Appl Math Optim (2016) 74:197–227
DOI 10.1007/s00245-015-9312-6

Discrete-Time Control for Systems of Interacting Objects with Unknown Random Disturbance Distributions: A Mean Field Approach

Carmen G. Higuera-Chan[1] · **Héctor Jasso-Fuentes**[2] · **J. Adolfo Minjárez-Sosa**[1]

Published online: 7 September 2015

Abstract We are concerned with stochastic control systems composed of a large number of N interacting objects sharing a common environment. The evolution of each object is determined by a stochastic difference equation where the random disturbance density ρ is unknown for the controller. We present the Markov control model (N-model) associated to the proportions of objects in each state, which is analyzed according to the mean field theory. Thus, combining convergence results as $N \to \infty$ (the mean field limit) with a suitable statistical estimation method for ρ, we construct the so-named eventually asymptotically optimal policies for the N-model under a discounted optimality criterion. A consumption-investment problem is analyzed to illustrate our results.

Keywords Systems of interacting objects · Mean field theory · Discounted criterion · Estimation and control

Mathematics Subject Classification 93E20 · 90C40 · 60K35

✉ J. Adolfo Minjárez-Sosa
aminjare@mat.uson.mx

Héctor Jasso-Fuentes
hjasso@math.cinvestav.mx

[1] Departamento de Matemáticas, Universidad de Sonora, Rosales s/n, Col. Centro, 83000 Hermosillo, Sonora, Mexico

[2] Departamento de Matemáticas, CINVESTAV-IPN, Apartado Postal 14-740, 07000 Mexico DF, Mexico

1 Introduction

This paper deals with a class of discrete-time controlled stochastic systems composed of a large number of N interacting objects which share a common environment. Denoting by $X_n^N(t)$ the state of the object n at time t, its evolution is determined by a difference equation, homogeneous in N, of the form

$$X_n^N(t+1) = F\left(X_n^N(t), C^N(t), a_t, \xi_t\right), \quad t = 0, 1, ..., \tag{1.1}$$

where F is a known function, $C^N(t)$ is the context of the environment, a_t is the control or action selected by a central controller, and ξ_t is the random disturbance. It is assumed that $\{\xi_t\}$ is an observable sequence of independent and identically distributed random variables with a density ρ which is *unknown* by the controller. In addition, at each stage, a cost resulting from the movement of the objects and the selected control is generated. In this sense, we propose a suitable Markov control model to study this class of systems, in which the controller aims to select actions to minimize a given discounted cost criterion.

The facts that N is too large ($N \sim \infty$) as well as the lack of knowledge of density ρ, lead to formulate an alternative scheme to analyze the corresponding optimal control problem. Indeed, our approach to follow will be framed in the context of the *mean field theory*, under which, instead of analyzing a single object, we focus on the number or proportion of objects occupying certain state at each stage. This defines a control model $\mathcal{M}_N$ whose states are precisely the proportions of objects evolving according to a suitable stochastic difference equation, depending on N. Then, by taking limit as N goes to infinity, we obtain the so-called mean field control model $\mathcal{M}$, whose states are probability measures, resulting of the limit of the aforementioned proportions, which in turn satisfy a *deterministic* difference equation. In this way $\mathcal{M}$ can be considered as an approximating model for $\mathcal{M}_N$, in the sense that any optimal control policy π^* associated to $\mathcal{M}$ can be used to control the original process (the $N-$system) on $\mathcal{M}_N$, and therefore the objective is to measure its optimality deviation. Clearly, the good performance of π^* on $\mathcal{M}_N$ depends of the accuracy of the mean field limit of $\mathcal{M}_N$ to $\mathcal{M}$ as $N \to \infty$.

Because the dynamic of the objects depends on the unknown density ρ, we also have the dependence on ρ of the mean field process. Thus, besides the analysis of the limit behavior as $N \to \infty$, the controller must implement a statistical estimation procedure for ρ in order to get some information about the dynamic of the objects. To this end, at each stage t, ρ is estimated from historical observations $\xi_0, \xi_1, ...\xi_t$, collected during the evolution of the system. Such estimation procedure is then combined with the minimization task for obtaining control policies. However, as is well-known, the discounted criterion depends strongly on the decisions selected in the first stages, precisely where the information on ρ is rather poor or deficient. This fact implies that, in general, under discounted criteria, procedures of estimation and control do not provide optimal policies (see, e.g., [10,11,13,19]). Thus, in this paper we seek optimality results in a weaker sense: the so-named eventually asymptotically optimality.

In the last years, mean field theory has become a useful tool to study systems composed of a large number of objects (particles or agents) under several scenarios: discrete- and continuous-time systems of interacting objects, mean field control problems, and mean field games; all of them according to different optimality criteria, and with applications, for instance, in statistical physics, finance, operations research, among others—see, e.g., [1–3,5–9,14–18,20] and the references therein.

In particular, the motivation of our results comes from the work [7], in which the authors consider the dynamic of each object to be represented by means of a known stochastic kernel K that depends on both, the environment and the actions selected by the controller. The main purpose of that work is to study, among other things, the speed of convergence of the N-system as $N \to \infty$ as well as to obtain bounds for the gap between the cost of the N-system and the corresponding one associated to the mean field model. In contrast, in this paper we assume that the N-system is modeled by the stochastic difference equation established in (1.1) where the density of the random disturbances becomes unknown for the controller. This constitutes the main feature of our model and the novelty of our paper. That is, our approach consists in to analyze estimation and control schemes on the mean field model $\mathcal{M}$, and then study the optimality, on the model $\mathcal{M}_N$, of the resulting policies. Hence, through a joint analysis of the mean field limit ($N \to \infty$) as well as of the estimation process ($t \to \infty$), we construct control policies that are nearly optimal for the control model $\mathcal{M}_N$, in a asymptotic sense as N goes to infinity. This class of policies are called *eventually asymptotically optimal policies*.

The paper is organized as follows. In Sect. 2 we present the system of N objects together with its corresponding Markov control model, whereas Sect. 3 is devoted to study the mean field control model we are concerned with. In both sections we provide optimality results ensuring the existence of minimizers based on the dynamic programming method. In Sect. 4 we introduce the estimation and control procedure in the mean field model to construct control policies. Finally, we conclude, in Sect. 5, with the analysis of the mean field convergence, providing, among other facts, the so-called eventually asymptotically optimal policies. Throughout the paper, we shall be developing a class of consumption-investment problem to illustrate our assumptions and results .

Notation As usual, $\mathbb{N}$ (respectively $\mathbb{N}_0$) denotes the set of positive (resp. nonnegative) integers; $\mathbb{R}$ (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 $\mathbb{M}(Z)$ be the set of finite signed measures on Z. If $Z \subset \mathbb{R}$ is finite, e.g. $Z = \{1, 2, ..., z\}$, we will identify any $p \in \mathbb{M}(Z)$ by the vector $p := (p(1), p(2), ..., p(z))$. In particular, consider $\mathbb{P}(Z) \subset \mathbb{M}(Z)$ the set of probability measures on Z. In this case, any $p \in \mathbb{P}(Z)$ can be expressed in terms of its probability distribution $\{p(i) : i \in Z\}$ where $p(i) \geq 0$, $i \in Z$, and $\sum_{i=1}^{z} p(i) = 1$. Observe that, under the topology of $\mathbb{R}$, $Z = \{1, 2, \cdots, z\}$ becomes a Borel set, and so is $\mathbb{P}(Z)$. As usual, $|\cdot|$ will denote the norm on $\mathbb{R}$.

We shall define the norm on $\mathbb{M}(Z) \times \mathbb{R}^d$, for Z finite under the corresponding L_∞ norm; that is, for each vector $(p, c) \in \mathbb{M}(Z) \times \mathbb{R}^d$:

$$\|(p, c)\|_\infty := \max\left\{\|p\|_\infty^1, \|c\|_\infty^2\right\},$$

where $\|p\|_\infty^1 := \max\{|p(1)|, ..., |p(z)|\}$, and $\|c\|_\infty^2 := \max\{|c_1|, ..., |c_d|\}$, with $c := (c_1, \cdots, c_d)$. Furthermore, for a given Borel space A, d_A will represent its associated metric. For all $(p, c, a), (p', c', a') \in \mathbb{P}(Z) \times \mathbb{R}^d \times A$ the corresponding L_∞−distance takes the form

$$\left\|(p, c, a) - (p', c', a')\right\|_\infty^3 := \max\left\{\left\|p - p'\right\|_\infty^1, \left\|c - c'\right\|_\infty^2, d_A(a, a')\right\},$$

whereas for a matrix $A_{n\times n}$, we will denote its corresponding norm $\|\cdot\|_\infty^0$ as

$$\|A\|_\infty^0 := \max_{i,j} |A_{ij}|.$$

Let Z and A be Borel spaces. A stochastic kernel $Q(\cdot|\cdot)$ is a function $Q : \mathcal{B}(Z) \times A \to [0, 1]$, such that $Q(\cdot|a)$ is a probability measure on $\mathcal{B}(Z)$ for each fixed $a \in A$, and $Q(B|\cdot)$ is a measurable function on A for each fixed $B \in \mathcal{B}(Z)$. Finally, $\mathbb{B}(Z)$ denotes the class of real-valued bounded functions on Z endowed with the supremum norm $\|v\| := \sup_{z\in Z} |v(z)|$, while $\mathbb{C}_b(Z)$ is the subspace of $\mathbb{B}(Z)$, consisting of all real-valued bounded continuous functions defined on Z.

We assume the existence of a fixed probability space $(\Omega, \mathcal{F}, P)$, and for the attribute a.s. we mean almost sure with respect to P.

2 The N-Objects Markov Control Model

We consider a discrete-time controlled stochastic system composed by a large number N of interacting objects defined as follows. Let $X_n^N(t)$, $n = 1, 2, \ldots, N$, $t \in \mathbb{N}_0$ be the state of the object n at time t taking values in a given set $S = \{1, 2, \ldots, s\} \subseteq \mathbb{N}$. There is a controller (or decision-maker) who, at each stage, can influence the behavior of the objects by means of actions or controls a_t selected from a given Borel set A. Moreover, the objects are assumed to share a common environment which also influences the behavior of the system. Let $C^N(t) \in \mathbb{R}^d$ be the context of the environment at time $t \in \mathbb{N}_0$. Once the environment is specified, the behavior as well as the evolution of the objects are considered to be independent each other. More specifically, the evolution of the process $\left\{X_n^N(t)\right\}_{t\in\mathbb{N}_0}$ is given according to the stochastic difference equation, homogeneous in N, defined in (1.1); that is,

$$X_n^N(t+1) = F\left(X_n^N(t), C^N(t), a_t, \xi_t\right), \quad t = 0, 1, \ldots, \tag{2.1}$$

where $F : S \times \mathbb{R}^d \times A \times \mathbb{R} \to S$ is a given (known) function and $\{\xi_t\}$ is a sequence of independent and identically distributed (i.i.d.) real random variables with a common density ρ which is *unknown* for the controller, and defined on the underlying probability space $(\Omega, \mathcal{F}, P)$. As a consequence of the above definitions, it is possible to define the transition law K_ρ of each object in terms of the function F, as follows: For all $n = 1, 2, \ldots, N$

$$\begin{aligned} K_{ij}^{\rho}(a,c) &:= P\Big[X_n^N(t+1) = j | X_n^N(t) = i, a_t = a, C^N(t) = c\Big] \\ &= \int_{\mathbb{R}} I_j\left[F(i,c,a,z)\right]\rho(z)dz, \;\; i,j \in S, \;\; (a,c) \in A \times \mathbb{R}^d. \end{aligned} \tag{2.2}$$

where I_B stands for the indicator function of the set B. This relation defines the transition law by means of the stochastic kernel $K_\rho = K_\rho(a,c) = \Big[K_{ij}^{\rho}(a,c)\Big]$. Notice that K_ρ represents the common conditional distribution of the states.

