

Original citation:

Kolokoltsov, V. N. (Vasilii Nikitich). (2012) Nonlinear markov games on a finite state space (mean-field and binary interactions). International Journal of Statistics and Probability, Volume 1 (Number 1). ISSN 1927-7032

Permanent WRAP url:

<http://wrap.warwick.ac.uk/64714>

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work of researchers of the University of Warwick available open access under the following conditions.

This article is made available under the Creative Commons Attribution 3.0 (CC BY 3.0) license and may be reused according to the conditions of the license. For more details see: <http://creativecommons.org/licenses/by/3.0/>

A note on versions:

The version presented in WRAP is the published version, or, version of record, and may be cited as it appears here.

For more information, please contact the WRAP Team at: publications@warwick.ac.uk



Nonlinear Markov Games on a Finite State Space (Mean-field and Binary Interactions)

Vassili N. Kolokoltsov

Department of Statistics, University of Warwick

Coventry CV4 7AL, UK

E-mail: v.kolokoltsov@warwick.ac.uk

Received: February 13, 2012 Accepted: February 27, 2012 Published: May 1, 2012

doi:10.5539/ijsp.v1n1p77 URL: <http://dx.doi.org/10.5539/ijsp.v1n1p77>

Supported by the AFOSR grant FA9550-09-1-0664 'Nonlinear Markov control processes and games' and by the Institute of Problems in Informatics of the Russian Academy of Science.

Abstract

Managing large complex stochastic systems, including competitive interests, when one or several players can control the behavior of a large number of particles (agents, mechanisms, vehicles, subsidiaries, species, police units, etc), say N_k for a player k , the complexity of the game-theoretical (or Markov decision) analysis can become immense as $N_k \rightarrow \infty$. However, under rather general assumptions, the limiting problem as all $N_k \rightarrow \infty$ can be described by a well manageable deterministic evolution. In this paper we analyze some simple situations of this kind proving the convergence of Nash-equilibria for finite games to equilibria of a limiting deterministic differential game.

Keywords: Markov control, Large complex systems, Dynamic law of large numbers, Differential games, Rates of convergence, Nonlinear Markov games

1. Introduction

A steady increase in complexity is one of the characteristic features of the modern technological development. It requires an appropriate (or better optimal) management of complex stochastic systems consisting of large number of interacting components (agents, mechanisms, vehicles, subsidiaries, species, police units, etc), which may have competitive or common interests. Carrying out a traditional Markov decision analysis for a large state space is often unfeasible. However, under rather general assumptions, the limiting problem as the number of components tends to infinity can be described by a well manageable deterministic evolution, which represents a performance of a dynamic law of large numbers (LLN). In general, this limiting deterministic evolution is measure-valued (it is an evolution of probability laws on the initial state space), and its probabilistic analysis has led to the notion of a nonlinear Markov process, see monograph (Kolokoltsov, 2010) and references therein. Its controlled version can be naturally called a nonlinear Markov control process or (in case of competitive interests) a nonlinear Markov game (Kolokoltsov, 2009). In case of finite initial state space, the corresponding space of measures is a finite-dimensional Euclidean space (more precisely its positive orthant \mathbf{R}_+^d), so that the limiting measure-valued evolution becomes a deterministic control process or a differential game in \mathbf{R}^d . This paper is devoted to the analysis of simplest situations of this kind, aiming at the identification of deterministic limit and proof of convergence with explicit rates. More precisely, we shall assume that there is a fixed number of players $\{1, \dots, K\}$ each controlling a stochastic system consisting of a large number $N_1, \dots, N_K \rightarrow \infty$ components respectively. These can be generals controlling armies, engineers controlling robot swarms, large banks managers controlling subsidiaries, etc. The components can interact between themselves and with agents of other groups. The limit $N_1, \dots, N_k \rightarrow \infty$ will be described by a differential game in \mathbf{R}_+^K .

The plan of the paper is as follows. In a preliminary Section 2 we set the stage by describing the dynamic law of large numbers for interacting Markov chains (without control). This topic is rather well developed by now, but we present it on the level of generality needed for what follows, including time nonhomogeneous chains (with discontinuous dependence on time) and rates of convergence resulting from Hölder continuity assumptions on the r.h.s. of the limiting ODE. Section 3 introduces control without competition. A strong progress for the analysis of such systems was made recently in (Gast, Gaujal & Le Boudec, 2010). We present it in a quite different form (see discussion in the last Section), which paves the road for the extension to competitive interests developed further. Main results are presented in Sections 3-5, where we discuss consecutively two player zero-sum games with mean -field interaction, two player zero-sum games with binary interaction and a K player noncooperative game. The proofs are given in Sections 6 and 7. The last section is devoted to a short review of relevant literature and to further perspectives.

The following (rather standard) notations for functional spaces will be used throughout the paper. For a closed subset Ω of a Euclidean space we shall denote by $C(\Omega)$ the Banach space of bounded continuous functions on Ω equipped with the usual sup-norm (which will be denoted simply $\|\cdot\|$ everywhere), and by $C^k(\Omega)$, $k \in \mathbf{N}$, the Banach space of k times continuously differentiable functions in the interior of Ω with f and all its derivatives up to and including order k having continuous and bounded extension to Ω , equipped with norm $\|f\|_{C^k(\Omega)}$ which is the sum of the sup-norms of f and all its derivatives up to and including order k . Finally, for $\alpha \in (0, 1]$, we denote by $C^{k,\alpha}(\Omega)$ the subspace of $C^k(\Omega)$ consisting of functions, whose k th order derivatives are Hölder continuous of index α . The Banach norm on this space is defined as the sum of the norm in $C^k(\Omega)$ plus the minimal Hölder constant. For an operator Φ in a Banach space B we shall denote by $\|\Phi\|_B$ the corresponding operator norm of Φ .

2. Preliminaries: LLN for Interacting Markov Chains

Let us first recall the basic setting of mean-field interacting particle systems with a finite number of types. Suppose our initial state space is a finite set $\{1, \dots, d\}$, which can be interpreted as the types of particles (say, possible opinions of individuals on a certain subject, or the levels of fitness in a military unit, or the types of robots in a robot swamp). Let $\{Q(t, x)\} = \{(Q_{ij})(t, x)\}$ be a family of $d \times d$ square Q -matrices or Kolmogorov matrices (i.e. non-diagonal elements of these matrices are non-negative and the elements of each row sum up to one) depending continuously on a vector x from the closed simplex

$$\Sigma_d = \{x = (x_1, \dots, x_d) \in \mathbf{R}_+^d : \sum_{j=1}^d x_j = 1\},$$

and piecewise continuously on time $t \geq 0$. For any x , the family $\{Q(\cdot, x)\}$ specifies a Markov chain on the state space $\{1, \dots, d\}$ with the generator

$$(Q(t, x)f)_n = \sum_{m \neq n} Q_{nm}(t, x)(f_m - f_n), \quad f = (f_1, \dots, f_d),$$

and with the intensity of jumps being

$$|Q_{ii}(t, x)| = -Q_{ii}(t, x) = \sum_{j \neq i} Q_{ij}(t, x).$$

In other words, the transition matrices $P(s, t, x) = (P_{ij}(s, t, x))_{i,j=1}^d$ of this chain satisfy the Kolmogorov forward equations

$$\frac{d}{dt} P_{ij}(s, t, x) = \sum_{l=1}^d Q_{lj}(t, x) P_{il}(s, t, x), \quad s \leq t.$$

Remark 1. Instead of piecewise continuous dependence on t we can assume that Q is uniformly bounded and depends measurably on t . Everything remains the same. This extension is relevant if one is interested in arbitrary discontinuous controls.

Suppose we have a large number of particles distributed arbitrary among the types $\{1, \dots, d\}$. More precisely our state space S is \mathbf{Z}_+^d , the set of sequences of d non-negative integers $N = (n_1, \dots, n_d)$, where each n_i specifies the number of particles in the state i . Let $|N|$ denote the total number of particles in state N : $|N| = n_1 + \dots + n_d$. For $i \neq j$ and a state N with $n_i > 0$ denote by N^{ij} the state obtained from N by removing one particle of type i and adding a particle of type j , that is n_i and n_j are changed to $n_i - 1$ and $n_j + 1$ respectively. The mean-field interacting particle system specified by the family $\{Q\}$ is defined as the Markov process on S specified by the generator

$$L_t f(N) = \sum_{i,j=1}^d n_i Q_{ij}(t, N/|N|) [f(N^{ij}) - f(N)]. \tag{1}$$

Probabilistic description of this process is as follows. Starting from any time and current state N one attaches to each particle a $|Q_{ii}(N/|N|)$ -exponential random waiting time (where i is the type of this particle). If the shortest of the waiting times τ turns out to be attached to a particle of type i , this particle jumps to a state j according to the distribution $(Q_{ij}/|Q_{ii}|)(N/|N|)$. Briefly, with this distribution and at rate $|Q_{ii}(N/|N|)$, any particle of type i can turn (migrate) to a type j . After any such transition the process starts again from the new state N^{ij} . Notice that since the number of particles $|N|$ is preserved by any jump, this process is in fact a Markov chain with a finite state space.

