Optimal control of interacting particle systems

Charles Bordenave, Venkat Anantharam

Optimal control of interacting particle systems

Charles Bordenave

CNRS & Université de Toulouse
Institut de Mathématiques de Toulouse
bordenave@math.univ-toulouse.fr

and

Venkat Anantharam

University of California, Berkeley
Department of Electrical Engineering and Computer Science
ananth@eecs.berkeley.edu

June 20, 2009

1 Introduction and motivation

The adaptive control of Markovian stochastic processes has a mature theory, either for optimal discounted reward [3, 2] than for average reward [1]. However, the analysis of controlled particles systems has not been studied closely. The aim of this paper is to fill this gap, and bring concepts of statistical physics into the field of optimal control.

The general framework is the following, N particles are in interaction and their interactions depend on a control parameter. The performance of the particle system is described by a reward associated to an average over the particles states and over time. With the appropriate hypothesis, if the control parameter is non-adaptive, (i.e. does not depend on the particle state), classical mean-field theory predicts that, as the number of particles grows large, the time evolution of the empirical measure of the particle states converges to the trajectory of a measure solution of an appropriate deterministic differential equation. Now consider, for each number of particles, N, an optimal adaptive control which maximizes the associated reward. It is now unclear whether or not the same convergence holds, as N grows large. More importantly, it is not clear whether or not this optimal control strategy and its associated reward converge to an optimal control strategy and its reward for the deterministic control problem associated to the limiting differential equation. In this paper, we deal with this issue.

For interacting particle systems, a phenomena which draws a tremendous attention in statistical physics is symmetry breaking. Loosely speaking, this phenomena occurs if a macroscopic state of the system is not symmetric with respect to the symmetry of the interactions in the system. In the framework of controlled particle systems, a control strategy is not symmetric, if enforcing this control breaks the initial symmetry of particles states. It is of prime interest to know if the optimal control strategies break the symmetry of the interactions. We will discuss when this type of symmetry breaking phenomena is expected through typical examples.

The main motivations of this work come from nanotechnologies and communication networks. Typically, in these engineering fields, a large number of particles or communication units are in interaction and a central controller may aim at optimizing a performance measure of the system via a control on the transitions rates of the particles or communication units. In many examples, the performance measure is simply the number of particles or communication units in a given state.

The remainder of the paper is organized as follows. In Section 2, we introduce our model and state our main results for discounted reward and average reward. Section 4 is dedicated to the proofs of our main results. In Section 3, we discuss some extensions of our model and of our results. Finally, Section 5 is a collection of examples and we illustrate the symmetry breaking phenomena for optimal control strategies.

2 Controlled particle systems

2.1 Model description

We consider N particles evolving in a finite state space $\mathcal{X}$ at discrete time $t \in \mathbb{N}$. The state of particle i at time t is $X_i^N(t) \in \mathcal{X}$ and the trajectory of the particle is $X_i^N = X_i^N(\cdot) \in \mathcal{X}^{\mathbb{N}}$. The state of the particle system at time t is described by the vector $X^N(t) = (X_1^N(t), \cdots, X_N^N(t))$.

Let $\mathcal{P}(\mathcal{X})$ (resp. $\mathcal{P}(\mathcal{X}^{\mathbb{N}})$) denotes the space of probability measures on $\mathcal{X}$ (resp. $\mathcal{X}^{\mathbb{N}}$). $\mathcal{P}_N(\mathcal{X}) = \{q \in \mathcal{P}(\mathcal{X}) : \forall x \in \mathcal{X}, Nq(\{x\}) \in \mathbb{N}\}$, the set of measures on $\mathcal{X}$ putting a mass in $\mathbb{N}/N$ at each element of $\mathcal{X}$. We define the empirical measures in $\mathcal{P}(\mathcal{X})$ and $\mathcal{P}(\mathcal{X}^{\mathbb{N}})$ respectively

$$\mu_t^N = \frac{1}{N}\sum_{i=1}^N \delta_{X_i^N(t)} \quad \text{and} \quad \mu^N = \frac{1}{N}\sum_{i=1}^N \delta_{X_i^N}.$$

We consider a sequence of random variables in $[0,1)$, $Z^N = (Z_0^N, Z_1^N, \cdots)$ thought as a source of external randomness in the system. Let $\mathcal{F}_t^N = \sigma(\mu_0^N, Z_0^N, \cdots, \mu_t^N, Z_t^N)$ be the filtration associated to the process (μ^N, Z^N) and $\overline{\mathcal{F}_t^N} = \sigma(X^N(0), Z_0^N \cdots, X^N(t), Z_t^N)$ the filtration associated to (X^N, Z^N). The evolution of the particle system depends on a $\mathcal{F}_t^N$-adapted control process $U^N = (U_t^N)_{t\in\mathbb{N}} \in \mathcal{U}^{\mathbb{N}}$, where $\mathcal{U}$ is a finite space. Given the past history $\overline{\mathcal{F}_t^N}$, $X^N(t+1)$ evolves according to the transition probability $\mathrm{P}(X^N(t+1) \in \cdot | \overline{\mathcal{F}_t^N}) = P_{U_t^N}^N(X^N(t), \cdot)$ where $P_u^N(X,Y)$ is a transition kernel on $\mathcal{X}^N$. We assume that for all $u \in \mathcal{U}$, P_u^N is exchangeable (that is $P_u^N(X,Y)$ is invariant by simultaneous permutations of the entries of X and Y). Then, we may define a projected kernel on $\mathcal{P}_N(\mathcal{X})$,

$$\mathcal{K}_u^N(q,p) = \sum_{Y\in\mathcal{X}^N : \frac{1}{N}\sum_{i=1}^N \delta_{Y_i} = p} P_u^N(X,Y),$$

where $p, q \in \mathcal{P}_N(\mathcal{X})$, and X is any vector in $\mathcal{X}^N$ such that $\sum_{i=1}^N \delta_{X_i} = qN$. Then, given $\overline{\mathcal{F}_t^N}$, μ_{t+1}^N evolves according the transition kernel on $\mathcal{P}_N(\mathcal{X})$,

$$\mathrm{P}(\mu_t^N \in \cdot \,|\, \overline{\mathcal{F}_t^N}) = \mathcal{K}_{U_t^N}^N(\mu_t^N, \cdot). \tag{1}$$

Similarly, we define for all $q \in \mathcal{P}_N(\mathcal{X})$ and x such that $q(\{x\}) > 0$, the projected kernel on $\mathcal{X}$

$$K_{u,q}^N(x,y) = \sum_{Y\in\mathcal{X}^N : Y_1 = y} P_u^N(X,Y),$$

where X is any vector satisfying $X_1 = x$ and $\sum_{i=1}^N \delta_{X_i} = qN$. Given the past history, $\overline{\mathcal{F}_t^N}$, at time $t+1$, particle i evolves according to the probability,

$$K^N_{U_t^N, \mu_t^N}(X_i^N(t), \cdot). \tag{2}$$

For a function f from $\mathcal{X}$ to $\mathbb{R}$ and a measure $q \in \mathcal{P}(\mathcal{X})$, we define:

$$\langle f, q\rangle = \sum_{x\in\mathcal{X}} f(x)q(x) = \int_{\mathcal{X}} f(x)q(dx).$$

Let r be a function from $\mathcal{X}\times\mathcal{U}$ to $\mathbb{R}$ which represents a reward, $0 < \beta < 1$ a discount parameter, and $T \geq 1$. If $u \in \mathcal{U}$ and $q \in \mathcal{P}(\mathcal{X})$, we define $r(q,u) = \langle r(\cdot,u), q\rangle = \int_{\mathcal{X}} r(x,u)q(dx)$. For any measure $\nu \in \mathcal{P}(\mathcal{X}^{\mathbb{N}})$ with marginals $(\nu_0, \nu_1, \cdots)$ and a sequence of control $V = (V_0, V_1, \cdots) \in \mathcal{U}^{\mathbb{N}}$, we define the discounted reward

$$J_\beta(\nu, V) = (1-\beta)\sum_{t=0}^{\infty} \beta^t r(\nu_t, V_t), \tag{3}$$

the finite average reward,

$$J_T(\nu, V) = \frac{1}{T}\sum_{t=0}^{T-1} r(\nu_t, V_t), \tag{4}$$

(which depends on the sequence (ν, V) only up to $T-1$), and the ergodic average reward

$$J_{av}(\nu, V) = \liminf_{T\to\infty} J_T(\nu, V). \tag{5}$$

A control process $U^N = (U_t^N), t \in \mathbb{N}$, is admissible if U_t^N is $\mathcal{F}_t^N$-measurable. A *strategy* $\pi = (\pi_t)_{t\in\mathbb{N}}$ for the N particles system is a sequence of functions π_t from $\mathcal{P}_N(\mathcal{X})^{t+1}$ to $\mathcal{U}$. By definition, setting $U_t^N = \pi_t(\mu_0^N, Z_0^N, \cdots, \mu_t^N, Z_t^N)$, there is an equivalence between admissible controls and strategies. A strategy $\pi = (\pi_t)_{t\in\mathbb{N}}$ is Markov if for all $t \in \mathbb{N}$, $\pi_t(q_0, z_0, \cdots, q_t, z_t)$ depends only on $(q_t, z_t) \in \mathcal{P}(\mathcal{X})\times[0,1)$. For a strategy π for the N particles system, we denote by (μ_π^N, U_π^N) the empirical measure and the control process associated to the strategy π and

$$J_\beta^{N,\pi}(\mu_0^N) = \mathrm{E}[J_\beta(\mu_\pi^N, U_\pi^N)],$$

and respectively for $J_T^{N,\pi}(\mu_0^N)$ and $J_{av}^{N,\pi}(\mu_0^N)$. The expectation E is with respect to the process Z^N and the randomness coming from the transitions in (1), note that the initial measure μ_0^N is not random here. From (1), it is well known, refer to Dynkin and Yushkevitch [5], that for any strategy π, there exists a Markov strategy σ such that $J_\beta^{N,\pi}(\mu_0^N) = J_\beta^{N,\sigma}(\mu_0^N)$ (and respectively for $J_T^{N,\pi}(\mu_0^N)$). In this paper, for each N we take interest to the supremum over all strategies of $J_\beta^{N,\pi}(\mu_0^N)$, $J_T^{N,\pi}(\mu_0^N)$ or $J_{av}^{N,\pi}(\mu_0^N)$. The aim being to prove a convergence of the N particles optimal control problem to an infinite particle optimal control problem. In optimal control theory the case of ergodic average reward is known to be much harder than the finite average and discounted reward cases.

Extra Notation and assumptions Let $\mathcal{Y}$ be a Polish space, for a random variable $Y \in \mathcal{Y}$, $\mathcal{L}(Y) \in \mathcal{P}(\mathcal{Y})$ will denote the distribution of Y. We denote by $\|\cdot\|$ the total variation norm on the measures on $\mathcal{X}$: $\|\nu\| = 1/2\sum_{x\in\mathcal{X}} |\nu(x)|$. We endow $\mathcal{X}^{\mathbb{N}}$ and $\mathcal{U}^{\mathbb{N}}$ with the topology associated to the metric $\|X - Y\|_\beta = \sum_{t=0}^{\infty} \beta^t \mathbf{1}_{X_t \neq Y_t}$. $\mathcal{X}^{\mathbb{N}}$ and $\mathcal{U}^{\mathbb{N}}$ are then Polish spaces.

A transition kernel on $\mathcal{X}$ is a linear mapping from $\mathcal{P}(\mathcal{X})$ to $\mathcal{P}(\mathcal{X})$. Thus a transition kernel K may be seen as a stochastic $\mathcal{X} \times \mathcal{X}$-matrix.

For a probability measure $\nu \in \mathcal{P}(\mathcal{Y})$ on a Polish space $\mathcal{Y}$, we define its support

$$\text{supp}(\nu) = \{x \in \mathcal{Y} : \text{ for all open sets } A \text{ such that } x \in A, \nu(A) > 0\}.$$

The only property of the support that we will use is that if A is a measurable sets outside the support then $\nu(A) = 0$.

Let Ψ be a finite measure on the product space $\mathcal{Y} \times \mathcal{Z}$, B a measurable set in $\mathcal{Z}$, the measure $\Psi(\cdot, B)$ will denote the measure on $\mathcal{Y}$: $A \mapsto \Psi(A \times B)$.

If $(\mathcal{A}_N), N \in \mathbb{N}$, is an infinite sequence of subsets of a set $\mathcal{Y}$, then we define $\limsup_N \mathcal{A}_N = \cap_{M \geq 1} \cup_{N \geq M} \mathcal{A}_M$.

The following extra assumptions are made:

A1. There exists a family of transitions kernels $\{K_{u,q}\}_{u \in \mathcal{U}, q \in \mathcal{P}(\mathcal{X})}$ on $\mathcal{X}$ such that, for all $u \in \mathcal{U}$, $x \in \mathcal{X}$

$$\lim_{N \to \infty} \sup_{q^N \in \mathcal{P}_N(\mathcal{X}) : q^N(\{x\}) > 0} \| K^N_{u,q^N}(x, \cdot) - K_{u,q^N}(x, \cdot) \| = 0.$$

A2. There exists $C > 0$, such that for all $u \in \mathcal{U}$, $x \in \mathcal{X}$, $p, q \in \mathcal{P}(\mathcal{X})$,

$$\| K_{u,q}(x, \cdot) - K_{u,p}(x, \cdot) \| \leq C \| p - q \|.$$

A3. There exists a sequence δ_N, with $\lim_N \delta_N = 0$, such that for all $u \in \mathcal{U}$, $x_1, x_2, y_1, y_2 \in \mathcal{X}$ and $X \in \mathcal{X}^N$ such that $X_1 = x_1$, $X_2 = x_2$,

$$\Big| K^N_{u, \frac{1}{N} \sum_{i=1}^N \delta_{X_i}}(x_1, y_1) K^N_{u, \frac{1}{N} \sum_{i=1}^N \delta_{X_i}}(x_2, y_2) - \sum_{Y \in \mathcal{X}^N : Y_1 = y_1, Y_2 = y_2} P^N_u(X, Y) \Big| \leq \delta_N.$$

Assumption A2 implies that the family of transition kernels $\{K_{u,q}\}_{u \in \mathcal{U}, q \in \mathcal{P}(\mathcal{X})}$ is measurable: the application $(u, q) \mapsto K_{u,q}$ from $\mathcal{U} \times \mathcal{P}(\mathcal{X})$ to the set of matrices of dimension $\mathcal{X} \times \mathcal{X}$ is Lipschitz for the standard matrix norm $\|A\|_1 = \max_{x \in \mathcal{X}} \sum_{y \in \mathcal{X}} |A(x, y)|$ and the distance on $\mathcal{P}(\mathcal{X}) \times \mathcal{U}$: $d((q, u), (p, v)) = \|q - p\| + \mathbf{1}_{u \neq v}$.