Throughout this work it is assumed that the objects are observable through their states, so that the controller can only determine the *number* of objects in each of the states $i \in S$. In this sense, the behavior of the system can be reformulated by means of the *proportions* of the objects at each state. Namely, let $M_i^N(t)$ be the proportion of objects in state $i \in S$ at time t defined as

$$M_i^N(t) := \frac{1}{N}\sum_{n=1}^{N} I_{\{X_n^N(t)=i\}}, \; i \in S.$$

Further, we denote by $\vec{M}^N(t)$ the vector whose components are the proportions; that is,

$$\vec{M}^N(t) = \Big(M_1^N(t), M_2^N(t), \ldots, M_s^N(t)\Big).$$

Observe that $\vec{M}^N(t) \in \mathbb{P}_N(S) := \{p \in \mathbb{P}(S) : Np(i) \in \mathbb{N}, \; \forall i \in S\} \subset \mathbb{P}(S)$, and it is easy to see that $\mathbb{P}_N(S)$ is a finite set.

In addition, we suppose that the context of the environment is a dynamical system whose evolution is determined by the difference equation:

$$C^N(t+1) = g\left(C^N(t), \vec{M}^N(t+1), a_t\right), \quad t \in \mathbb{N}_0, \tag{2.3}$$

where $g : \mathbb{R}^d \times \mathbb{P}(S) \times A \to \mathbb{R}^d$ is a known function.

Let us assume now the evolution of $\vec{M}^N(\cdot)$ in a recursive way through a difference equation. Clearly, such an evolution is strongly dependent on the transition law K_ρ of the objects, and as a consequence on the unknown density ρ. Hence, we assume the existence of a measurable function $G_\rho^N : \mathbb{P}_N(S) \times \mathbb{R}^d \times A \times \mathbb{R}^N \to \mathbb{P}_N(S)$ such that

$$\vec{M}^N(t+1) = G_\rho^N\left(\vec{M}^N(t), C^N(t), a_t, \vec{w}_t\right), \tag{2.4}$$

where $\{\vec{w}_t\}$ is a sequence of i.i.d. random vectors on $\mathbb{R}^N$, with common distribution θ.

For ease notation, we denote $\mathbb{Y}_N := \mathbb{P}_N(S)\times\mathbb{R}^d$, and let $H_\rho^N : \mathbb{Y}_N\times A\times\mathbb{R}^N \to \mathbb{Y}_N$ be the function defined as

$$H_\rho^N(y,a,w) := \left(G_\rho^N(y,a,w), g\left(c, G_\rho^N(y,a,w), a\right)\right). \tag{2.5}$$

Then, denoting $y^N(t) := \left(\vec{M}^N(t), C^N(t)\right)$, according to (2.3) and (2.4), H_ρ^N defines the dynamic of the process $\left\{y^N(t)\right\}$; that is,

$$\begin{aligned} y^N(t+1) &= \left(G_\rho^N\left(y^N(t), a_t, \vec{w}_t\right), g\left(C^N(t), \vec{M}^N(t+1), a_t\right)\right) \\ &= \left(G_\rho^N\left(y^N(t), a_t, \vec{w}_t\right), g\left(C^N(t), G_\rho^N(y^N(t), a_t, \vec{w}_t), a_t\right)\right) \\ &= H_\rho^N\left(y^N(t), a_t, \vec{w}_t\right). \end{aligned} \tag{2.6}$$

Finally, a cost depending on the proportion of the objects, on the environment, and on the selected control, is generated at each stage. This cost will be represented by the measurable function $r : \mathbb{P}(S)\times\mathbb{R}^d\times A \to \mathbb{R}$.

Let us consider the space $\mathbb{Y} := \mathbb{P}(S)\times\mathbb{R}^d$. Observe that $\mathbb{Y}_N := \mathbb{P}_N(S)\times\mathbb{R}^d \subseteq \mathbb{Y}$ and the one-stage cost can be then redefined as $r : \mathbb{Y}\times A \to \mathbb{R}$.

2.1 Formulation of the N-Markov Control Model (N-MCM)

We define the discrete-time Markov control model associated to the system of N objects previously introduced (in short N-MCM) as follows:

$$\mathcal{M}_N := \left(\mathbb{Y}_N, A, H_\rho^N, \theta, r\right). \tag{2.7}$$

The model $\mathcal{M}_N$ describes the performance of the system in the following sense: at time t, the controller observes the state $y = y^N(t) = (\vec{M}^N(t), C^N(t)) \in \mathbb{Y}_N$ which is composed by both the proportions of the objects and the context of the environment, and then he/she selects a control $a = a_t \in A$. As consequence the following happens: (1) a cost $r(y,a)$ is incurred, and (2) the system moves to a new state $y' = y^N(t+1) = (\vec{M}^N(t+1), C^N(t+1))$ according to the transition law

$$\begin{aligned} Q_\rho(B|y,a) &:= P\left[y^N(t+1)\in B | y^N(t) = y, a_t = a\right] \\ &= \int_{\mathbb{R}^N} I_B\left[H_\rho^N(y,a,w)\right]\theta(dw), \end{aligned}$$

with H_ρ^N as in (2.5). Once the transition to the state y' occurs, the procedure is repeated. In addition, we will assume that the one-stage costs are accumulated during the evolution of the system in an *infinite* horizon by using a given discounted cost criterion, and therefore the actions selected by the controller are directed to minimize the *total expected discounted cost* introduced in (2.22) below.

In order to ensure the existence of minimizers, we impose the following continuity and compactness conditions on some elements of $\mathcal{M}_N$.

Assumption 2.1 (a) The control space A is a compact metric Borel space, whose metric is denoted by d_A.

(b) The function g in (2.3) is a Lipschitz function with constant L_g; that is, for $c, c' \in \mathbb{R}^d$, $\vec{m}, \vec{m}' \in \mathbb{P}(S)$, $a, a' \in A$,

$$\left\| g(c, \vec{m}, a) - g(c', \vec{m}', a') \right\|_\infty^2 \leq L_g \max \left\{ \left\| c - c' \right\|_\infty^2, \left\| \vec{m} - \vec{m}' \right\|_\infty^1, d_A(a, a') \right\}. \tag{2.8}$$

Without lost of generality, we assume that $L_g \geq 1$.

(c) The mapping $a \longmapsto H_\rho^N(y, a, w)$ defined in (2.5) is continuous, for all $y \in \mathbb{Y}_N$ and $w \in \mathbb{R}^N$.

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

$$|r(y, a)| \leq R \, \forall (y, a) \in \mathbb{Y} \times A,$$

and for every $a, a' \in A$, and $y, y' \in \mathbb{Y}$,

$$\sup_{(a, a') \in A \times A} |r(y, a) - r(y', a')| \leq L_r \left\| y - y' \right\|_\infty .$$

2.2 A consumption-Investment Model with Controlled Subsidy/Fee

We consider a consumption-investment system composed by N "small" investors (i.e., economic agents whose actions do not influence the market prices) which invest among various assets with different return rates, but that also consume some specific product. There is a central controller, for instance the government or a public body, who provides a subsidy to assist the investors or imposes a fee that the investors must pay. For simplicity, we shall consider only two assets for the investors: one of them is a risk-free asset with fixed rate τ, and the other a risk asset with a stochastic return rate ξ_t taking values in a bounded set $Z \subseteq \mathbb{R}$. The fraction associated to the wealth to be invested in the risky asset is a function $\varphi_1 : \mathbb{R}^d \rightarrow [0, 1]$ that depends on the context of the environment; this environment might be, for example, uncertainty of the investors, type of markets that investors are trading, frecuency of transactions, etc. In an analogous way, the quantity $(1 - \varphi_1)$ will represent the fraction of wealth to be invested in the risk-free asset. On the other hand, we will assume that each investor consumes a quantity $\varphi_2 : \mathbb{R}^d \rightarrow \mathbb{R}_+$ that is also a bounded function dependent on the context of the environment.

In the spirit of our assumption, since the state space S is denumerable, we shall assume that the use of cents is negligible. Hence, let a_t be the decision of the central controller at time t which is assumed to satisfy $a_t \in \{0, \pm 1, ... \pm a^*\} =: A$ for some $a^* \geq 0$. That is

$$a_t = \text{fee of size } -a_t \text{ (if } a_t < 0) \text{ or subsidy of size } a_t \text{ (if } a_t > 0), \text{ at time } t.$$

Denoting by $X_n^N(t) \in \{0, 1, \cdots, s\} = S$ the wealth of the investor n at time t, we can represent this process by means of the following difference equation

$$\begin{aligned} X_n^N(t+1) = \operatorname{int}\Big\{\Big[(1-\varphi_1(C^N(t)))(1+\tau) + \varphi_1(C^N(t))\xi_t\Big] \\ \times \Big[X_n^N(t) - \varphi_2(C^N(t)) + a_t\Big]\Big\}, \end{aligned} \tag{2.9}$$

where int$\{x\}$ is the integer part of x. It is assumed that $s \in \mathbb{N}_0$ is sufficiently large, and the functions φ_m, $m = 1, 2$, satisfy the Lipschitz conditions with constants L_{φ_m}, $m = 1, 2$, respectively, taking values in appropriate sets such that the following holds true

$$F(i, c, a, z) := \operatorname{int}\Big\{ [(1-\varphi_1(c))(1+\tau) + \varphi_1(c)z]\,[i - \varphi_2(c) - a] \Big\} \in S. \tag{2.10}$$

Furthermore, using the Lipschitz properties of φ_1 and φ_2, we can deduce that F is in fact a Lipschitz function in the following sense:

$$\big|F(i, c, a, z) - F(i, c', a', z)\big| \leq L_F \max\Big\{\|c - c'\|_\infty^2,\ |a - a'|\Big\}, \tag{2.11}$$

where

$$\begin{aligned} L_F = 1 + (1 + \tau + \max\nolimits_{z\in Z}|z|)\left(L_{\varphi_1}s + \bar{L}_{\varphi_1}L_{\varphi_2} + \bar{L}_{\varphi_2}L_{\varphi_1} + a^* L_{\varphi_1} + L_{\varphi_2}\right) \\ +(1+\tau)(1+L_{\varphi_2}) \end{aligned}$$

and $\bar{L}_{\varphi_m}$ represents some (uniform) bound of φ_m, $m = 1, 2$.

Assuming that ρ is the density of the random rate ξ_t, the transition law turns out to be

$$K_{ij}^\rho(a, c) = \int_{\mathbb{R}} I_j\,[F(i, c, a, z)]\,\rho(z)dz, \tag{2.12}$$

for each $i, j \in S$ and $(a, c) \in A \times \mathbb{R}^d$. Further, since F is an $S-$valued function and $S := \{0, 1, \cdots, s\}$ is finite, it is easy to see that, for all $i, j \in S$, $a, a' \in A$, $c, c' \in \mathbb{R}^d$, the indicator function satisfies

$$\begin{aligned} \big|I_j[F(i, c, a, z)] - I_j[F(i, c', a', z)]\big| &\leq \big|F(i, c, a, z) - F(i, c', a', z)\big| \\ &\leq L_F \max\Big\{\big\|c - c'\big\|_\infty^2, d_A(a, a')\Big\}, \end{aligned}$$

where the last inequality is due to Lipschitz property of F given in (2.11). Hence, from (2.2),

$$\begin{aligned} \Big|K_{ij}^\rho(a, c) - K_{ij}^\rho(a', c')\Big| &\leq \int \big|I_j[F(i, c, a, z)] - I_j[F(i, c', a', z)]\big|\,\rho(z)dz \\ &\leq L_F \max\Big\{\big\|c - c'\big\|_\infty^2, d_A(a, a')\Big\}, \end{aligned} \tag{2.13}$$

which implies that K_ρ is Lipschitz.

On the other hand, for each $i \in S$, the evolution of the proportions $M_i(t)$ of the investors can be seen in a recursive way as follows (see [7])

$$M_i^N(t+1) = \frac{1}{N} \sum_{k=0}^{s} \sum_{n=1}^{NM_k^N(t)} I_{\{A_{ki}^\rho(a_t, C^N(t))\}}(w_n^k(t)), \tag{2.14}$$

where $w_n^k(t)$ are i.i.d. random variables uniformly distributed on [0, 1],

$$A_{ki}^\rho(a,c) := \left[\Gamma_{ki}^\rho(a,c), \Gamma_{ki+1}^\rho(a,c)\right] \subseteq [0,1], \tag{2.15}$$

and

$$\Gamma_{ki}^\rho(a,c) := \sum_{l=0}^{i-1} K_{kl}^\rho(a,c), \;\; k, i \in S. \tag{2.16}$$

For each $i \in S$ and $t \in \mathbb{N}_0$, we denote

$$\vec{w}^i(t) := \left(w_1^i(t), \cdots, w_{NM_i^N}^i(t)\right)$$

and

$$\vec{w}_t := \left(\vec{w}^0(t), \cdots, \vec{w}^s(t)\right).$$

It is worth noting that $\sum_{i=0}^{s} NM_i^N(t) = N$, thus $\vec{w}_t \in [0,1]^N$. This assertion implies that the number of (uniform) random variables involved in the dynamic (2.14) coincides with the number N of small agents; a fact that is presented in a general way through the dynamic (2.4).