Remark 2. Yet another way of describing the chain generated by L_t is via the forward Kolmogorov (or master) equation for its transition probabilities $P_{MN}(s, t)$:

$$\frac{d}{dt} P_{MN}(s, t) = \sum_{i,j=1}^d (n_i + 1) Q_{ij}(t, \frac{N^{ji}}{|N|}) P_{MN^{ji}}(s, t) - \sum_{i,j=1}^d n_i Q_{ij}(t, \frac{N^{ij}}{|N|}) P_{MN}(s, t), \quad s \leq t.$$

To shorten the formulas, we shall denote the inverse number of particles by h , that is $h = 1/|N|$. Normalizing the states to $N/|N| \in \Sigma_d^h$, where Σ_d^h is a subset of Σ_d with coordinates proportional to h , leads to the generator of the form

$$L_t^h f(N/|N|) = \sum_{i=1}^d \sum_{j=1}^d \frac{n_i}{|N|} |N| Q_{ij}(t, N/|N|) [f(N^{ij}/|N|) - f(N/|N|)], \quad (2)$$

or equivalently

$$L_t^h f(x) = \sum_{i=1}^d \sum_{j=1}^d x_i Q_{ij}(t, x) \frac{1}{h} [f(x - he_i + he_j) - f(x)], \quad x \in h\mathbf{Z}_+^d, \quad (3)$$

where e_1, \dots, e_d denotes the standard basis in \mathbf{R}^d . With some abuse of notation, let us denote by $hN^{t,h}$ the corresponding Markov chain. The transition operators of this chain will be denoted by $\Psi_{s,t}^h$:

$$\Psi_{s,t}^h f(hN) = \mathbf{E}_{s,hN} f(hN(t, h)), \quad s \leq t, \quad (4)$$

where $\mathbf{E}_{s,x}$ denotes the expectation of the chain started at x at time s . These operators are known to form a propagator, i.e. they satisfy the chain rule (or Chapman-Kolmogorov equation)

$$\Psi_{s,t}^h \Psi_{t,r}^h = \Psi_{s,r}^h, \quad s \leq t \leq r.$$

We shall be interested in the asymptotic behavior of these chains as $h \rightarrow 0$. To this end, let us observe that, for $f \in C^1(\Sigma_d)$,

$$\lim_{|N| \rightarrow \infty, N/|N| \rightarrow x} |N| [f(N^{ij}/|N|) - f(N/|N|)] = \frac{\partial f}{\partial x_j}(x) - \frac{\partial f}{\partial x_i}(x),$$

so that

$$\lim_{|N| \rightarrow \infty, N/|N| \rightarrow x} L_t^h f(N/|N|) = \Lambda_t f(x),$$

where

$$\Lambda_t f(x) = \sum_{i=1}^d \sum_{j \neq i} x_i Q_{ij}(t, x) \left[\frac{\partial f}{\partial x_j} - \frac{\partial f}{\partial x_i} \right](x) = \sum_{k=1}^d \sum_{i \neq k} [x_i Q_{ik}(t, x) - x_k Q_{ki}(t, x)] \frac{\partial f}{\partial x_k}(x). \quad (5)$$

The limiting operator $\Lambda_t f$ is a first-order PDO with characteristics solving the equation

$$\dot{x}_k = \sum_{i \neq k} [x_i Q_{ik}(t, x) - x_k Q_{ki}(t, x)] = \sum_{i=1}^d x_i Q_{ik}(t, x), \quad k = 1, \dots, d, \quad (6)$$

called the *kinetic equations* for the process of interaction described above. The characteristics specify the dynamics of the deterministic time-nonhomogeneous Markov Feller process in Σ_d defined via the generator Λ_t . The corresponding transition operators act on $C(\Sigma_d)$ as

$$\Phi_{s,t} f(x) = f(X_{s,x}(t)), \quad s \leq t, \quad (7)$$

where $X_{s,x}(t)$ is the solution to (6) with the initial condition x at time s . These operators form a Feller propagator (i.e. $\Phi_{s,t}$ depend strongly continuous on s, t and satisfy the chain rule $\Phi_{s,t} \Phi_{t,r} = \Phi_{s,r}$, $s \leq t \leq r$). Of course in case of Q that do not depend on time t explicitly, $\Phi_{s,t}$ depend only on the difference $t - s$ and the operators $\Phi_t = \Phi_{0,t}$ form a Feller semigroup.

Remark 3. It is easy to see that if $x_k \neq 0$, then $(X_{s,x}(t))_k \neq 0$ for any $t \geq s$. Hence the boundary of Σ_d is not attainable for this semigroup, but, depending on Q , it can be glueing or not. For instance, if all elements of Q never vanish, then the points $X_{s,x}(t)$ never belong to the boundary of Σ_d for $t > s$, even if the initial point x does so.

The convergence of the Markov chains with generators of type (2) to a deterministic evolution and various versions of this result are well known, see e.g. (Darling & Norris, 2008; Kolokoltsov, 2010; Benaïm & Le Boudec, 2008) and references therein.

We present here an extension (for time-nonhomogeneous chains with discontinuous time dependence) of a result from (Kolokoltsov, 2011, Sect. 5.11), on a level of generality which allows us to get the corresponding convergence results for controlled problems as more or less straightforward corollaries.

Theorem 1. (i) Let all the elements $Q_{ij}(t, \cdot)$ belong to $C^{1,\alpha}(\Sigma)$, $\alpha \in (0, 1]$, with norms uniformly bounded in t . Then, if for some $s > 0$ and $x \in \mathbf{R}^d$, the initial data hN_s converge to x in \mathbf{R}^d , as $h \rightarrow 0$, the Markov chains $hN(t, h)$ with the initial data

hN_s (generated by L_t^h and with transitions $\Psi_{s,t}$) converge in distribution and in probability to the deterministic characteristic $X_{s,x}(t)$. For the corresponding converging propagators of transition operators the following rates of convergence hold:

$$\sup_{0 \leq s \leq t \leq T} \sup_{N \in \mathbb{Z}_+^d: |N|=1/h} [\Psi_{s,t}^h f(hN) - \Phi_{s,t} f(hN)] \leq C(T)(t-s)h^\alpha \|f\|_{C^{1,\alpha}(\Sigma_d)}, \tag{8}$$

for $f \in C^{1,\alpha}(\Sigma)$ and

$$\sup_{0 \leq s \leq t \leq T} [\mathbf{E}_{s,hN} f(hN(t,h)) - f(X_{s,x}(t))] \leq C(T) \left((t-s)h^\alpha \|f\|_{C^{1,\alpha}(\Sigma_d)} + \|f\|_{C^1(\Sigma_d)} |hN - x| \right), \tag{9}$$

where $C(T)$ depends only on the supremum in t of $C^{1,\alpha}(\Sigma)$ -norm of the functions $Q(t, x)$.

(ii) Assuming a weaker regularity condition, namely that $Q_{ij}(t, \cdot)$ belong to $C^1(\Sigma)$ uniformly in t , the convergence of Markov chains $hN(t, h)$ in distribution and in probability to the deterministic characteristics still holds, but instead of (8), we have weaker rates in terms of the modulus of continuity w_h of ∇f and Q (see (39) for the definition):

$$\begin{aligned} & \sup_{0 \leq s \leq t \leq T} \sup_{N \in \mathbb{Z}_+^d: |N|=1/h} [\Psi_{s,t}^h f(hN) - \Phi_{s,t} f(hN)] \\ & \leq C(T)(t-s) \left(w_{hC(T)}(\nabla f) + w_{hC(T)}(\nabla Q) \|f\|_{C^1(\Sigma_d)} \right), 0 \end{aligned} \tag{10}$$

where $C(T)$ depends on the $C^1(\Sigma)$ -norm of Q . A similar modification of (9) holds.

Our objective is to extend this result to interacting and competitively controlled families of Markov chains.

3. Mean Field Markov Control

As a warm-up, let us start with mean-field controlled Markov chains without competition. Suppose we are given a family of Q -matrices $\{Q(t, u, x)\} = \{(Q_{ij})(t, u, x), i, j = 1, \dots, d\}$, depending on $x \in \Sigma_d, t \geq 0$ and a parameter u from a metric space interpreted as control. The main assumption will be that $Q \in C^{1,\alpha}(\Sigma_d)$ as a function of x with the norm bounded uniformly in t, u , and Q depends continuously on t and u .

Any given bounded measurable curve $u(t), t \in [0, T]$, defines a Markov chain on Σ_d^h with the time-dependent family of generators of type (2), that is

$$L_{t,u(t)} f \left(\frac{N}{|N|} \right) = \sum_{i,j}^d n_i Q_{ij} \left(t, u(t), \frac{N}{|N|} \right) \left[f \left(\frac{N^{ij}}{|N|} \right) - f \left(\frac{N}{|N|} \right) \right], \tag{11}$$

or equivalently

$$L_{t,u(t)}^h f(x) = \sum_{i=1}^d \sum_{j=1}^d x_i Q_{ij}(t, u(t), x) \frac{1}{h} [f(x - h e_i + h e_j) - f(x)]. \tag{12}$$

For simplicity (and effectively without loss of generality), we shall stick further to controls $u(\cdot)$ from the class $C_{pc}[0, T]$ of piecewise-continuous curves (with a finite number of discontinuities).