In words, assumption A3 implies that at any time t, the evolution of two particles becomes asymptotically independent given the control U^N_t and the empirical measure μ^N_t.

Let $x \in \mathcal{X}$, note that assumptions A1 and A2 imply that if $q^N \in \mathcal{P}_N(\mathcal{X})$ converges to $q \in \mathcal{P}(\mathcal{X})$ in total variation with $q^N(\{x\}) > 0$ (but possibly $q(\{x\}) = 0$) then $\| K^N_{u,q^N}(x, \cdot) - K_{u,q}(x, \cdot) \| \leq \| K^N_{u,q^N}(x, \cdot) - K_{u,q^N}(x, \cdot) \| + \| K_{u,q^N}(x, \cdot) - K_{u,q}(x, \cdot) \|$ goes to 0 as N goes to infinity.

2.2 Discounted reward

Dynamic programming recursion. Consider an initial condition $\mu^N_0 \in \mathcal{P}_N(\mathcal{X})$. We define the optimal discounted reward with initial condition μ^N_0 as

$$\mathcal{J}^N_\beta(\mu^N_0) = \sup_\pi J^{N,\pi}_\beta(\mu^N_0) = \sup_\pi \mathrm{E} J_\beta(\mu^N_\pi, U^N_\pi), \tag{6}$$

where the supremum is over all N particles strategies π and, as above, (μ^N_π, U^N_π) the empirical measure and the control process associated to the strategy π. For each μ^N_0, the existence of an

optimal Markov strategy π_* with associated empirical measure and control process $(\mu^N_{\pi_*}, U^N_{\pi_*})$ follows from the classical theory of fully observed dynamic programming on a finite state space, refer to Bertsekas [2], chapter 1. Indeed, this problem may be restated as a fully observed Dynamic Programming (DP) recursion (or Bellman Equation). DP theory predicts that Equation (1) implies that $\mathcal{J}^N_\beta$ is uniquely defined by the recursion, for all $q \in \mathcal{P}_N(\mathcal{X})$,

$$\mathcal{J}^N_\beta(q) = \max_{u\in\mathcal{U}} \Big\{ (1-\beta) r(q,u) + \beta \sum_{p\in\mathcal{P}_N(\mathcal{X})} \mathcal{J}^N_\beta(p) \mathcal{K}^N_u(q,p) \Big\},$$

and $(U^N_{\pi_*})_t \in \mathcal{U}^N_\beta((\mu^N_{\pi_*})_t)$ where, for $q \in \mathcal{P}_N(\mathcal{X})$,

$$\mathcal{U}^N_\beta(q) = \Big\{ u \in \mathcal{U} : \mathcal{J}^N_\beta(q) = (1-\beta) r(q,u) + \beta \sum_{p\in\mathcal{P}_N(\mathcal{X})} \mathcal{J}^N_\beta(p) \mathcal{K}^N_u(q,p) \Big\}$$

is the set of controls reaching the maximum in the DP recursion (see Bertsekas [2], chapter 1). Hence if $\mathcal{U}^N_\beta(\mu^N_t)$ is not reduced to a singleton, the solution to the dynamic programming optimization (6) is not unique. The process Z^N may be used to draw randomly a control in $\mathcal{U}^N_\beta(\mu^N_t)$. For example, we index the controls by integers: $\mathcal{U} = \{u_1, \cdots, u_{|\mathcal{U}|}\}$ and if $\mathcal{U}^N_\beta(\mu^N_t) = \{u_{i_1}, \cdots, u_{i_n}\}$, with $i_1 < \cdots < i_n$, the strategy π_* picks u_{i_ℓ}, $1 \le \ell \le n$, if $\ell - 1 \le n Z^N_t < \ell$. This strategy π_* is discount optimal and Markovian.

Mean-Field approximation. We define the mapping F from $\mathcal{P}(\mathcal{X}) \times \mathcal{U}$ to $\mathcal{P}(\mathcal{X})$:

$$F: \quad (q,u) \mapsto qK_{u,q}. \tag{7}$$

Now let $(\mu, U) \in \mathcal{P}(\mathcal{X}^{\mathbb{N}}) \times \mathcal{U}^{\mathbb{N}}$ be a pair of empirical measure and control sequence and let $q_0 \in \mathcal{P}(\mathcal{X})$. We say that (μ, U) is a solution of (7) with initial condition q_0 if $\mu_0 = q_0$, for all $t \in \mathbb{N}$,

$$\mu_{t+1} = F(\mu_t, U_t)$$

For each arbitrary sequence $U = (U_t)_{t\in\mathbb{N}}$, there exists a pair (μ, U) solution of (7) with initial condition q_0. This pair is not unique but if (μ, U) and (ν, U) are solutions of (7) with initial condition q_0 then for all $t \in \mathbb{N}$, $\mu_t = \nu_t$. We may thus define without ambiguity the associated discounted reward $J^U_\beta(q_0) = J_\beta(\mu, U)$, and

$$\mathcal{J}_\beta(q_0) = \sup_{U\in\mathcal{U}^{\mathbb{N}}} J^U_\beta(q_0). \tag{8}$$

The problem may be restated as a fully observed DP recursion: Equations (7), (8) implies that $\mathcal{J}_\beta$ is uniquely defined by the recursion, for all $q \in \mathcal{P}(\mathcal{X})$,

$$\mathcal{J}_\beta(q) = \max_{u\in\mathcal{U}} \Big\{ (1-\beta) r(q,u) + \beta \mathcal{J}_\beta(qK_{u,q}) \Big\}.$$

We define

$$\mathcal{U}_\beta(q) = \Big\{ u \in \mathcal{U} : \mathcal{J}_\beta(q) = (1-\beta) r(q,u) + \beta \mathcal{J}_\beta(qK_{u,q}) \Big\}.$$

Then if (μ_*, U_*) solves (7) and for all $t \ge 0$, $(U_*)_t \in \mathcal{U}_\beta((\mu_*)_t)$ then $\mathcal{J}_\beta(q_0) = J_\beta(\mu_*, U_*)$. The next theorem implies the convergence of the N particles DP problem to the DP problem (7)-(8). We consider a given sequence $(\mu^N_0)_{N\in\mathbb{N}}$ which converges as N goes to infinity.

Theorem 1 *Assume that assumptions A1-A3 hold, and the empirical measure μ_0^N converges to q_0. Then,*

$$\lim_{N\to\infty} \mathcal{J}_\beta^N(\mu_0^N) = \mathcal{J}_\beta(q_0),$$

and

$$\limsup_{N\to\infty} \mathcal{U}_\beta^N(\mu_0^N) \subseteq \mathcal{U}_\beta(q_0).$$

This result is interesting because it states the convergence of DP problem without any assumption on the reward function r or on the sets $\mathcal{U}_\beta$ of optimal strategies.

We may also define the sets, for all $u \in \mathcal{U}$,

$$\mathcal{P}_\beta^N(u) = \{q \in \mathcal{P}_N(\mathcal{X}) : u \in \mathcal{U}_\beta^N(q)\} \quad \text{and} \quad \mathcal{P}_\beta(u) = \{q \in \mathcal{P}(\mathcal{X}) : u \in \mathcal{U}_\beta(q)\}.$$

Theorem 1 implies that for all $u \in \mathcal{U}$,

$$\limsup_{N\to\infty} \mathcal{P}_\beta^N(u) \subseteq \mathcal{P}_\beta(u).$$

Indeed, let $u \in \mathcal{U}$ and $q \in \limsup_N \mathcal{P}_\beta^N(u)$, then for an increasing subsequence $(N_k), k \in \mathbb{N}$, $q \in \mathcal{P}_\beta^{N_k}(u)$ or equivalently, $u \in \mathcal{U}_\beta^{N_k}(q)$. Thus, from Theorem 1, $u \in \mathcal{U}_\beta(q)$ and it follows that $q \in \mathcal{P}_\beta(u)$.

2.3 Finite average reward

Dynamic programming recursion. Consider an initial condition $\mu_0^N \in \mathcal{P}_N(\mathcal{X})$, the finite average optimal reward is defined as the supremum over all N-particles strategies of the average reward:

$$\mathcal{J}_T^N(\mu_0^N) = \sup_\pi J_T^{N,\pi}(\mu_0^N). \tag{9}$$

Again, this problem may be restated as a fully observed Dynamic Programming (DP) recursion (see Bertsekas [3], chapter 2). DP theory predicts that Equation (1) implies that $\mathcal{J}_T^N$ is uniquely defined by the recursion, for all $q \in \mathcal{P}_N(\mathcal{X})$ and $T \in \mathbb{N}$:

$$\mathcal{J}_T^N(q) = \frac{1}{T} \max_{u\in\mathcal{U}} \Big\{ r(q,u) + (T-1) \sum_{p\in\mathcal{P}_N(\mathcal{X})} \mathcal{J}_{T-1}^N(p)\mathcal{K}_u^N(q,p) \Big\},$$

where by convention, $\mathcal{J}_0^N(q) = 0$. For all $T \in \mathbb{N}$, we define:

$$\mathcal{U}_T^N(q) = \Big\{ u \in \mathcal{U} : T\mathcal{J}_T^N(q) = r(q,u) + (T-1) \sum_{p\in\mathcal{P}_N(\mathcal{X})} \mathcal{J}_{T-1}^N(p)\mathcal{K}_u^N(q,p) \Big\}$$

Any strategy π_* with associated process $(\mu_{\pi_*}^N, U_{\pi_*}^N)$ such that for all $0 \le t < T$, $(U_{\pi_*}^N)_t$ is in the set $\mathcal{U}_{T-t}^N((\mu_{\pi_*}^N)_t)$ is optimal.

Mean-Field approximation. Let $(\mu, U) \in \mathcal{P}(\mathcal{X}^{\mathbb{N}}) \times \mathcal{U}^{\mathbb{N}}$ be a solution of (7) with initial condition q_0. The associated finite average reward is $J_T(\mu, U) = J_T^U(q_0)$. The finite average optimal reward is defined as

$$\mathcal{J}_T(q_0) = \sup_{U\in\mathcal{U}^{\mathbb{N}}} J_T^U(q_0) \tag{10}$$

$\mathcal{J}_T$ is uniquely defined by the recursion, $\mathcal{J}_0 \equiv 0$ and for all $q \in \mathcal{P}(\mathcal{X})$ and $T \in \mathbb{N}$:

$$\mathcal{J}_T(q) = \frac{1}{T} \max_{u \in \mathcal{U}} \Big\{ r(q,u) + (T-1)\mathcal{J}_{T-1}(qK_{u,q}) \Big\},$$

We define,

$$\mathcal{U}_T(q) = \Big\{ u \in \mathcal{U} : T\mathcal{J}_T(q) = r(q,u) + (T-1)\mathcal{J}_{T-1}(qK_{u,q}) \Big\}$$

Then any pair $(\mu, U) \in \mathcal{P}(\mathcal{X}^{\mathbb{N}}) \times \mathcal{U}^{\mathbb{N}}$ solving (7) and satisfying for all $0 \leq t < T-1$, $U_t \in \mathcal{U}_{T-t}(\mu_t)$ is optimal: $\mathcal{J}_T(\mu_0) = J_T(\mu, U)$. Note again that even if $\mathcal{J}$ is deterministic, U may not be uniquely defined. The next result is the analog to Theorem 1.

Theorem 2 *Assume that assumptions A1-A3 hold, and that the empirical measure μ_0^N converges to q_0. Then,*

$$\lim_{N \to \infty} \mathcal{J}_T^N(\mu_0^N) = \mathcal{J}_T(q_0),$$

and for all $1 \leq t \leq T$,

$$\limsup_{N \to \infty} \mathcal{U}_t^N(\mu_0^N) \subseteq \mathcal{U}_t(q_0).$$

As in the discounted reward case, we may define the sets, for all $u \in \mathcal{U}$, $1 \leq t \leq T$,

$$\mathcal{P}_t^N(u) = \{q \in \mathcal{P}_N(\mathcal{X}) : u \in \mathcal{U}_t^N(q)\} \quad \text{and} \quad \mathcal{P}_t(u) = \{q \in \mathcal{P}(\mathcal{X}) : u \in \mathcal{U}_t(q)\}.$$

Theorem 1 implies that for all $u \in \mathcal{U}$,

$$\limsup_{N \to \infty} \mathcal{P}_t^N(u) \subseteq \mathcal{P}_t(u).$$

2.4 Ergodic average reward

Ergodic occupation measure. We now optimize over all admissible strategies the average reward:

$$\mathcal{J}_{av}^N(\mu_0^N) = \sup_{\pi} J_{av}^{N,\pi}(\mu_0^N) = \sup_{\pi} \mathrm{E} \liminf_{T \to \infty} J_T(\mu_\pi^N, U_\pi^N). \tag{11}$$

We add an extra irreducibility assumption,

A4. For all $N \in \mathbb{N}$, $p, q \in \mathcal{P}_N(\mathcal{X})$, $\mathcal{J}_{av}^N(p) = \mathcal{J}_{av}^N(q)$.

Note that assumption A4 holds if for all $N \in \mathbb{N}$, $p, q \in \mathcal{P}_N(\mathcal{X})$ there exist $k \in \mathbb{N}$ and $((p_1, u_1), \cdots, (p_k, u_k)) \in (\mathcal{P}_N(\mathcal{X}) \times \mathcal{U})^k$ such that $p_1 = p$, $p_k = q$ and $1 \leq i < k$, $\mathcal{K}_{u_i}^N(p_i, p_{i+1}) > 0$.