Let us now rewrite the above expressions as in (2.6); namely, we define

$$G_{\rho,i}^N\left(y^N(t), a_t, \vec{w}_t\right) := \frac{1}{N} \sum_{k=0}^{s} \sum_{n=1}^{NM_k^N(t)} I_{\{A_{ki}^\rho(C^N(t), a_t)\}}(w_n^k), \;\; i \in S.$$

This function G_ρ^N takes the following vectorial form

$$G_\rho^N(y,a,w) = \left(G_{\rho,0}^N(y,a,w), \ldots, G_{\rho,s}^N(y,a,w)\right), \;\; (y,a,w) \in \mathbb{Y}_N \times A \times [0,1]^N, \tag{2.17}$$

yielding to the following expression

$$\vec{M}^N(t+1) = G_\rho^N\left(\vec{M}^N(t), C^N(t), a_t, \vec{w}_t\right). \tag{2.18}$$

In addition, recalling that $\mathbb{P}(S)$ denotes the space of probability measures on S, we assume that $g : \mathbb{R}^d \times \mathbb{P}(S) \times A \to \mathbb{R}$ is an arbitrary function satisfying Assumption 2.1(b), such that the context of the environment satisfies

$$C^N(t+1) = g\left(C^N(t), \vec{M}^N(t+1), a_t\right), \quad t \in \mathbb{N}_0. \tag{2.19}$$

Then, (2.18) and (2.19) define the function

$$H_\rho^N(y, a, w) := \left(G_\rho^N(y, a, w), g(c, G_\rho^N(y, a, w), a)\right), \tag{2.20}$$

which determines the dynamic of the process $\left\{y^N(t)\right\}$ similar to (2.6).

Finally, since the action space A is denumerable, the continuity of $a \longmapsto H_\rho^N(\cdot, a, \cdot)$, required in Assumption 2.1(c), trivially holds.

Remark 2.2 In the case when $A \subset \mathbb{R}$ is an arbitrary compact set, say $A = [-a^*, a^*]$ for some $a^* \geq 0$, the continuity of the function H_ρ^N, can be verified as follows. For $i, j \in S$, $w \in [0, 1]$, $c \in \mathbb{R}^d$, and $a \in [-a^*, a^*]$, let $\delta_w(A_{ij}^\rho(c, a))$ be the Dirac measure corresponding to the indicator function $I_{\{A_{ij}^\rho(c,a)\}}(w)$ (see 2.15, 2.16). Now take a sequence $\{a_k\} \in [-a^*, a^*]$ such that $a_k \to a \in [-a^*, a^*]$, which is possible because $[-a^*, a^*]$ is a compact set. Since $a \longmapsto K_{ij}^\rho(a, c)$ is continuous for all $i, j \in S$ and $c \in \mathbb{R}^d$, so is the mapping $a \longmapsto \Gamma_{ij}^\rho(a, c)$. Hence, $A_{ij}^\rho(c, a_k) \to A_{ij}^\rho(c, a)$ as $k \to \infty$ in the set sense. Therefore, due to the fact that $\delta_w(\cdot)$ is a probability measure (so it is continuous), we conclude that $\delta_w(A_{ij}^\rho(c, a_k) \to \delta_w(A_{ij}^\rho(c, a))$, as $k \to \infty$. This fact and the continuity of the function g given in Assumption 2.1(b) yield the continuity of the map $a \longmapsto H_\rho^N(\cdot, a, \cdot)$.

2.3 Optimality in the *N*-MCM

In this subsection we introduce the elements that define the optimal control problem as well as the results regarding existence of optimal policies respect to the discounted criterion, associated to the N-MCM (2.7).

Control policies The actions applied by the controller are selected according to rules known as control policies, which are defined as follows. Let $\mathbb{H}_0^N := \mathbb{Y}_N$ and $\mathbb{H}_t^N := \left(\mathbb{Y}_N \times A \times \mathbb{R} \times \mathbb{R}^N\right)^t \times \mathbb{Y}_N, t \geq 1$, be the space of histories up to time t. An element h_t^N of $\mathbb{H}_t^N$ is written as

$$h_t^N = \left(y^N(0), a_0, \xi_0, \vec{w}_0, \ldots, y^N(t-1), a_{t-1}, \xi_{t-1}, \vec{w}_{t-1}, y^N(t)\right),$$

where $y^N(t) = \left(\vec{M}^N(t), C^N(t)\right)$. A control policy is a sequence $\pi^N = \left\{\pi_t^N\right\}$ of stochastic kernels π_t^N on A given $\mathbb{H}_t^N$ such that $\pi_t^N\left(A|h_t^N\right) = 1$ for all $h_t^N \in \mathbb{H}_t^N$, $t \in \mathbb{N}_0$. We denote by Π^N the set of all control policies.

Now, let $\mathbb{F}$ be the set consisting of all measurable functions $f : \mathbb{Y} \to A$ and $\mathbb{F}^N := \mathbb{F}|_{\mathbb{Y}^\mathbb{N}}$ be the restriction of $\mathbb{F}$ over $\mathbb{Y}^N$. A policy $\pi^N \in \Pi^N$ is said to be a (deterministic) Markov policy if there exists a sequence $\left\{f_t^N\right\} \subseteq \mathbb{F}^N$ such that for all $t \in \mathbb{N}_0$ and $h_t^N \in \mathbb{H}_t^N$, $\pi_t^N\left(\cdot|h_t^N\right) = \delta_{f_t^N(y^N(t))}(\cdot)$. In this case π^N takes the form $\pi^N = \left\{f_t^N\right\}$. In particular, if $f_t^N \equiv f^N$ for some $f^N \in \mathbb{F}$ and for all $t \in \mathbb{N}_0$, we say that π^N is a *stationary* policy. We denote by Π_M^N the set of all Markov policies, and following a standard convention, we shall use the same notation of $\mathbb{F}^N$ to denote the set of stationary policies.

Remark 2.3 (a) We denote by Π_M the set of deterministic Markov policies when we use $\mathbb{F}$ instead of $\mathbb{F}^N$ in the above definition; that is, Π_M is the family of sequences of functions $\{f_t\} \subset \mathbb{F}$. Observe that any policy $\pi = \{f_t\} \in \Pi_M$ whose elements f_t are restricted to $\mathbb{Y}_N$ turns out to be an element of Π^N.

(b) Under standard arguments (see, e.g., [12]), for each $\pi^N \in \Pi^N$ and initial state $y^N(0) = y \in \mathbb{Y}_N$, there exists a probability space $\left(\Omega', \mathcal{F}', P_y^{\pi^N}\right)$ consisting in $\Omega' := \left(\mathbb{Y}_N \times A \times \mathbb{R} \times \mathbb{R}^N\right)^\infty$, $\mathcal{F}'$ its respective $\sigma-$algebra, and a probability measure $P_y^{\pi^N}$ satisfying the following properties: For each $t \in \mathbb{N}_0$

(i) $P_y^{\pi^N}(y^N(0) \in B) = \delta_y(B), \quad B \in \mathcal{B}(\mathbb{Y}_N)$,

(ii) $P_y^{\pi^N}(a_t \in C|h_t^N) = \pi_t^N(C|h_t^N), \quad C \in \mathcal{B}(A)$,

(iii) (Like-Markov property):

$$\begin{aligned} P_y^{\pi^N}\left[y^N(t+1) \in B|h_t^N, a_t\right] &= Q_\rho\left(B|y^N(t), a_t\right) \\ &= \int_{\mathbb{R}^N} I_B\left[H_\rho^N\left(y^N(t), a_t, w\right)\right]\theta(dw), \\ &\quad B \in \mathcal{B}(\mathbb{Y}_N). \end{aligned} \tag{2.21}$$

The discounted optimality criterion For each control policy $\pi^N \in \Pi^N$and initial state $y^N(0) = y \in \mathbb{Y}_N$, we define the total expected discounted cost as

$$V^N(\pi^N, y) := E_y^{\pi^N}\sum_{t=0}^{\infty}\alpha^t r\left(y^N(t), a_t\right), \tag{2.22}$$

where $\alpha \in (0, 1)$ is the so-called discount factor and $E_y^{\pi^N}$ denotes the expectation operator with respect to the probability measure $P_y^{\pi^N}$ induced by the policy π^N given $y^N(0) = y$. We say that π_*^N is optimal for the N-MCM if and only if

$$V_*^N(y) := \inf_{\pi^N \in \Pi^N} V^N\left(\pi^N, y\right) = V^N\left(\pi_*^N, y\right), \quad y \in \mathbb{Y}_N. \tag{2.23}$$

In this case, V_*^N is said to be the $N-$*value function*.

Under the conditions imposed on the N-MCM $\mathcal{M}_N$, we can state the following well known result that provides a characterization on the optimal policies and on the N-value function in terms of the solution of a certain functional equation so-called the N- optimality equation (see, e.g., [11,21]):

Proposition 2.4 *(a) The $N-$value function V_*^N satisfies the $N-$optimality equation*

$$V_*^N(y) = \min_{a\in A}\left\{r(y,a) + \alpha\int_{\mathbb{R}^N} V_*^N\left[H_\rho^N(y,a,w)\right]\theta(dw)\right\},\quad y\in\mathbb{Y}_N. \quad (2.24)$$

In addition,

$$\left|V_*^N(y)\right| \le \frac{R}{1-\alpha},\quad y\in\mathbb{Y}_N,$$

with R being the (uniform) bound of the one-stage cost r defined in Assumption 2.1(d), and α the discount factor in (2.22).

(b) There exists $f_*^N \in \mathbb{F}^N$ such that $f_*^N(y)\in A$ attains the minimum in (2.24), i.e.,

$$V_*^N(y) = r(y,f_*^N) + \alpha\int_{\mathbb{R}^N} V_*^N\left[H_\rho^N\left(y,f_*^N,w\right)\right]\theta(dw),\quad y\in\mathbb{Y}_N, \quad (2.25)$$

and furthermore, the stationary policy $\pi_*^N = \left\{f_*^N\right\} \in \Pi_M^N$ is optimal for the control model $\mathcal{M}_N$.

Proposition 2.4 provides a flexible framework for the optimality analysis of the interacting objects system. However, from the practical point of view, its usefulness is seriously limited either because N is too large ($N\sim\infty$) or for the lack of knowledge of density ρ. Indeed, to analyze equations (2.24) and (2.25), we first need to deal with a multiple integral of dimension N which could be considerably difficult to calculate, besides that the dynamics of the system depends heavily on the unknown density ρ. Both situations will be discussed in the following sections in order to overcome these obstacles. Specifically, we first introduce a suitable control model $\mathcal{M}$ that represents the "limit model" of $\mathcal{M}_N$ as $N\to\infty$; this new model is referred to as the *mean field control model*, and of course also depends on the unknown density ρ. Hence, we pose the mean field control problem which is independent of N, but dependent on ρ. Then, in Sect. 4, an statistical estimation and control procedure is proposed to construct nearly optimal policies for the control model $\mathcal{M}_N$ in an asymptotic sense when $N\to\infty$. In other words, $\mathcal{M}$ is used as an approximating model for $\mathcal{M}_N$, and the hope is that optimal policies in $\mathcal{M}$ have a good performance in $\mathcal{M}_N$, whenever the model $\mathcal{M}$ gives a good approximation to the model $\mathcal{M}_N$.

