Optimal Mean field Limits: From discrete to continuous optimization

Nicolas Gast, Bruno Gaujal and Jean-Yves Le Boudec

Grenoble University INRIA EPFL

Warwick – May, 2012

Outline

Empirical Measure and Control

We consider a system composed of N *objects*. Each object has a state from the finite set $\mathcal{S} = \{1 \dots S\}$. Time is discrete and the state of the object n at step $k \in \mathbb{N}$ is denoted $X_n^N(k)$. The actions of the central controller form a compact metric space.

$M^N(k)$ is the empirical measure of the objects $\left(X_1^N(k) \dots X_N^N(k)\right)$ at time k:

$$M_n^N(k) \stackrel{\text{def}}{=} \frac{1}{N} \sum_{n=1}^{N} \delta_{X_n^N(k)}, \qquad (1)$$

We assume that

(A0) Objects are observable only through their states

A direct consequence is:

Theorem

- *For any given sequence of actions, the process $M^N(t)$ is a Markov chain*
- *There exists an optimal policy $\pi = (\pi_0, \pi_1, \dots, \pi_k, \dots)$ where π_k is a deterministic function $\mathcal{P}(\mathcal{S}) \to \mathcal{A}$.*

Value function

The controller focusses on a finite-time horizon $[0; H^N]$. If the system has an occupancy measure $M^N(k)$ at time step $k \in [0; H^N]$ and if the controller chooses the action $A^N(k)$, she gets an *instantaneous reward* $r^N(M^N(k), A^N(k))$. At time H^N, she gets a *final reward* $r_f(M^N(H^N))$. The value of a policy π is the expected gain over the horizon $[0; H^N]$ starting from m_0 when applying the policy π. It is defined by

$$V_\pi^N(m) \stackrel{\text{def}}{=} \mathbb{E}\Big(\sum_{k=0}^{H^N-1} r^N(M_\pi^N(k), \pi(M_\pi^N(k))) + r_f(M_\pi^N(H^N)) \Big| M_\pi^N(0) = m \Big). \tag{2}$$

The goal of the controller is to find an optimal policy that maximizes the value. We denote by $V_*^N(m)$ the optimal value when starting from m:

$$V_*^N(m) = \sup_\pi V_\pi^N(m) \tag{3}$$

Scaling Time and Space

The drift

$$F^N(m,a) \stackrel{\text{def}}{=} \mathbb{E}\big(M^N(k+1) - M^N(k) \\ |M^N(k) = m, A^N(k) = a\big). \quad (4)$$

goes to 0 at speed $I(N)$ when N goes to infinity and $F^N/I(N)$ converges to a Lipschitz continuous function f.

We define the continuous time process $(\hat{M}^N(t))_{t\in\mathbb{R}^+}$ as the affine interpolation of $M^N(k)$, rescaled by the intensity function, i.e. $\hat{M}^N$ is affine on the intervals $[kI(N), (k+1)I(N)]$, $k \in \mathbb{N}$ and

$$\hat{M}^N(kI(N)) = M^N(k).$$

We assume that the time horizon and the reward per time slot scale accordingly, i.e. we impose

$$\begin{aligned} H^N &= \left\lfloor \frac{T}{I(N)} \right\rfloor \\ r^N(m,a) &= I(N)r(m,a) \end{aligned}$$

Mean Field Limit

An action function $\alpha : [0; T] \to \mathcal{A}$ is a piecewise Lipschitz continuous function that associates to each time t an action $\alpha(t)$. For an action function α and an initial condition m_0, we consider the following ordinary integral equation for $m(t)$, $t \in \mathbb{R}^+$:

$$m(t) - m(0) = \int_0^t f(m(s), \alpha(s)) ds. \quad (5)$$

We call ϕ_t, $t \in \mathbb{R}^+$, the corresponding semi-flow: the unique solution of Eq.(5) is

$$m(t) = \phi_t(m_0, \alpha). \quad (6)$$

Its value is

$$v_\alpha(m_0) \stackrel{\text{def}}{=} \int_0^T r\left(\phi_s(m_0, \alpha), \alpha(s)\right) ds + r_f(\phi_T(m_0, \alpha)).$$