With this assumption, the *convex analytic approach* gives a convenient way to describe a control policy. Let $(\mu^N, U^N) \in \mathcal{P}(\mathcal{X}^{\mathbb{N}}) \times \mathcal{U}^{\mathbb{N}}$ be an admissible controlled process, following Borkar (Chap. 11 in [7]), for each $T \in \mathbb{N}$, we define the *ergodic occupation measure* on $\mathcal{P}_N(\mathcal{X}) \times \mathcal{U}$,

$$\Psi_T^N := \Psi_T^N(\mu^N, U^N) = \frac{1}{T} \sum_{t=0}^{T-1} \delta_{(\mu_t^N, U_t^N)}.$$

Note that, by definition, $J_{av}(\mu^N, U^N) = \liminf_T \langle \Psi_T^N, r \rangle$. Now, since $\mathcal{P}_N(\mathcal{X}) \times \mathcal{U}$ is a finite space, the sequence $(\Psi_T^N)_{T \in \mathbb{N}}$ has limit points. The following sample path result holds (for a proof, see Lemma 5.1 in Arapostathis et al. [1]),

Lemma 3 (Borkar) *Almost-surely, any limit point ψ^N of (Ψ_T^N) in $\mathcal{P}(\mathcal{P}_N(\mathcal{X})\times\mathcal{U})$ satisfies,*

$$\forall p\in\mathcal{P}_N(\mathcal{X}),\qquad \sum_{q\in\mathcal{P}_N(\mathcal{X})}\sum_{u\in\mathcal{U}}\psi^N(q,u)\mathcal{K}_u^N(q,p)=\psi^N(p,\mathcal{U}), \tag{12}$$

Note that the limit point ψ^N is a sample path limit point. Now, reciprocally let $\psi^N\in\mathcal{P}(\mathcal{P}_N(\mathcal{X})\times\mathcal{U})$ satisfying (12). Using the extra randomness Z^N, we define a Markov strategy $\overline{\pi}$ and its associated controlled process $(\overline{\mu}^N,\overline{U}^N)$ such that the law of $\overline{U}_t^N$ given $\overline{\mu}_t^N$ is $\psi^N(\overline{\mu}_t^N,\cdot)/\psi^N(\overline{\mu}_t^N,\mathcal{U})$. Then $(\overline{\mu}_t^N,\overline{U}_t^N)_{t\in\mathbb{N}}$ is a Markov chain on $\mathcal{P}(\mathcal{X})\times\mathcal{U}$ with transition kernel

$$\mathrm{P}((\overline{\mu}_1^N,\overline{U}_1^N)=(q,u)|(\overline{\mu}_0^N,\overline{U}_0^N)=(p,v))=\mathcal{K}_v^N(p,q)\frac{\psi^N(q,u)}{\psi^N(q,\mathcal{U})}.$$

If ψ^N satisfies (12), we check easily that ψ^N is a stationary distribution of this Markov chain. Therefore if $\mathcal{L}(\overline{\mu}_0^N,\overline{U}_0^N)=\psi^N$, then for all $T\geq 1$,

$$\mathrm{E}J_T(\overline{\mu}^N,\overline{U}^N)=\langle\psi^N,r\rangle.$$

Now we define $\mathcal{S}^N=\{\psi^N\in\mathcal{P}(\mathcal{P}_N(\mathcal{X})\times\mathcal{U}):(12)\text{ holds true}\}$. From what precedes we have $\mathcal{J}_{av}^N\geq\inf_{\psi^N\in\mathcal{S}^N}\langle\psi^N,r\rangle$. The next lemma implies that an equality holds.

Lemma 4 (Borkar) *Let π_* be an ergodic average optimal strategy with associated process $(\mu_{\pi_*}^N,U_{\pi_*}^N)$ which maximizes (11) with initial condition μ_0^N. Then, almost surely, the following holds*

- $\lim_{T\to\infty}J_T(\mu_{\pi_*}^N,U_{\pi_*}^N)=\mathcal{J}_{av}^N$,
- *for any limit point ψ^N of $\Psi_T^N(\mu_{\pi_*}^N,U_{\pi_*}^N)$, $\mathcal{J}_{av}^N=\langle\psi^N,r\rangle$.*

Proof. Let A be the event of probability one such that the conclusion of Lemma 3 holds for the process $(\mu_{\pi_*}^N,U_{\pi_*}^N)$. We define the event $B=A\cap\{\liminf_T J_T(\mu_{\pi_*}^N,U_{\pi_*}^N)\geq\mathcal{J}_{av}^N\}$. Since $\mathcal{J}_{av}^N=\mathrm{E}\liminf_T J_T(\mu_{\pi_*}^N,U_{\pi_*}^N)$, we have $\mathrm{P}(B)>0$ and B is not empty. On the event B, we may extract a subsequence T_k such that $\lim_{T_k}J_{T_k}(\mu_{\pi_*}^N,U_{\pi_*}^N)=\liminf_T J_T(\mu_{\pi_*}^N,U_{\pi_*}^N)\geq\mathcal{J}_{av}^N$. Up to extracting another sequence from (T_k) we may also assume that $\Psi_{T_k}^N(\mu_{\pi_*}^N,U_{\pi_*}^N)$ converges to ψ'^N and then

$$\langle\psi'^N,r\rangle\geq\mathcal{J}_{av}^N.$$

Now, from what precedes, there exists a Markov strategy $\overline{\pi}'$ with associated controlled process $(\overline{\mu}'^N,\overline{U}'^N)$ such that $\mathrm{E}J_{av}(\overline{\mu}'^N,\overline{U}'^N)=\langle\psi'^N,r\rangle$. However, by assumption A4., $\mathcal{J}_{av}^N\geq\mathrm{E}J_{av}(\overline{\mu}'^N,\overline{U}'^N)=\langle\psi'^N,r\rangle$, and we deduce that $\mathcal{J}_{av}^N=\langle\psi'^N,r\rangle$ and thus $\mathrm{P}(B)=1$. We have proved so far that

$$\text{a.s -}\liminf_T J_T(\mu_{\pi_*}^N,U_{\pi_*}^N)=\mathcal{J}_{av}^N=\langle\psi'^N,r\rangle.$$

Now, still on the event B, let ψ^N be any limit point of $\Psi_T^N(\mu_{\pi_*}^N,U_{\pi_*}^N)$, thanks to our choice of ψ'^N, $\langle\psi^N,r\rangle\geq\langle\psi'^N,r\rangle=\mathcal{J}_{av}$. However, again, there exists a Markov strategy $\overline{\pi}$ with associated controlled process $(\overline{\mu}^N,\overline{U}^N)$ such that $\mathrm{E}J_{av}(\overline{\mu}^N,\overline{U}'^N)=\langle\psi^N,r\rangle$. Then, by assumpton A4. $\mathrm{E}J_{av}(\overline{\mu}^N,\overline{U}^N)\leq\mathcal{J}_{av}^N$, and we deduce that $\langle\psi^N,r\rangle=\langle\psi'^N,r\rangle$. □

We define the set of optimal ergodic occupation measures:

$$\mathcal{S}_{av}^N=\{\Psi^N\in\mathcal{P}(\mathcal{P}_N(\mathcal{X})\times\mathcal{U}):(12)\text{ holds true and }\langle\Psi^N,r\rangle=\mathcal{J}_{av}^N\}.$$

$\mathcal{S}_{av}^N$ is a non-empty closed convex set. It is also possible to describe $\mathcal{J}_{av}^N$ as a fixed point of DP recursion, see Arapostathis et al. [1]. In this paper, we will however not use this representation of $\mathcal{J}_{av}^N$. The Borkar's representation via the ergodic occupation measure has appeared to be more convenient to state limit theorems.

Mean-Field approximation. Let $q_0 \in \mathcal{P}(\mathcal{X})$. For each sequence $U = (U_t)_{t\in\mathbb{N}}$, there exists a pair (μ, U) solution of (7) with initial condition q_0. We may consider the associated ergodic average reward $J_{av}^U(q_0) = J_{av}(\mu, U)$, and we define:

$$\mathcal{J}_{av}(q_0) = \sup_{U\in\mathcal{U}^{\mathbb{N}}} J_{av}^U(q_0). \tag{13}$$

We formulate for this deterministic average cost problem an irreducibility assumption, similar to assumption A4:

A5. For all $p, q \in \mathcal{P}(\mathcal{X})$, $\mathcal{J}_{av}(p) = \mathcal{J}_{av}(q)$.

Again, for a pair (μ, U), solution of (7), we define the ergodic accumulation measure on $\mathcal{P}(\mathcal{X})\times\mathcal{U}$ by

$$\Psi_T := \Psi_T(\mu, U) = \frac{1}{T}\sum_{t=0}^{T-1} \delta_{(\mu_t, U_t)},$$

$\mathcal{P}(\mathcal{P}(\mathcal{X}) \times \mathcal{U})$ is compact and any limit point Ψ of $(\Psi_T)_{T\in\mathbb{N}}$ satisfies for all measurable sets $A \subset \mathcal{P}(\mathcal{X})$, such that $\Psi(\partial A,\mathcal{U}) = 0$ and for all $u \in \mathcal{U}$, $\Psi(\partial\{q : qK_{u,q} \in A\}, u) = 0$,

$$\sum_{u\in\mathcal{U}} \Psi(\{q : qK_{u,q} \in A\}, u) = \Psi(A,\mathcal{U}). \tag{14}$$

Again, Ψ may be interpreted as a stationary distribution of a Markov chain. Indeed, let α_u be Radon-Nikodym derivative of $\Psi(\cdot, u)$ with respect to $\Psi(\cdot,\mathcal{U})$. Ψ is a stationary distribution of the Markov chain $(\overline{\mu}_t, \overline{U}_t)_{t\in\mathbb{N}}$ with transition kernel

$$\mathrm{P}((\overline{\mu}_1, \overline{U}_1) = (q, u)|(\overline{\mu}_0, \overline{U}_0) = (p, v)) = \alpha_u(q)\mathbf{1}(pK_{v,p} = q). \tag{15}$$

Note that $J_{av}(\mu, U) = \liminf_T\langle\Psi_T, r\rangle$. Note that this Markov transition kernel is defined only on the support of $\Psi(\cdot,\mathcal{U})$. In order to define a transition kernel of $\mathcal{P}(\mathcal{X}) \times \mathcal{U}$, we need to extend on $\mathcal{P}(\mathcal{X})$, for all $u \in \mathcal{U}$, the mapping: $q \mapsto \alpha_u(q)$, in a measurable way. Even if this extension is obviously not unique, Ψ is always an invariant measure of the Markov chain.

Now, note that the supremum in (13) is reached for some $U_* \in \mathcal{U}^{\mathbb{N}}$ (depending on q_0) and a pair (μ_*, U_*) solution of (7). Then, if Ψ is a limit point of the occupation measure $\Psi_T(\mu_*, U_*)$, by assumption A5, we also have

$$\mathcal{J}_{av} = \langle\Psi, r\rangle.$$

We may define the set of optimal ergodic occupation measures:

$$\mathcal{S}_{av} = \{\Psi \in \mathcal{P}(\mathcal{P}(\mathcal{X}) \times \mathcal{U}) : (14) \text{ holds true and } \langle\Psi, r\rangle = \mathcal{J}_{av}\}.$$

A stationary solution (μ, U) of (7) is a solution of (7) such that for all $t \geq 0$, $\mathcal{L}(\mu_0, U_0) = \mathcal{L}(\mu_t, U_t)$ and (14) holds true for $\mathcal{L}(\mu_0, U_0)$. A stationary policy $(\alpha_u)_{u\in\mathcal{U}}$ is a set of measurable mappings from $\mathcal{P}(\mathcal{X})$ to $[0, 1]$ such that for all $q \in \mathcal{P}(\mathcal{X})$, $\sum_{u\in\mathcal{U}} \alpha_u(q) = 1$.

A stationary policy $(\alpha_u)_{u\in\mathcal{U}}$ is continuous if for all $u \in \mathcal{U}$, the mapping $q \mapsto \alpha_u(q)$ is continuous. We define $\mathcal{C}_r$ as the set of continuous stationary policies such that if Ψ_1, $\Psi_2 \in \mathcal{P}(\mathcal{P}(\mathcal{X}) \times \mathcal{U})$ are invariant probability measures of the Markov transition kernel (15) with $(\alpha_u)_{u\in\mathcal{U}} \in \mathcal{C}_r$, then $\langle\Psi_1, r\rangle = \langle\Psi_2, r\rangle$. Finally we add a key assumption on $\mathcal{S}_{av}$.

A6. There exists $\Psi \in \mathcal{S}_{av}$ such that Ψ is an invariant measure of a Markov transition kernel (15) with $(\alpha_u)_{u\in\mathcal{U}} \in \mathcal{C}_r$,

Theorem 5 *Assume that assumptions A1-A6 hold then*

$$\lim_{N\to\infty} \mathcal{J}_{av}^N = \mathcal{J}_{av},$$

and

$$\overline{\limsup_{N\to\infty} \mathcal{S}_{av}^N} \subseteq \mathcal{S}_{av},$$

(that is if $\Psi^N \in \mathcal{S}_{av}^N$ for all N then any limit point of Ψ^N is in $\mathcal{S}_{av}$).

On Assumptions A6. Assumption A6 holds if there exists (q,u) such that $\delta_{(q,u)} \in \mathcal{S}_{av}$, and q is the globally stable fixed point of the mapping from $\mathcal{P}(\mathcal{X})$ to $\mathcal{P}(\mathcal{X})$, $p \mapsto pK_{u,p}$. Similarly, assume that there exists Ψ in $\mathcal{S}_{av}$ such that $\Psi = 1/M \sum_{i=1}^M \delta_{(q^i,u^i)}$ with $q^i K_{u^i,q^i} = q^{i+1}$ (and $q^{M+1} = q^1$). Since $\mathcal{P}(\mathcal{X})$ is separable, there exists a stationary continuous policy $(\alpha_u)_{u\in\mathcal{U}}$ such that $\alpha_u(q^i) = \mathbf{1}(u^i = u)$. If Ψ is the unique invariant measure of the Markov transition kernel (15) for this choice of $(\alpha_u)_{u\in\mathcal{U}}$ then assumption A6 is satisfied.