Again for $f \in C^1(\Sigma_d)$,

$$\lim_{h=1/|N| \rightarrow 0, N/|N| \rightarrow x} L_{t,u(t)}^h f(N/|N|) = \Lambda_{t,u(t)} f(x),$$

where

$$\Lambda_{t,u(t)} f(x) = \sum_{k=1}^d \sum_{i \neq k} [x_i Q_{ik}(t, u(t), x) - x_k Q_{ki}(t, u(t), x)] \frac{\partial f}{\partial x_k}(x), \tag{13}$$

with the corresponding controlled characteristics governed by the equations

$$\dot{x}_k = \sum_{i \neq k} [x_i Q_{ik}(t, u(t), x) - x_k Q_{ki}(t, u(t), x)] = \sum_{i=1}^d x_i Q_{ik}(t, u(t), x), \quad k = 1, \dots, d. \tag{14}$$

For a given $T > 0$ and continuous functions J (current payoff) and V_T (terminal payoff), let $\Gamma(T, h)$ denote the problem of a centralized controller of the chain with $|N| = 1/h$ particles, aiming at maximizing the payoff

$$\int_0^T J \left(s, u(s), \frac{N(s, h)}{|N|} \right) ds + V_T \left(\frac{N(T, h)}{|N|} \right). \tag{15}$$

The optimal payoff will be denoted by $V^h(t, x)$:

$$V^h(t, x) = \sup_{u(\cdot) \in C_{pc}[t, T]} \mathbf{E}_{t,x}^{u(\cdot)} \left[\int_t^T (J(s, u(s), hN(s, h))) ds + V_T(hN(T, h)) \right], \quad (16)$$

where $\mathbf{E}_{t,x}^{u(\cdot)}$ denotes the expectation with respect to the Markov chain on Σ_d^h generated by (11) and started at $x = hN$ at time t .

We are aiming at approximating $V^h(t, x)$ by the optimal payoff

$$V(t, x) = \sup_{u(\cdot) \in C_{pc}[t, T]} \left[\int_t^T J(s, u(s), X_{t,x}(s)) ds + V_T(X_{t,x}(T)) \right] \quad (17)$$

for the controlled dynamics (14).

We can also obtain approximate optimal synthesis for problems $\Gamma(T, h)$ with large $|N| = 1/h$, at least if regular enough synthesis is available for the limiting system. Let us recall that a function $\gamma(t, x)$ is called an optimal synthesis (or an adaptive policy) for the problem $\Gamma(T, h)$ if

$$V^h(t, x) = \mathbf{E}_{t,x}^\gamma \left[\int_t^T (J(s, \gamma(s, hN(s, h)), hN(s)) ds + V_T(hN(T, h)) \right] \quad (18)$$

for all $t \leq T$ and $x \in \Sigma_d^h$, where $\mathbf{E}_{t,x}^\gamma$ denotes the expectation with respect to the Markov chain on Σ_d^h generated by (11) with $u(t) = \gamma(t, x)$ and starting at $x = hN$ at time t . A function $\gamma(t, x)$ is called an ϵ -optimal synthesis or an ϵ -adaptive policy, if the r.h.s. of (18) differs from its l.h.s. by not more than ϵ . Similarly an optimal synthesis or an adaptive policy are defined for the limiting deterministic system.

Theorem 2. (i) Assume that Q, J depend continuously on t, u and $Q, J \in C^{1,\alpha}(\Sigma_d)$, $\alpha \in (0, 1]$, as functions of x , with the norms bounded uniformly in t, u , and finally $V_T \in C^{1,\alpha}(\Sigma_d)$. Then

$$\begin{aligned} & \sup_{0 \leq t \leq T} [V^h(t, hN) - V(t, x)] \\ & \leq C(T)((T-t)h^\alpha + |hN - x|) \left(\|V_T\|_{C^{1,\alpha}(\Sigma_d)} + \sup_{s,u} \|J(t, u, \cdot)\|_{C^{1,\alpha}(\Sigma_d)} \right), \end{aligned} \quad (19)$$

with $C(T)$ depending only on the bounds of the norms of Q in $C^{1,\alpha}(\Sigma_d)$. Moreover, if $u(t)$ is an ϵ -optimal control for deterministic dynamics (14), that is the payoff obtained by using $u(\cdot)$ differs by ϵ from $V(t, x)$, then $u(\cdot)$ is also an $(\epsilon + C(T)h^\alpha)$ -optimal control for $|N| = 1/h$ particle system.

(ii) Suppose additionally that u belong to a convex subset of a Euclidean space and that $Q(t, u, x)$ depends Lipschitz continuously on u . Let $\epsilon \geq 0$, and let $\gamma(t, x)$ be a Lipschitz continuous function of x uniformly in t that represents an ϵ -optimal synthesis for the limiting deterministic control problem. Then, for any $\delta > 0$, there exists h_0 such that, for $h \leq h_0$, $\gamma(t, x)$ is an $(\epsilon + \delta)$ -optimal synthesis for the approximate optimal problem $\Gamma(T, h)$ on Σ_d^h .

Remark 4. As in Theorem (ii), there is a version of Theorem with $Q \in C^1(\Sigma_d)$ and $V_T \in C^1(\Sigma_d)$. We omit the detail. The same remark concerns other theorems given below.

Notice finally that by the standard dynamic programming, see e.g. (Fleming & Soner, 2006; McEneaney, 2006), the optimal payoff $V(t, x)$ given by (17) represents the unique viscosity solution of the HJB-Isaacs equation

$$\frac{\partial V}{\partial t}(t, x) + \max_u \left[J(t, u, x) + \sum_{i,k=1}^d x_i Q_{ik}(u, x) \frac{\partial V}{\partial x_k}(t, x) \right] = 0, \quad (20)$$

and the optimal payoff $V^h(t, x)$ given by (16) solves the HJB equation

$$\frac{\partial V^h}{\partial t}(t, x) + \max_u [J(t, u, x) + L_{t,u}^h V^h(t, x)] = 0. \quad (21)$$

Thus, as a corollary of Theorem , we have proved the convergence of the solutions of the Cauchy problem for equation (21) to the viscosity solution of (20).

4. Two Players with Mean-field Interaction

Let us turn to a game-theoretic setting starting with a simplest model of two competing mean-field interacting Markov chains. Suppose we are given two families of Q -matrices $\{Q(t, u, x) = (Q_{ij})(u, x)\}$ and $\{P(t, v, x) = (P_{ij})(v, x)\}$, $i, j = 1, \dots, d$, depending on $x \in \Sigma_d$ and parameters u and v from two subsets U and V of Euclidean spaces. Any given bounded measurable curves $u(t), v(t), t \in [0, T]$, define a Markov chain on $\Sigma_d^{1/|N|} \times \Sigma_d^{1/|M|}$, specified by the generator

$$L_{t,u(t),v(t)}f\left(\frac{N}{|N|}, \frac{M}{|M|}\right) = \sum_{i,j}^d n_i Q_{ij}(t, u(t), \frac{N}{|N|}) [f\left(\frac{N^{ij}}{|N|}, \frac{M}{|M|}\right) - f\left(\frac{N}{|N|}, \frac{M}{|M|}\right)] + \sum_{i,j}^d m_i P_{ij}(t, v(t), \frac{M}{|M|}) [f\left(\frac{N}{|N|}, \frac{M^{ij}}{|M|}\right) - f\left(\frac{N}{|N|}, \frac{M}{|M|}\right)], \tag{22}$$

where $N = (n_1, \dots, n_d), M = (m_1, \dots, m_d)$.

We shall assume for simplicity that $|N| = |M| = 1/h$.

Then (22) rewrites as

$$L_{t,u(t),v(t)}^h f(x, y) = \sum_{i=1}^d \sum_{j=1}^d x_i Q_{ij}(t, u(t), x) \frac{1}{h} [f(x - he_i + he_j, y) - f(x, y)] + \sum_{i=1}^d \sum_{j=1}^d y_i P_{ij}(t, v(t), y) \frac{1}{h} [f(x, y - he_i + he_j) - f(x, y)], \quad x, y \in h\mathbf{Z}_+^d. \tag{23}$$

For $f \in C^1(\Sigma_d \times \Sigma_d)$,

$$\lim_{h \rightarrow 0, N/|N| \rightarrow x, M/|M| \rightarrow y} L_{t,u(t),v(t)}^h f(N/|N|, M/|M|) = \Lambda_{t,u(t),v(t)} f(x, y),$$

where

$$\Lambda_{t,u(t),v(t)} f(x, y) = \sum_{k=1}^d \sum_{i \neq k} [x_i Q_{ik}(t, u(t), x) - x_k Q_{ki}(t, u(t), x)] \frac{\partial f}{\partial x_k}(x) + \sum_{k=1}^d \sum_{i \neq k} [y_i P_{ik}(t, v(t), y) - y_k P_{ki}(t, v(t), y)] \frac{\partial f}{\partial y_k}(y). \tag{24}$$