We also define the optimal value of the deterministic limit $v_*(m_0)$:

$$v_*(m_0) = \sup_\alpha v_\alpha(m_0),$$

Technical Assumptions

(A1) (Transition probabilities) the number of objects changing at time k satisfies

$$\begin{aligned} \mathbb{E}\left(\Delta_\pi^N(k) \middle| M_\pi^N(k) = m\right) &\leq NI_1(N) \\ \mathbb{E}\left(\Delta_\pi^N(k)^2 \middle| M_\pi^N(k) = m\right) &\leq N^2 I(N) I_2(N) \end{aligned}$$

(A2) (Convergence of the Drift) f bounded on $\mathcal{P}(\mathcal{S}) \times \mathcal{A}$ and $\lim_{N\to\infty} I(N) = \lim_{N\to\infty} I_0(N) = 0$ such that $\left\| \frac{1}{I(N)} F^N(m,a) - f(m,a) \right\| \leq I_0(N)$

(A3) (Lipschitz Continuity) $F^N, (f), r$ are Lipschitz continuous in m and (a).

Technical Assumptions(II)

To make things more concrete, here is a simple but useful case where all assumptions are true.

- There are constants c_1 and c_2 such that the expectation of the number of objects that perform a transition in one time slot is $\leq c_1$ and its standard deviation is $\leq c_2$,
- and $F^N(m, a)$ can be written under the form $\frac{1}{N}\varphi(m, a, 1/N)$ where φ is a continuous function on $\Delta_S \times \mathcal{A} \times [0, \epsilon)$ for some neighborhood Δ_S of $\mathcal{P}(\mathcal{S})$ and some $\epsilon > 0$, continuously differentiable with respect to m.

In this case we can choose $I(N) = 1/N$, $I_0(N) = c_0/N$ (where c_0 is an upper bound to the norm of the differential $\frac{\partial\varphi}{\partial m}$), $I_1(N) = c_1/N$ and $I_2(N) = (c_1^2 + c_2^2)/N$.

Main results(I)

Theorem (1: Convergence for action functions)

Under (A0-A3), let α is a piecewise Lipschitz continuous action function on $[0; T]$, of constant K_α, with p jumps. Let $\hat{M}^N_\alpha(t)$ be the linear interpolation of the discrete time process M^N_α. Then for all $\epsilon > 0$:

$$\mathbb{P}\left\{ \sup_{0 \le t \le T} \left\| \hat{M}^N_\alpha(t) - \phi_t(m_0, \alpha) \right\| > \left[\left\| M^N(0) - m_0 \right\| + I'_0(N, \alpha) T + \epsilon \right] e^{L_1 T} \right\} \le \frac{J(N, T)}{\epsilon^2} \quad (7)$$

and

$$\left| V^N_\alpha \left(M^N(0) \right) - v_\alpha(m_0) \right| \le B' \left(N, \left\| M^N(0) - m_0 \right\| \right) \quad (8)$$

where J, I'_0 and B' are constants and satisfy $\lim_{N\to\infty} I'_0(N, \alpha) = \lim_{N\to\infty} J(N, T) = 0$ and $\lim_{N\to\infty, \delta\to 0} B'(N, \delta) = 0$. In particular, if $\lim_{N\to\infty} M^N_\pi(0) = m_0$ almost surely [resp. in probability] then $\lim_{N\to\infty} V^N_\alpha \left(M^N(0) \right) = v_\alpha(m_0)$ almost surely [resp. in probability].

Main results (II)

Consider the system with N objects under policy π. The process M_π^N is defined on some probability space Ω. To each $\omega \in \Omega$ corresponds a trajectory $M_\pi^N(\omega)$, and for each $\omega \in \Omega$, we define an action function $A_\pi^N(\omega)$.