3 Auxiliary results and model extensions

3.1 Phase transition on the average reward

In this paragraph, we discuss what happens when the conclusions of Theorem 5 fail to hold. We first start with a general lemma

Lemma 6 *For all $N \in \mathbb{N}$ and $q \in \mathcal{P}_N(\mathcal{X})$,*

$$\mathcal{J}_{av}^N(q) = \lim_{T\to\infty} \mathcal{J}_T^N(q) = \lim_{\beta\uparrow 1} \mathcal{J}_\beta^N(q).$$

For all $q \in \mathcal{P}(\mathcal{X})$,

$$\mathcal{J}_{av}(q) \le \liminf_{T\to\infty} \mathcal{J}_T(q).$$

Proof. Fix $q \in \mathcal{P}^N(\mathcal{X})$ and note that for each N, the space $\mathcal{P}_N(\mathcal{X})$ is finite. A well-known result of Blackwell [4] implies that $\lim_{\beta\uparrow 1} \mathcal{J}_\beta^N(q) = \mathcal{J}_{av}^N(q)$. More precisely (see the proof of Theorem 4.3 in [1]), there exists a mapping f from $\mathcal{P}_N(\mathcal{X})$ to $\mathcal{U}$ and $0 < \beta_0 < 1$, such that for all $q \in \mathcal{P}_N(\mathcal{X})$ and $\beta \in (\beta_0, 1)$, $\mathcal{J}_\beta^N(q) = \mathrm{E}J_\beta(\mu^N, U_f^N)$ and $\mathcal{J}_{av}^N(q) = \mathrm{E}J_{av}(\mu^N, U_f^N)$, where U_f^N is the adapted sequence obtained by setting $(U_f^N)_t = f(\mu_t^N)$. Then a Tauberian Theorem of Hardy and Littlewood (see Theorem 2.3 in [9]) implies that $\lim_{T\to\infty} \mathrm{E}J_T(\mu^N, U_f^N) = \mathrm{E}J_{av}(\mu^N, U_f^N) = \mathcal{J}_{av}^N(q)$. Since by definition, $\mathrm{E}J_T(\mu^N, U_f^N) \le \mathcal{J}_T^N(q)$, we obtain: $\mathcal{J}_{av}^N(q) \le \liminf_{T\to\infty} \mathcal{J}_T^N(q)$.

Reciprocally, let $(T_k), k \in \mathbb{N}$, be an increasing sequence such that $\lim_k \mathcal{J}_{T_k}^N(q) = \limsup_T \mathcal{J}_T^N(q)$. The family of sets $\{\mathcal{U}_T^N(q)\}_{q\in\mathcal{P}_N(\mathcal{X})}$ is included in a finite set, hence there exists a subsequence T_{k_n} and a mapping g from $\mathcal{P}_N(\mathcal{X})$ to $\mathcal{U}$ such that for all $n \in \mathbb{N}$ and $q \in \mathcal{P}_N(\mathcal{X})$, $g(q) \in \mathcal{U}_{T_{k_n}}^N(q)$. We have $\mathcal{J}_{T_{k_n}}(q) = \mathrm{E}J_{T_{k_n}}(\mu^N, U_g^N)$ where U_g^N is the adapted sequence obtained by setting $(U_g^N)_t = g(\mu_t^N)$. Now, (μ^N, U_g^N) is a Markov chain on a finite state space with initial condition $(q, g(q))$, thus $\lim_{T\to\infty} \mathrm{E}J_T(\mu^N, U_g^N) = \mathrm{E}J_{av}(\mu^N, U_g^N)$. Since by definition $\mathrm{E}J_{av}(\mu^N, U_g^N) \le \mathcal{J}_{av}^N$, we deduce that $\limsup_T \mathcal{J}^N(q) = \lim_n \mathrm{E}J_{T_{k_n}}(\mu^N, U_g^N) \le \mathcal{J}_{av}^N$.

It remains to consider the recursion (7), let $U_*(q)$ be an optimal sequence for the average ergodic reward with initial condition q: $\mathcal{J}_{av}(q) = J_{av}(\mu, U_*(q))$. Then by definition for all $T \geq 1$, $J_T(\mu, U^*(q)) \leq \mathcal{J}_T(q)$, hence letting T tend to infinity, we get: $J_{av}(q) \leq \liminf_T \mathcal{J}_T(q)$. □

We now assume that assumptions A1-A4 hold. Theorem 2 states that for all $T \in \mathbb{N}$ and $q \in \cup_{N\in\mathbb{N}} \mathcal{P}_N(\mathcal{X})$:

$$\lim_{N\to\infty} \mathcal{J}_T^N(q) = \mathcal{J}_T(q).$$

(and similarly with $\mathcal{J}_\beta^N$ using Theorem 1). Assume that $\mathcal{J}_{av}(q) = \lim_{T\to\infty} \mathcal{J}_T(q)$, then we obtain,

$$\lim_{T\to\infty} \lim_{N\to\infty} \mathcal{J}_T^N(q) = \mathcal{J}_{av}(q).$$

Theorem 5 gives sufficient conditions under which the inversion of limits in N and T holds:

$$\lim_{T\to\infty} \lim_{N\to\infty} \mathcal{J}_T^N(q) = \mathcal{J}_{av}(q) \stackrel{?}{=} \lim_{N\to\infty} \mathcal{J}_{av}^N = \lim_{N\to\infty} \lim_{T\to\infty} \mathcal{J}_T^N(q).$$

As we will see through a well-known example, this inversion fails in general. Note in particular that a necessary condition is that $\mathcal{J}_{av}(q)$ does not depend on q (i.e. assumption A4). When this inversion does not hold, a phase transition occurs: the limit behavior of the average reward depends on the initial condition. Without the somewhat restrictive assumptions A5 and A6 of Theorem 5, we have the following result:

Corollary 7 *If assumptions A1-A4 holds,*

$$\limsup_{N\to\infty} \mathcal{J}_{av}^N \leq \sup_{q\in\mathcal{P}(\mathcal{X})} \mathcal{J}_{av}(q).$$

The proof of this corollary is contained in the first half of the forthcoming Lemma 20.

3.2 Partial information

In Theorems 1 and 5, we have assumed that the control U_t^N was $\mathcal{F}_t^N$-measurable where $\mathcal{F}_t^N$ is the σ-field generated by $(\mu_0^N, \cdots, \mu_t^N)$. In optimal control words, we have assumed that the system was fully observed.

Assume now that the control U_t^N is constrained to be measurable with respect to $\mathcal{G}_t^N$, the σ-field generated by $(\mu_0^N, F(\mu_1^N), \cdots, F(\mu_t^N))$ where F is a mapping from $\mathcal{P}(\mathcal{X})$ to an observation space $\mathcal{Z}$. Then for each N the problem of optimal control is partially observed. However, with the assumptions of Theorem 1, as the number of particles N goes large, we have proved the convergence of the fully observed problem to a deterministic optimal control problem. In particular, along the proof of this result, we have shown that a deterministic optimal control (depending on the initial state μ_0^N) achieves asymptotically the optimum. Hence as a Corollary, since μ_0^N is $\mathcal{G}_t$-measurable we have the following:

Corollary 8 *Assume that assumptions A1-A3 hold, and that for all N, and the empirical measure μ_0^N converges to a deterministic limit q_0. Then the statements of Theorems 1, 2 also hold for the $\mathcal{G}_t^N$-partially observed problem.*

Note that the crucial assumption is that the initial state μ_0^N is fully observed and neither F nor the observation space $\mathcal{Z}$ play any role. This assumption is fulfilled in many potential applications. If μ_0^N is not fully observed, then the convergence of the optimal control is more complicated and we will not address this problem here. The same remark also applies to the average reward case where the initial value may not play any role.

3.3 Propagation of chaos

The propagation of chaos is an important concept in mean field theory. This phenomena appears if the trajectories of the particles are asymptotically independent, see Sznitman [10].

Discounted reward. Let $\mu_0^N \in \mathcal{P}_N(\mathcal{X})$ converging to $q_0 \in \mathcal{P}(\mathcal{X})$ as N goes to infinity. For all N, we consider a controlled process (μ^N, U^N) which is optimal for the discounted reward, i.e. $\mathrm{E}J_\beta(\mu^N, U^N) = \mathcal{J}_\beta^N(\mu_0^N)$. Under assumptions A1-A3, Theorem 1 states the convergence of $\mathrm{E}J_\beta(\mu^N, U^N)$ to $\mathcal{J}_\beta(q_0)$. However, this theorem does not state any result on the convergence of the trajectories of the particles. Notice that, since $\mathcal{U}$ is finite, the sequence U^N is tight in $\mathcal{U}^{\mathbb{N}}$, the next result states the convergence of the trajectories of the particles along any converging subsequence of U^N.

Corollary 9 *Let (μ^N, U^N) be as above. Assume that assumptions A1-A3, that for all N, $(X_1^N(0), \cdots, X_N^N(0))$ is exchangeable, and that the empirical measure μ_0^N converges to a deterministic limit q_0. Finally assume that (U^N) converges weakly to $U \in \mathcal{U}^{\mathbb{N}}$. Then there exists $\mu \in \mathcal{P}(\mathcal{X}^{\mathbb{N}})$ such that (μ, U) solves (7) with initial condition $\mu_0 = q_0$, and for all subsets $I \subset \mathbb{N}$ of finite cardinal $|I|$,*

$$\lim_{N\to\infty} \mathcal{L}\left((X_i^N)_{i\in I}\right) = \mu^{\otimes|I|} \quad \textit{weakly in } \mathcal{P}((\mathcal{X}^{\mathbb{N}})^{|I|}). \tag{16}$$

The measure μ appearing in the statement of Corollary 9 is explicitly defined in the proof: μ is the distribution of a non-homogeneous Markov chain with transition kernel at time t, K_{U_t,μ_t} and $\mu_0 = q_0$. Note that Corollary 9 assumes that the control sequence U^N converges. For example if there exists a sequence such that $q_{i+1} = F(q_i, u_i)$ and $\mathcal{U}_\beta(q_i) = \{u_i\}$ then by Theorem 1, if μ_0^N converges to q_0, (U^N) converges to $(u_i)_{i\in\mathbb{N}}$.

Finite average reward. If the controlled process (μ^N, U^N) is optimal for the discounted reward, i.e. $\mathrm{E}J_T(\mu^N, U^N) = \mathcal{J}_T^N(\mu_0^N)$, the statement of Corollary 9 holds for the finite average reward case without change.

Ergodic average reward. We now consider the extra assumption

A7. There exists $q_* \in \mathcal{P}(\mathcal{X})$ such that for all $\Psi \in \mathcal{S}_{av}$: $\Psi(\cdot, \mathcal{U}) = \delta_{q_*}$.

With this assumption, any optimal stationary solution (μ, U) of (7) satisfies for all $t \in \mathbb{N}$, $\mu_t = q_*$. We have the following propagation of chaos.

Corollary 10 *Assume that assumptions A1-A7 hold, and that for all N, $(X_i^N(t))_{1\le i\le N}$, $t \in \mathbb{N}$, is a stationary exchangeable solution of the ergodic average cost problem. Then for all subsets $I \subset \mathbb{N}$ of finite cardinal $|I|$,*

$$\lim_{N\to\infty} \mathcal{L}\left((X_i^N(0))_{i\in I}\right) = q_*^{\otimes|I|} \quad \textit{weakly in } \mathcal{P}(\mathcal{X}^{|I|}). \tag{17}$$

3.4 More general state and control space

If $\mathcal{U}$ or $\mathcal{X}$ are countable and not necessarily finite then the proofs of Theorem 1 and 5 extend provided that $\mathcal{L}(\mu^N, U^N)$ is tight in $\mathcal{P}(\mathcal{P}(\mathcal{X}^{\mathbb{N}}) \times \mathcal{U}^{\mathbb{N}})$.

3.5 Continuous time

We have considered so far a discrete time $t \in \mathbb{N}$. We could define similarly an interacting particle systems, $(\mu_t^N)_{t\in\mathbb{R}_+}$, evolving in continuous time governed by a control process $U^N = (U_t^N)_{t\in\mathbb{R}_+}$ and a family of Markovian generator $\mathcal{L}_u^N(q,p)$ (in place of the transition kernel $\mathcal{K}_u^N(p,q)$). If the generator associated to a single particle is $L_{u,q}^N(x,y)$ (in place of $K_{u,q}(x,y)$) and if $L_{u,q}^N$ converges to $L_{u,q}$, then the mean-field approximation is

$$\frac{d\mu_t}{dt} = \mu_t L_{U_t,\mu_t} - \mu_t.$$

There are obvious continuous analog of assumptions A1-A3. If one can prove that the process (μ^N, U^N) is tight then the proof of Theorem 1 extends to continuous time (using a non-linear martingale problem formulation, see e.g. Graham [8]). Therefore the main technical issue is proving that the optimal control process $(U_t^N)_{t\in\mathbb{R}_+}$ is tight. We do not have a proof of a claim of this type and postpone this issue to future work.

4 Proofs of main results

4.1 Proof of Theorem 1 and Corollary 9.

We consider a sequence of discount optimal strategies π_*^N with associated process $(\mu_*^N, U_*^N) \in \mathcal{P}(\mathcal{X}^{\mathbb{N}}) \times \mathcal{U}^{\mathbb{N}})$. In order to simplify notation, the optimal control sequence (μ_*^N, U_*^N) is simply denoted by (μ^N, U^N). The following proposition will imply both Theorem 1 and Corollary 9. We consider for all $N \in \mathbb{N}$, a vector of initial condition $(X_1^N(0), \cdots, X_N^N(0))$ which is exchangeable, and that satisfies $\mu_0^N = \frac{1}{N}\sum_{i=1}^N \delta_{X_i^N(0)}$ (considering a uniform random permutation of $\{1, \cdots, N\}$ it is possible).

Proposition 11 *With the assumptions of Theorem 1, if the initial condition $(X_1^N(0), \cdots, X_N^N(0))$ is exchangeable, then any limit point of $\mathcal{L}(\mu^N, U^N)$ in $\mathcal{P}(\mathcal{P}(\mathcal{X}^{\mathbb{N}})\times\mathcal{U}^{\mathbb{N}})$ has its support included in the set of the solutions (μ, U) of (7) with initial condition $\mu_0 = q_0$ such that $J_\beta(\mu, U) = \mathcal{J}_\beta(q_0)$.*

Proof of Proposition 11 *Step 1 : Tightness.* We prove the tightness of $\mathcal{L}(\mu^N, U^N)$ in $\mathcal{P}(\mathcal{P}(\mathcal{X}^{\mathbb{N}}) \times \mathcal{U}^{\mathbb{N}})$. The control sequence $\mathcal{L}(U^N(\cdot))$ is obviously tight in $\mathcal{P}(\mathcal{U}^{\mathbb{N}})$ since $\mathcal{U}$ is finite. Note also that the sequence $\mathcal{L}(\mu^N)$ is also tight in $\mathcal{P}(\mathcal{P}(\mathcal{X}^{\mathbb{N}}))$. Indeed, thanks again to Sznitman [10] Proposition 2.2, we only have to prove that $\mathcal{L}(X_1^N(\cdot))$ is tight in $\mathcal{P}(\mathcal{X}^{\mathbb{N}})$. This follows immediately from the finiteness of $\mathcal{X}$.
Step 2 : Martingale formulation.