3 The Mean Field Control Model

Recall the set $\mathbb{Y} = \mathbb{P}(S)\times\mathbb{R}^d$. We consider a general controlled deterministic system $\{(\vec{m}(t),c(t))\}\in\mathbb{Y}$ that depends implicitly on the distribution ρ in (2.2) and whose dynamic is governed by means of the following difference equations

$$\vec{m}(t+1) = G_\rho\left(\vec{m}(t),c(t),a_t\right); \quad (3.1)$$

$$c(t+1) = g\left(c(t),\vec{m}(t+1),a_t\right), \quad (3.2)$$

where $(\vec{m}(0), c(0)) = (\vec{m}, c) \in \mathbb{Y}$ represents the initial condition, $a_t \in A$ is the control (or action) selected at time t, $g : \mathbb{R}^d \times \mathbb{P}(S) \times A \to \mathbb{R}^d$ is the function defined in (2.3), and $G_\rho : \mathbb{P}(S) \times \mathbb{R}^d \times A \to \mathbb{P}(S)$ is a known Lipschitz function (dependent on ρ) with constant L_G; that is, for $\vec{m}, \vec{m}' \in \mathbb{P}(S)$, $c, c' \in \mathbb{R}^d$, and $a, a' \in A$,

$$\left\| G_\rho(\vec{m}, c, a) - G_\rho(\vec{m}', c', a') \right\|_\infty^1 \leq L_G \max \left\{ \left\| \vec{m} - \vec{m}' \right\|_\infty^1, \left\| c - c' \right\|_\infty^2, d_A(a, a') \right\}. \tag{3.3}$$

Due to the deterministic nature of the process (3.1)–(3.2), it is evident that the dynamic is completely determined by the sequence of actions $\{a_t\} \subset A$ and by the initial condition $(\vec{m}, c) \in \mathbb{Y}$. Furthermore, we will assume (see Assumption 5.1 below) that the process $y(t) := (\vec{m}(t), c(t))$ represents the mean field limit; that is, $y(t)$ will be the limit process of $y^N(t) := (\vec{M}^N(t), C^N(t))$ in (2.6) as N goes to infinity.

Let $H_\rho : \mathbb{Y} \times A \to \mathbb{Y}$ be the function that defines the dynamic of the process $\{(\vec{m}(t), c(t))\}$; that is,

$$H_\rho(y, a) := \big(G_\rho(\vec{m}, c, a), g(c, G_\rho(\vec{m}, c, a), a)\big), \quad y = (\vec{m}, c) \in \mathbb{Y}, \ a \in A. \tag{3.4}$$

From (3.1) and (3.2), we can write

$$\begin{aligned} y(t+1) &= \big(G_\rho(\vec{m}(t), c(t), a_t), g(c(t), \vec{m}(t+1), a_t)\big) \\ &=: H_\rho(y(t), a_t), \quad t \geq 0, \end{aligned} \tag{3.5}$$

with $y(0) = (\vec{m}, c) \in \mathbb{P}(S) \times \mathbb{R}^d$. A straightforward calculation yields that the function H_ρ is a Lipschitz function (recall G_ρ and g are Lipschitz functions). Specifically, for $(y, a), (y', a') \in \mathbb{Y} \times A$,

$$\left\| H_\rho(y, a) - H_\rho(y', a') \right\|_\infty \leq L_{H_\rho} \max \left\{ \left\| y - y' \right\|_\infty, d_A(a, a') \right\}, \tag{3.6}$$

where $L_{H_\rho} = \max \left\{ L_g, L_g L_G \right\}$. Using the same one-stage cost r defined for the N-MCM (2.7), we can then define the mean field control model as

$$\mathcal{M} = (\mathbb{Y}, A, H_\rho, r),$$

which has a similar interpretation as the $N-$MCM $\mathcal{M}_N$.

Example 3.1 (**Consumption-investment problem**) Carrying on with our example, we define the controlled deterministic system $\{(\vec{m}(t), c(t))\} \in \mathbb{Y}$ as (see [7])

$$\vec{m}(t+1) = \vec{m}(t) K_\rho(a_t, c(t)) \tag{3.7}$$

$$c(t+1) = g(c(t), \vec{m}(t+1), a_t). \tag{3.8}$$

where K_ρ becomes the matrix $[K_{ij}^\rho]$, whose elements turn out to be stochastic kernels defined in (2.12) and $g : \mathbb{R}^d \times \mathbb{P}(S) \times A \to \mathbb{R}^d$ is the function defined in (2.19). Observe that $\vec{m}(t+1)$ is the vector with components

$$m_j(t+1) = \sum_{i=1}^{s} m_i(t) K_{ij}^{\rho}(a_t, c(t)),$$

where $\vec{m}(0) = m \in \mathbb{P}(S)$. In this case the function G_ρ in (3.1) takes the form

$$G_\rho(\vec{m}, c, a) = \vec{m} K_\rho(a, c), \quad (\vec{m}, c) \in \mathbb{Y}, a \in A, \tag{3.9}$$

and, since the kernel K_ρ is Lipschitz (see (2.13)), so is G_ρ, as was stated in (3.3), with some constant L_G.

In Sect. 5 we will show that (3.7)-(3.8) are in fact the limit processes of (2.18)–(2.19). □

3.1 Optimality in the Mean Field

In this subsection we present a well-known theory regarding optimality results for the controlled system (3.1)–(3.2) when using the deterministic discounted criterion (3.10). Basically these results show characterizations on the optimal policies and on the corresponding value function in the sense that these optimal quantities become solutions of a given functional equation associated to the mean field control model.

As is well-known (see, e.g., [4]), for the deterministic controlled systems, a control policy π is a sequence of decision rules (or selectors) $\pi = \{f_t\} \subset \mathbb{F}$. Therefore, according to the Remark 2.3(a), we can naturally consider the set Π_M as the set of all control policies for the model $\mathcal{M}$. Hence, given a control policy $\pi \in \Pi_M$ together with the initial condition $y(0) = y \in \mathbb{Y}$, we define the total discounted cost for the mean field model as

$$v(\pi, y) = \sum_{t=0}^{\infty} \alpha^t r\left(y(t), a_t\right). \tag{3.10}$$

Then, the mean field optimal control problem is to find a policy $\pi_* \in \Pi_M$ such that

$$v_*(y) := \inf_{\pi \in \Pi_M} v(\pi, y) = v(\pi_*, y), \quad y \in \mathbb{Y}, \tag{3.11}$$

where v_* is the *mean field value function* and π_* is said to be an *optimal policy* for the mean field control model $\mathcal{M}$.

Observe that from the continuity of the function H_ρ [see (3.6)], the compactness of the control space A, and the continuity of the one-stage cost r, we can state the following result regarding the value function (see, e.g., [11,21]).

Proposition 3.2 *(a) The value function v_* satisfies the mean field optimality equation*

$$v_*(y) = \min_{a \in A} \left\{ r(y, a) + \alpha v_* \left[H_\rho\left(y, a\right) \right] \right\}, \quad y \in \mathbb{Y}. \tag{3.12}$$

Equivalently,

$$\min_{a \in A} \Phi(y, a) = 0, \quad y \in \mathbb{Y},$$

where

$$\Phi(y,a) := r(y,a) + \alpha v_* \left[H_\rho(y,a) \right] - v_*(y), \tag{3.13}$$

is the so-called discrepancy function. In addition,

$$|v_*(y)| \leq \frac{R}{1-\alpha}, \quad y \in \mathbb{Y}.$$

(b) There exists $f^* \in \mathbb{F}$ such that $f^*(y) \in A$ attains the minimum in (3.12), i.e.,

$$v_*(y) = r(y, f^*) + \alpha v_* \left[H_\rho\left(y, f^*\right) \right], \quad y \in \mathbb{Y}, \tag{3.14}$$

and furthermore, the stationary policy $\pi^* = \{f^*\} \in \Pi_M$ is optimal for the control model $\mathcal{M}$.

Remark 3.3 Let $\{(y_t, a_t)\}$ be a sequence of state-action pairs corresponding to the application of a stationary policy $\pi^* = \{f^*\} \in \Pi_M$. Observe that by the optimality principle and dynamic programming arguments, π^* is an optimal policy if, and only if $\Phi(y_t, f^*(y_t)) = 0$, for all $t \in \mathbb{N}_0$.

Although the optimal value function and the optimal policy are well characterized through Proposition 3.2 and Remark 3.3, the information about the density ρ plays an important role in equations (3.12)–(3.14), and as a consequence, the optimality equation and its minimizers are highly dependent on the density ρ. However, under certain conditions, when this density is unknown, as is our case, suitable estimation-control procedures can be applied in order to find optimal policies. This point is studied in the next section.

4 Estimation and Control in the Mean Field

The main problem we address in this paper is to obtain optimality results under the assumption that the density ρ in (2.2), and as consequence the function H_ρ in (3.4)–(3.5), are unknown. In this scenario, assuming observability of the random disturbances $\xi_0, \xi_1, ...,$ the controller has to appeal to a combination of statistical estimation methods and control procedures on the mean field model $\mathcal{M}$, in order to gain some insights on the evolution of the system. That is, before choosing the action a_t at time t, the controller gets an estimate ρ_t of ρ —thus gets also an estimate $H_t = H_{\rho_t}$ of the function H_ρ—, then, the decisions of the controller are adapted to this estimate, obtaining a control $a_t = a_t(\rho_t)$.

To fix ideas, let us consider $\xi_0, \xi_1, ..., \xi_{k-1}$ be independent realizations of a random variable with the unknown density ρ observed up to time $k-1$, and let $\rho_k(\cdot) := \rho_k(\cdot; \xi_0, \xi_1, ..., \xi_{k-1})$ be a density which is an estimator such that, as $k \to \infty$

$$\int_{\mathbb{R}} |\rho_k(z) - \rho(z)|\, dz \to 0 \text{ a.s} \tag{4.1}$$

and

$$\sup_{(y,a)\in\mathbb{Y}\times A}\left\|G_{\rho_k}(y,a)-G_\rho(y,a)\right\|_\infty^1 \to 0 \text{ a.s,} \tag{4.2}$$

where $y=(\vec{m},c)$, and for each $k\in\mathbb{N}$, G_{ρ_k} is the function defining the dynamic of the process $\{\vec{m}(t)\}$ [see (3.1)] when the density ρ_k is used instead of ρ. Thus, G_{ρ_k} defines a new estimated process which is generated by the function [see (3.4), (3.5)]

$$H_k(y,a):=\left(G_{\rho_k}(y,a),g(c,G_{\rho_k}(y,a),a)\right),\quad y=(\vec{m},c)\in\mathbb{Y},\ a\in A.$$

It is easy to see that

$$\sup_{(y,a)\in\mathbb{Y}\times A}\left\|H_k(y,a)-H_\rho(y,a)\right\|_\infty \to 0 \text{ a.s., as } k\to\infty. \tag{4.3}$$

Indeed, since g is a Lipschitz function, we have that, for all $y=(\vec{m},c)\in\mathbb{Y},\ a\in A$,

$$\left\|g(c,G_{\rho_k}(y,a),a)-g(c,G_\rho(y,a),a)\right\|_\infty^2 \le L_g\left\|G_{\rho_k}(y,a)-G_\rho(y,a)\right\|_\infty^1. \tag{4.4}$$

Then, combining (4.2) and (4.4), we get

$$\sup_{(y,a)\in\mathbb{Y}\times A}\left\|g(c,G_{\rho_k}(y,a),a)-g(c,G_\rho(y,a),a)\right\|_\infty^2 \to 0 \text{ a.s., as } k\to\infty.$$

Thus, we can easily see that (4.3) holds. Moreover, for each $\pi\in\Pi_M$ and $y\in\mathbb{Y}$, from (4.3) together with a simple use of the dominated convergence theorem, we can conclude

$$E_y^\pi\left[\sup_{(x,a)\in\mathbb{Y}\times A}\left\|H_k(x,a)-H_\rho(x,a)\right\|_\infty\right]\to 0,\ \text{as } k\to\infty, \tag{4.5}$$

because ρ_k does not depend on π and y.