Theorem (2: Uniform convergence of the value)

Let A_π^N be the random action function associated with M_π^N, as defined earlier. Under Assumptions (A0) to (A3),

$$\left| V_\pi^N \left(M^N(0) \right) - \mathbb{E} \left[v_{A_\pi^N}(m_0) \right] \right| \leq B\left(N, \left\| M^N(0) - m_0 \right\| \right)$$

where B is such that $\lim_{N\to\infty,\delta\to 0} B(N,\delta) = 0$; in particular, if $\lim_{N\to\infty} M_\pi^N(0) = m_0$ almost surely [resp. in probability] then $\left| V_\pi^N \left(M^N(0) \right) - \mathbb{E} \left[v_{A_\pi^N}(m_0) \right] \right| \to 0$ almost surely [resp. in probability].

Main results(III)

Corollary (Asymptotically Optimal Policy)

If α_* *is an optimal action function for the limiting system and if* $\lim_{N\to\infty} M^N(0) = m_0$ *almost surely [resp. in probability], then we have:*

$$\lim_{N\to\infty} \left|V^N_{\alpha_*} - V^N_*\right| = \left|V^N_* - v_*\right| = 0,$$

almost surely [resp. in probability].

In other words, an optimal action function for the limiting system is asymptotically optimal for the system with N objects.

Main ingredient of the proof: coupling

Consider the system with N objects under policy π. The process M_π^N is defined on some probability space Ω. To each $\omega \in \Omega$ corresponds a trajectory $M_\pi^N(\omega)$, and for each $\omega \in \Omega$, we define an action function $A_\pi^N(\omega)$. This random function is piecewise constant on each interval $[kI(N), (k+1)I(N))$ $(k \in \mathbb{N})$ and is such that $A_\pi^N(\omega)(kI(N)) \stackrel{\text{def}}{=} \pi_k(M^N(k))$ is the action taken by the controller of the system with N objects at time slot k, under policy π. For every ω, $\phi_t(m_0, A_\pi^N(\omega))$ is the solution of the limiting system with action function $A_\pi^N(\omega)$, i.e.

$$\phi_t(m_0, A_\pi^N(\omega)) = m_0 + \int_0^t f(\phi_s(m_0, A_\pi^N(\omega)), A_\pi^N(\omega)(s))ds.$$

Main ingredient of the proof: coupling (II)

Let $\epsilon > 0$ and $\alpha(.)$ be an action function such that $v_\alpha(m_0) \geq v_*(m_0) - \epsilon$ Th. 1 shows that $\lim_{N\to\infty} V_\alpha^N(M^N(0)) = v_\alpha(m_0) \geq v_*(m_0) - \epsilon$ a.s. This shows that $\liminf_{N\to\infty} V_*^N(M^N(0)) \geq \lim_{N\to\infty} V_\alpha^N(M^N(0)) \geq v_*(m_0) - \epsilon$; this holds for every $\epsilon > 0$ thus $\liminf_{N\to\infty} V_*^N(M^N(0)) \geq v_*(m_0)$ a.s.

Now, let $B(N, \delta)$ be as in Th. 2 , $\epsilon > 0$ and π^N such that
$V_*^N(M^N(0)) \leq V_{\pi^N}^N(M^N(0)) + \epsilon$.
$V_{\pi^N}^N(M^N(0)) \leq \mathbb{E}\left(v_{A_{\pi^N}^N}(m_0)\right) + B(N, \delta^N) \leq v_*(m_0) + B(N, \delta^N)$ where $\delta^N \stackrel{\text{def}}{=} \left\|M^N(0) - m_0\right\|$. Thus $V_*^N(M^N(0)) \leq v_*(m_0) + B(N, \delta^N) + \epsilon$. If further $\delta^N \to 0$ a.s. it follows that $\limsup_{N\to\infty} V_*^N(M^N(0)) \leq v_*(m_0) + \epsilon$ a.s. for every $\epsilon > 0$, thus $\limsup_{N\to\infty} V_*^N(M^N(0)) \leq v_*(m_0)$ a.s.

