MEAN FIELD FOR MARKOV DECISION PROCESSES: FROM DISCRETE TO CONTINUOUS OPTIMIZATION

Jean-Yves Le Boudec,
Nicolas Gast,
Bruno Gaujal
July 26, 2011
1. Mean Field Interaction Model

2. Mean Field Interaction Model with Central Control

3. Convergence and Asymptotically Optimal Policy
1

MEAN FIELD INTERACTION MODEL
Mean Field Interaction Model

- Time is discrete
- $N$ objects, $N$ large
- Object $n$ has state $X_n(t)$
- $(X_1^N(t), \ldots, X_N^N(t))$ is Markov
- Objects are observable only through their state

“Occupancy measure” $M^N(t) = \text{distribution of object states at time } t$

Example [Khouzani 2010]:
$M^N(t) = (S(t), I(t), R(t), D(t))$
with
$S(t) + I(t) + R(t) + D(t) = 1$
$S(t) = \text{proportion of nodes in state `S’}$

![Diagram](image)
Mean Field Interaction Model

- Time is discrete
- $N$ objects, $N$ large
- Object $n$ has state $X_n(t)$
- $(X_1^N(t), \ldots, X_N^N(t))$ is Markov
- Objects are observable only through their state

- "Occupancy measure" $M^N(t) = \text{distribution of object states at time } t$

- Theorem [Gast (2011)]
  $M^N(t)$ is Markov

- Called "Mean Field Interaction Models" in the Performance Evaluation community
  [McDonald(2007), Benaïm and Le Boudec(2008)]
Intensity $I(N)$

- $I(N) = \text{expected number of transitions per object per time unit}$

- A mean field limit occurs when we re-scale time by $I(N)$
i.e. we consider $X^N(t/I(N))$

- $I(N) = O(1)$: mean field limit is in discrete time
  [Le Boudec et al (2007)]

- $I(N) = O(1/N)$: mean field limit is in continuous time
  [Benaïm and Le Boudec (2008)]
Virus Infection [Khouzani 2010]

- \( N \) nodes, homogeneous, pairwise meetings
- One interaction per time slot, \( I(N) = 1/N \); mean field limit is an ODE
- Occupancy measure is \( M(t) = (S(t), I(t), R(t), D(t)) \) with \( S(t) + I(t) + R(t) + D(t) = 1 \)
  \( S(t) = \) proportion of nodes in state ‘S’

Mean field limit

\[ \alpha = 0.1 \]

\[ \alpha = 0.7 \]

\( N = 100, q = b = 0.1, \beta = 0.6 \)

\( S + R \) or \( S \)

dead nodes
The Mean Field Limit

Under very general conditions (given later) the occupancy measure converges, in law, to a deterministic process, \( m(t) \), called the mean field limit

\[
M^N \left( \frac{t}{I(N)} \right) \to m(t)
\]

Finite State Space => ODE
Sufficient Conditions for Convergence

- [Kurtz 1970], see also [Bordenav et al 2008], [Graham 2000]
- Sufficient condition verifiable by inspection:

[Benaïm and Le Boudec(2008), Ioannidis and Marbach(2009)]

- Let $W^N(k)$ be the number of objects that do a transition in time slot $k$. Note that $\mathbb{E} (W^N(k)) = NI(N)$, where $I(N) \overset{\text{def.}}{=} \text{intensity}$. Assume

$$\mathbb{E} \left( W^N(k)^2 \right) \leq \beta(N) \quad \text{with} \quad \lim_{N \to \infty} I(N) \beta(N) = 0$$

Example: $I(N) = 1/N$
- Second moment of number of objects affected in one timeslot $= o(N)$

- Similar result when mean field limit is in discrete time [Le Boudec et al 2007]
The Importance of Being Spatial

- Mobile node state = (c, t)
  \( c = 1 \ldots 16 \) (position)
  \( t \in \mathbb{R}^+ \) (age of gossip)

- Time is continuous, \( I(N) = 1 \)

- Occupancy measure is
  \[ F_c(z, t) = \text{proportion of nodes that at location } c \text{ and have age } \leq z \]

[Age of Gossip, Chaintreau et al. (2009)]
2

MEAN FIELD INTERACTION MODEL WITH CENTRAL CONTROL
Central controller

Action state $\text{A}$ (metric, compact)

