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

Jean-Yves Le Boudec,
Nicolas Gast,
Bruno Gaujal

July 26, 2011

INRIA
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Contents

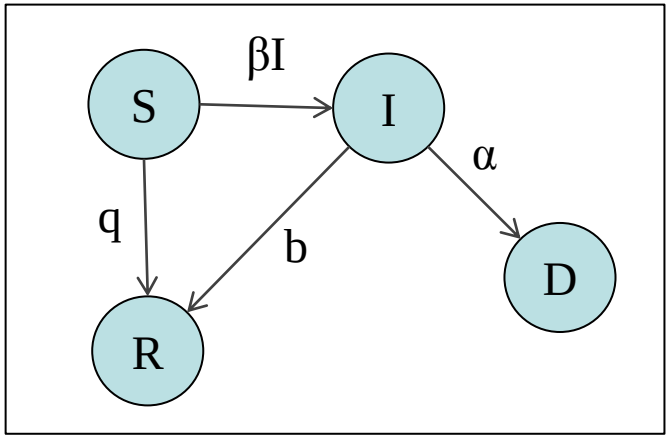
1

MEAN FIELD INTERACTION MODEL

Mean Field Interaction Model

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

- "Occupancy measure" $M^N(t)$ = distribution of object states at time t
- 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) = 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)* = 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 $X^N(t/I(N))$

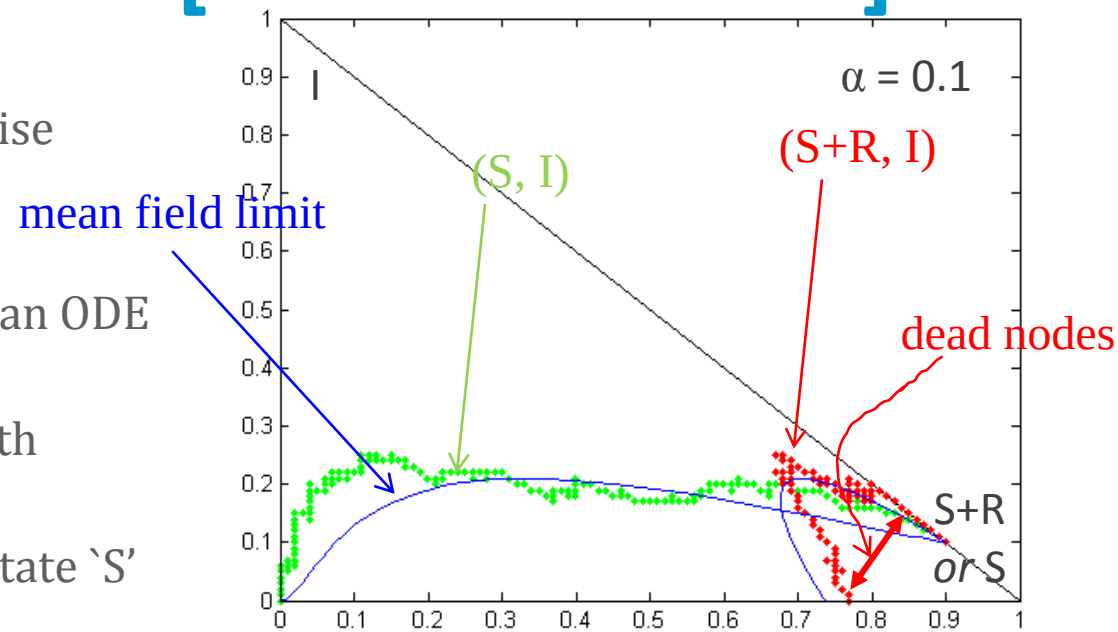
- I(N) = O(1): mean field limit is in discrete time
  [Le Boudec et al (2007)]