The corresponding controlled characteristics are governed by the equations

$$\dot{x}_k = \sum_{i \neq k} [x_i Q_{ik}(t, u(t), x) - x_k Q_{ki}(t, u(t), x)] = \sum_{i=1}^d x_i Q_{ik}(t, u(t), x), \quad k = 1, \dots, d, \tag{25}$$

$$\dot{y}_k = \sum_{i \neq k} [y_i P_{ik}(t, v(t), y) - y_k P_{ki}(t, v(t), y)] = \sum_{i=1}^d y_i P_{ik}(t, v(t), y), \quad k = 1, \dots, d. \tag{26}$$

For a given $T > 0$, let us denote by $\Gamma(T, h)$ the stochastic game with the dynamics specified by the generator (22) and with the objective of the player I (controlling Q via u) to maximize the payoff

$$\int_0^T J\left(s, u(s), v(s), \frac{N(s, h)}{|N|}, \frac{M(s, h)}{|M|}\right) ds + V_T\left(\frac{N(T, h)}{|N|}, \frac{M(T, h)}{|M|}\right) \tag{27}$$

for given functions J (current payoff) and V_T (terminal payoff), and with the objective of player II (controlling P via v) to minimize this payoff (zero-sum game). As previously we want to approximate it by the deterministic zero-sum differential game $\Gamma(T)$, defined by dynamics (25), (26) and the payoff of player I given by

$$\int_0^T J(s, u(s), v(s), X_{t,x}(s), Y_{t,y}(s)) ds + V_T(X_{t,x}(T), Y_{t,y}(T)). \tag{28}$$

Recall the basic notions of the upper and lower values for a game $\Gamma(T)$, see e.g. (Fleming & Soner, 2006) or (Malaeyev, 2000). As above, we shall use controls $u(\cdot)$ and $v(\cdot)$ from the classes $C_{pc}([0, T]; U)$ and $C_{pc}([0, T]; V)$ of piecewise-continuous curves with values in U and V respectively. A progressive strategy of player I is defined as a mapping β

from $C_{pc}([0, T]; V)$ to $C_{pc}([0, T]; U)$ such that if $v_1(\cdot)$ and $v_2(\cdot)$ coincide on some initial interval $[0, t]$, $t < T$, then so do $u_1 = \beta(v_1(\cdot))$ and $u_2 = \beta(v_2(\cdot))$. Similarly progressive strategies are defined for player *II*. Let us denote the sets of progressive strategies for players *I* and *II* by $S_p([0, T]; U)$ and $S_p([0, T]; V)$. Then the upper and the lower values for the game $\Gamma(T)$ are defined as

$$\begin{aligned}
 V_+(t, x, y) &= \sup_{\beta \in S_p([0, T]; U)} \inf_{v(\cdot) \in C_{pc}([0, T]; V)} \left[\int_t^T J(s, (\beta(v))(s), v(s), X_{t,x}(s), Y_{t,x}(s)) ds + V_T(X_{t,x}(T), Y_{t,x}(T)) \right], \\
 V_-(t, x, y) &= \inf_{\beta \in S_p([0, T]; V)} \sup_{u(\cdot) \in C_{pc}([0, T]; U)} \left[\int_t^T J(s, u(s), (\beta(u))(s), X_{t,x}(s), Y_{t,x}(s)) ds + V_T(X_{t,x}(T), Y_{t,x}(T)) \right].
 \end{aligned}
 \tag{29}$$

If the so called Isaac’s condition holds, that is, for any p_k, q_k ,

$$\begin{aligned}
 &\max_u \min_v \left[J(t, u, v, x, y) + \sum_{i,k=1}^d x_i Q_{ik}(t, v, x) q_k + \sum_{i,k=1}^d y_i P_{ik}(t, v, x) p_k \right] \\
 &= \min_v \max_u \left[J(t, u, v, x, y) + \sum_{i,k=1}^d x_i Q_{ik}(t, v, x) q_k + \sum_{i,k=1}^d y_i P_{ik}(t, v, x) p_k \right],
 \end{aligned}
 \tag{30}$$

then the upper and lower values coincide: $V_+(t, x, y) = V_-(t, x, y)$.

Similarly the upper and the lower values $V_+^h(t, x, y)$ and $V_-^h(t, x, y)$ for the stochastic game $\Gamma(T, h)$ are defined.

Theorem 3. Assume that Q, P, J depend continuously on t, u and $Q, P, J, V_T \in C^{1,\alpha}(\Sigma_d)$, $\alpha \in (0, 1]$, as functions of x , with the norms bounded uniformly in t, u, v . Then

$$\begin{aligned}
 &\sup_{0 \leq t \leq T} [V_{\pm}^h(t, hN) - V_{\pm}(t, x)] \\
 &\leq C(T)((T - t)h^\alpha + |hN - x|) \left(\|V_T\|_{C^{1,\alpha}(\Sigma_d)} + \sup_{s,u} \|J(t, u, v, \cdot)\|_{C^{1,\alpha}(\Sigma_d)} \right),
 \end{aligned}
 \tag{31}$$

with $C(T)$ depending only on the bounds of the norms of Q in $C^{1,\alpha}(\Sigma_d)$. Moreover, if $\beta \in S_p([0, T]; U)$ and $v(\cdot) \in C_{pc}([0, T]; V)$ are ϵ -optimal for the minimax problem (29), then this pair is also $(\epsilon + C(T)h^\alpha)$ -optimal for the corresponding stochastic game $\Gamma(T, h)$.

As in Theorem (ii), one can also approximate optimal (equilibrium) adaptive polices for $\Gamma(T, h)$, if regular enough (i.e. Lipschitz continuous) equilibrium adaptive policies exist for the limiting game $\Gamma(T)$. In fact, as is known from differential games, see e.g. (Fleming & Soner, 2006; Malaejev, 2000) or (Petrosjan & Zenkevich, 1996), the upper value $V_+(t, x, y)$ represents the unique viscosity solution of the upper Isaac’s equation

$$\frac{\partial V_+}{\partial t}(t, x, y) + \min_v \max_u [J(t, u, v, x, y) + \Lambda_{t,u,v} V_+(t, x, y)], \quad V_+(T, x, y) = V_T(x, y),
 \tag{32}$$

and $V_-(t, x, y)$ of the lower Isaac’s equation (with min and max placed in a different order). Similar equations are satisfied by the values of stochastic games $V_{\pm}^h(t, x, y)$, see e.g. (Fleming & Souganidis, 1989). Now, if V^* is a solution to the Cauchy problem (32) and there exist Lipschitz continuous functions $v^*(t, x, y)$ and $u^*(t, v, x, y)$ such that

$$\begin{aligned}
 u^*(t, v, x, y) &\in \operatorname{argmax}[J(t, u, v, x, y) + \Lambda_{t,u,v} V^*(t, x, y)], \\
 v^*(t, x, y) &\in \operatorname{argmin} \max_v [J(t, u, v, x, y) + \Lambda_{t,u,v} V^*(t, x, y)],
 \end{aligned}$$

then V^* is a saddle point for the differential game $\Gamma^+(T)$ giving the information advantage to maximizing player *I*, see e.g. (Fleming & Soner, 2006, Theorem 3.1). Analogously to Theorem (ii), we can conclude by Theorem that the policies $v^*(t, x, y)$ and $u^*(t, v, x, y)$ represent ϵ -equilibria for the corresponding stochastic game $\Gamma^+(T, h)$.

5. Two Players with Binary Interaction

In a slightly different setting one can assume that changes in a competitive control process occur as a result of group interactions, and are not determined just by the overall mean field distribution. Let us discuss a simple situation with binary interaction.

As in the previous section, assume we have two groups of d states (of objects or agents) controlled by players I and II respectively. Suppose now that any particle from a state i of the first group can interact with any particle from a state j of the second group (binary interaction) producing changes i to l and j to r with certain rates $Q_{ij}^{lr}(t, u, v)$ that may depend on controls u and v of the players. Assuming, as usual, that our particles are indistinguishable (any particle from a state is selected for interaction with equal probability), leads to the process, generated by the operators

$$L_{t,u(t),v(t)}f(N, M) = \sum_{i,j,l,r=1}^d n_i m_j Q_{ij}^{lr}(t, u(t), v(t), \frac{N}{|N|}, \frac{M}{|M|}) [f(N^{il}, M^{jr}) - f(N, M)].$$

Again let us assume for simplicity that $|M| = |N|$ and define $h = 1/|N| = 1/|M|$. To get a reasonable scaling limit, it is necessary to scale time by factor h leading to the generators

$$L_{t,u(t),v(t)}^h f(\frac{N}{|N|}, \frac{M}{|M|}) = h \sum_{i,j,l,r=1}^d n_i m_j Q_{ij}^{lr}(t, u(t), v(t), \frac{N}{|N|}, \frac{M}{|M|}) [f(N^{il}, M^{jr}) - f(N, M)], \tag{33}$$

which, for $x = hN, y = hM$ and $h \rightarrow 0$, tends to

$$\Lambda_{t,u(t),v(t)}f(x, y) = \sum_{i,j,l,r=1}^d x_i y_j Q_{ij}^{lr}(t, u(t), v(t), x, y) \left[\frac{\partial f}{\partial x_l} + \frac{\partial f}{\partial y_r} - \frac{\partial f}{\partial x_i} - \frac{\partial f}{\partial y_j} \right] (x, y). \tag{34}$$

The corresponding kinetic equations (characteristics of this first order partial differential operator) have the form

$$\begin{aligned} \dot{x}_k &= \sum_{i,j,r=1}^d y_j [x_i Q_{ij}^{kr}(t, u(t), v(t)) - x_k Q_{kj}^{ir}(t, u(t), v(t))], \\ \dot{y}_k &= \sum_{i,j,l=1}^d x_i [y_j Q_{ij}^{lk}(t, u(t), v(t)) - y_k Q_{ik}^{lj}(t, u(t), v(t))], \end{aligned}$$

As in the previous section, we are interested in the zero-sum stochastic game, which will again be denoted by $\Gamma(T, h)$, with the dynamics specified by generator (33) and with the objective of the player I (controlling Q via u) to maximize the payoff of the same type (27), and in an approximation of this game by the limiting deterministic zero-sum differential game $\Gamma(T)$, defined by the payoff (28) of player I .