Let $\{v_k\}$ be a sequence of functions $v_k:\mathbb{Y}\to\mathbb{R}$ in $\mathbb{C}_b(\mathbb{Y})$ to be defined as follows:

$$\begin{aligned} v_0 &\equiv 0;\\ v_k(y) &= \min_{a\in A}\left\{r(y,a)+\alpha v_{k-1}\left[H_k(y,a)\right]\right\},\ \ k\in\mathbb{N},\ \ y\in\mathbb{Y}. \end{aligned} \tag{4.6}$$

Then, noting that the function $(y,a)\to H_k(y,a)$, $k\in\mathbb{N}$, is continuous and that A is compact, from standard measurable selection theorems (see, e.g., Proposition D5(a) in [12]), for each $k\in\mathbb{N}$, there exists $\hat{f}_k\in\mathbb{F}$ (dependent on ρ_k), such that

$$v_k(y)=r(y,\hat{f}_k)+\alpha v_{k-1}\left[H_k\left(y,\hat{f}_k\right)\right],\ \ y\in\mathbb{Y}. \tag{4.7}$$

We define the control policy $\hat{\pi}=\left\{\hat{f}_k\right\}\in\Pi_M$. Observe that this policy is completely computable for the controller, and therefore, according to our objective, we are interested in to study its optimality. However, it is worth noting that the discounted criterion strongly depends on the decisions selected in the early stages, right where the statistical estimation process yields poor information about the unknown dynamic. This leads to thinking that, in general, it is not possible to ensure that $\hat{\pi}$ is an optimal policy in the usual sense for the mean field model. Hence, we need to use the following weaker optimality criterion to analyze its optimality, which is motivated by the comment in the Remark 3.3 (see, e.g., [10,11,13,19] for further information about this optimality criterion).

Definition 4.1 We say that a policy $\pi \in \Pi_M$ is eventually optimal for the mean field control model (or simply eventually optimal) if and only if, for any initial condition $y(0) = y \in \mathbb{Y}$,

$$\lim_{t\to\infty} E_y^{\pi} \Phi(y(t), a_t) = 0, \quad y \in \mathbb{Y},$$

where Φ is the discrepancy function defined in (3.13).

Before establishing the result, we need to impose the following technical requirement.

Assumption 4.2 The constant L_{H_ρ} defined in (3.6) satisfies $\alpha L_{H_\rho} < 1$.

Theorem 4.3 *Under Assumptions 2.1, and 4.2, the policy $\hat{\pi}$ obtained by means of the iterative method described in (4.7), is eventually optimal.*

The proof of this theorem is based on several lemmas, so it will be presented at the end of the section.

Example 4.4 (**Consumption-investment problem**) For the estimator ρ_k, we define, similarly as (2.12), the estimated transition kernel $K_k(a,c) = \left[K_{ij}^k(a,c)\right]$ with components [see (2.10)]

$$K_{ij}^k(a,c) =: \int_{\mathbb{R}} I_j[F(i,c,a,z)]\rho_k(z)dz, \quad i, j \in S,\ (a,c) \in A \times \mathbb{R}.$$

Also, we define

$$G_{\rho_k}(\vec{m}, c, a) := \vec{m} K_k(a,c), \quad (\vec{m}, c) \in \mathbb{Y}, a \in A,$$

and

$$H_k(y,a) := (\vec{m} K_k(a,c), g(c, \vec{m} K_k(a,c), a)), \quad y = (\vec{m}, c) \in \mathbb{Y}, a \in A.$$

Observe that for all $i, j \in S,\ (a,c) \in A \times \mathbb{R}^d$,

$$\left| K_{ij}^k(a,c) - K_{ij}^{\rho}(a,c) \right| \le \int_{\mathbb{R}} |\rho_k(z) - \rho(z)|\, dz.$$

Therefore, according to (4.1)

$$\sup_{(a,c)\in A\times\mathbb{R}^d}\left\|K_k(a,c)-K_\rho(a,c)\right\|_\infty^0\to 0\text{ a.s, as }t\to\infty,\tag{4.8}$$

which, in turn, implies (see (3.9))

$$\sup_{(y,a)\in\mathbb{Y}\times A}\left\|G_{\rho_k}(y,a)-G_\rho(y,a)\right\|_\infty^1=\sup_{(y,a)\in\mathbb{Y}\times A}\left\|\vec{m}K_k(a,c)-\vec{m}K_\rho(a,c)\right\|_\infty^1$$
$$\to 0\text{ a.s., as }k\to\infty.$$

□

The remainder of this section is focused in the proof of Theorem 4.3.

Let $\{u_t\}\subset\mathbb{C}_b(\mathbb{Y})$ be the mean field value iteration functions defined as:

$$u_0\equiv 0;\tag{4.9}$$
$$u_t(y)=\min_{a\in A}\left\{r(y,a)+\alpha u_{t-1}\left[H_\rho\left(y,a\right)\right]\right\},\quad t\in\mathbb{N},\ y\in\mathbb{Y}.\tag{4.10}$$

As shown in [4,11,21], our hypotheses lead to

$$v_*(y)=\lim_{t\to\infty}u_t(y),\quad y\in\mathbb{Y},\tag{4.11}$$

where v_* is the mean field value function satisfying (3.12).

Lemma 4.5 *Suppose that Assumption 2.1 holds. Then:*

(a) For each $t\in\mathbb{N}_0$, the functions u_t generated by means of the iterations (4.9)–(4.10) are Lipschitz continuous with constant

$$L_{u_t}:=L_r\sum_{l=0}^{t-1}\left(\alpha L_{H_\rho}\right)^l.\tag{4.12}$$

(b) In addition, if Assumption 4.2 holds, then the mean field value function v_* is Lipschitz continuous with constant

$$L_{v_*}=\frac{L_r}{1-\alpha L_{H_\rho}},\tag{4.13}$$

where L_r and L_{H_ρ} are the Lipschitz constants in Assumption 2.1(d) and (3.6), respectively.

Proof (a) We proceed by induction. First, from (4.9), clearly part (a) holds for $t=0$. Now we assume that u_t is a Lipschitz function with constant given in (4.12). Then, for $y,y'\in\mathbb{Y}$, from (4.10) we have

$$\left|u_{t+1}(y)-u_{t+1}(y')\right|\leq\sup_{a\in A}\left\{\left|r(y,a)-r(y',a)\right|+\alpha\left|u_t\left[H_\rho(y,a)\right]-u_t\left[H_\rho(y',a)\right]\right|\right\}.$$

Thus, since r and H_ρ are Lipschitz functions (see Assumption 2.1(d) and (3.6)), as long as (4.12) is used, we get

$$\begin{aligned}&\left|u_{t+1}(y)-u_{t+1}(y')\right| \leq L_r \left\|y-y'\right\|_\infty + \alpha L_{u_t} L_{H_\rho}\left\|y-y'\right\|_\infty\\ &\leq \left(L_r+\alpha L_{H_\rho} L_r \sum_{l=0}^{t-1}\left(\alpha L_{H_\rho}\right)^l\right)\left\|y-y'\right\|_\infty \leq L_r\left(1+\sum_{l=0}^{t-1}\left(\alpha L_{H_\rho}\right)^{l+1}\right)\left\|y-y'\right\|_\infty\\ &= L_r \sum_{l=0}^{t}\left(\alpha L_{H_\rho}\right)^l \left\|y-y'\right\|_\infty .\end{aligned}$$

Therefore, u_{t+1} is a Lipschitz function with constant

$$L_{u_{t+1}} := L_r \sum_{l=0}^{t}\left(\alpha L_{H_\rho}\right)^l .$$

This fact proves part (a).

(b) For $y, y' \in \mathbb{Y}$, adding and subtracting the terms $u_t(y)$ and $u_t(y')$ to $|v_*(y) - v_*(y')|$, we obtain

$$\begin{aligned}\left|v_*(y)-v_*(y')\right| &\leq \left|v_*(y)-u_t(y)\right| + \left|u_t(y)-u_t(y')\right| + \left|u_t(y')-v_*(y')\right|\\ &\leq \left|v_*(y)-u_t(y)\right| + L_{u_t}\left\|y-y'\right\|_\infty + \left|u_t(y')-v_*(y')\right|, \quad \forall t \in \mathbb{N}_0,\end{aligned} \tag{4.14}$$

where the last inequality is due to part (a). Now observe that under Assumption 4.2

$$\lim_{t\to\infty} L_{u_t} = L_r \sum_{l=0}^{\infty}\left(\alpha L_{H_\rho}\right)^l = \frac{L_r}{1-\alpha L_{H_\rho}}. \tag{4.15}$$

Therefore, letting $t \to \infty$ in (4.14), we have that (4.11) together (4.15) yield

$$\left|v_*(y)-v_*(y')\right| \leq \frac{L_r}{1-\alpha L_{H_\rho}}\left\|y-y'\right\|_\infty , \qquad y, y' \in \mathbb{Y},$$

that is, v_* is a Lipschitz continuous function. □

Lemma 4.6 *Let v_k be the family of functions generated by the iterations (4.6) and v_* the value function in (3.11) (see (3.12)). Then, under Assumptions 2.1 and 4.2, for each $\pi \in \Pi_M$ and $y \in \mathbb{Y}$, $E_y^\pi \left\|v_* - v_k\right\| \to 0$, as $k \to \infty$.*

Proof From (3.12) and (4.6), we have, for each $k \in \mathbb{N}$ and $y \in \mathbb{Y}$,

$$\begin{aligned}&\left|v_*(y)-v_k(y)\right| \leq \alpha \sup_{a\in A}\left|v_*\left[H_\rho(y,a)\right]-v_{k-1}\left[H_k(y,a)\right]\right|\\ &\leq \alpha \sup_{a\in A}\left|v_*\left[H_\rho(y,a)\right]-v_*\left[H_k(y,a)\right]\right| + \alpha \sup_{a\in A}\left|v_*\left[H_k(y,a)\right]-v_{k-1}\left[H_k(y,a)\right]\right| ,\end{aligned}$$

where in the last inequality we have added and subtracted the term $v_*\left[H_k(y,a)\right]$. Hence, from Lemma 4.5 and the fact that $v_*, v_k \in \mathbb{B}(\mathbb{Y})\ \forall k \in \mathbb{N}$,

$$0 \le \|v_* - v_k\| \le L_{v_*} \sup_{(y,a)\in\mathbb{Y}\times A} \left\| H_\rho(y,a) - H_k(y,a)\right\|_\infty + \alpha \|v_* - v_{k-1}\|, \quad (4.16)$$

which implies

$$E_y^\pi \|v_* - v_k\| \le L_{v_*} E_y^\pi \left[\sup_{(y,a)\in\mathbb{Y}\times A} \left\| H_\rho(y,a) - H_k(y,a)\right\|_\infty \right] + \alpha E_y^\pi \|v_* - v_{k-1}\|, \quad (4.17)$$

for each $\pi \in \Pi_M$ and $y \in \mathbb{Y}$. Let $l := \limsup_{k\to\infty} E_y^\pi \|v_* - v_k\| < \infty$. Hence, letting $k \to \infty$ in (4.17), and using the convergence in (4.5), we get $l \le \alpha l$. Finally, since $\alpha < 1$, we can deduce that $\lim_{k\to\infty} E_y^\pi \|v_\infty - v_k\| = 0$, which proves the result. □

Proof of Theorem 4.3 We define, for each $k \in \mathbb{N}$, the approximate discrepancy function $\Phi_k : \mathbb{Y} \times A \to \mathbb{R}$ as

$$\Phi_k(y,a) := r(y,a) + \alpha v_{k-1}\left[H_k(y,a)\right] - v_k(y), \quad (y,a) \in \mathbb{Y} \times A.$$

Now observe that, for each $k \in \mathbb{N}$ and $(y,a) \in \mathbb{Y} \times A$,

$$|\Phi(y,a) - \Phi_k(y,a)| \le \left| v_*\left[H_\rho(y,a)\right] - v_{k-1}\left[H_k(y,a)\right]\right| + |v_*(y) - v_k(y)|.$$

Then, from Lemma 4.6, letting $k \to \infty$ we get

$$E_y^{\hat\pi}\left[\sup_{(y,a)\in\mathbb{Y}\times A} |\Phi(y,a) - \Phi_k(y,a)| \right] \to 0. \quad (4.18)$$

On the other hand, observing that $\Phi_k(y, \hat f_k(y)) = 0$, $y \in \mathbb{Y}$ when using the control policy generated by (4.7), we have

$$\begin{aligned} 0 \le \Phi(y(k), \hat f_k(y(k))) &= \left| \Phi(y(k), \hat f_k(y(k))) - \Phi_k(y(k), \hat f_k(y(k)))\right| \\ &\le \sup_{(y,a)\in\mathbb{Y}\times A} |\Phi(y,a) - \Phi_k(y,a)|, \end{aligned}$$

Thus, from (4.18), we obtain

$$\lim_{k\to\infty} E_y^{\hat\pi} \Phi(y(k), a_k) = 0.$$

□

5 Mean Field Convergence

In this section we study the performance of the eventually optimal policy $\hat{\pi}$ obtained in Sect. 4; that is, we are interested in to analyze the optimality deviation of $\hat{\pi}$ when it is used to control the process $\left\{y^N(t)\right\}$. Clearly, such an optimality deviation must be measured in terms of the difference between the corresponding optimal value functions V_*^N and v_* of the models $\mathcal{M}_N$ and $\mathcal{M}$ respectively, and moreover, as was pointed out in Sect. 2, it must be analyzed in an asymptotic sense as N goes to infinity. To this end, we impose the following assumption which concerns with the convergence of the trajectories $y^N(\cdot)$ to the trajectories $y(\cdot)$ defined in (2.6) and (3.5), respectively, in the sense of (5.1) below.

