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 $X_n(t)$
- $(X^N_1(t), ..., X^N_N(t))$ 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 $S(t) + I(t) + R(t) + D(t) = 1$
  $S(t) = \text{proportion of nodes in state ‘S’}$
Mean Field Interaction Model

- Time is discrete
- \( N \) objects, \( N \) large
- Object \( n \) has state \( X_n(t) \)
- \( (X^N_1(t), \ldots, X^N_N(t)) \) is Markov
- Objects are observable only through their state

- “Occupancy measure”
  \( M^N(t) = \) 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'

\[
\begin{align*}
\beta I & \quad \text{beta}\hat{q} \\
q & \quad \text{gamma}
\end{align*}
\]

\[
\begin{align*}
\alpha & = 0.1 \\
\beta & = 0.6 \\
N & = 100, q = b = 0.1
\end{align*}
\]

\[
\begin{align*}
\alpha & = 0.7
\end{align*}
\]
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

$$M^N \left( \frac{t}{I(N)} \right) \rightarrow m(t)$$

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 ... 16 (position)
  t ∈ 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)]

Qqplots simulation vs mean field

no class  16 classes
2

MEAN FIELD INTERACTION MODEL WITH CENTRAL CONTROL
Markov Decision Process

- Central controller
- **Action state** $A$ (metric, compact)
- Running reward depends on state and action
- **Goal**: maximize expected reward over horizon $T$

$\textbf{Policy } \pi \textbf{ selects action at every time slot}$

- Optimal policy can be assumed $\textbf{Markovian}$
  $(X^N_1(t), \ldots, X^N_N(t)) \rightarrow \text{action}$

- Controller observes only object states
  $\Rightarrow \pi$ depends on $M^N(t)$ only

$$V^N_{\pi}(m) \overset{\text{def}}{=} \mathbb{E} \left( \sum_{k=0}^{[NT]} r^N \left( M^N_{\pi}(k), \pi(M^N_{\pi}(k)) \right) \left| M^N_{\pi}(0) = m \right. \right)$$
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:

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

- Goal is to maximize

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

\[
v^*_0(m_0) = \sup_\alpha v_\alpha(m_0)
\]

$m_0$ is initial condition

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

- Variants: terminal values, infinite horizon with discount
Optimal function $\alpha(t)$ can be obtained with Pontryagin’s maximum principle or Hamilton Jacobi Bellman equation.
3

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

$$\lim_{N \to \infty} V^N_* \left( M^N(0) \right) = v_* (m_0)$$

- **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 $\alpha^*$ to system with $N$ 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
$v_\alpha$: Optimal value of the limiting system.
$V^N_\alpha$: Optimal value reward
$V^N_\alpha$: Expected value when applying $\alpha_\star$
Expected value for the heuristic

Number of objects $N$
Conclusions

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


[Khouzani 2010]