Theorem 4. Assume that Q, J depend continuously on t, u, v and $Q, J, V_T \in C^{1,\alpha}(\Sigma_d)$, $\alpha \in (0, 1]$, as functions of x , with the norms bounded uniformly in t, u, v . Then the same estimate (31) holds for the difference of upper and lower values of limiting and approximating games.

Moreover, literally the same approximations (as in the game of the previous section) hold for optimal strategies and policies.

6. Several Players with Coupled Mean-field Interaction

Suppose now that there are K players, each one controlling a mean-field interacting Markov chain, which are coupled via the joint mean-field distribution. Namely, suppose we are given K families of Q -matrices $\{Q^k = (Q_{ij}^k)(t, u_k, \mu_1, \dots, \mu_K)\}$, $k = 1, \dots, K, i, j = 1, \dots, d$, depending on a parameter u_k from a metric space U_k (control space of the player k) and K vectors $\mu_k \in \Sigma_d$. Any bounded measurable vector-valued curve $u(t) = (u_1(t), \dots, u_K(t))$, $t \in [0, T]$, and natural numbers $|N_1|, \dots, |N_K|$ define a Markov chain on $S_{1/|N_1|} \times \dots \times S_{1/|N_K|}$ (each S_k consists of sequences of d negative integers $N_k = (n_1^k, \dots, n_d^k)$ totting up to a given number $|N_k|$), specified by the generator

$$\begin{aligned} L_{t,u(t)}f(\frac{N_1}{|N_1|}, \dots, \frac{N_K}{|N_K|}) &= \sum_{k=1}^K \sum_{i_k, j_k=1}^d n_{i_k}^k Q_{i_k j_k}^k \left(t, u_k(t), \frac{N_1}{|N_1|}, \dots, \frac{N_K}{|N_K|} \right) \\ &\times \left[f(\frac{N_1}{|N_1|}, \dots, \frac{N_{k-1}}{|N_{k-1}|}, \frac{N_k^{i_k j_k}}{|N_k|}, \frac{N_{k+1}}{|N_{k+1}|}, \dots, \frac{N_K}{|N_K|}) - f(\frac{N_1}{|N_1|}, \dots, \frac{N_K}{|N_K|}) \right], \end{aligned} \tag{35}$$

where $N_k = (n_1^k, \dots, n_d^k)$ with elements totting up to $|N_k|$. For

$$x_k = \frac{N_k}{|N_k|} = \left(\frac{n_1^k}{|N_1|}, \dots, \frac{n_d^k}{|N_k|} \right) \in h\mathbf{Z}_+^d, \quad k = 1, \dots, K,$$

this rewrites as

$$\begin{aligned} I_{t,u(t)}^h f(x_1, \dots, x_K) &= \frac{1}{h} \sum_{k=1}^K \sum_{i_k, j_k=1}^d x_k^{i_k} Q_{i_k j_k}^k(t, u_k(t), x_1, \dots, x_K) \\ &\times [f(x_1, \dots, x_{k-1}, x_k - he_{i_k} + he_{j_k}, x_{k+1}, \dots, x_K) - f(x_1, \dots, x_K)], \end{aligned} \tag{36}$$

where we denote by h the maximum of all $|N_k|^{-1}$.

For $f \in C^1(\Sigma_d \times \dots \times \Sigma_d)$,

$$\lim_{h \rightarrow 0, N_k/|N_k| \rightarrow x_k, k=1, \dots, K} I_{t,u(t)}^h f\left(\frac{N_1}{|N_1|}, \dots, \frac{N_K}{|N_K|}\right) = \Lambda_{t,u(t)} f(x_1, \dots, x_K),$$

where

$$\Lambda_{t,u(t)} f(x) = \sum_{k=1}^K \sum_{j,i=1}^d x_i Q_{ij}(t, u_k(t), x_1, \dots, x_K) \frac{\partial f}{\partial x_j^k}(x_1, \dots, x_K). \tag{37}$$

The corresponding controlled characteristics are governed by the equations

$$\frac{d}{dt} x_k^j = \sum_{i=1}^d x_i Q_{ij}(t, u_k(t), x_1, \dots, x_K), \quad j = 1, \dots, d, \quad k = 1, \dots, K. \tag{38}$$

For a given $T > 0$, suppose the objective of player k is to maximize the payoff

$$\begin{aligned} &V_k^h(t, x_1, \dots, x_K, u(\cdot)) \\ &= \mathbf{E}_{t,x_1, \dots, x_K}^{u(\cdot)} \int_0^T J_k\left(s, u_k(s), \frac{N_1(s, h)}{|N_1|}, \dots, \frac{N_K(s, h)}{|N_K|}\right) ds + V_T^k\left(\frac{N_1(T, h)}{|N_1|}, \dots, \frac{N_K(T, h)}{|N_K|}\right) \end{aligned}$$

with given functions J_k (current payoffs) and V_T^k (terminal payoffs). Let us denote by $\Gamma_K(T, h)$ the corresponding stochastic game. In the limit $h \rightarrow 0$ with all ratios $N_k/|N_j|$ uniformly bounded we get the deterministic differential game $\Gamma_K(T)$ of K players with separated dynamics (38) and the payoffs

$$V(t, x_1, \dots, x_K, u(\cdot)) = \int_0^T J_k(s, u_k(s), x_1(s), \dots, x_K(s)) ds + V_T^k(x_1(T), \dots, x_K(T)).$$

K -player differential games are much less understood as two-player zero-sum games, see e.g. (Friedman, 1971; Malafeyev, 2000; Olsder, 2001; Ramasubraanian, 2007; Tolwinski, Haurie & Leitmann, 1986) for informative discussions, including links with viscosity solutions of the systems of HJB equations. It is not our objective here to contribute to this development. We just want to stress that, as our method shows, most of the natural equilibria of various kinds that can be analyzed for the limiting differential game $\Gamma_K(T)$ do approximate the corresponding equilibria for games $\Gamma_K(T, h)$. As simplest examples, let us consider open-loop equilibria and K -player analogs of upper and lower values of zero-sum games. A vector curve $u^*(t) = (u_1^*(t), \dots, u_K^*(t))$ is called an open-loop Nash equilibrium for the game $\Gamma_K(T)$, if for any $k = 1, \dots, K$,

$$V(t, x_1, \dots, x_K, u^*(\cdot)) = \max_{u_k(\cdot)} V(t, x_1, \dots, x_K, u_1^*(\cdot), \dots, u_{k-1}^*(\cdot), u_k(\cdot), u_{k+1}^*(\cdot), \dots, u_K^*(\cdot)).$$

Similarly open-loop Nash equilibria are defined for the games $\Gamma_K(T, h)$.

On the other hand, for any permutation π of K players one can define a vector-value V_π arising from information discrimination specified by π . That is, the set of strategies of player $\pi(1)$ is just $C_{pc}([0, T]; U_{\pi(1)})$, and the progressive strategies for a player $\pi(k) \neq \pi(1)$ are defined as mapping

$$\beta_k : \times_{i=1}^{k-1} C_{pc}([0, T]; U_{\pi_i}) \rightarrow C_{pc}([0, T]; U_{\pi_k})$$

such that if $u^1 = (u_{\pi(1)}(\cdot), \dots, u_{\pi(k-1)}(\cdot))$ and $u^2 = (u_{\pi(1)}(\cdot), \dots, u_{\pi(k-1)}(\cdot))$ coincide on some initial interval $[0, t]$, $t < T$, then so do $\beta_k(u^1)$ and $\beta_k(u^2)$. Let us denote the sets of progressive strategies for players $\pi(k)$ by $S_p([0, T]; U_{\pi(k)})$.

The discriminated vector-value $V_\pi(t, x_1, \dots, x_K) = (V_\pi^1, \dots, V_\pi^K)(t, x_1, \dots, x_K)$ of the game $\Gamma_K^\pi(T)$ is defined as a Nash-equilibrium payoff of the game (in normal form) with these strategy spaces.