Infinite horizon with discounted costs

Under a policy π, the expected discounted value starting from $M^N(0) = m$ is:

$$W_\pi^N(m) = \mathbb{E}\left(\sum_{k=0}^{\infty} \delta^{kI(N)} r(M_\pi^N(k), \pi_k(M_\pi^N(k))) \middle| M_\pi^N(0) = m\right)$$

Similarly, the discounted cost can be defined for the infinite system:

$$w_\alpha(m) = \int_0^\infty \delta^s r\left(\phi_s(m, \alpha), \alpha(s)\right) ds.$$

Theorem

Under hypothesis (A1,A2,A3) and if $M_\pi^N(0) \xrightarrow{\mathcal{P}} m_0$, then:

$$\lim_{N\to\infty} W_*^N\left(M_\pi^N(0)\right) = \sup_\pi W_\pi^N\left(M_\pi^N(0)\right) = \sup_\alpha w_\alpha(m) = w_*(m_0)$$

HJB Equation and Dynamic Programming

The optimal value can be computed by a discrete dynamic programming algorithm by setting $U^N(m,T) = r_f(m)$ and

$$U^N(m,t) = \sup_{a\in\mathcal{A}} \mathbb{E}\Big[r^N(m,a) + U^N(M^N(t+I(N)), t+I(N)) \Big| \bar{M}^N(t) = m, A^N(t) = a\Big].$$

Then, the optimal cost over horizon $[0; T/I(N)]$ is $V_*^N(m) = U(m,0)$. Similarly, if we denote by $u(m,t)$ the optimal cost over horizon $[t;T]$ for the limiting system, $u(m,t)$ satisfies the classical Hamilton-Jacobi-Bellman equation:

$$\frac{\partial u(m,t)}{\partial t} + \max_a \{\nabla u(m,t).f(m,a) + r(m,a)\} = 0. \qquad (9)$$

Algorithm

– From the original system with N objects, construct the occupancy measure M^N and its kernel Γ^N and let $M^N(0)$ be the initial occupancy measure;
– Compute the limit f of the drift of Γ^N;
Solve the HJB equation (9) on $[0, HI(N)]$. This provides an optimal control function $\alpha_*(M_0^N, t)$;
– Construct a discrete control π for the discrete system: the action to be taken under state $M^N(k)$ at step k is

$$\pi(M^N(k), k) \stackrel{\text{def}}{=} \alpha_*(\phi_{kI(N)}(M^N(0), \alpha)).$$

– Return π

Algorithm 2

The policy π constructed by Algorithm 1 is static in the sense that it does not depend on the state $M^N(k)$ but only on the initial state $M^N(0)$, and the deterministic estimation of $M^N(k)$ provided by the differential equation. One can construct a more adaptive policy by updating the starting point of the differential equation at each step.

– $M := M^N(0)$; $k := 0$

– Repeat until $k = H$

    $\alpha_k^*(M, \cdot) :=$ solution of HJB over $[kI(N), HI(N)]$ starting in M

    $\pi'(M, k) := \alpha_k^*(\phi_{kI(N)}(M, \alpha_k))$

    M is changed by applying kernel $\Gamma_{\pi'}^N$

    k:= k+1

– Return π'

Infection Strategy of a Viral Worm

A *susceptible* (S) node is a mobile wireless device, not contaminated by the worm but prone to infection. A node is *infective* (I) if it is contaminated by the worm. An infective node spreads the worm to a susceptible node whenever they meet, with probability β. The worm can also choose to kill an infective node, i.e., render it completely dysfunctional - such nodes are denoted *dead* (D). A functional node that is immune to the worm is referred to as *recovered* (R).
The goal of the worm is to maximize the damages done to the network by choosing the rate $\alpha(t)$ at which it kills node at time t.

$$\mathbb{E}\left(D_\pi(T) + \frac{1}{NT}\sum_{k=1}^{NT} g(I_\pi(k))\right).$$

Infection Strategy of a Viral Worm (II)

the dynamics of this population process converges to the solution of the following differential equations.

$$
\begin{aligned}
\frac{dS}{dt} &= -\beta IS - qS \\
\frac{dI}{dt} &= \beta IS - bI - \alpha(t)I \\
\frac{dD}{dt} &= \alpha(t)I \\
\frac{dR}{dt} &= bI + qS,
\end{aligned} \tag{10}
$$