Observe that according to the Propositions 2.4 and 3.2, as well as the definition of the policy $\hat{\pi}$, we can restrict our analysis to the class of Markov policies Π_M.

Assumption 5.1 We assume:

(a) $(\vec{M}^N(0), C^N(0)) = (\vec{m}(0), c(0)) = (\vec{m}_0, c_0) = y \in \mathbb{Y}_N$, for all $N \in \mathbb{N}$.

(b) For any $y \in \mathbb{Y}_N$, $T \in \mathbb{N}$, and $\varepsilon > 0$, there exist positive constants K and λ such that

$$\sup_{\pi\in\Pi_M} P_y^\pi \left\{ \sup_{0\le t\le T} \left\| y^N(t) - y(t) \right\|_\infty \ge \gamma_T(\varepsilon) \right\} \le KTe^{-\lambda N\varepsilon^2}, \tag{5.1}$$

where $\gamma_T(\varepsilon) \to 0$ as $\varepsilon \to 0$.

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

$$a_t^{\pi,N} := f_t(y^N(t)) \text{ and } a_t^\pi := f_t(y(t))$$

the actions at time t corresponding to the application of the policy π under the process $\left\{y^N(t)\right\}$ and $\{y(t)\}$, respectively.

Now, following similar ideas to those of [7], we show that the example we have been working satisfies Assumption 5.1(b).

Example 5.2 (**Consumption-investment problem**) Recall the relations (2.12)–(2.16). Let $\pi = \{f_t\} \in \Pi_M$ be an arbitrary policy and $y \in \mathbb{Y}_N \subset \mathbb{Y}$ be the initial state. We denote

$$B_{inj}^{N\rho}(t) := I_{\left\{A_{ij}^\rho\left(a_t^{\pi,N}, C^N(t)\right)\right\}}(w_n^i(t)), \quad i, j \in S, n \in \mathbb{N},$$

where $C^N(t)$ is as in (2.3) and $w_n^i(t)$ are i.i.d. random variables uniformly distributed on [0, 1]. Observe that, for each $t \in \mathbb{N}_0$, $\left\{B_{inj}^{N\rho}(t)\right\}_{inj}$ are i.i.d. Bernoulli random variables with mean

$$E_y^\pi \left[B_{inj}^{N\rho}(t) | a_t^{\pi,N} = a, C^N(t) = c\right] = K_{ij}^\rho(a, c)$$
$$= I_j[F(i, c, a, z)]\rho(z)dz, \quad i, j \in S, \ (a, c) \in A \times \mathbb{R}^d.$$

Then, for a fixed $\varepsilon > 0$, by Hoeffding's inequality, we have

$$P_y^\pi \left[\left| \sum_{n=1}^{NM_i^N(t)} B_{inj}^{N\rho}(t) - NM_i^N(t) K_{ij}^\rho \left(a_t^{\pi,N}, C^N(t) \right) \right| < N\varepsilon \right] > 1 - 2e^{-2N\varepsilon^2}.$$

Consider the set $\bar{\Omega} = \left\{ \omega \in \Omega' \left| \sum_{n=1}^{NM_i^N(t)} B_{inj}^{N\rho}(t) - NM_i^N(t) K_{ij}^\rho \left(a_t^{\pi,N}, C^N(t) \right) \right| < N\varepsilon \right\} \subset \Omega'$ (see Remark 2.3(b)), and let ε_t be a positive number such that $\left\| y^N(t) - y(t) \right\|_\infty \leq \varepsilon_t$; that is,

$$\left\| \vec{M}^N(t) - \vec{m}(t) \right\|_\infty^1 \leq \varepsilon_t \text{ and } \left\| C^N(t) - c(t) \right\|_\infty^2 \leq \varepsilon_t. \tag{5.2}$$

Thus, from (2.14), (3.7), and (5.2), we have that the following relations hold true on $\bar{\Omega}$:

$$\begin{aligned} \left| M_j^N(t+1) - m_j(t+1) \right| &= \left| \sum_{i=0}^s \frac{1}{N} \left[\sum_{n=1}^{NM_i^N(t)} B_{inj}^{N\rho}(t) - Nm_i(t) K_{ij}^\rho \left(a_t^{\pi,N}, c(t) \right) \right] \right| \\ &\leq \sum_{i=0}^s \frac{1}{N} \left| \sum_{n=1}^{NM_i^N(t)} B_{inj}^{N\rho}(t) - Nm_i(t) K_{ij}^\rho \left(a_t^{\pi,N}, c(t) \right) \right| \\ &\leq \sum_{i=0}^s \frac{1}{N} \left| \sum_{n=1}^{NM_i^N(t)} B_{inj}^{N\rho}(t) - NM_i^N(t) K_{ij}^\rho \left(a_t^{\pi,N}, C^N(t) \right) \right| \\ &\quad + \sum_{i=0}^s \left| M_i^N(t) - m_i(t) \right| K_{ij}^\rho \left(a_t^{\pi,N}, C^N(t) \right) \\ &\quad + \sum_{i=0}^s m_i(t) \left| K_{ij}^\rho \left(a_t^{\pi,N}, C^N(t) \right) - K_{ij}^\rho \left(a_t^{\pi,N}, c(t) \right) \right| \\ &\quad < (s+1)\varepsilon + (s+1)\varepsilon_t + L_K \varepsilon_t. \end{aligned} \tag{5.3}$$

Hence, since the right-hand of this last inequality does not depend on j, we have

$$\left\| \vec{M}^N(t+1) - \vec{m}(t+1) \right\|_\infty^1 \leq (s+1)\varepsilon + (s+1)\varepsilon_t + L_K \varepsilon_t.$$

On the other hand, since g is a Lipschitz function (see Assumption 2.1), expressions (2.19) and (3.2)) together with (5.2) and (5.3) lead to

$$
\begin{aligned}
\left\| C^N(t+1) - c(t+1) \right\|_\infty^2 &\leq \left\| g(C^N(t), \vec{M}^N(t+1), a_t^{\pi,N}) \right. \\
&\quad \left. - g(c(t), \vec{m}(t+1), a_t^{\pi,N}) \right\|_\infty^2 \\
&< L_g \max\{\varepsilon_t, (s+1)\varepsilon + (s+1)\varepsilon_t + L_K\varepsilon_t\} = L_g\left((s+1)\varepsilon + (s+1)\varepsilon_t + L_K\varepsilon_t\right),
\end{aligned}
$$

which implies that on the set $\bar{\Omega}$ (recall $L_g \geq 1$)

$$
\left\| y^N(t+1) - y(t+1) \right\|_\infty < L_g\left((s+1)\varepsilon + (s+1)\varepsilon_t + L_K\varepsilon_t\right).
$$

Considering now $\left\| y^N(0) - y(0) \right\|_\infty = \varepsilon_0 = 0$ (see Assumption 5.1(a)) and applying an inductive procedure, a straightforward calculation yields that, on the set $\bar{\Omega}$,

$$
\left\| y^N(t+1) - y(t+1) \right\|_\infty < L_g(s+1)\varepsilon\beta_t, \quad t \in \mathbb{N}_0,
$$

where $\{\beta_t\}$ is a increasing sequence. Then, for a fixed $T \in \mathbb{N}$,

$$
\left\| y^N(t+1) - y(t+1) \right\|_\infty < L_g(s+1)\varepsilon\beta_T, \quad \forall 0 \leq t \leq T
$$

on the set $\bar{\Omega}$. Therefore, under the policy $\pi \in \Pi_M$,

$$
P_y^\pi \left[\sup_{0 \leq t \leq T} \left\| y^N(t+1) - y(t+1) \right\|_\infty < L_g(s+1)\varepsilon\beta_T \right] \geq 1 - 2e^{-2N\varepsilon^2},
$$

which, letting $\gamma_T(\varepsilon) := L_g(s+1)\varepsilon\beta_T$, $K = \lambda = 2$, implies

$$
\sup_{\pi \in \Pi_M} P_y^\pi \left\{ \sup_{0 \leq t \leq T} \left\| y^N(t) - y(t) \right\|_\infty \geq \gamma_T(\varepsilon) \right\} \leq KTe^{-\lambda N\varepsilon^2}.
$$

Finally, we observe that $\gamma_T(\varepsilon) \to 0$ as $\varepsilon \to 0$. □

Now we introduce the following additional notation: For any $T \in \mathbb{N}$ we denote

$$
Y_T := \sup_{0 \leq t \leq T} \| y^N(t) - y(t) \|_\infty \tag{5.4}
$$

and

$$
\mathcal{K}(T) := (L_g)^T \max\{L_g, diam(A)\}, \tag{5.5}
$$

where $L_g \geq 1$ is the Lipschitz constant in Assumption 2.1 (b) and $diam(A) := \sup_{(a,a') \in A \times A} d(a, a')$.

Recall that for any given $t \in \mathbb{N}_0$,

$$
\| y^N(t) - y(t) \|_\infty = \max\left\{ \| \vec{M}^N(t) - \vec{m}(t) \|_\infty^1, \ \| C^N(t) - c(t) \|_\infty^2 \right\}. \tag{5.6}
$$

We are now in conditions to set our main results. Firstly, we provide a bound for the gap between the value functions V_*^N and v_*, which in turns defines an approximation scheme as $N \to \infty$. Next we show that the control policy $\hat{\pi}$ is eventually optimal on the control model $\mathcal{M}_N$ in an asymptotic sense.

Theorem 5.3 *Under the Assumptions 2.1, 4.2, and 5.1, the following statements hold true:*

(a) For each $T \in \mathbb{N}$, $0 \le t \le T$, and $y \in \mathbb{Y}_N$,

$$\sup_{\varphi \in \Pi_M} E_y^{\varphi} \left| V_*^N(y^N(t)) - v_*(y(t)) \right| \le \frac{2R\alpha^T}{1-\alpha} + L_r \frac{1-\alpha^T}{1-\alpha} \times \left[KTe^{-\lambda N \varepsilon^2}(1+\mathcal{K}(T)) + \gamma_T(\varepsilon) \right]. \tag{5.7}$$

(b) The control policy $\hat{\pi} \in \Pi_M$ estimated in (4.7) is eventually asymptotically optimal for the N Markov control model $\mathcal{M}_N$, as $N \to \infty$; that is

$$\lim_{t\to\infty} \lim_{N\to\infty} E_y^{\hat{\pi}} \Phi^N(y^N(t), \hat{f}_t) = 0, \tag{5.8}$$

where

$$\Phi^N(y^N, a) := r(y^N, a) + \alpha \int_{\mathbb{R}^N} V_*^N \left[H_\rho^N \left(y^N, a, w \right) \right] \theta(dw) - V_*^N(y^N), \;\; y^N \in \mathbb{Y}_N \tag{5.9}$$

is the discrepancy function in the $N-$MCM $\mathcal{M}_N$ (see also (3.13)).

In the remainder of this section we will assume that Assumptions 2.1, 4.2, and 5.1 hold true. Based in this fact, the proof of Theorem 5.3 will be a consequence of the following propositions.