Theorem 5. Assume that Q_k, J_k depend continuously on all parameters and $Q_k, J_k, V_T^K \in C^{1,\alpha}(\Sigma_d)$, $\alpha \in (0, 1]$, as functions of x_1, \dots, x_k , with the norms bounded uniformly in t, u, v . Then the same estimate (31) holds for the difference of payoffs in open-loop Nash equilibria of limiting and approximating games, as well as for the difference of payoffs in Nash equilibria of the limiting and approximating discriminatory games specified by any permutation π .

There does not seem to exist any general results on the regularity of adaptive policies for N player game. On the other hand, in many examples, see (Case, 1967), and in some sense in a general position, see (Malafeyev, 2000), the state space can be decomposed into a finite number of open sets, where equilibrium adaptive policies are smooth, with these sets being separated by lower dimensional manifolds, where switching occurs. Under this condition, one could possibly prove the convergence of adaptive policies for large $|N|$ -approximations to adaptive policies of the limiting deterministic differential games, but such analysis is beyond the present contribution.

7. Auxiliary Lemmas

We shall need two simple lemmas (possibly well known for experts), one on the bounds for semigroups arising from deterministic processes, and another on the coupling of jump-type Markov processes.

Let us recall that the modulus of continuity for a continuous function ϕ on Σ_d is defined as follows:

$$w_h(\phi) = \sup\{|\phi(x_1) - \phi(x_2)| : |x_1 - x_2| \leq h\}, \quad h > 0. \tag{39}$$

Below we denote by ∇ the derivative operator with respect to variable $x \in \Sigma_d$.

Lemma 1. (i) Let $Q(t, x) \in C^1(\Sigma_d)$ as a function of x uniformly in t and depends measurably on t . Let $\Phi_{s,t}$ denote the linear operators (7), where $X_{s,x}(t)$ is the solution to (6) with the initial condition x at time s . Then $\Phi_{s,t}$ preserve the space $C^1(\Sigma_d)$ and

$$\|\Phi_{s,t}\|_{C^1(\Sigma_d)} \leq e^{\omega(t-s)}, \tag{40}$$

or more explicitly

$$\|\Phi_{s,t}f\|_{C^1(\Sigma_d)} \leq e^{\omega(t-s)}\|f\|_{C^1(\Sigma_d)},$$

where ω is the supremum over t of the norms of the functions $xQ(t, x)$ (as functions of x) in $C^1(\Sigma_d)$. Moreover, let \tilde{w}_h denote the supremum over t of the modulus of continuity of the functions $\nabla(xQ(t, x))$ (as functions of x). Then

$$w_h(\nabla(\Phi_{s,t}f)) \leq e^{\omega(t-s)} \left[w_h e^{\omega(t-s)} (\nabla f) + \tilde{w}_h e^{\omega(t-s)} \|f\|_{C^1(\Sigma_d)} \right]. \tag{41}$$

(ii) Suppose additionally that $Q(t, \cdot) \in C^{1,\alpha}(\Sigma_d)$, $\alpha \in (0, 1]$, uniformly in t , and let ω_α denote any uniform upper bound for the corresponding Hölder constants. Then the space $C^{1,\alpha}(\Sigma_d)$ is preserved by $\Phi_{s,t}$ and

$$\|\Phi_{s,t}\|_{C^{1,\alpha}(\Sigma_d)} \leq [1 + (t - s)\omega_\alpha]e^{\omega(t-s)}. \tag{42}$$

Proof. (i) The derivative of $X_{s,x}(t)$ with respect to x solves the linear ODE

$$\frac{d}{dt} \frac{\partial X_{s,x}(t)}{\partial x} = \frac{\partial}{\partial y} (yQ(t, y)|_{y=X_{s,x}(t)}) \frac{\partial X_{s,x}(t)}{\partial x}$$

implying

$$\left\| \frac{\partial X_{s,x}(t)}{\partial x} \right\| \leq e^{\omega(t-s)}, \tag{43}$$

and hence (40). The proof of (41) is the same as of (42) below, and therefore is omitted.

(ii) The function

$$\phi_{s,t}(x_1, x_2) = \left. \frac{\partial X_{s,x}(t)}{\partial x} \right|_{x=x_1} - \left. \frac{\partial X_{s,x}(t)}{\partial x} \right|_{x=x_2}$$

satisfies the linear ODE

$$\frac{d}{dt} \phi_{s,t}(x_1, x_2) = \frac{\partial}{\partial y} (yQ(t, y)|_{y=X_{s,x_1}(t)}) \left. \frac{\partial X_{s,x}(t)}{\partial x} \right|_{x=x_1} - \frac{\partial}{\partial y} (yQ(t, y)|_{y=X_{s,x_2}(t)}) \left. \frac{\partial X_{s,x}(t)}{\partial x} \right|_{x=x_2}.$$

The r.h.s. of this equation rewrites as

$$\left(\frac{\partial}{\partial y} (yQ(t, y)|_{y=X_{s,x_1}(t)}) - \frac{\partial}{\partial y} (yQ(t, y)|_{y=X_{s,x_2}(t)}) \right) \left. \frac{\partial X_{s,x}(t)}{\partial x} \right|_{x=x_1}$$

$$+ \frac{\partial}{\partial y} (yQ(t, y))|_{y=X_{s,x_2}(t)} \left(\frac{\partial X_{s,x}(t)}{\partial x} \Big|_{x=x_1} - \frac{\partial X_{s,x}(t)}{\partial x} \Big|_{x=x_2} \right)$$

and hence, by (40), is bounded in magnitude by the expression

$$\omega_\alpha e^{\omega(t-s)} |x_1 - x_2|^\alpha + \omega |\phi_{s,t}(x_1, x_2)|.$$

By Gronwall's lemma,

$$|\phi_{s,t}(x_1, x_2)| \leq (t - s)\omega_\alpha e^{\omega(t-s)} |x_1 - x_2|^\alpha.$$

Hence the Hölder constant (of index α) for the function $\partial X_{s,x}(t)/\partial x$, as a function of x , is bounded by $\omega_\alpha e^{\omega(t-s)}$ implying (42), because

$$\begin{aligned} & \frac{\partial}{\partial x} f(X_{s,x}(t))|_{x=x_1} - \frac{\partial}{\partial x} f(X_{s,x}(t))|_{x=x_2} \\ &= \left[\frac{\partial f}{\partial x}(X_{s,x_1}(t)) - \frac{\partial f}{\partial x}(X_{s,x_2}(t)) \right] \frac{\partial X_{s,x}(t)}{\partial x} \Big|_{x=x_1} \\ &+ \frac{\partial f}{\partial x}(X_{s,x_2}(t)) \left(\frac{\partial X_{s,x}(t)}{\partial x} \Big|_{x=x_1} - \frac{\partial X_{s,x}(t)}{\partial x} \Big|_{x=x_2} \right). \end{aligned}$$

The next lemma will be used only for the proof of the second part of Theorem 2. It is needed to compare the effects of applying different adaptive policies.

Lemma 2. *Let $Z_z(t)$ and $W_w(t)$ be two jump-type Markov processes in \mathbf{R}^d (z and w stand for the initial points) with integral generators*

$$L^Z f(x) = \int (f(x+y) - f(x))\nu(x, y)M(dy), \quad L^W f(x) = \int (f(x+y) - f(x))\mu(x, y)M(dy),$$

with a certain probability measure M on \mathbf{R}^d and uniformly bounded non-negative functions $\nu(x, y)$, $\mu(x, y)$ such that $\int |y|\nu(x, y)M(dy)$ and $\int |y|\mu(x, y)M(dy)$ are bounded for any x and

$$\begin{aligned} & \int |y|\nu(x_1, y) - \nu(x_2, y)|M(dy) \leq \varkappa|x_1 - x_2|, \\ & \int |y|\nu(x, y) - \mu(x, y)|M(dy) \leq \epsilon \end{aligned}$$

for a $\varkappa > 0$ and an $\epsilon > 0$.

Then there exists a Markov process $X_{z,w}(t)$ on $\mathbf{R}^d \times \mathbf{R}^d$ that couples $Z_z(t)$ and $W_w(t)$ in the sense that the distribution of the first (respectively second) coordinate of $X_{z,w}(t)$ coincides with the distribution of $Z_z(t)$ (respectively $W_w(t)$), and such that, with respect to this coupling,

$$\mathbf{E}|Z_z(t) - W_w(t)| \leq e^{\varkappa t}(|z - w| + t\epsilon). \tag{44}$$

Proof. Let $X_{z,w}(t)$ be specified by the so called marching coupling, that is by the generator

$$\begin{aligned} Lf(x_1, x_2) &= \int [f(x_1 + y, x_2 + y) - f(x_1, x_2)]m(x_1, x_2, y)M(dy) \\ &+ \int [f(x_1 + y, x_2) - f(x_1, x_2)][\nu(x_1, y) - m(x_1, x_2, y)]M(dy) \\ &+ \int [f(x_1, x_2 + y) - f(x_1, x_2)][\mu(x_2, y) - m(x_1, x_2, y)]M(dy). \end{aligned}$$

where

$$m(x_1, x_2, y) = \min(\nu(x_1, y), \mu(x_2, y)).$$

Clearly, if f does not depend on the second argument, then $Lf(x_1, x_2) = L^Z f(x_1)$, and if f does not depend on the first argument, then $Lf(x_1, x_2) = L^W f(x_2)$, so that $X_{z,w}(t)$ is really a coupling of $Z_z(t)$ and $Z_w(t)$. Moreover, as one sees by inspection, for $f(x_1, x_2) = |x_1 - x_2|$,

$$Lf(x_1, x_2) \leq \int |y|[\max(\nu(x_1, y), \mu(x_2, y)) - m(x_1, x_2, y)]M(dy).$$