Let f be a function from $\mathcal{X}$ to $\mathbb{R}$, $t \geq 1$ and $f^y(x) = f(y) - f(x)$. We have

$$\begin{aligned} f(X_i^N(t)) - f(X_i^N(0)) &= \sum_{s=0}^{t-1} f(X_i^N(s+1)) - f(X_i^N(s)) \\ &= \sum_{y\in\mathcal{X}} \sum_{s=0}^{t-1} f^y(X_i^N(s)) \left(\mathbf{1}_{X_i^N(s+1)=y} - \mathrm{P}(X_i^N(s+1) = y | \overline{\mathcal{F}_s^N}) \right) \\ &\quad + \sum_{y\in\mathcal{X}} \sum_{s=0}^{t-1} f^y(X_i^N(s)) \mathrm{P}(X_i^N(s+1) = y | \overline{\mathcal{F}_s^N}), \end{aligned} \tag{18}$$

Then we define

$$M_i^{f,N}(t) \quad = \quad \sum_{y\in\mathcal{X}}\sum_{s=0}^{t-1} f^y(X_i^N(s))\left(\mathbf{1}_{X_i^N(s+1)=y} - K^N_{U_s^N,\mu_s^N}(X_i^N(s),y)\right) \tag{19}$$

and the operator $\mathcal{G}^N f$ defined by, for $(x,u,q)\in\mathcal{X}\times\mathcal{U}\times\mathcal{P}(\mathcal{X})$,

$$\mathcal{G}^N f(x,u,q) = \sum_{y\in\mathcal{X}} f^y(x) K^N_{u,q}(x,y).$$

So that, we may rewrite Equation (18) as

$$f(X_i^N(t)) - f(X_i^N(0)) \quad = \quad M_i^{f,N}(t) + \sum_{s=0}^{t-1} \mathcal{G}^N f(X_i^N(s), U_s^N, \mu_s^N).$$

Lemma 12 *$M_i^{f,N}(t)$ defined at (19) is a $\overline{\mathcal{F}^N}$-martingale. There exists $C>0$ such that for all $i\neq j$, $\mathrm{E}M_i^{f,N}(t)M_j^{f,N}(t) \leq Ct\|f\|_\infty^2\delta_N$ and $\mathrm{E}M_i^{f,N}(t)M_i^{f,N}(t)\leq Ct\|f\|_\infty^2$.*

Proof. We define:

$$A_i^{N,y}(t) = \{X_i^N(t+1)=y\}$$

Recall that

$$\mathrm{P}(A_i^{N,y}(t)|\overline{\mathcal{F}_t^N}) = K^N_{U_t^N,\mu_t^N}(X_i^N(t),y),$$

and we can rewrite Equation (19) as:

$$M_i^{f,N}(t) = \sum_{s=0}^{t-1}\sum_{y\in\mathcal{X}} f^y(X_i^N(s))\left(\mathbf{1}_{A_i^{N,y}(s)} - \mathrm{E}[\mathbf{1}_{A_i^{N,y}(s)}|\overline{\mathcal{F}_s^N}]\right).$$

Thus, $M_i^{f,N}(t)$ is a square-integrable martingale by the Dynkin's formula. Now assumption A3 implies that for all $i\neq j$, $y,y'\in\mathcal{X}$,

$$\left|\,\mathrm{P}(A_i^{N,y}(t)A_j^{N,y'}(t)|\overline{\mathcal{F}_t^N}) - \mathrm{P}(A_i^{N,y}(t)|\overline{\mathcal{F}_t^N})\mathrm{P}(A_j^{N,y'}(t)|\overline{\mathcal{F}_t^N})\,\right| \leq \delta_N. \tag{20}$$

We need to compute $\mathrm{E}[M_1^{f,N}(t)M_2^{f,N}(t)]$. Since $(M_i^{f,N}(t))_{t\in\mathbb{N}}$ is a martingale this product is equal to:

$$\begin{aligned}\mathrm{E}[M_1^{f,N}(t)M_2^{f,N}(t)] \quad &= \quad \sum_{s=0}^{t-1}\sum_{y,y'\in\mathcal{X}} \mathrm{E} f^y(X_1^N(s))\left(\mathbf{1}_{A_1^{N,y}(s)} - \mathrm{E}[\mathbf{1}_{A_1^{N,y}(s)}|\overline{\mathcal{F}_s^N}]\right)\\ &\qquad \times f^{y'}(X_2^N(s))\left(\mathbf{1}_{A_2^{N,y'}(s)} - \mathrm{E}[\mathbf{1}_{A_2^{N,y'}(s)}|\overline{\mathcal{F}_s^N}]\right).\end{aligned}$$

Now, let

$$\begin{aligned} I_s^N \quad &= \quad \sum_{y,y'\in\mathcal{X}} \mathrm{E}\left[f^y(X_1^N(s))(\mathbf{1}_{A_1^{N,y}(s)} - \mathrm{E}[\mathbf{1}_{A_1^{N,y}(s)}|\overline{\mathcal{F}_s^N}])f^{y'}(X_2^N(s))(\mathbf{1}_{A_2^{N,y'}(s)} - \mathrm{E}[\mathbf{1}_{A_2^{N,y'}(s)}|\overline{\mathcal{F}_s^N}])\right]\\ &= \quad \sum_{y,y'\in\mathcal{X}} \mathrm{E}\Big[f^y(X_1^N(s))f^{y'}(X_2^N(s))\\ &\qquad\qquad \times\left(\mathrm{E}[\mathbf{1}_{A_1^{N,y}(s)}\mathbf{1}_{A_2^{N,y'}(s)}|\overline{\mathcal{F}_s^N}] - \mathrm{E}[\mathbf{1}_{A_1^{N,y}(s)}|\overline{\mathcal{F}_s^N}]\mathrm{E}[\mathbf{1}_{A_2^{N,y'}(s)}|\overline{\mathcal{F}_s^N}]\right)\Big].\end{aligned}$$

Hence, using (20),

$$|I_s^N| \quad \leq \quad 4\|f\|_\infty^2 |\mathcal{X}|^2 \delta_N,$$

and the lemma follows. □

Now let $\mathcal{G}f(x,u,q) = \sum_{y\in\mathcal{X}} f^y(x) K_{u,q}(x,y)$. By assumption A1,

$$|\mathcal{G}^N f(X_i^N(s), U_s^N, \mu_s^N) - \mathcal{G}f(X_i^N(s), U_s^N, \mu_s^N)| \leq 2\|f\|_\infty \sup_{(q^N,u,x)} \|K^N_{u,q^N}(x,\cdot) - K_{u,q^N}(x,\cdot)\| \leq \|f\|_\infty \epsilon_N.$$

for some sequence (ϵ_N) tending to 0 as N goes to infinity and where the supremum is over all triples $(q^N, u, x) \in \mathcal{P}_N(\mathcal{X}) \times \mathcal{U} \times \mathcal{X}$ such that $q^N(\{x\}) > 0$. It follows that

$$M_i^{f,N}(t) = f(X_i^N(t)) - f(X_i^N(0)) - \sum_{s=0}^{t-1} \mathcal{G}f(X_i^N(s), U_s^N, \mu_s^N) + \varepsilon_i^{f,N}(t), \tag{21}$$

with $|\varepsilon_i^{f,N}(t)| \leq t\|f\|_\infty \epsilon_N$.

Step 3 : Accumulation to the infinite control problem.

In this paragraph, we finish the proof of Theorem 9 and prove that any weak limit of (μ^N, U^N) is a solution of the infinite control problem.

Let Π^∞ be an limit point of $\mathcal{L}(\mu^N, U^N)$ and let $(\nu, V) \in \mathcal{P}(\mathcal{X}^{\mathbb{N}}) \times \mathcal{U}^{\mathbb{N}}$ be a random variable with law Π^∞.

Lemma 13 *Π^∞-a.s. for all $t \geq 0$,*

$$\nu_{t+1} = \nu_t K_{V_t, \nu_t}.$$

Proof. Let (μ^N, U^N) be a weak converging subsequence to (ν, V). Summing (21) for all i, if $t \geq 1$, we obtain

$$\langle f, \mu_t^N\rangle - \langle f, \mu_0^N\rangle - \sum_{s=0}^{t-1} (\langle f, \mu_s^N K_{U_s^N, \mu_s^N}\rangle - \langle f, \mu_s^N\rangle) = \frac{1}{N}\sum_{i=1}^N M_i^{f,N}(t) + \varepsilon_i^{f,N}(t). \tag{22}$$

By assumption A2, the mapping $\varphi : (\nu, V) \mapsto \langle f, \nu_t\rangle - \langle f, \nu_0\rangle - \sum_{s=0}^{t-1}(\langle f, \nu_s K_{V_s,\nu_s}\rangle - \langle f, \nu_s\rangle)$ is Lipschitz: for some $C > 0$,

$$|\varphi(\nu, V) - \varphi(\nu', V')| \leq C\|f\|_\infty \sum_{s=0}^{t} \|V_s - V_s'\| + \|\nu_s - \nu_s'\|.$$

Hence, φ is continuous and $\varphi(\mu^N, U^N)$ converges weakly to $\varphi(\nu, V)$.

We have checked that $|\varepsilon_i^{f,N}(t)| \leq t\|f\|_\infty \epsilon_N$ where ϵ_N tends to 0. Also, by Lemma 12, $\frac{1}{N}\sum_{i=1}^N M_i^{f,N}(t)$ converges in L^2 to 0, hence, by Fatou's Lemma, Equation (22) gives:

$$\mathrm{E}\left| \langle f, \nu_t\rangle - \langle f, \nu_0\rangle - \sum_{s=0}^{t-1}(\langle f, \nu_s K_{V_s,\nu_s}\rangle - \langle f, \nu_s\rangle) \right|^2 \leq \lim_{N\to\infty} 2t^2\|f\|_\infty^2 \epsilon_N^2 + 2\mathrm{E}\left| \frac{1}{N}\sum_{i=1}^N M_i^{f,N}(t) \right|^2 = 0.$$

We conclude by induction that for all $t \geq 0$, Π^∞-a.s. $\nu_{t+1} = \nu_t K_{V_t,\nu_t}$.

We now define the mapping from $\mathcal{P}(\mathcal{X})^{\mathbb{N}} \times \mathcal{U}^{\mathbb{N}}$ to $\mathbb{R}^+$:

$$J_\beta(\nu, V) = (1-\beta)\sum_{t=0}^{\infty} \beta^t r(\nu_t, V_t).$$

J_β is continuous for the product topology, indeed:

Lemma 14 *There exists $C > 0$ such that for all $\nu, \nu' \in \mathcal{P}(\mathcal{X}^{\mathbb{N}})$, $V, V' \in \mathcal{U}^{\mathbb{N}}$,*

$$|J_\beta(\nu, V) - J_\beta(\nu', V')| \leq C \left(\|\nu - \nu'\|_\beta + \|V - V'\|_\beta \right),$$

where $\|\nu - \nu'\|_\beta = \sum_{t=0}^{\infty} \beta^t \|\nu_t - \nu'_t\|$.

Proof. Let $C = \max_{(x,u)\in\mathcal{X}\times\mathcal{U}} |r(x,u)| = \|r\|_\infty$. We write,

$$\begin{aligned} |J_\beta(\nu, V) - J(\nu', V')| &\leq |J_\beta(\nu, V) - J(\nu', V)| + |J_\beta(\nu', V) - J_\beta(\nu', V')| \\ &\leq (1-\beta) \sum_{t=0}^{\infty} \beta^t C \|\nu_t - \nu'_t\| + (1-\beta) \sum_{t=0}^{\infty} \beta^t C \mathbf{1}_{V_t = V'_t} \\ &\leq C \left(\|\nu - \nu'\|_\beta + \|V - V'\|_\beta \right). \end{aligned}$$

□

We now check the continuity of $\mathcal{J}_\beta(q)$.

Lemma 15 *The mapping $q \mapsto \mathcal{J}_\beta(q)$ is uniformly continuous.*

Proof. First, fix $U = (U_t)_{t\in\mathbb{N}}$ a control sequence and consider two initial conditions μ_0 and $\tilde{\mu}_0$ and their trajectory $(\mu_t)_{t\in\mathbb{N}}$ and $(\tilde{\mu}_t)_{t\in\mathbb{N}}$ obtained by the recursion (7). By assumption A2, we have: $\|\mu_1 - \tilde{\mu}_1\| = \|\mu_0 K_{U_0,\mu_0} - \tilde{\mu}_0 K_{U_0,\tilde{\mu}_0}\| \leq \|\mu_0 K_{U_0,\mu_0} - \tilde{\mu}_0 K_{U_0,\mu_0}\| + \|\tilde{\mu}_0 K_{U_0,\mu_0} - \tilde{\mu}_0 K_{U_0,\tilde{\mu}_0}\| \leq (C+1)\|\mu_0 - \tilde{\mu}_0\| = C_1 \|\mu_0 - \tilde{\mu}_0\|$. By recursion we deduce that:

$$\|\mu_t - \tilde{\mu}_t\| \leq C_1^t \|\mu_0 - \tilde{\mu}_0\|. \tag{23}$$

Fix $\epsilon > 0$ and let $J_{\beta,T}(\nu, V) = (1-\beta) \sum_{t=0}^{T} \beta^t r(\nu_t, V_t)$. There exists $T \in \mathbb{N}$, such that for all $(\nu, V) \in \mathcal{P}(\mathcal{X})^{\mathbb{N}} \times \mathcal{U}^{\mathbb{N}}$,

$$|J_{\beta,T}(\nu, V) - J_\beta(\nu, V)| \leq \epsilon. \tag{24}$$

Now let $q, \tilde{q} \in \mathcal{P}(\mathcal{X})$, we apply an optimal control strategy $U = (U_t)_{t\in\mathbb{N}}$ of q both to q and $\tilde{q}$. We obtain two trajectories $(\mu_t)_{t\in\mathbb{N}}$ and $(\tilde{\mu}_t)_{t\in\mathbb{N}}$ with initial condition $\mu_0 = q$ and $\tilde{\mu}_0 = \tilde{q}$. By construction:

$$\mathcal{J}_\beta(q) = J_\beta(\mu, U) \quad \text{and} \quad \mathcal{J}_\beta(\tilde{q}) \geq J_\beta(\tilde{\mu}, U).$$