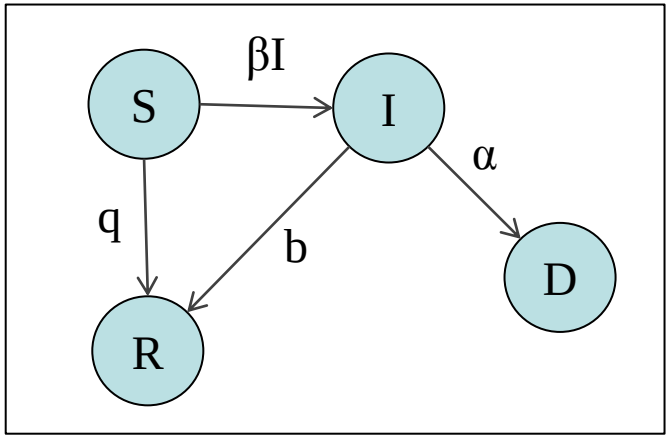
  I(N) = O(1/N): mean field limit is in continuous time
  [Benaïm and Le Boudec (2008)]

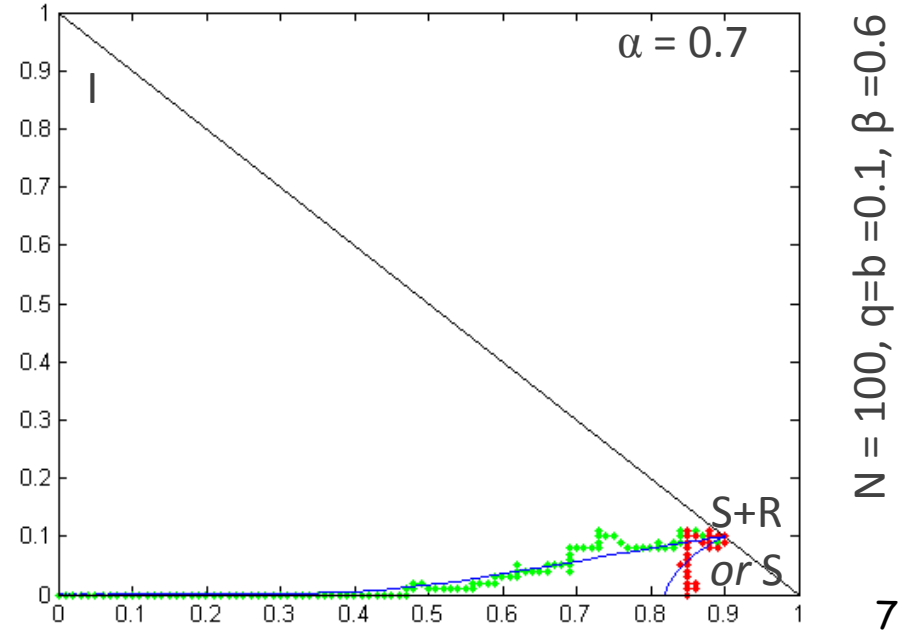
Virus Infection [Khouzani 2010]

- *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'