Consequently,

$$Lf(x_1, x_2) \leq \int |y| |\nu(x_1, y) - \mu(x_2, y)| M(dy) \leq \varkappa |x_1 - x_2| + \epsilon.$$

By Dynkin’s formula, the process

$$f(X_{x_1, x_2}(t)) - \int_0^t Lf(X_{x_1, x_2}(s)) ds$$

is a martingale. Hence

$$\begin{aligned} & \mathbf{E}|X_{x_1, x_2}^1(t) - X_{x_1, x_2}^2(t)| \\ & \leq |x_1 - x_2| + \varkappa \int_0^t \mathbf{E}|X_{x_1, x_2}^1(s) - X_{x_1, x_2}^2(s)| ds + t\epsilon \end{aligned}$$

(where $X = (X^1, X^2)$), which by Gronwall’s lemma implies

$$\mathbf{E}|X_{x_1, x_2}^1(t) - X_{x_1, x_2}^2(t)| \leq e^{\varkappa t} (|x_1 - x_2| + t\epsilon),$$

that is (44).

8. Proof of the Theorems

Proof of Theorem 1.

Let us start by noting that the transition operators $\Psi_{s,t}^h$ of the Markov chains $hN^{t,h}$ satisfy (outside the finite number of discontinuity points of $Q(t, x)$) the standard Kolmogorov equations

$$\frac{d}{dt} \Psi_{s,t}^h f = \Psi_{s,t}^h L_t f, \quad \frac{d}{ds} \Psi_{s,t}^h f = -L_s \Psi_{s,t}^h f, \quad 0 \leq s \leq t, \tag{45}$$

for any function f on Σ_d^h . Similarly, the transition operators $\Phi_{s,t}$ of the deterministic Markov process $X_{s,x}(t)$ satisfy the equations

$$\frac{d}{dt} \Phi_{s,t} f = \Phi_{s,t}^h \Lambda_t f, \quad \frac{d}{ds} \Phi_{s,t} f = -\Lambda_s \Phi_{s,t} f, \quad 0 \leq s \leq t \tag{46}$$

(again outside the discontinuity points of $Q(t, x)$), for any $f \in C^1(\Sigma_d)$, which is of course easily checked.

To compare these propagators, we shall use the following standard trick. We write

$$\Psi_{s,t}^h f - \Phi_{s,t} f = \Psi_{s,t}^h \Phi_{r,t} |_{r=s} f = \int_s^t \frac{d}{dr} (\Psi_{s,r}^h \Phi_{r,t}) f dr = \int_s^t \Psi_{s,r}^h (L_r^h - \Lambda_r) \Phi_{r,t} f dr,$$

if only $\Phi_{r,t} f \in C^1(\Sigma_d)$ for any r . Let us apply this equation to an $f \in C^{1,\alpha}(\Sigma_d)$. By Lemma 1 (ii), $\Phi_{r,t} f \in C^{1,\alpha}(\Sigma_d)$ for all r with a uniform bound, so that the above equation applies. Moreover, from (3) and (5) we conclude that

$$\|(L_r^h - \Lambda_r) f\| \leq C(T) h^\alpha \|f\|_{C^{1,\alpha}(\Sigma_d)}, \tag{47}$$

with a constant $C(T)$ depending on the sup-norm of $Q(t, x)$ only. Consequently, by (42),

$$\|\Psi_{s,t}^h f - \Phi_{s,t} f\| \leq (t - s) C(T) h^\alpha [1 + (t - s) \omega_\alpha] e^{\omega(t-s)} \|f\|_{C^{1,\alpha}(\Sigma_d)},$$

implying (8). The estimate (9) is a straightforward corollary, taking into account (40). The convergence of the chains in distribution follows from the convergence of transition operators by a well known general result, see e.g. (Kallenberg, 2002), or (Kolokoltsov, 2011). Moreover, weak convergence of random variables to a constant implies their convergence in probability.

Finally, (10) is proved analogously, if instead of (47), one uses the estimate

$$\|(L_r^h - \Lambda_r) f\| \leq C(T) w_h(\nabla f) \tag{48}$$

together with (41).

Proof of Theorem 2.

(i) By Theorem 1 (and taking into account that the integral of the function J can be approximated by Riemannian sums), for any $u(\cdot) \in C_{pc}(t, T)$,

$$V^h(t, x, u(\cdot)) = \mathbf{E}_{t,x}^{u(\cdot)} \left[\int_t^T (J(s, u(s), hN(s, h)) ds + V_T(hN(T, h)) \right]$$

differs from

$$V(t, x, u(\cdot)) = \int_t^T J(s, u(s), X_{t,x}(s)) ds + V_T(X_{t,x}(T))$$

by

$$C(T)((T-t)h^\alpha + |hN-x|) \left(\|V_T\|_{C^{1,\alpha}(\Sigma_d)} + \sup_{s,u} \|J(t, u, \cdot)\|_{C^{1,\alpha}(\Sigma_d)} \right).$$

Hence the same estimate holds for the difference of the suprema of these two functions of $u(\cdot)$.

(ii) First of all approximating \tilde{J} by a smooth function (and noting that all payoffs will be then uniformly approximated) reduces the problem to the case of a smooth J . Next, we can approximate $\gamma(t, x)$ arbitrary close by a smooth function $\tilde{\gamma}(t, x)$ in x , so that the corresponding $V(t, x, \gamma)$ and $V(t, x, \tilde{\gamma})$ differ by arbitrary small amounts. By Theorem 1, $V^h(t, x, \tilde{\gamma})$ will converge to $V(t, x, \tilde{\gamma})$ as $h \rightarrow 0$, and hence $\tilde{\gamma}$ becomes an $(\epsilon + \delta)$ -optimal policy of $\Gamma(T, h)$ for small enough h . It remains to compare $V^h(t, x, \tilde{\gamma})$ and $V^h(t, x, \gamma)$. But they are close by Lemma 2.

Proof of Theorems 3, 4, 5. It is the same as the proof of Theorem 2, being based on Theorem 1 and an evident observation that if two families of functions are uniformly close, then so are also their minimax values.

9. Conclusion and Bibliographical Comments

The work on deterministic LLN limits for interacting particles and its representation in terms of nonlinear kinetic equations goes back to (Leontovich, 1935) and (Bogolyubov, 1946). A deduction of kinetic equations in a very general setting of k th order (binary, ternary, etc) interaction can be found in (Maslov & Tariverdiev, 1982) or (Belavkin & Kolokoltsov, 2003). In particular, for the corresponding limit in a game-theoretic setting (replicator dynamics), we can refer to (Benaïm & Weibull, 2003) or the last section of (Kolokoltsov & Malafeyev, 2010).

For the introduction to nonlinear Markov processes we refer to (Kolokoltsov, 2007, 2010) and references therein, or, for more physical point of view, to (Frank, 2008) or (Zak, 2002).

Mean-field control is a rapidly expanding area, see e.g. (Andersson & Djehiche, 2003; Buckdahn, B. Djehiche, J. Li, & S. Peng, 2009) and references therein, for diffusion based models, and (Le Boudec, McDonald, & Munding, 2007; Gast & B. Gaujal, 2009; Gast, Gaujal, & Le Boudec, 2010; Bordenave, McDonald, & Proutiere, 2007; Benaïm & Le Boudec, 2008; Milutinovic & Lima, 2006) for discrete models, more engineering application oriented. In these papers one can find various concrete applications (from robot swarms to transportation theory and networks), which are also relevant to the mathematical models discussed in the present paper.

The closest to our setting seems to be the recent work (Gast, Gaujal, & Le Boudec, 2010), which is devoted to a convergence result similar to our Theorem 2. However, in (Gast, Gaujal, & Le Boudec, 2010) a continuous time mean-field control model is obtained as a limit of discrete-time models, and we work directly with continuous time models. More essentially, under a slightly stronger regularity assumptions on the model (basically continuous differentiability of coefficients instead of Lipschitz continuity) we obtain explicit rate of convergence for the averages over empirical measures, and in some cases even for adaptive control policies. Moreover, our main objective was to develop a general framework to treat competitive control problems presented in Sections 4-6. Even further, we developed a method to treat not only the simplest mean-field type interactions, but more involved binary or ternary ones.

As was mentioned, our initial state space was finite, resulting in the corresponding measure-valued limit being a finite-dimensional differential game. In more general setting (which will be discussed elsewhere), for an arbitrary initial state space of a single particle (in our large ensemble), the corresponding limit becomes truly measure-valued controlled nonlinear evolution (controlled nonlinear Markov process) specified by kinetic equations of rather general type. Other possible extensions can include models with variable number of particles (robots becoming out of order, or military units destroyed, etc).