From Equations (23), (24), we obtain:

$$\mathcal{J}_\beta(q) \leq \mathcal{J}_\beta(\tilde{q}) + \epsilon + \|q - \tilde{q}\|(1-\beta)\|r\|_\infty \sum_{t=0}^{T} \beta^t C_1^t.$$

and reciprocally by inverting q and $\tilde{q}$. Thus,

$$|\mathcal{J}_\beta(q) - \mathcal{J}_\beta(\tilde{q})| \leq \epsilon + \|q - \tilde{q}\|(1-\beta)\|r\|_\infty \sum_{t=0}^{T} \beta^t C_1^t,$$

and the continuity follows directly. □

Lemma 16 *If $\tilde{\mu}_0 \in \mathcal{P}_N(\mathcal{X})$, let $(\tilde{\mu}, \tilde{U}) \in \mathcal{P}(\mathcal{X}^{\mathbb{N}}) \times \mathcal{U}$ be a solution of (7), with initial condition $\tilde{\mu}_0 \in \mathcal{P}_N(\mathcal{X})$ and $\tilde{\mu}^N \in \mathcal{P}(\mathcal{X}^{\mathbb{N}})$ be the empirical measure of the trajectories of the particles in a N particles system with initial condition $\tilde{\mu}_0$ and control sequence $\tilde{U}$. Then*

$$\lim_{N\to\infty} \sup_{\tilde{\mu}_0 \in \mathcal{P}_N(\mathcal{X})} \mathrm{E} \left| J_\beta(\tilde{\mu}, \tilde{U}) - J_\beta(\tilde{\mu}^N, \tilde{U}) \right| = 0.$$

Proof. Fix $\epsilon > 0$, as in the proof of Lemma 15, let $J_{\beta,T}(\nu, V) = (1-\beta)\sum_{t=0}^{T} \beta^t r(\nu_t, V_t)$. There exists $T \in \mathbb{N}$, such that for all $(\nu, V) \in \mathcal{P}(\mathcal{X}^{\mathbb{N}}) \times \mathcal{U}^{\mathbb{N}}$,

$$|J_{\beta,T}(\nu, V) - J_\beta(\nu, V)| \leq \epsilon/2.$$

It remains to check that there exists a sequence $(\alpha_N)_{N\in\mathbb{N}}$ with $\lim_N \alpha_N = 0$ such that, for all $\tilde{\mu}_0 \in \mathcal{P}_N(\mathcal{X})$,

$$\left| J_{\beta,T}(\tilde{\mu}, \tilde{U}) - \mathrm{E} J_{\beta,T}(\tilde{\mu}^N, \tilde{U}) \right| \leq \alpha_N. \tag{25}$$

The proof of (25) is an extension of the proofs of Lemmas 13 and 15. Notice first that Lemma 13 was proved for the sequence of processes (μ^N, U^N) obtained from the strategies π_*^N. However, no assumption on π_*^N were used and Lemma 13 holds for any limit point of sequence of processes $(\mu^N_{\pi^N}, U^N_{\pi^N})$ obtained from the strategies π^N. We may thus apply this result to the strategy which enforces at any time t, the control $\tilde{U}_t$. Equation (22) for $t = 1$ reads $\langle f, \tilde{\mu}_1^N\rangle - \langle f, \tilde{\mu}_0 K_{\tilde{U}_0,\tilde{\mu}_0}\rangle = \frac{1}{N}\sum_{i=1}^N M_i^{f,N}(1) + \varepsilon_i^{f,N}(1)$. Hence by Lemma 12,

$$\mathrm{E}\left| \langle f, \tilde{\mu}_1^N\rangle - \langle f, \tilde{\mu}_0 K_{\tilde{U}_0,\tilde{\mu}_0}\rangle \right|^2 \leq 2\epsilon_N^2 \|f\|_\infty + 2C\|f\|_\infty^2/N + 2C\|f\|_\infty^2 \delta_N$$

Since $\mathcal{X}$ is finite, we deduce that there exists a sequence γ_N with $\lim_N \gamma_N = 0$ such that:

$$\mathrm{E}\|\tilde{\mu}_1^N - \tilde{\mu}_0 K_{\tilde{U}_0,\tilde{\mu}_0}\| \leq \gamma_N,$$

and the sequence γ_N does not depend on the pair $(\tilde{\mu}_0, \tilde{U}_0)$. Notice also that for $t \geq 1$,

$$\|\tilde{\mu}_{t+1}^N - \tilde{\mu}_{t+1}\| \leq \|\tilde{\mu}_{t+1}^N - \tilde{\mu}_t^N K_{\tilde{U}_t,\tilde{\mu}_t^N}\| + \|\tilde{\mu}_t^N K_{\tilde{U}_t,\tilde{\mu}_t^N} - \tilde{\mu}_t K_{\tilde{U}_t,\tilde{\mu}_t}\|.$$

By assumption A2, as in the proof of Lemma 15, we have: $\|\tilde{\mu}_t K_{\tilde{U}_t,\tilde{\mu}_t} - \tilde{\mu}_t^N K_{\tilde{U}_t,\tilde{\mu}_t^N}\| \leq C_1 \|\tilde{\mu}_t - \tilde{\mu}_t^N\|$, with $C_1 > 1$. It follows that:

$$\mathrm{E}\|\tilde{\mu}_{t+1}^N - \tilde{\mu}_{t+1}\| \leq \gamma_N + C_1 \mathrm{E}\|\tilde{\mu}_t^N - \tilde{\mu}_t\|.$$

So that:

$$\mathrm{E}\|\tilde{\mu}_t^N - \tilde{\mu}_t\| \leq \gamma_N \frac{C_1^t - 1}{C_1 - 1},$$

and

$$\mathrm{E}\left| J_{\beta,T}(\tilde{\mu}, \tilde{U}) - J_{\beta,T}(\tilde{\mu}^N, \tilde{U}) \right| \leq (1-\beta)\gamma_N \|r\|_\infty \sum_{t=0}^{T} \beta^t \frac{C_1^t - 1}{C_1 - 1}.$$

Equation (25) follows. □

The next lemma concludes the proof of Proposition 11.

Lemma 17 *Π^∞-a.s., for all t, $V_t \in \mathcal{U}_\beta(\nu_t)$ and $\mathcal{J}_\beta(q_0) = J_\beta(\nu, V)$.*

Proof. The projection map $\nu \mapsto \nu_0$ is continuous. In particular, this implies that $\nu_0 = q_0$, Π^∞-a.s. Hence, Lemma 13 implies, Π^∞-a.s.

$$J_\beta(\nu, V) \leq \mathcal{J}_\beta(q_0) \tag{26}$$

Π^∞ is an limit point of $\mathcal{L}(\mu^N, U^N)$, thus, up to extracting a converging subsequence of N, Lemma 14 implies that $J_\beta(\mu^N, U^N)$ converges weakly to $J_\beta(\nu, V)$. Since, by construction,

$\mathrm{E}J_\beta(\mu^N, U^N) = \mathcal{J}_\beta^N(\mu_0^N)$, we deduce that for this converging subsequence, $\lim_N \mathcal{J}_\beta^N(\mu_0^N) = \mathrm{E}J_\beta(\nu, V) \leq \mathcal{J}_\beta(q_0)$.

The other way around, we define an admissible control strategy by induction for the N particles system (this strategy is however not Markovian). Let $\overline{\mu}_0^N = \mu_0^N$ and draw with Z_0^N a control $\overline{U}_0^N \in \mathcal{U}_\beta(\overline{\mu}_0^N)$. Then set $\overline{\mu}_1^N = F(\overline{\mu}_0^N, \overline{U}_0^N)$ and draw with Z_1^N a control $\overline{U}_1^N \in \mathcal{U}_\beta^N(\overline{\mu}_1^N)$ (see page 5 for details on how to draw with Z_1^N an element in $\mathcal{U}_\beta^N(\overline{\mu}_1^N)$). By induction, we define a sequence of controls $\overline{U}^N = (\overline{U}_0^N, \overline{U}_1^N, \cdots)$, such that $\overline{U}_t^N \in \sigma(\mu_0^N, Z_0^N, Z_1^N, \cdots, Z_t^N) \subseteq \mathcal{F}_t^N$. Define $\tilde{\mu}^N$ as the empirical measure of the particle system with N particles when the control applied at time t is $\overline{U}_t^N$ and the initial condition is μ_0^N. This control process is admissible and it is sub-optimal, i.e.

$$\mathrm{E}[J_\beta(\tilde{\mu}^N, \overline{U}^N)] \leq \mathcal{J}_\beta^N(\mu_0^N). \tag{27}$$

From Lemma 16, there exists a sequence $(\gamma_N)_{N\in\mathbb{N}}$ with $\lim_N \gamma_N = 0$ such that

$$\mathrm{E}[\big|\ J_\beta(\overline{\mu}^N, \overline{U}^N) - J_\beta(\tilde{\mu}^N, \overline{U}^N)\ \big|] \leq \gamma_N.$$

Hence Equation (27) implies

$$\mathcal{J}_\beta(\mu_0^N) \leq \mathcal{J}_\beta^N(\mu_0^N) + \gamma_N.$$

Then, by Lemma 15,

$$\mathcal{J}_\beta(q_0) \leq \liminf_N \mathcal{J}_\beta^N(\mu_0^N).$$

So finally, with Equation (26),

$$\Pi^\infty\text{-a.s.} \quad \mathcal{J}_\beta(q_0) = J_\beta(\nu, V).$$

It follows that Π^∞-a.s. the pair (ν, V) is a solution of the DP recursion given by (7), (8). □

Proof of Theorem 1. Note that the sequence $(\mathcal{J}_\beta^N(\mu_0^N))_{N\in\mathbb{N}}$ is bounded, indeed $|\mathcal{J}_\beta^N(\mu_0^N)| \leq \|r\|_\infty$. Therefore, in order to prove that $\lim_N \mathcal{J}_\beta^N(\mu_0^N) = \mathcal{J}_\beta(q_0)$ it suffices to show that along any increasing subsequence of integers N, we may extract a subsequence (N_k) such that $\mathcal{J}_\beta^{N_k}(\mu_0^{N_k})$ converges to $\mathcal{J}_\beta(q_0)$. Now, along an increasing subsequence of integers N, let $(\mu^N, U^N) \in \mathcal{P}(\mathcal{X}^\mathbb{N}) \times \mathcal{U}^\mathbb{N}$ be a discount optimal control process with initial condition μ_0^N: $\mathcal{J}_\beta^N(\mu_0^N) = \mathrm{E}J_\beta(\mu^N, U^N)$. From Step 1 in the proof of Proposition 11, $\mathcal{L}(\mu^N, U^N)$ is tight in $\mathcal{P}(\mathcal{P}(\mathcal{X}^\mathbb{N}) \times \mathcal{U}^\mathbb{N})$. Let $(N_k)_{k\in\mathbb{N}}$ be a converging subsequence, $\mathcal{L}(\mu^{N_k}, U^{N_k})$ converges weakly to $\Pi^\infty = \mathcal{L}(\nu, V)$. From Proposition 11, Π^∞-a.s., $J_\beta(\nu, V) = \mathcal{J}_\beta(q_0)$. The mapping $J_\beta : (\mu, U) \mapsto J_\beta(\mu, U)$ is continuous, hence $\lim_{k\to\infty} \mathcal{J}_\beta^{N_k}(\mu_0^{N_k}) = \lim_{k\to\infty} \mathrm{E}J_\beta(\mu^{N_k}, U^{N_k}) = \mathrm{E}^\infty J_\beta(\nu, V) = \mathcal{J}_\beta(q_0)$, where E^∞ denotes the expectation with respect to the law Π^∞. Therefore, we have proved that $\lim_N \mathcal{J}_\beta^N(\mu_0^N) = \mathcal{J}_\beta(q_0)$.

Similarly, let $u \in \limsup \mathcal{U}_\beta^N(\mu_0^N)$, then along an increasing subsequence of integers N, $u \in \mathcal{U}_\beta^N(\mu_0^N)$. Along this increasing subsequence, there exists a discount optimal control process $(\mu^N, U^N) \in \mathcal{P}(\mathcal{X}^\mathbb{N}) \times \mathcal{U}^\mathbb{N}$ with initial condition μ_0^N and initial control $U_0^N = u$. From Step 1 in the proof of Proposition 11, $\mathcal{L}(\mu^N, U^N)$ is tight in $\mathcal{P}(\mathcal{P}(\mathcal{X}^\mathbb{N}) \times \mathcal{U}^\mathbb{N})$. Let $(N_k)_{k\in\mathbb{N}}$ be a converging subsequence, $\mathcal{L}(\mu^{N_k}, U^{N_k})$ converges weakly to $\Pi^\infty = \mathcal{L}(\nu, V)$. From Lemma 17, Π^∞-a.s., $V_0 \in \mathcal{U}_\beta(\nu_0)$. The projection mapping $(\mu, U) \mapsto (\mu_0, U_0)$ is continuous, hence Π^∞-a.s. $\nu_0 = q_0$ and $V_0 = u$, and we obtain $u \in \mathcal{U}_\beta(q_0)$.

Proof of Corollary 9. As in the statement of Corollary 9, we now assume that U^N converges weakly to U. It implies that Π^∞-a.s. $V = U$. From Proposition 2.2. in Sznitman [10], it is sufficient (and nearly equivalent) to prove only, for some $\mu \in \mathcal{P}(\mathcal{X}^{\mathbb{N}})$,

$$\lim_{N\to\infty} \mathcal{L}(\mu^N) = \delta_\mu \quad \text{weakly in } \mathcal{P}(\mathcal{P}(\mathcal{X}^{\mathbb{N}})). \tag{28}$$

Thus we need to prove that Π^∞-a.s. $\nu = \mu$. The proof is an extension of the above arguments and follows a standard technique. Let $X = (X_t) \in \mathcal{X}^{\mathbb{N}}$ be a canonical trajectory. We prove the following

Lemma 18 *Π^∞-a.s., ν satisfies a discrete martingale problem: for all functions f on $\mathcal{X}$,*

$$M_t^f := f(X_t) - f(X_0) - \sum_{s=0}^{t-1} \mathcal{G}f(X_s, U_s, \nu_s) \tag{29}$$

is a ν-martingale with initial condition $\nu_0 = q_0$.