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*

$$M^N\left(\frac{t}{l(N)}\right) \to m(t)$$

- Finite State Space => ODE

Sufficient Conditions for Convergence

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

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

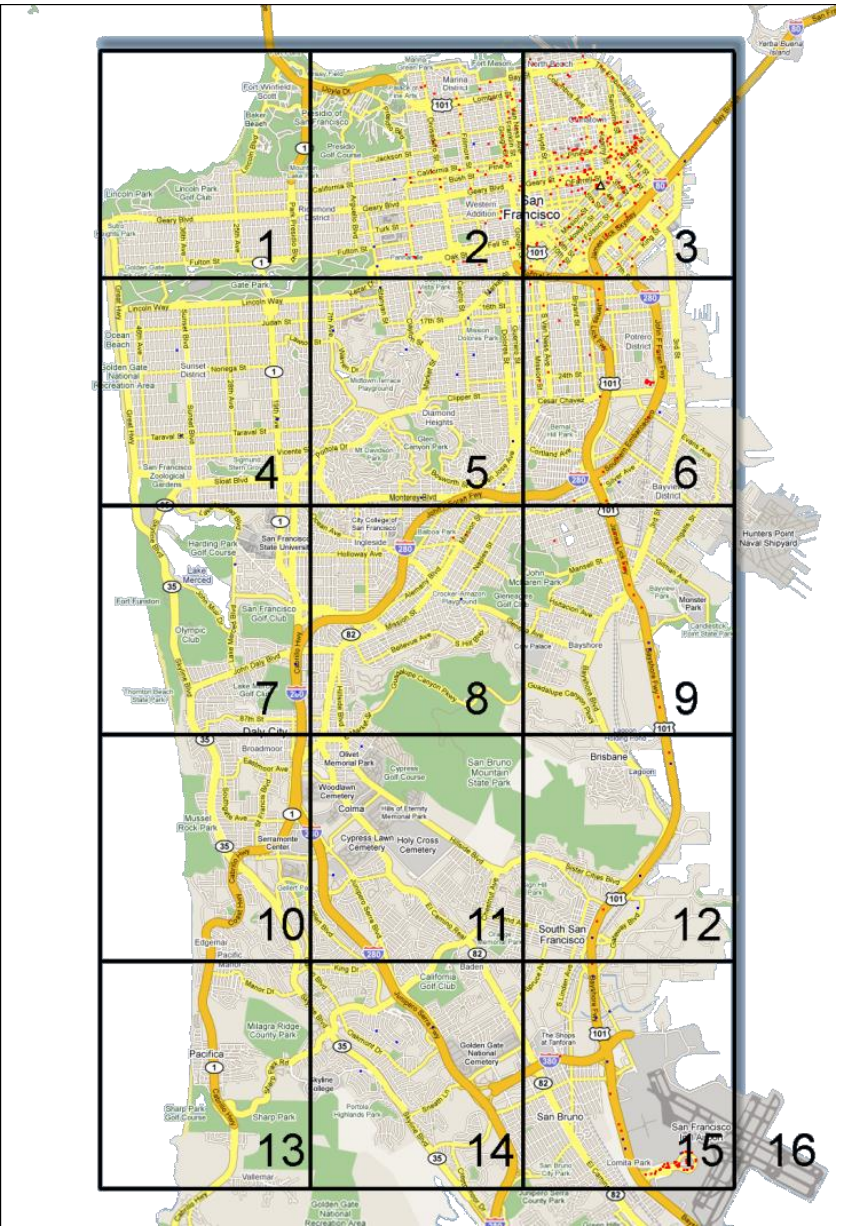
- Let $W^N(k)$ be the number of objects that do a transition in time slot k. Note that $\mathbb{E}\left(W^N(k)\right) = NI(N)$, where $I(N) \stackrel{\text{def}}{=}$intensity. Assume

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

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
  $F_c(z,t)$ = proportion of nodes that at location c and have age ≤ z

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

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

- ***Policy π*** selects action at every time slot
- Optimal policy can be assumed ***Markovian*** $(X^N_1(t), \ldots, X^N_N(t))$ -> *action*
- Controller observes only object states

=> *π* depends on $M^N(t)$ only

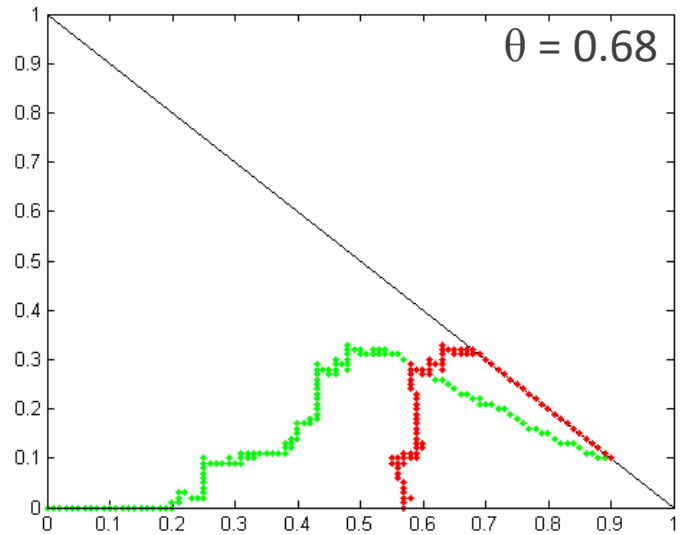
$$V_\pi^N(m) \stackrel{\text{def}}{=} \mathbb{E}\left(\sum_{k=0}^{\lfloor H^N \rfloor} r^N\left(M_\pi^N(k), \pi(M_\pi^N(k))\right) \middle| M_\pi^N(0) = m \right)$$