A slightly different class of models describe the situations when each agent in a large group pursue his/her own interests. These class of models is often referred to as mean-field games. Related equilibrium concept was called the Nash certainty equivalence principle, see (Lasry & Lions, 2006; Achdou & Capuzzo-Dolcetta, 2010; Gomes, Mohr, & Souza, 2010; Huang, Caines, & Malhamé, 2003; Huang, Malhamé, & Caines, 2006; Huang, 2010; Kolokoltsov, Li, & Yang, 2011) and references therein. Unlike our centralized setting above, analysis of these models does not lead to the control of distributions in the limit, but rather to certain consistency condition on homogeneous individual controls (Nash certainty equivalence). However, recent paper (Huang, Malhamé, & Caines, 2007) aims at linking individual and centralized controls.

The main objective of this paper was to approximate discrete stochastic systems with large number of particles by a simpler continuous state limit. One can also look at these results from an opposite point of view: as approximating

differential games by discrete Markov chains. This point of view would relate our results with Kushner (2002), Kushner and Dupuis (2001).

References

- Andersson, D., & Djehiche, B. (2011). A maximum principle for SDEs of mean-field type. *Appl. Math. Optim.*, *63*, 341-356. <http://dx.doi.org/10.1007/s00245-010-9123-8>
- Belavkin, V. P., & Kolokoltsov, V. N. (2003). On general kinetic equation for many particle systems with interaction, fragmentation and coagulation. *Proc. Royal Soc. Lond. A*, *459*, 727-748. <http://dx.doi.org/10.1098/rspa.2002.1026>
- Bogolyubov, N. N. (1946). Problems of the dynamic theory in statistical physics. Moscow, Nauka (in Russian).
- Bordenave, C., McDonald, D., & Proutiere A. (2007). A particle system in interaction with a rapidly varying environment: Mean field limits and applications.
- Buckdahn, R., Djehiche, B., Li, J., & Peng, S. (2009). Mean-field backward stochastic differential equations: a limit approach. *Ann. Prob.*, *37*, 1524-1565. <http://dx.doi.org/10.1214/08-AOP442>
- Case, J. H. (1967). Toward a Theory of Many Player Differential Games. *SIAM Journal on Control*, *7*, 179-197. <http://dx.doi.org/10.1137/0307013>
- Darling, R. W. R., & Norris, J. R. (2008). Differential equation approximations for Markov chains. *Probab. Surv.*, *5*, 3779.
- Fleming, W. H., & Soner, H. M. (2006). Controlled Markov Processes and Viscosity Solutions. Sec. Edition, NY: Springer.
- Fleming, W. H., & Souganidis, P. E. (1989). On the existence of value functions of two-player, zero-sum stochastic differential games. *Indiana Univ. Math. J.*, *38*, 293-314. <http://dx.doi.org/10.1512/iumj.1989.38.38015>
- Frank, T. D. (2008). Nonlinear Markov processes. *Phys. Lett. A*, *372*, 4553-4555. <http://dx.doi.org/10.1016/j.physleta.2008.04.027>
- Friedman, A. (1971). *Differential Games*. New York: Wiley-Interscience.
- Gast, N., & Gaujal, B. (2009). A Mean Field Approach for Optimization in Particle Systems and Applications. Proc. 4th Intern. ICST Conf. on Performance Evaluation Methodologies and Tools.
- Gast, N., Gaujal, B., & Le Boudec, J.-Y. (2010). Mean field for Markov Decision Processes: from Discrete to Continuous Optimization. INRIA report No. 7239.
- Gomes, D. A., Mohr, J., & Souza, R. R. (2010). Discrete time, finite state space mean field games. *J. Math. Pures Appl.*, *9*, 308-328. <http://dx.doi.org/10.1016/j.matpur.2009.10.010>
- Huang, M., Caines, E., & Malhamé, R. P. (2003). Individual and mass behaviour in large population stochastic wireless power control problems: centralized and Nash equilibrium solutions. Proc. the 42nd IEEE Conference on Decision and Control, Maui, Hawaii, 98-103.
- Huang, M., Malhamé, R. P., & Caines, P. E. (2006). Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. *Communications in information and systems*, *6*, 221-252.
- Huang, M. (2010). Large-population LQG games involving a major player: the Nash certainty equivalence principle. *SIAM J. Control Optim.*, *48*, 3318-3353. <http://dx.doi.org/10.1137/080735370>
- Huang, M, Caines, P. E. & Malhamé, R. P. (2007). Large-population cost-coupled LQG problems with nonuniform agents: individual-mass behavior and decentralized ϵ -Nash equilibria. *IEEE Trans. Automat. Control*, *52*, 1560-1571. <http://dx.doi.org/10.1109/TAC.2007.904450>
- Kallenberg, O. (2002). Foundations of Modern Probability. 2ed., Springer.
- Kolokoltsov, V. N. (2007). Nonlinear Markov Semigroups and Interacting Lévy Type Processes. *Journ. Stat. Physics*, *126*, 585-642. <http://dx.doi.org/10.1007/s10955-006-9211-y>
- Kolokoltsov, V. N. (2009). Nonlinear Markov games. Talk on 'Adversarial and Stochastic Elements in Autonomous Systems' Workshop, Washington DC, <http://maeresearch.ucsd.edu/mceneaney/ASEAS/Kolokoltsov.pdf>
- Kolokoltsov, V. N. (2010). *Nonlinear Markov processes and kinetic equations*. Cambridge University Press.
- Kolokoltsov, V. N. (2011). Markov processes, semigroups and generators. De Gruyter, Studies in Mathematics v. 38.

- Kolokoltsov, V. N., Li, J., & Yang, W. (2011) Mean Field Games and Nonlinear Markov Processes.
- Kolokoltsov, V. N., & Malafeyev, O. A. (2010). Understanding Game Theory. World Scientific, Singapore.
- Kushner, H. J. (2002). Numerical approximations for stochastic differential games. *SIAM Journal of Control and Optimization*, 41, 457-486. <http://dx.doi.org/10.1137/S0363012901389457>
- Kushner, H. J., & Dupuis, P. (2001). Numerical methods for stochastic control problems in continuous time. 2ed., New York: Springer.
- Lasry, J.-M., & Lions, L. Jeux Champ Moyen. *C. R. Math. Acad. Sci. Paris*, 343, 619-625 and 679-684.
- Le Boudec, J.-Y., McDonald, D., & Munding, J. (2007). A Generic Mean Field Convergence Result for Systems of Interacting Objects. *QEST 2007 (4th International Conference on Quantitative Evaluation of SysTems)*, 3-18.
- Leontovich, M. A. (1935). Main equations of the kinetic theory from the point of view of random processes. *Journal of Experimental and Theoretical Physics*, 5, 211-231.
- Malafeyev, O. A. (2000). Controlled conflict systems (In Russian). St. Petersburg State University Publication, ISBN 5-288-01769-7.
- Maslov, V. P., & Tariverdiev, C. E. (1982). Asymptotics of the Kolmogorov-Feller equation for systems with the large number of particles. *Itogi Nauki i Techniki. Teoriya veroyatnosti*, v.19, VINITI, Moscow (in Russian), 85-125.
- Benaïm, M., & Le Boudec, J.-Y. (2008). A class of mean field interaction models for computer and communication systems. *Performance Evaluation*, 65, 823-838. <http://dx.doi.org/10.1016/j.peva.2008.03.005>
- Benaïm, M., & Weibull, J. (2003). Deterministic approximation of stochastic evolution in games. *Econometrica*, 71, 873-903. <http://dx.doi.org/10.1111/1468-0262.00429>
- McEneaney, W. (2006). *Max-plus methods for nonlinear control and estimation. Systems and Control: Foundations and Applications*. Boston, MA: Birkhäuser Boston, Inc.
- Milutinovic, D., & Lima, P. (2006). Modeling and Optimal Centralized Control of a Large-Size Robotic Population. *Robotics, IEEE Transactions on Robotics*, 22, 1280-1285. <http://dx.doi.org/10.1109/TRO.2006.882941>
- Olsder, G. J. (2001). On open- and closed-loop bang-bang control in nonzero-sum differential games. *SIAM J. Control Optim.*, 40, 1087-1106. <http://dx.doi.org/10.1137/S0363012900373252>
- Petrosjan, L. A., & Zenkevich, N. A. (1996). Game Theory, World Scientific, Singapore.
- Ramasubraanian, S. (2007). A d -person Differential Game with State Space Constraints. *Appl. Math. Optim.*, 56, 312-342. <http://dx.doi.org/10.1007/s00245-007-9011-z>
- Song, Q. S. (2008). Convergence of Markov chain approximation on generalized HJB equation and its applications. *Automatica J. IFAC*, 44, 761-766. <http://dx.doi.org/10.1016/j.automatica.2007.07.014>
- Tolwinski, B., Haurie, A., & Leitmann, G. (1986). Cooperative Equilibria in Differential Games. *J. Math. Anal. Appl.*, 119, 182-202. [http://dx.doi.org/10.1016/0022-247X\(86\)90152-6](http://dx.doi.org/10.1016/0022-247X(86)90152-6)
- Zak, M. (2000). Quantum Evolution as a Nonlinear Markov Process. *Foundations of Physics Letters*, 15, 229-243. <http://dx.doi.org/10.1023/A:1021079403550>