Running reward depends on state and action

Goal: maximize expected reward over horizon $T$

Policy $\pi$ selects action at every time slot

Optimal policy can be assumed Markovian $(X_{N_1}^N(t), ..., X_{N_N}^N(t))$ -> action

Controller observes only object states

$\Rightarrow \pi$ depends on $M_N^N(t)$ only

$$V_{\pi}^N(m) \overset{\text{def}}{=} \mathbb{E} \left( \sum_{k=0}^{H_N^N} r_N^N \left( M_{\pi}^N(k), \pi(M_{\pi}^N(k)) \right) \bigg| M_{\pi}^N(0) = m \right)$$
**Example**

**Policy** $\pi$: set $\alpha = 1$ when $R + S > \theta$

Value = $\frac{1}{NT} \sum_{k=1}^{NT} D^N(k) \approx D^N(NT')$

$$r^N(S, I, R, D, \pi) = \frac{1}{N} D$$
Optimal Control

Optimal Control Problem

Find a policy $\pi$ that achieves (or approaches) the supremum in

$$V^*_N(m) = \sup_{\pi} V^N_{\pi}(m)$$

$m$ is the initial condition of occupancy measure

- Can be found by iterative methods
- State space explosion (for $m$)
Can We Replace MDP By Mean Field Limit?

- Assume the mean field model converges to fluid limit for every action
  - E.g. mean and std dev of transitions per time slot is $O(1)$

- Can we replace MDP by optimal control of mean field limit?
Controlled ODE

- Mean field limit is an ODE
- Control = action function \( \alpha(t) \)
- Example:

\[
\begin{align*}
    \text{if } t > t_0 \quad & \alpha(t) = 1 \quad \text{else } \alpha(t) = 0 \\
    \frac{\partial S}{\partial t} &= -\beta IS - qS \\
    \frac{\partial I}{\partial t} &= \beta IS - bI - \alpha(t)I \\
    \frac{\partial D}{\partial t} &= \alpha(t)I \\
    \frac{\partial R}{\partial t} &= bI + qS.
\end{align*}
\]

- Goal is to maximize

\[
v_\alpha(m_0) \overset{\text{def}}{=} \int_0^T r(\phi_s(m_0, \alpha), \alpha(s)) \, ds
\]

\[
v_*(m_0) = \sup_{\alpha} v_\alpha(m_0).
\]

\( m_0 \) is initial condition

\( r(S, I, R, D, \alpha) = D \)

- Variants: terminal values, infinite horizon with discount
Optimal Control for Fluid Limit

Optimal function $\alpha(t)$ can be obtained with Pontryagin’s maximum principle or Hamilton Jacobi Bellman equation.

$t_0 = 1$

$t_0 = 5.6$

$t_0 = 25$
3

CONVERGENCE,
ASYMPTOTICALLY OPTIMAL POLICY
Theorem [Gast 2011]
Under reasonable regularity and scaling assumptions:

\[
\lim_{N \to \infty} V^*_N \left( M^N(0) \right) = v_* \left( m_0 \right)
\]

- Optimal value for system with \( N \) objects (MDP)
- Optimal value for fluid limit
**Convergence Theorem**

- **Theorem** [Gast 2011]
  Under reasonable regularity and scaling assumptions:

  \[
  \lim_{N \to \infty} V^N_*(M^N(0)) = v_*(m_0)
  \]

- Does this give us an asymptotically optimal policy?

  Optimal policy of system with \( N \) objects may not converge
Asymptotically Optimal Policy

- Let $\alpha^*$ be an optimal policy for mean field limit

- Define the following control for the system with $N$ objects
  - At time slot $k$, pick same action as optimal fluid limit would take at time $t = k I(N)$

- This defines a time dependent policy.

- Let $V_{\alpha^*}^N = \text{value function when applying } \alpha^* \text{ to system with } N \text{ objects}$

**Theorem** [Gast 2011]

$$\lim_{N \to \infty} |V_{\alpha^*}^N - V_*^N| = 0$$

Optimal value for system with $N$ objects (MDP)

Value of this policy
\[ v_* : \text{Optimal value of the limiting system.} \]
\[ V_*^N : \text{Optimal value reward} \]
\[ V_{\alpha_*}^N : \text{Expected value when applying } \alpha_* \]
\[ \text{Expected value for the heuristic} \]
Conclusions

- Optimal control on mean field limit is justified
- A practical, asymptotically optimal policy can be derived
Questions?