Proposition 5.4 *(a) For each $\pi \in \Pi_M$ and $T \in \mathbb{N}$,*

$$Y_T := \sup_{0\le t\le T} \|y^N(t) - y(t)\|_\infty \le \mathcal{K}(T). \tag{5.10}$$

(b) For each $y \in \mathbb{Y}_N$ and $T \in \mathbb{N}$,

$$\sup\nolimits_{\pi\in\Pi_M} E_y^{\pi} \left[\sup_{0\le t\le T} \|y^N(t) - y(t)\|_\infty \right] \le KTe^{-\lambda N \varepsilon^2}(1+\mathcal{K}(T)) + \gamma_T(\varepsilon). \tag{5.11}$$

Proof **(a)** To obtain (5.10) it is sufficient to prove that for each $t \in \mathbb{N}_0$ and $\pi \in \Pi_M$

$$\|y^N(t) - y(t)\|_\infty \le (L_g)^{t-1} \max\{L_g, diam(A)\}. \tag{5.12}$$

We then focus to get (5.12). Notice that under Assumption 5.1(a) we have $a_0^{\pi,N} = a_0^{\pi} =: a_0 \in A$ and $\|y^N(0) - y(0)\|_\infty = 0$. On the other hand, since $\vec{M}^N(t)$ and $\vec{m}(t)$

are probability measures, it follows that $\|\vec{M}^N(t)-\vec{m}(t)\|^1_\infty \leq 1$, for all $t \in \mathbb{N}_0$. Hence, because $L_g \geq 1$, the proof reduces to analyze the norm $\|\cdot\|^2_\infty$ in (5.6). In particular, (5.12) will be proved if we show that

$$\|C^N(t)-c(t)\|^2_\infty \leq (L_g)^{t-1}\max\{L_g, diam(A)\}, \quad \forall t \in \mathbb{N}_0. \tag{5.13}$$

To this end, we proceed by induction. First, observe that from (2.3) and (3.2) we obtain

$$\begin{aligned}\|C^N(1)-c(1)\|^2_\infty &= \|g(c_0, \vec{M}^N(1), a_0) - g(c_0, \vec{m}(1), a_0)\|^2_\infty \\ &\leq L_g\|\vec{M}^N(1)-\vec{m}(1)\|^1_\infty \leq L_g \quad \text{(by (2.8))}.\end{aligned}$$

Also,

$$\begin{aligned}\|C^N(2)-c(2)\|^2_\infty &= \|g(C^N(1), \vec{M}^N(2), a_1^{\pi,N}) - g(c(1), \vec{m}(1), a_1^\pi)\|^2_\infty \\ &\leq L_g \max\Big\{\|C^N(1)-c(1)\|^2_\infty,\ \|\vec{M}^N(2)-\vec{m}(2)\|^1_\infty, \\ &\quad d_A(a_1^{\pi,N}, a_1^\pi)\Big\} \\ &\leq L_g \max\big\{L_g,\ 1,\ diam(A)\big\} = L_g \max\big\{L_g, diam(A)\big\}.\end{aligned}$$

Now, assume that (5.13) holds for some $t \in \mathbb{N}$. Then

$$\begin{aligned}&\|C^N(t+1)-c(t+1)\|^2_\infty = \|g(C^N(t), \vec{M}^N(t+1), a_t^{\pi,N}) - g(c(t), \vec{m}(t+1), a_t^\pi)\|^2_\infty \\ &\leq L_g \max\Big\{\|C^N(t)-c(t)\|^2_\infty,\ \|\vec{M}^N(t+1)-m(t+1)\|^1_\infty,\ d_A(a_t^{\pi,N}, a_t^\pi)\Big\} \quad \text{(by (2.8))} \\ &\qquad \leq L_g \max\big\{(L_g)^{t-1}\max\{L_g, diam(A)\}, 1, diam(A)\big\} \quad \text{(by (5.13))} \\ &\qquad \leq (L_g)^t \max\{L_g, diam(A)\}.\end{aligned}$$

This proves (5.13), which in turns yields (5.12) and (5.10).
(b) Observe that for each $y \in \mathbb{Y}_N$, $\pi \in \Pi_M$, $T \in \mathbb{N}$, and $\varepsilon > 0$, the expectation in (5.11) satisfies (see (5.4))

$$\begin{aligned}&E_y^\pi[Y_T] = E_y^\pi\left[Y_T I_{\{Y_T \geq \gamma_T(\varepsilon)\}} + Y_T I_{\{Y_T < \gamma_T(\varepsilon)\}}\right] \\ &\leq E_y^\pi\left[Y_T I_{\{Y_T \geq \gamma_T(\varepsilon)\}}\right] + \gamma_T(\varepsilon) P_y^\pi(Y_T < \gamma_T(\varepsilon)) \leq E_y^\pi\left[Y_T I_{\{T_T \geq \gamma_T(\varepsilon)\}}\right] + \gamma_T(\varepsilon).\end{aligned} \tag{5.14}$$

On the other hand, by (5.10) as well as the non negativeness of Y_T, we have

$$\frac{Y_T}{1+\mathcal{K}(T)} \leq \frac{Y_T}{1+Y_T} \leq 1,$$

which implies

$$\frac{Y_T}{1+\mathcal{K}(T)} I_{\{Y_T \geq \gamma_T(\varepsilon)\}} \leq I_{\{Y_T \geq \gamma_T(\varepsilon)\}}.$$

This fact together with the definition of Y_T and Assumption 5.1(b) give

$$\frac{1}{1+\mathcal{K}(T)}E_y^\pi[Y_T I_{\{Y_T\geq\gamma_T(\varepsilon)\}}] \leq P_y^\pi(Y_T \geq \gamma_T(\varepsilon)) \leq KTe^{-\lambda N\varepsilon^2}, \;\; \pi \in \Pi_M.$$

Finally, from (5.14) we get

$$E_y^\pi\left[Y_T\right] \leq KTe^{-\lambda N\varepsilon^2}(1+\mathcal{K}(T)) + \gamma_T(\varepsilon), \quad \pi \in \Pi_M, \tag{5.15}$$

and by taking supremum over $\pi \in \Pi_M$ in (5.15) we prove the part (b). □

The next results are related with the finite horizon discounted cost criteria for the $N-$MCM $\mathcal{M}_N$ and for the mean field control model $\mathcal{M}$. For any $\pi \in \Pi_M$, $y \in \mathbb{Y}_N \subset \mathbb{Y}$, and $T \in \mathbb{N}$, we define

$$V_T^N(\pi, y) := E_y^\pi\left[\sum_{k=0}^{T-1}\alpha^k r(y^N(k), a_k)\right] \quad \text{and} \quad v_T(\pi, y) := \sum_{k=0}^{T-1}\alpha^k r(y^N(k), a_k).$$

Proposition 5.5 *Let L_r and R be the constants in Assumption 2.1(d). Then, for each $y \in \mathbb{Y}_N$, $\varepsilon > 0$, $T \in \mathbb{N}$, and $0 \leq t \leq T$, the following statements hold true:*

(a)

$$\sup_{\pi\in\Pi} E_y^\pi\left|r(y^N(t), a_t^{\pi,N}) - r(y(t), a_t^\pi)\right| \leq L_r\left(KTe^{-\lambda N\varepsilon^2}(1+\mathcal{K}(T)) + \gamma_T(\varepsilon)\right); \tag{5.16}$$

(b)

$$\sup_{\varphi\in\Pi} E_y^\varphi\left[\sup_{\pi\in\Pi}\left|V_T^N(\pi, y^N(t)) - v_T(\pi, y(t))\right|\right] \leq L_r\frac{1-\alpha^T}{1-\alpha} \times\left[KTe^{-\lambda N\varepsilon^2}(1+\mathcal{K}(T)) + \gamma_T(\varepsilon)\right]; \tag{5.17}$$

(c)

$$\sup_{\varphi\in\Pi} E_y^\varphi\left[\sup_{\pi\in\Pi}\left|V^N(\pi, y^N(t)) - V_T^N(\pi, y^N(t))\right|\right] \leq \frac{R\alpha^T}{1-\alpha}; \tag{5.18}$$

(d)

$$\sup_{\varphi\in\Pi} E_y^\varphi\left[\sup_{\pi\in\Pi}\left|v(\pi, y(t)) - v_T(\pi, y(t))\right|\right] \leq \frac{R\alpha^T}{1-\alpha}. \tag{5.19}$$

Proof **(a)** Let us fix any $\pi \in \Pi_M$ and $T \in \mathbb{N}$. Then, Assumption 2.1(d) together with Proposition 5.4, lead to the following relations

$$E_y^\pi\left|r(y^N(t), a_t^{\pi,N}) - r(y(t), a_t^\pi)\right| \leq L_r E_y^\pi\left[\|y^N(t) - y(t)\|_\infty\right] \tag{5.20}$$

$$\leq L_r E_y^\pi\left[\sup_{0\leq t\leq T}\|y^N(t) - y(t)\|_\infty\right] \leq L_r\left(KTe^{-\lambda N\varepsilon^2}(1+\mathcal{K}(T)) + \gamma_T(\varepsilon)\right).$$

This implies the part (a).
(b) For each $\pi \in \Pi_M$,

$$\begin{aligned}|V_T(\pi, y^N(t)) - v_T(\pi, y(t))| &= \left| E^{\pi}_{y^N(t)} \left\{ \sum_{k=0}^{T-1} \alpha^k r(y^N(k), a_k^{\pi,N}) \right.\right.\\ &\qquad \left.\left. - \sum_{k=0}^{T-1} \alpha^k r(y(k), a_k^{\pi}) \right\} \right|\\ &\le \sum_{k=0}^{T-1} \alpha^k E^{\pi}_{y^N(t)} \left| r(y^N(k), a_k^{\pi,N}) - r(y(t), a_k^{\pi}) \right|\\ &\le L_r \frac{1-\alpha^T}{1-\alpha} \left[KTe^{-\lambda N \varepsilon^2}(1+\mathcal{K}(T)) + \gamma_T(\varepsilon) \right],\end{aligned}$$

where the last inequality follows from (5.20). This gives

$$\sup_{\pi\in\Pi} \left| V_T^N(\pi, y^N(t)) - v_T(\pi, y(t)) \right| \le L_r \frac{1-\alpha^T}{1-\alpha} \left[KTe^{-\lambda N\varepsilon^2}(1+\mathcal{K}(T)) + \gamma_T(\varepsilon) \right], \quad \forall t \in \mathbb{N}_0.$$

Taking expectation E_y^{φ} in both sides of the above expression, and then taking supremum over $\varphi \in \Pi_M$, we obtain (5.17).
(c) For each $\pi \in \Pi_M$, we have

$$\begin{aligned}&\left| V^N(\pi, y^N(t)) - V_T^N(\pi, y^N(t)) \right|\\ &\le \left| E^{\pi}_{y^N(t)} \left\{ \sum_{k=0}^{\infty} \alpha^k r(y^N(k), a_k^{\pi,N}) - \sum_{k=0}^{T-1} \alpha^k r(y^N(k), a_k^{\pi,N}) \right\} \right|\\ &\le \sum_{k=T}^{\infty} \alpha^k E^{\pi}_{y^N(t)} |r(y^N(k), a_k^{\pi,N})| \le R \sum_{k=T}^{\infty} \alpha^k \le \frac{R\alpha^T}{1-\alpha}.\end{aligned}$$

Hence, easily we can see that (5.18) holds.
(d) It follows by using the same arguments of (c). □

5.1 Proof of Theorem 5.3(a)

Let $\pi_*^N = \{f_*^N\} \in \Pi_M^N$ be an optimal stationary policy for the $N-$MCM $\mathcal{M}_N$ (see Proposition 2.4(b)), and for an arbitrary selector $\tilde{f} \in \mathbb{F}$, we define the stationary policy $\bar{\pi} = \{\bar{f}\} \in \Pi_M$, where $\bar{f} : \mathbb{Y} \to A$ is given by

$$\bar{f}(y) = f_*^N(y) I_{\mathbb{Y}_N}(y) + \tilde{f}(y) I_{[\mathbb{Y}_N]^c}(y).$$

In addition, let $\varphi \in \Pi_M$ be an arbitrary policy and let us denote $y_\varphi^N(t) = y^N(t) \in \mathbb{Y}_N$ and $y_\varphi(t) := y(t) \in \mathbb{Y}$. Observe that for each $t \in \mathbb{N}_0$,

$$V_*^N(y^N(t)) = V^N(\pi_*^N, y^N(t)) = V^N(\bar{\pi}, y^N(t)) \leq \sup_{\pi \in \Pi_M} V^N(\pi, y^N(t)).$$

