stochastic dynamic programming*, Stochastic Processes and their Applications **32** (1989), 141–150.
[2] D. Blackwell, *Discrete Dynamic Programming*, The Annals of Mathematical Statistics **33** (1963), 719–726.
[3] J. B. Conway, *A Course in Functional Analysis*, Springer New York, NY, 2007.
[4] E. B. Dynkin and A. A. Yushkevich, *Controlled Markov Processes*, Springer, New York, 1979.
[5] B. A. Escobedo-Trujillo and H. Jasso-Fuentes and J. López-Barrientos, *Blackwell-Nash equilibria in zero-sum stochastic differential games*, in: XII Symposium of Probability and Stochastic Processes, D. Hernández-Hernández, J. C. Pardo and V. Rivero (eds), Springer International Publishing, Cham, 2018, pp. 169–193.
[6] B. A. Escobedo-Trujillo, J. D. López-Barrientos, C. G. Higuera-Chan and F. A. Alaffita-Hernández, *Robust statistic estimation of constrained optimal control problems of pollution accumulation (Part I)*, Mathematics **11** (2023): 923.
[7] E. A. Feinberg, A. Jaśkiewicz and A. S. Nowak, *Constrained discounted Markov decision processes with Borel state spaces*, Automatica **111** (2020): 108582.
[8] E. A. Feinberg, P. O. Kasyanov and N. V. Zadoianchuk, *Average cost Markov decision processes with weakly continuous transition probabilities*, Mathematics of Operations Research **37** (2012), 591–607.
[9] M. A. García-Meza, E. V. Gromova and J. D. López-Barrientos *Stable marketing cooperation in a differential game for an oligopoly*, International Game Theory Review **20** (2018): 1750028.
[10] N. Gast and B. Gaujal, *A mean field approach for optimization in discrete time*, Discrete Event Dynamic Systems **21** (2011), 63–101.
[11] N. Gast, B. Gaujal and J.-Y. Le Boudec, *Mean field for Markov decision processes: From discrete to continuous optimization*, IEEE Transactions on Automatic Control **57** (2012), 2266–2280.
[12] G. H. Hardy, *Divergent Series*, Clarendon Press, Oxford, 1949.
[13] O. Hernández-Lerma and J. B. Lasserre, *Discrete-time Markov control processes: Basic optimality criteria*, vol. 30, Springer Science & Business Media, 1996.
[14] O. Hernández-Lerma and J. B. Lasserre, *Policy iteration for average cost Markov control processes on Borel spaces*,j Acta Applicandae Mathematicae **47** (1997), 125–154.

[15] O. Hernández-Lerma, L. R. Laura-Guarachi and S. Mendoza-Palacios, *A survey of average cost problems in deterministic discrete-time control systems*, Journal of Mathematical Analysis and Applications **522** (2023): 126906.

[16] O. Hernández-Lerma, L. R. Laura-Guarachi, S. Mendoza-Palacios and D. González-Sánchez, *An Introduction to Optimal Control Theory: The dynamic programming approach*, Springer Cham, 2023.

[17] C. G. Higuera-Chan, *Approximation and mean field control of systems of large populations*, in: Advances in Probability and Mathematical statistics, D. Hernández-Hernández, Springer Nature Switzerland, 2021.

[18] 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*, Applied Mathematics and Optimization **74** (2016), 197–227.

[19] C. G. Higuera-Chan, H. Jasso-Fuentes and J. A. Minjárez-Sosa, *Control systems of interacting objects modeled as a game against nature under a mean-field approach*, Journal of Dynamics and Games **4** (2017), 59–74.

[20] C. G. Higuera-Chan and A. Minjárez-Sosa, *A mean field approach for discounted zero-sum games in a class of systems of interacting objects*, Dynamic Games and Applications **11** (2021), 512–537.

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

[22] J.-Y. Le Boudec and D. McDonald and J. Mundinger, *A generic mean field convergence result for systems of interacting objects*, in: 4th Int. Conf. Quantitative Evaluation of Systems, IEEE, Edinburgh, 2007.

[23] J. D. López-Barrientos ,*Policy iteration algorithms for zero-sum stochastic differential games with long-run average payoff criteria*, Journal of the Operations Research Society of China **2** (2014), 395–421.

[24] J. Neveu, *Mathematical Foundations of the Calculus of Probability*, Holden-day, 1965.

[25] A. S. Nowak, *On the weak topology on a space of probability measures induced by policies*, Bulletin of the Polish Academy of Sciences. Mathematics **36** (1988), 181–186.

[26] M. Schäl, *Conditions for optimality in dynamic programming and for the limit of n-stage optimal policies to be optimal*, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **32** (1975), 179–196.

[27] M. Schäl, *On dynamic programming: Compactness of the space of policies*, Stochastic Processes and their Applications **3** (1975), 345–364

[28] Ó. Vega-Amaya, *On the vanishing discount factor approach for Markov decision processes with weakly continuous transition probabilities*, Journal of Mathematical Analysis and Applications **426** (2015), 978–985.

Manuscript received February 27 2023
revised March 22 2023

José Daniel López-Barrientos
Facultad de Ciencias Físico Matemáticas de la Benemérita Universidad Autónoma de Puebla, Av San Claudio, Cd Universitaria, Jardines de San Manuel, 72572 Puebla, Pue
E-mail address: jose.lopezbar@fcfm.buap.mx

J. M. Mendoza Madrid
Researcher assistant at Universidad Anáhuac México
E-mail address: Mendoza.manolo1922@gmail.com

Paola Friné González-Vega
Reinsurance officer at Petróleos Mexicanos
E-mail address: paolafrine@gmail.com