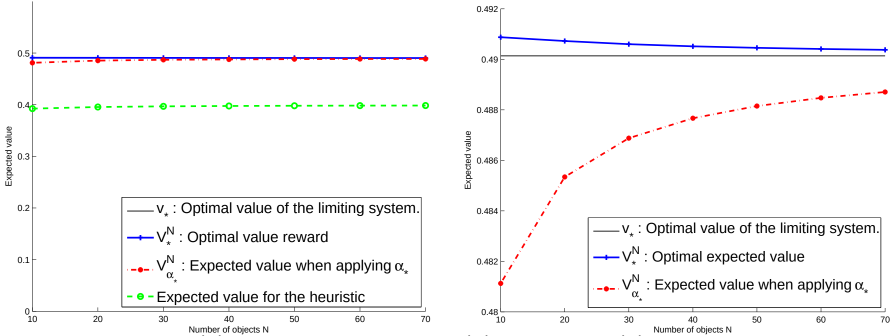
where $\alpha(t)$ is the action taken by the worm at time t.

Infection Strategy of a Viral Worm (III)

In the continuous control problem, the objective of the worm is to find an action function α such that the damage function $D(T) + \frac{1}{T}\int_0^T g(I(t))dt$ is maximized under the constraint $0 \leq \alpha(t) \leq \alpha_{\max}$ (where f is convex). In [Khousani, Sarkar, Altamn, 2010], this problem is shown to have a solution and the Pontryagin maximum principle is used to show that the optimal action function α_* is of bang-bang type: there exists $t_1 \in [0 \ldots T)$ s.t.

$$\alpha_*(t) = \begin{cases} 0 & \text{for } 0 < t < t_1 \\ \alpha_{\max} & \text{for } t_1 < t < T \end{cases} \qquad (11)$$

Infection Strategy of a Viral Worm (III)



(a)

(b) Same as (a) with $y-$axis zoomed around 0.49

Figure: Damage caused by the worm for various infection policies as a function of the size of the system N.

Utility provider pricing

We consider a system made of a utility and N users; users can be either in state S (subscribed) or U (unsubscribed). The utility fixes their price $\alpha \in [0, 1]$.
Each customer revises her status independently. If she is in state U [resp. S], with probability $s(\alpha)$ [resp. $a(\alpha)$] she moves to the other state; $s(\alpha)$ is the probability of a new subscription, and $a(\alpha)$ is the probability of attrition.

An equivalent model is that at every time step (which size decreases as $1/N$), one customer is chosen randomly

Utility provider pricing (II)

This problem can be seen as a Markovian system made of N objects (users) and one controller (the provider). The intensity is $I(N) = 1/N$. if $x(t)$ is the fraction of objects in state S at time t and $\alpha(t) \in [0;1]$ is the action taken by the provider at time t, the mean field limit of the system is:

$$\begin{aligned} \frac{dx}{dt} &= -x(t)a(\alpha(t)) + (1 - x(t))s(\alpha(t)) \\ &= s(\alpha(t)) - x(s(\alpha(t)) + a(\alpha(t)) \end{aligned} \quad (12)$$

and the rescaled profit over a time horizon T is $\int_0^T x(t)\alpha(t)dt$. Call $u_*(t,x)$ the optimal benefit over the interval $[t, T]$ if there is a proportion x of subscribers at time t. The Hamilton-Jaccobi-Bellman equation is

$$\frac{\partial}{\partial t}u_*(t,x) + H\left(x, \frac{\partial}{\partial x}u_*(t,x)\right) = 0 \quad (13)$$

with

$$H(x,p) = \max_{\alpha \in [0,1]} \left[p(s(\alpha) - x(s(\alpha) + a(\alpha)) + \alpha x\right]$$

Utility provider pricing (III)

Consider the case where $\alpha \in \{0, 1\}$ and $s(0) = a(1) = 1$ and $s(1) = a(0) = 0$. The ODE becomes

$$\frac{dx}{dt} = -x(t)\alpha(t) + (1 - x(t))(1 - \alpha(t)) = 1 - x(t) - \alpha(t), \qquad (14)$$

and $H(x, p) = \max(x(1-p), (1-x)p)$. The optimal policy is $\alpha = 1$ if $x > 1/2$ or $x > 1 - \exp(-(T - t))$, and 0 otherwise.