Proof. It suffices to prove that for all $t \geq 1$ and all functions g from $\mathcal{X}^t$ to $\mathbb{R}$,

$$\Pi^\infty\text{-a.s.} \quad \langle \nu, M_t^f g(X_0, \cdots, X_{t-1})\rangle = \langle \nu, M_{t-1}^f g(X_0, \cdots, X_{t-1})\rangle.$$

(recall that $\langle \nu, g(X_0, \cdots, X_{t-1})\rangle = \int_{\mathcal{X}^{\mathbb{N}}} g(X_1, \cdots, X_{t-1})\nu(dX)$). We define the function Φ from $\mathcal{P}(\mathcal{X}^{\mathbb{N}})$ to $\mathbb{R}$, defined by

$$\Phi(\nu) = \langle \nu, (M_t^f - M_{t-1}^f) g(X_0, \cdots, X_{t-1})\rangle.$$

Let $g_i^N = g(X_i^N(0), \cdots, X_i^N(t-1))$, applying (21) to f at time t and $t-1$, we get:

$$\begin{gathered}\left(f(X_i^N(t)) - f(X_i^N(t-1) - \mathcal{G}f(X_i^N(t-1), U_{t-1}^N, \mu_{t-1}^N)\right) g_i^N = \\ \left(M_i^{f,N}(t) - M_i^{f,N}(t-1) + \varepsilon_i^{f,N}(t) - \varepsilon_i^{f,N}(t-1)\right) g_i^N,\end{gathered}$$

and summing over all i, we deduce

$$\begin{aligned}
\mathrm{E}|\Phi(\mu^N)| &= \mathrm{E}\frac{1}{N}\Big|\sum_{i=1}^N \left(f(X_i^N(t)) - f(X_i^N(t-1) - \mathcal{G}f(X_i^N(t-1), U_{t-1}^N, \mu_{t-1}^N)\right) g_i^N \Big| \\
&= \mathrm{E}\frac{1}{N}\Big|\sum_{i=1}^N \left(M_i^{f,N}(t) - M_i^{f,N}(t-1) + \varepsilon_i^{f,N}(t) - \varepsilon_i^{f,N}(t-1)\right) g_i^N \Big| \\
&\leq \mathrm{E}\Big|\frac{1}{N}\sum_{i=1}^N \left(M_i^{f,N}(t) - M_i^{f,N}(t-1)\right) g_N^i \Big| + \mathrm{E}\Big|\frac{1}{N}\sum_{i=1}^N \left(\varepsilon_i^{f,N}(t) - \varepsilon_i^{f,N}(t-1)\right) g_i^N \Big| \\
&\leq \mathrm{I} + \mathrm{II}.
\end{aligned} \tag{30}$$

From Step Two, $|\varepsilon_i^{f,N}(t)| \leq t\|f\|_\infty \epsilon_N$ hence, $\mathrm{II} \leq 2t\|g\|_\infty\|f\|_\infty\epsilon_N$ which tends to 0 as N goes to infinity. Using exchangeability, Lemma 12 and Cauchy-Schwartz inequality,

$$\mathrm{I}^2 \leq \frac{\|g\|_\infty^2}{N}\mathrm{E}|M_i^{f,N}(t) - M_i^{f,N}(t-1)|^2 + \frac{N-1}{N}\mathrm{E}(M_1^{f,N}(t) - M_1^{f,N}(t-1))(M_2^{f,N}(t) - M_2^{f,N}(t-1)) g_1^N g_2^N.$$

Thus from Lemma 12, I tends to 0 as N goes to infinity. By Fatou's Lemma, we obtain:

$$\mathrm{E}|\Phi(\nu)| \leq \liminf_{N\to\infty} \mathrm{E}|\Phi(\mu^N)| = 0.$$

Therefore Π^∞-a.s. $\langle \nu, (M_t^f - M_{t-1}^f) g(X_0, \cdots, X_{t-1}) \rangle = 0$. □

The solution of the discrete martingale problem (29) is unique and is known (see for example Problem 16 p 264 in Ethier and Kurtz [6]). The unique measure $\mu \in \mathcal{P}(\mathcal{X}^{\mathbb{N}})$ such that $\mu_0 = q_0$ and which satisfies (29) for all functions f is the law of the process (X_t) defined by recursion with $\mathcal{L}(X_0) = \mu_0 = q_0$ and transition probabilities:

$$\mathrm{P}(X_{t+1} \in \cdot | (X_0, U_0, \cdots, X_t, U_t)) = K_{U_t, \mu_t}(X_t, \cdot).$$

We have proved that Π^∞-a.s. $\nu = \mu$. Hence, we have proved that $\lim_{N \to \infty} \mathcal{L}(\mu^N) = \delta_\mu$ weakly in $\mathcal{P}(\mathcal{P}(\mathcal{X}^{\mathbb{N}}))$. □

4.2 Proof of Theorem 2

The proof of Theorem 1 extends without any difficulty to the finite average reward and leads to Theorem 2.

4.3 Proofs of Theorem 5 and Corollary 10

As in Lemma 4, we consider a stationary solution (μ^N, U^N) of the DP average cost problem. There exists $\Psi^N \in \mathcal{S}_{av}^N$ such that $\mathcal{L}(\mu_0^N) = \Psi^N(\cdot, \mathcal{U})$ and the law of U_t^N given $(\mu_t^N, \overline{\mathcal{F}_{t-1}^N})$ is $\Psi^N(\mu_t^N, \cdot)/\Psi^N(\mu_t^N, \mathcal{U})$. The following proposition implies both Theorem 5 and Corollary 10. We consider for all $N \in \mathbb{N}$, a vector of initial condition $(X_1^N(0), \cdots, X_N^N(0))$ which is exchangeable, and that satisfies $\mu_0^N = \frac{1}{N} \sum_{i=1}^N \delta_{X_i^N(0)}$.

Proposition 19 *Under the assumptions of Theorem 5, if $(X_1^N(0), \cdots, X_N^N(0))$ is exchangeable, then any limit point of $\mathcal{L}(\mu^N, U^N)$ in $\mathcal{P}(\mathcal{P}(\mathcal{X}^{\mathbb{N}}) \times \mathcal{U}^{\mathbb{N}})$ has its support included in $\mathcal{S}_{av}$.*

We first prove Theorem 5 and Corollary 10 using Proposition 19. Note that if a subsequence of $\mathcal{L}(\mu^N, U^N)$ converges in $\mathcal{P}(\mathcal{P}(\mathcal{X}^{\mathbb{N}}) \times \mathcal{U}^{\mathbb{N}})$, then $\mathcal{L}(\mu_0^N, U_0^N) = \Psi^N$ also converges. By continuity we deduce that $\mathcal{J}_{av}^N = \langle \Psi^N, r \rangle = \mathrm{E} r(\mu_0^N, U_0^N)$ converges to $\mathcal{J}_{av}$ and since $\Psi^N \in \mathcal{S}_{av}^N$ it implies also that $\limsup_N \mathcal{S}_{av}^N \subseteq \mathcal{S}_{av}$. If in addition Assumption A7 holds, then there exists a unique stationary measure q_* solution of (7). Hence, since the mapping $\nu \mapsto \nu_t$ is continuous, by Proposition 19, for all t, $\lim_N \mu_t^N = q_*$ weakly in $\mathcal{P}(\mathcal{X})$. From Proposition 2.2 in Sznitman [10] it implies Corollary 10.

Proof of Proposition 19 The proof follows step by step the proof of Theorem 9. Let $(\mu^N, U^N) \in \mathcal{P}(\mathcal{X}^{\mathbb{N}}) \times \mathcal{U}^{\mathbb{N}}$ be a stationary solution of the average cost problem with N particles.

We may apply Steps 1 and 2 of the proof of Theorem 9 to the process (μ^N, U^N). We deduce that its law is tight in $\mathcal{P}(\mathcal{P}(\mathcal{X}^{\mathbb{N}}) \times \mathcal{U}^{\mathbb{N}})$ and Lemma 13 holds to any limit point of its law. The difficulty comes from deriving an analog of Lemma 17 to the mapping from $\mathcal{P}(\mathcal{X})^{\mathbb{N}} \times \mathcal{U}^{\mathbb{N}}$ to $\mathbb{R}^+$: $J_{av}(\nu, V)$ which is certainly not continuous for the topology induced by the distance $\|\nu - \nu'\|_\beta + \|V - V'\|_\beta$. Borkar's ergodic occupation measure solves this difficulty. Let Π^∞ be an limit point of $\mathcal{L}(\mu^N, U^N)$ and let $(\nu, V) \in \mathcal{P}(\mathcal{X}^{\mathbb{N}}) \times \mathcal{U}^{\mathbb{N}}$ be a random variable with law Π^∞. Proposition 19 is a consequence of Lemma 13 and the next Lemma.

Lemma 20 *Π^∞-a.s. $J_{av}(\nu, V) = \mathcal{J}_{av}$*

Proof. Let $\Pi_t^\infty \in \mathcal{P}(\mathcal{P}(\mathcal{X}) \times \mathcal{U})$ be the distribution of (ν_t, V_t) and let A be a measurable set in $\mathcal{P}(\mathcal{X})$ such that $\Pi_0^\infty(\partial A, \mathcal{U}) = \Pi_1^\infty(\partial A, \mathcal{U}) = 0$. Equation (12) reads:

$$\mathrm{P}(\mu_1^N \in A) = \mathrm{P}(\mu_0^N \in A).$$

Along a converging subsequence, we deduce that: $\mathrm{P}(\nu_1 \in A) = \mathrm{P}(\nu_0 \in A)$. Since Π^∞-a.s. $\nu_{t+1} = \nu_t K_{V_t,\nu_t}$, this last equation can be restated as $\mathrm{P}(\nu_0 K_{V_0,\nu_0} \in A) = \mathrm{P}(\nu_0 \in A)$. Thus if $\Psi' = \Pi_0^\infty$, we obtain

$$\sum_{u \in \mathcal{U}} \Psi'(\{q : qK_{u,q} \in A\}, u) = \Psi'(A, \mathcal{U}).$$

Hence, by the definition of $\mathcal{S}_{av}$, Π^∞-a.s.,

$$\langle \Psi', r \rangle \leq \mathcal{J}_{av}.$$

Moreover, note that the mapping $(\nu', V') \mapsto r(\nu_0', V_0')$ is continuous, so that along a converging subsequence:

$$\langle \Psi', r \rangle = \lim_{N \to \infty} \mathrm{E} r(\mu_0^N, U_0^N) = \lim_{N \to \infty} \mathcal{J}_{av}^N.$$

The other way around, let $\Psi \in \mathcal{S}_{av}$ as in Assumption A6. For N particle system, using the extra randomness Z^N, we may define a Markov strategy π and associated process (μ^N, U^N) such that $\mathrm{P}(U_t^N = u | \mu_t^N, \mu_{t-1}^N, \cdots, \mu_0^N) = \alpha_u(\mu_t^N)$. (μ^N, U^N) is then a Markov chain on the finite state space $\mathcal{P}_N(\mathcal{X}) \times \mathcal{U}$. Let Π_0^N be a stationary distribution of this Markov chain (note that it is not necessarily unique), and (μ^N, U^N) the stationary process with initial distribution Π_0^N. As usual, Π^N denotes the law of the process (μ^N, U^N) and Π_t^N the law of (μ_t^N, U_t^N). By construction, for all $t \in \mathbb{N}$, $\Pi_t^N = \Pi_0^N$ and Equation (12) holds for Π_0^N and it follows:

$$\langle \Pi_0^N, r \rangle \leq \mathcal{J}_{av}^N.$$

As in Step 1 of the proof of Theorem 1, Π^N is tight in $\mathcal{P}(\mathcal{P}(\mathcal{X}^\mathbb{N}) \times \mathcal{U}^\mathbb{N})$. Let Π^∞ be an limit point of $\mathcal{L}(\mu^N, U^N)$. The projection mapping $\Pi \mapsto \Pi_t$ is continuous, hence for all $t \in \mathbb{N}$, $\Pi_t^\infty = \Pi_0^\infty$. Now, as above Lemma 13 holds: if $\mathcal{L}(\nu, V) = \Pi^\infty$, then Π^∞-a.s., $\nu_{t+1} = \nu_t K_{V_t,\nu_t}$. Moreover, the continuity of $q \mapsto \alpha_u(q)$ implies also $\mathrm{P}(V_t = u | \nu_t, \nu_{t-1}, \cdots, \nu_0) = \alpha_u(\nu_t)$. Hence, Π_0^∞ is an invariant measure of the Markov transition kernel (15). By assumption A6, $(\alpha_u)_{u \in \mathcal{U}} \in \mathcal{C}_r$ and thus $\langle \Pi_0^\infty, r \rangle = \langle \Psi, r \rangle = \mathcal{J}_{av}$. We obtain that along a converging subsequence,

$$\lim_{N \to \infty} \langle \Pi_0^N, r \rangle = \langle \Pi_0^\infty, r \rangle = \mathcal{J}_{av},$$

and it concludes the proof. □

5 Symmetry breaking in controlled particle systems

5.1 Phase transition in epidemic models

We consider the following simplistic epidemic model on N particles. The particles are agents in a network and a virus is spreading throughout the network. Time is slotted, and a central controller controls the global activity in the network. The state of a particle is $\mathcal{X} = \{0, 1\}$, in state 0, the particle is healthy whereas in state 1 the particle is infected. For a given control parameter, $u \in \mathcal{U} \subseteq (0, 1)$, at a given time slot, an infected particle becomes healthy with probability $0 < h(u) < 1$ independently of everything else. A healthy particle, with probability u, communicates in the network with a particle picked uniformly among the $N - 1$ remaining particles independently of everything else. If the randomly picked particle is infected then the healthy particle becomes infected otherwise it remains healthy. We consider the ergodic average

reward for the reward function $r(x,u) = 1 - x$. This model fits in our framework and, with $q \in \mathcal{P}_N(\mathcal{X})$,

$$\begin{aligned} K^N_{u,q}(1,0) &= h(u), \\ K^N_{u,q}(0,1) &= \frac{N}{N-1} uq(1), \end{aligned}$$

and we obtain a limit kernel $K_{u,q}$. Note that for all $u \in \mathcal{U}$ and $N \in \mathbb{N}$, the corresponding Markov chain is irreducible, and the state where all particles are healthy is absorbing, therefore, for all $N \in \mathbb{N}$:

$$\mathcal{J}^N_{av} = 1.$$

As N goes to infinity, the limit empirical measure evolves according to the recursion $\mu_{t+1} = \mu_t K_{u_t,\mu_t}$. Hence if $\alpha_t = \mu_t(\{1\})$ is the proportion of infected particles, we have

$$\alpha_{t+1} = \alpha_t(1 + u_t(1 - \alpha_t) - h(u_t)).$$

The fixed points of this equation are $\alpha^{(1)} = 0$ and $\alpha^{(2)} = 1 - h(u)/u$. If there exists $u \in \mathcal{U}$ such that $h(u) \geq u$, then there is a unique attracting fixed point to the equation $\mu_{t+1} = \mu_t K_{u,\mu_t}$: the measure δ_0 with associated reward 1. We then have $\mathcal{J}^*(q) = 1$ for all $q \in \mathcal{P}(\mathcal{X})$. Otherwise, if $\gamma = \max_{u \in \mathcal{U}} h(u)/u < 1$, then there exists another locally stable fixed point: the measure $(h(u)/u)\delta_0 + (1 - h(u)/u)\delta_1$ with associated reward $h(u)/u$. A quick calculation shows then that $\mathcal{J}_{av}(q) = 1$ if $q(1) < \gamma$ and $\mathcal{J}_{av}(q) = \gamma$, otherwise.