Hence,

$$V_*^N(y^N(t)) - v_*(y(t)) \leq \sup_{\pi \in \Pi_M} V^N(\pi, y^N(t)) - \inf_{\pi \in \Pi_M} v(\pi, y(t))$$

which, in turns implies

$$\left| V_*^N(y^N(t)) - v_*(y(t)) \right| \leq \sup_{\pi \in \Pi_M} \left| V^N(\pi, y^N(t)) - v(\pi, y(t)) \right|, \quad t \in \mathbb{N}_0.$$

Therefore, for each $y \in \mathbb{Y}_N$ and $0 \leq t \leq T$,

$$\begin{aligned}
E_y^\varphi \left| V_*^N(y^N(t)) - v_*(y(t)) \right| &\leq E_y^\varphi \left[\sup_{\pi \in \Pi_M} \left| V^N(\pi, y^N(t)) - v(\pi, y(t)) \right| \right] \\
&\leq E_y^\varphi \left[\sup_{\pi \in \Pi_M} \left\{ \left| V^N(\pi, y^N(t)) - V_T^N(\pi, y^N(t)) \right| + \left| V_T^N(\pi, y^N(t)) - v_T(\pi, y(t)) \right| \right. \right. \\
&\quad \left. \left. + \left| v_T(\pi, y(t)) - v(\pi, y(t)) \right| \right\} \right] \\
&\leq E_y^\varphi \left[\sup_{\pi \in \Pi_M} \left| V^N(\pi, y^N(t)) - V_T^N(\pi, y^N(t)) \right| \right] \\
&\quad + E_y^\varphi \left[\sup_{\pi \in \Pi_M} \left| V_T^N(\pi, y^N(t)) - v_T(\pi, y(t)) \right| \right] \\
&\quad + E_y^\varphi \left[\sup_{\pi \in \Pi_M} \left| v_T(\pi, y(t)) - v(\pi, y(t)) \right| \right] \\
&\leq \frac{2R\alpha^T}{1-\alpha} + L_r \frac{1-\alpha^T}{1-\alpha} \left[KTe^{-\lambda N \varepsilon^2}(1 + \mathcal{K}(T)) + \gamma_T(\varepsilon) \right],
\end{aligned}$$

where the last inequality is due to Proposition 5.5. Finally, by taking supremum over $\varphi \in \Pi_M$, we obtain (5.7). □

5.2 Proof of Theorem 5.3(b)

For ease notation, we let $\hat{a}_t^N := a_t^{\hat{\pi}, N}$ and $\hat{a}_t := a_t^{\hat{\pi}}$. Then, consider $\left\{(y^N(t), \hat{a}_t^N)\right\} \in \mathbb{Y}_N \times A$ and $\left\{(y(t), \hat{a}_t)\right\} \in \mathbb{Y} \times A$ the sequences of state-action pairs corresponding to application of the policy $\hat{\pi}$ (see (4.7)). For each $t \in \mathbb{N}_0$, we define the random variable

$$\Delta_t^N := \left|\Phi^N(y^N(t), \hat{a}_t^N) - \Phi(y(t), \hat{a}_t)\right|.$$

Then, from the definition of the discrepancy functions Φ^N and Φ given in (5.9) and (3.13), respectively, we have for each $t \in \mathbb{N}_0$,

$$\begin{aligned}
\Delta_t^N \leq & \left|r(y^N(t), \hat{a}_t^N) - r(y(t), \hat{a}_t)\right| + \left|V_*^N(y^N(t)) - v_*(y(t))\right| \\
& + \alpha \left|\int_{\mathbb{R}^N} V_*^N\left[H_\rho^N\left(y^N(t), \hat{a}_t^N, w\right)\right]\theta(dw) - v_*(H_\rho\left(y(t), \hat{a}_t\right))\right| \\
\leq & \left|r(y^N(t), \hat{a}_t^N) - r(y(t), \hat{a}_t)\right| + \left|V_*^N(y^N(t)) - v_*(y(t))\right| \\
& + \left|\int_{\mathbb{R}^N} \left\{V_*^N\left[H_\rho^N\left(y^N(t), \hat{a}_t^N, w\right)\right] - v_*(y(t+1))\right\}\theta(dw)\right| \quad \text{(by (3.5))} \\
= & \left|r(y^N(t), \hat{a}_t^N) - r(y(t), \hat{a}_t)\right| + \left|V_*^N(y^N(t)) - v_*(y(t))\right| \\
& + \left|E_y^{\hat{\pi}}\left[V_*^N(y^N(t+1)) - v_*(y(t+1)) \mid h_t^N, \hat{a}_t^N\right]\right| \quad \text{(by (5.21))} \\
\leq & \left|r(y^N(t), \hat{a}_t^N) - r(y(t), \hat{a}_t)\right| + \left|V_*^N(y^N(t)) - v_*(y(t))\right| \\
& + E_y^{\hat{\pi}}\left[\left|V_*^N(y^N(t+1)) - v_*(y(t+1))\right| \mid h_t^N, \hat{a}_t^N\right].
\end{aligned} \tag{5.21}$$

Taking expectation $E_y^{\hat{\pi}}$ in (5.21), and using properties of conditional expectation we get

$$\begin{aligned}
E_y^{\hat{\pi}}\left[\Delta_t^N\right] \leq & E_y^{\hat{\pi}}\left|r(y^N(t), \hat{a}_t^N) - r(y(t), \hat{a}_t)\right| + E_y^{\hat{\pi}}\left|V_*^N(y^N(t)) - v_*(y(t))\right| \\
& + E_y^{\hat{\pi}}\left|V_*^N(y^N(t+1)) - v_*(y(t+1))\right|.
\end{aligned}$$

Furthermore, Proposition 5.5 and Theorem 5.3 yield

$$\begin{aligned}
E_y^{\hat{\pi}}\left[\Delta_t^N\right] \leq & L_r\left[KTe^{-\lambda N\varepsilon^2}(1 + \mathcal{K}(T) + \gamma_T(\varepsilon))\right] + \frac{4R\alpha^T}{1-\alpha} \\
& + 2L_r\frac{1-\alpha^T}{1-\alpha}\left[KTe^{-\lambda N\varepsilon^2}(1 + \mathcal{K}(T)) + \gamma_T(\varepsilon)\right],
\end{aligned}$$

for any arbitrary $\varepsilon > 0$ and $T > t$.

Also, observe that

$$\begin{aligned}
E_y^{\hat{\pi}}\left[\Phi^N(y^N(t), \hat{a}_t^N)\right] \leq & E_y^{\hat{\pi}}\left[|\Phi^N(y^N(t), \hat{a}_t^N) - \Phi(y(t), \hat{a}_t)|\right] + E_y^{\hat{\pi}}\left[\Phi(y(t), \hat{a}_t)\right] \\
= & E_y^{\hat{\pi}}[\Delta_t^N] + E_y^{\hat{\pi}}\left[\Phi(y(t), \hat{a}_t)\right] \leq L_r\left[KTe^{-\lambda N\varepsilon^2}(1 + \mathcal{K}(T)) + \gamma_T(\varepsilon)\right] + \frac{4R\alpha^T}{1-\alpha} \\
& + 2L_r\frac{1-\alpha^T}{1-\alpha}\left[KTe^{-\lambda N\varepsilon^2}(1 + \mathcal{K}(T)) + \gamma_T(\varepsilon)\right] + \quad E_y^{\hat{\pi}}\left[\Phi(y(t), \hat{a}_t)\right].
\end{aligned}$$

Thus, taking limit as $N \to \infty$ we obtain

$$0 \leq \lim_{N\to\infty} E_y^{\hat{\pi}} \left[\Phi^N(y^N(t), \hat{a}_t^N) \right] \leq L_r \gamma_T(\varepsilon) + \frac{4R\alpha^T}{1-\alpha} + 2L_r \frac{1-\alpha^T}{1-\alpha} \gamma_T(\varepsilon) + E_y^{\hat{\pi}} \left[\Phi(y(t), \hat{f}_t(y(t))) \right]. \tag{5.22}$$

Finally, as ε and T are arbitrary, by letting $t \to \infty$ in (5.22), a simple use of Theorem 4.3 shows that

$$\lim_{t\to\infty} \lim_{N\to\infty} E_y^{\hat{\pi}} \left[\Phi^N(y^N(t), \hat{a}_t^N) \right] = 0$$

which proves the desired result. □

Acknowledgments Work partially supported by Consejo Nacional de Ciencia y Tecnologia (CONACyT) under Grants CB2015/254306 and CB2015/238045.

References

1. Achdou, I., Capuzzo-Dolcetta, I.: Mean field games: numerical methods. SIAM J. Numer. Anal. **48**, 1136–1162 (2010)
2. Aoki, M.: New macroeconomic modeling approaches. Hierarchical dynamics and mean field approximation. J. Econ. Dyn. Control **18**, 865–877 (1994)
3. Bensoussan, A., Frehse, J., Yam, P.: Mean Field Games and Mean Field Control Theory. Springer, New York (2010)
4. Bertsekas, D.P.: Dynamic Programming: Deterministic and Stochastic Models. Prentice-Hall, Englewood Cliffs (1987)
5. Budhiraja, A., Dupuis, P., Fischer, M.: Large deviation properties of weakly interacting processes via weak convergence methods. Ann. Probab. **40**, 74–102 (2012)
6. Carmona, R., Fouque, J.P., Vestal, D.: Interacting particle systems for the computation of rare credit portfolio losses. Financ. Stoch. **13**, 613–633 (2009)
7. Gast, N., Gaujal, B.: A mean field approach for optimization in discrete time. Discret. Event Dyn. Syst. **21**, 63–101 (2011)
8. Gast, N., Gaujal, B., Le Boudec, J.Y.: Mean field for Markov decision processes: from discrete to continuous optimization. IEEE Trans. Autom. Control **57**, 2266–2280 (2012)
9. Gomes, D.A., Mohr, J., Souza, R.R.: Discrete time, finite state space mean field games. J. Math. Pures Appl. **93**, 308–328 (2010)
10. Gordienko, E.I., Minjárez-Sosa, J.A.: Adaptive control for discrete-time Markov processes with unbounded costs: discounted criterion. Kybernetika **34**, 217–234 (1998)
11. Hernández-Lerma, O.: Adaptive Markov Control Processes. Springer, New York (1989)
12. Hernández-Lerma, O., Lasserre, J.B.: Discrete-Time Markov Control Processes: Basic Optimality Criteria. Springer, New York (1996)
13. Hilgert, N., Minjárez-Sosa, J.A.: Adaptive control of stochastic systems with unknown disturbance distribution: discounted criterion. Math. Method Oper. Res. **63**, 443–460 (2006)
14. Huang, M.: Large-population LQG games involving a major player: the Nash certainty equivalence principle. SIAM J. Control Optim. **48**, 3318–3353 (2010)
15. Kolokoltsov, V.N., Troeva, M., Yang, W.: On the rate of convergence for the mean-field approximation of controlled diffusions with large number of players. Dyn. Games Appl. **4**, 208–230 (2014)
16. Lachapelle, A., Salomon, J., Turinici, G.: Computation of mean field equilibria in economics. Math. Models Methods Appl. Sci. **20**, 567–588 (2010)
17. Le Boudec, J.Y., McDonald, D., Mundinger, J.: A generic mean field convergence result for systems of interacting objects. In: 4th International Conference Quantitative Evaluation of Systems (2007). ISBN:0-7695-2883-X/07

18. Lasry, J.M., Lions, P.L.: Mean field games. Jap. J. Math. **2**, 229–260 (2007)
19. Minjárez-Sosa, J.A., Vega-Amaya, O.: Asymptotically optimal strategies for adaptive zero-sum discounted Markov games. SIAM J. Control Optim. **48**, 1405–1421 (2009)
20. Peyrard, N., Sabbadin, R.: Mean field approximation of the policy iteration algorithm for graph-based Markov decision processes. In: Biewka, G., Coradeschi, S., Perini, A., Traverso, P. (eds.) Frontiers in Artificial Intelligence and Applications, pp. 595–599. IOS Press, Amsterdam (2006)
21. Puterman, M.L.: Markov Decision Processes. Discrete Stochastic Dynamic Programming. Wiley, New York (1994)