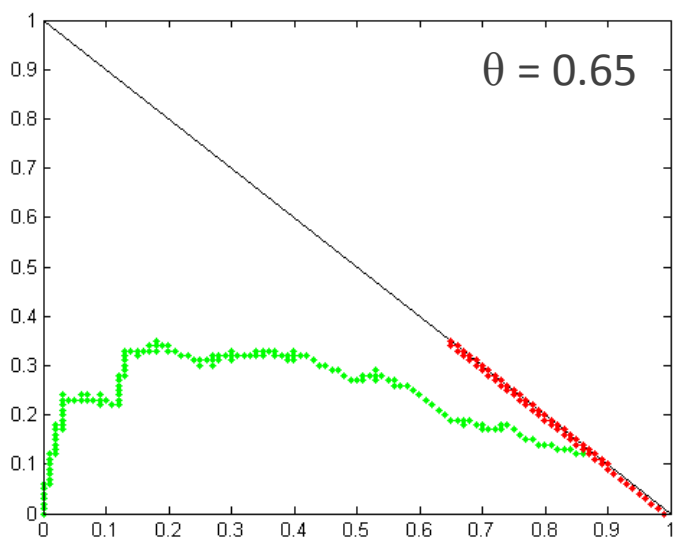
Example

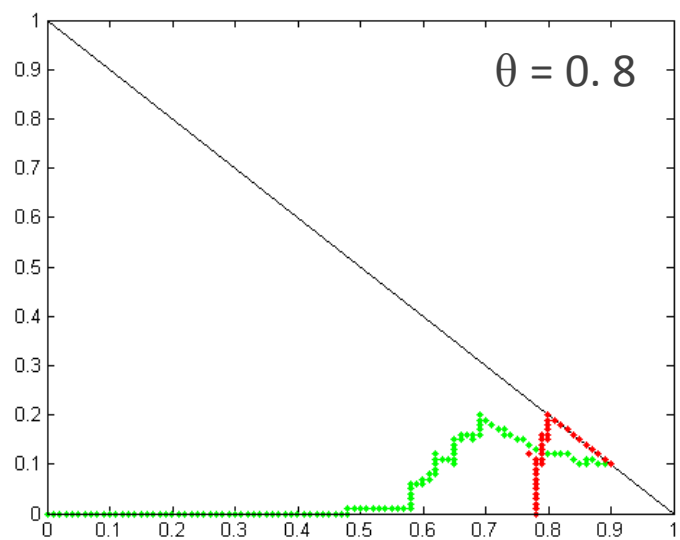
Policy π: set α=1 when R+S > θ

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 $\boldsymbol{\pi}$ that achieves (or approaches) the supremum in