As a conclusion, depending on the value of γ, there is a phase transition in this simplistic epidemic model. If $\gamma \geq 1$, then for all $q \in \cup_{N \in \mathbb{N}} \mathcal{P}_N(\mathcal{X})$, $\lim_T \lim_N \mathcal{J}^N_T(q) = \lim_N \lim_T \mathcal{J}^N_T(q) = 1$, whereas, if $\gamma < 1$, this exchange of limits fails for some initial conditions q.

5.2 Non-uniform initial condition - Uniform interactions

The state of a particle is $\mathcal{X} = \{0,1\} \times \{1,2\}$ corresponding to an energy state 0 or 1 and a class 1 or 2. The control parameter is $\mathcal{U} = \{u = (u(1), u(2)) \in [0,1]^2 : u(1) + u(2) = 1\}$, the control $u(1)$ is applied to class-1 particles and $u(2)$ to class-2 particles. The transition kernel $K_{u,q}$ for a single particle is described as follows

$$\begin{aligned} &(1,c) \to (1,c) \quad \text{with probability} \quad K_{u,q}((1,c),(1,c)) = \varphi(u(1)q(1,1) + u(2)q(1,2)), \\ &(0,c) \to (0,c) \quad \text{with probability} \quad K_{u,q}((0,c),(0,c)) = 1. \end{aligned}$$

For some function $0 < \varphi(\cdot) < 1$. All other transitions have probability 0. The energy state 0 is an absorbing energy state. We define the recursion:

$$\mu_{t+1} = \mu_t K_{u_t,\mu_t}.$$

The reward function is

$$r(x,u) = x.$$

The aim is to maximize $\sum_{t \geq 0} \beta^t r(\mu_t, u_t)$. If $\alpha_t(c) = \mu_t(1,c)$, and $\alpha_t = \alpha_t(1) + \alpha_t(2)$. The reward is simply

$$\sum_{t \geq 0} \beta^t \alpha_t,$$

and:

$$\begin{aligned}\alpha_{t+1} &= \alpha_t \varphi(u_t(1)\alpha_t(1) + u_t(2)\alpha_t(2)) \\ \alpha_{t+1}(c) &= \alpha_t(c)\varphi(u_t(1)\alpha_t(1) + u_t(2)\alpha_t(2))\end{aligned}$$

For example assume φ is convex and monotone, then the optimal control is

$$\begin{aligned}\alpha_{t+1} &= \alpha_t \max(\varphi(\alpha_t(1)), \varphi(\alpha_t(2))) \\ \alpha_{t+1}(c) &= \alpha_t(c) \max(\varphi(\alpha_t(1)), \varphi(\alpha_t(2))).\end{aligned}$$

The associated reward should be compared with the reward of the system with a unique class: $u = (1/2, 1/2)$. However, even if the particle system has uniform interactions, if $\alpha_0(1) \neq \alpha_0(2)$, it is possible to benefit from the control over two classes.

5.3 Uniform initial condition - Symmetric interactions

Same as above but the transition kernel $K_{u,q}$ for one particle is

$$\begin{aligned}(1,1) \to (1,1) \quad &\text{with probability} \quad K_{u,q}((1,1),(1,1)) = \varphi(u(1)q(1,1), u(2)q(1,2)), \\ (1,2) \to (1,2) \quad &\text{with probability} \quad K_{u,q}((1,2),(1,2)) = \varphi(u(2)q(1,2), u(1)q(1,1)) \\ (0,c) \to (0,c) \quad &\text{with probability} \quad K_{u,q}((0,c),(0,c)) = 1.\end{aligned}$$

For some function $0 < \varphi(\cdot,\cdot) < 1$. With the above notations, we have

$$\begin{aligned}\alpha_{t+1}(c) &= \alpha_t(1)\varphi(u_t(1)\alpha_t(1), u_t(2)\alpha_t(2)) \\ \alpha_{t+1}(c) &= \alpha_t(2)\varphi(u_t(2)\alpha_t(2), u_t(1)\alpha_t(1))\end{aligned}$$

Assume that $\alpha_0(1) = \alpha_0(2) = \alpha_0$: uniform initial conditions. Then, we may benefit from the control over two classes if:

$$\max_{u \in \mathcal{U}: u_1 + u_2 = 1} \varphi(u(1)\alpha_0, u(2)\alpha_0) + \varphi(u(2)\alpha_0, u(1)\alpha_0) > \varphi(\alpha_0/2, \alpha_0/2).$$

5.4 Uniform initial condition - Symmetric interactions - Average reward

Same as above, we add an extra transition

$$(0,c) \to (1,c) \quad \text{with probability} \quad K_{u,q}((0,c),(1,c)) = \delta.$$

The evolution equations are:

$$\begin{aligned}\alpha_{t+1}(1) &= \alpha_t(1)\varphi(u_t(1)\alpha_t(1), u_t(2)\alpha_t(2)) + \delta\mu_t(0,1) \\ \alpha_{t+1}(2) &= \alpha_t(2)\varphi(u_t(2)\alpha_t(2), u_t(1)\alpha_t(1)) + \delta\mu_t(0,2).\end{aligned}$$

We may then consider the average reward optimization. To this end, we need to compute the fixed points of these evolution equations, we might expect to find optimal configurations such that $\alpha(1) \neq \alpha(2)$ if φ is not symmetric in its first and second variable.

5.5 Space-time particle system

The particles have a position on the d-dimensional unit torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$. The state of the particle is in $\mathcal{X} = \{0,1\} \times \mathbb{T}^d$, corresponding to an energy state 0 or 1 and a position on the torus. The energy state 0 is a ground state and 1 is an excited state.

The interactions between particles depend on their relative distance on $\mathbb{T}^d$. This interaction is captured by an influence function I on $\mathbb{T}^d$, $I(z)$ being the influence of z over a particle located at 0. We will assume that the influence is a non-increasing function of $|z|$, ($|\cdot|$ denotes the norm on the unit torus). Typical examples are $I(z) = \mathbf{1}(|z| \leq r)$, the interaction with range r, or $I(z) = (1+|z|)^{-\alpha}$ long range interaction with exponent $\alpha > 0$.

The control set $\mathcal{U}$ is a subset of the set of measurable non-negative functions on $\mathbb{T}^d$ such that $\int_{\mathbb{T}^d} u(z)dz = 1$. $u(z)$ is thought as the density of the effort for particles located at z.

For simplicity, the particles are assumed to have a fixed position, so that for all $z \neq z' \in \mathbb{T}^d$, $K_{u,q}((x,z),(x,z')) = 0$. We define $K^z_{u,q}$ as the transition kernel of a particle located at $z \in \mathbb{T}^d$, we assume that $K^z_{u,q}$ is as follows

$$\begin{array}{lll} (1,z) \to (1,z) & \text{with probability} & K^z_{u,q}(1,1) = \varphi\left(\int_{\mathbb{T}^d} I(z-\xi)u(\xi)q(1,d\xi)\right), \\ (0,z) \to (0,z) & \text{with probability} & K^z_{u,q}(0,0) = 1. \end{array}$$

For some function $0 < \varphi < 1$. Note in particular that 0 is an absorbing energy state.

The recursion satisfied by the empirical measure is:

$$\mu_{t+1}(\cdot, dz) = \mu_t(\cdot, dz) K^z_{u_t,\mu_t}.$$

The initial state is uniform: $\mu_0(1,dz) = \alpha_0 dz$, where $0 < \alpha_0 \leq 1$ is the density of particles in energy state 1 at time 0. In particular for all $t \geq 0$, $\mu_t(1,dz)$ has a density $p_t(z)$ with respect to the Lebesgue measure, $p_0(z) = \alpha_0$ and

$$p_{t+1}(z) = p_t(z)\varphi\left(\int_{\mathbb{T}^d} I(z-\xi)u_t(\xi)p_t(\xi)d\xi\right).$$

The reward function is simply taken as $r((x,z),u) = x$, so that the associated discounted reward is

$$\sum_{t\geq 0} \beta^t \int_{\mathbb{T}^d} p_t(z)dz.$$

We assume also that the control space $\mathcal{U}$ is finite: the torus is divided into $(R_i)_{1\leq i\leq M}$ regions and due to the coarseness of the control, $u(z)$ is constrained to be constant on each region $(R_i)_{1\leq i\leq M}$ and the possible values of $u(z)$ are finite. We assume also that the system is only partially observed: the control is a function of the integrated measure $\int_{\mathbb{T}^d} \mu(\cdot, dz)$ and the initial state is known to be spatially uniform. Hence the density α_0 is known and we are in the framework of Corollary 8. So finally, if $u_t(i)$ is the control applied in region R_i, the density evolution is:

$$p^{(M)}_{t+1}(z) = p^{(M)}_t(z)\varphi\left(\sum_{i=1}^M u_t(i)\int_{R_i} I(z-\xi)p^{(M)}_t(\xi)d\xi\right). \tag{31}$$

If $M = 1$, then $u_t(1) = 1$ and due to the complete symmetry of the system,

$$p^{(1)}_{t+1}(z) = p^{(1)}_{t+1}(0) = p^{(1)}_t(0)\varphi\left(p^{(1)}_t(0)\int_{\mathbb{T}^d} I(\xi)d\xi\right).$$

If $M \geq 2$ is it possible to achieve a higher average density $\int_{\mathbb{T}^d} p_t^{(M)}(z)dz$? If this is possible, then the controller enforces a non-symmetric control putting more effort in some regions and will break the symmetry of the system. The other way around, does there exists a partition of $\mathbb{T}^d$ into M regions such that a non-symmetric control achieves a better performance ? Indeed, in many applications, the controller may be free to choose to break the system into a few regions.

For example, assume that the controller may divide the system into up to M regions $\{R_i\}_{1\leq i\leq M}$. If $\int_{\mathbb{T}^d} p_1^{(M)}(z)dz > p_1^{(1)}(0) = \alpha_0\varphi(\alpha_0 \int_{\mathbb{T}^d} I(\xi)d\xi)$, the optimal control will break the symmetry of the system (at least for β small enough). This will happen if there exists a partition into M regions $\{R_i\}_{1\leq i\leq M}$, $(u_i)_{1\leq i\leq M} \in \mathbb{R}_+^M$, such that $\sum_{i=1}^M u_i|R_i| = 1$ and

$$\int_{\mathbb{T}^d} \varphi\left(\alpha_0 \sum_{i=1}^M u_i \int_{R_i} I(z-\xi)d\xi\right) dz > \varphi\left(\alpha_0 \int_{\mathbb{T}^d} I(\xi)d\xi\right).$$

If φ is linear, then we may check easily that the right hand side always equals the left hand side. Then the general answer is no: for linear φ the optimal control does not break the symmetry of the system, irrespectively of the influence function I. Notice that a linear function φ will arise in pair-wise interaction.

If φ is non-linear, the expected answer is yes. Indeed, for any partition $\{R_i\}_{1\leq i\leq M}$,

$$\sup_{\substack{(u_i)_{1\leq i\leq M}\in\mathbb{R}_+^M,\\ \sum_{i=1}^M u_i|R_i|=1}} \int_{\mathbb{T}^d} \varphi\left(\alpha_0 \sum_{i=1}^M u_i \int_{R_i} I(z-\xi)d\xi\right) dz,$$

will not in general be reached for $u_i|R_i| = 1/M$. For example if φ is convex, from Jensen's inequality, for any partition $\{R_i\}_{1\leq i\leq M}$ and (u_i) as above:

$$\int_{\mathbb{T}^d} \varphi\left(\alpha_0 \sum_{i=1}^M u_i \int_{R_i} I(z-\xi)d\xi\right) dz \geq \varphi\left(\alpha_0 \int_{\mathbb{T}^d} I(\xi)d\xi\right),$$

and the symmetric control is the worst possible control ! The optimal control lies on the boundary of the constraint set:

$$\sup_{\substack{(u_i)_{1\leq i\leq M}\in\mathbb{R}_+^M,\\ \sum_{i=1}^M u_i|R_i|=1}} \int_{\mathbb{T}^d} \varphi\left(\alpha_0 \sum_{i=1}^M u_i \int_{R_i} I(z-\xi)d\xi\right) dz = \max_{1\leq i\leq M} \int_{\mathbb{T}^d} \varphi\left(\frac{\alpha_0}{|R_i|} \int_{R_i} I(z-\xi)d\xi\right) dz.$$

Convex and non-linear functions φ will naturally appear in k-particle interactions with $k \geq 3$.

Acknowledgment

Research of the authors was supported by NSF grants CCF-0500023, CCF-0635372, and CNS-062716.

References

[1] A. Arapostathis, V. Borkar, E. Fernández-Gaucherand, M. Ghosh, and S. Marcus. Discrete-time controlled Markov processes with average cost criterion: a survey. *SIAM J. Control Optim.*, 31(2):282–344, 1993.

[2] D. Bertsekas. *Dynamic programming and optimal control. Vol. II.* Athena Scientific, Belmont, MA, second edition, 2001.

[3] D. Bertsekas. *Dynamic programming and optimal control. Vol. I.* Athena Scientific, Belmont, MA, third edition, 2005.

[4] D. Blackwell. Discrete dynamic programming. *Ann. Math. Statist.*, 33:719–726, 1962.

[5] E. B. Dynkin and A. A. Yushkevich. *Controlled Markov processes*, volume 235 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1979.

[6] S. Ethier and T. Kurtz. *Markov processes*. Wiley, New York, 1986.

[7] E. Feinberg and A. Shwartz, editors. *Handbook of Markov decision processes*. International Series in Operations Research & Management Science, 40. Kluwer Academic Publishers, Boston, MA, 2002. Methods and applications.

[8] C. Graham. Chaoticity on path space for a queueing network with selection of the shortest queue among several. *J. Appl. Prob.*, 37:198–211, 2000.

[9] R. Sznajder and J. A. Filar. Some comments on a theorem of Hardy and Littlewood. *J. Optim. Theory Appl.*, 75(1):201–208, 1992.

[10] A.S. Sznitman. *Propagation of chaos, in Ecole d'été de probabilités de Saint-Flour XIX*. lecture notes in Maths 1464. Springer, Berlin, 1991.