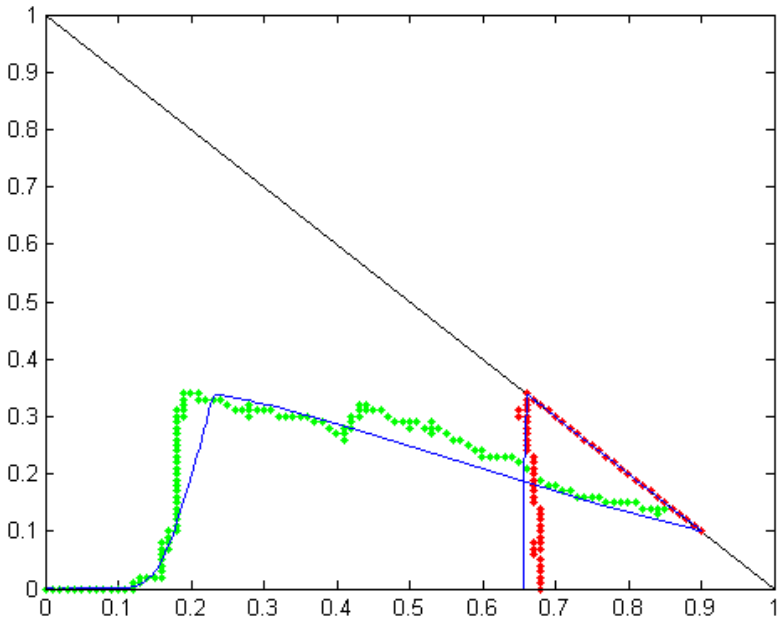
$$V_*^N(m) = \sup_{\pi} V_\pi^N(m)$$

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 α(t)
- Example:

if $t > t_0$ $\alpha(t) = 1$ **else** $\alpha(t) = 0$

$$\begin{aligned}
\frac{\partial S}{\partial t} &= -\beta I S - qS \\
\frac{\partial I}{\partial t} &= \beta I S - bI - \alpha(t) I \\
\frac{\partial D}{\partial t} &= \alpha(t) I \\
\frac{\partial R}{\partial t} &= bI + qS.
\end{aligned}$$

- Goal is to maximize

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

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

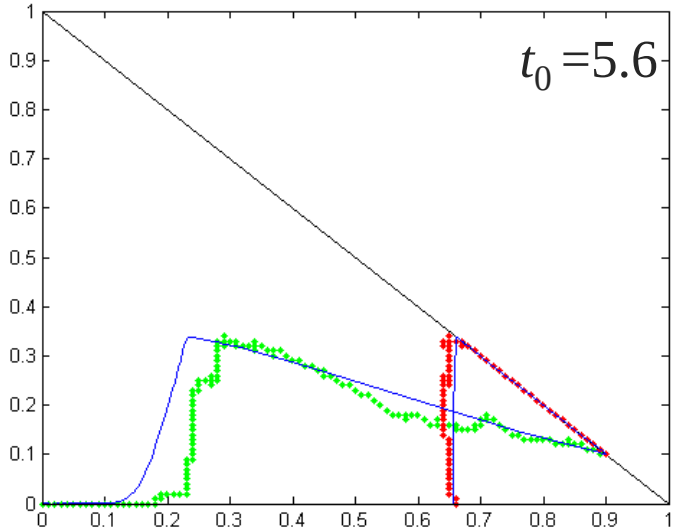
m_0 is initial condition

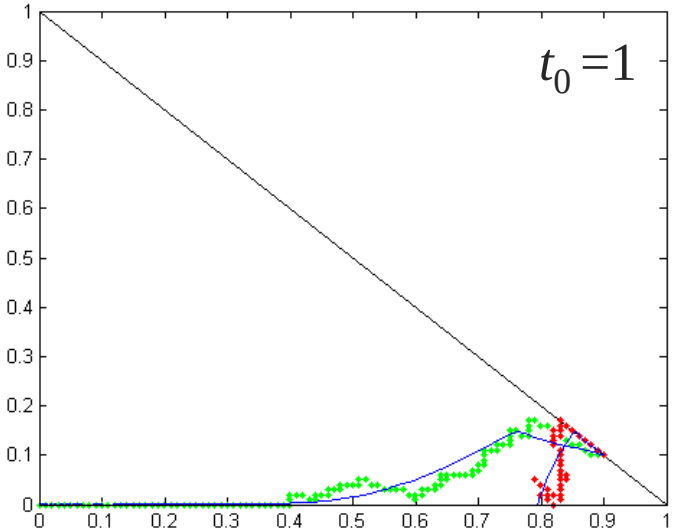
$r(S, I, R, D, \alpha) = D$

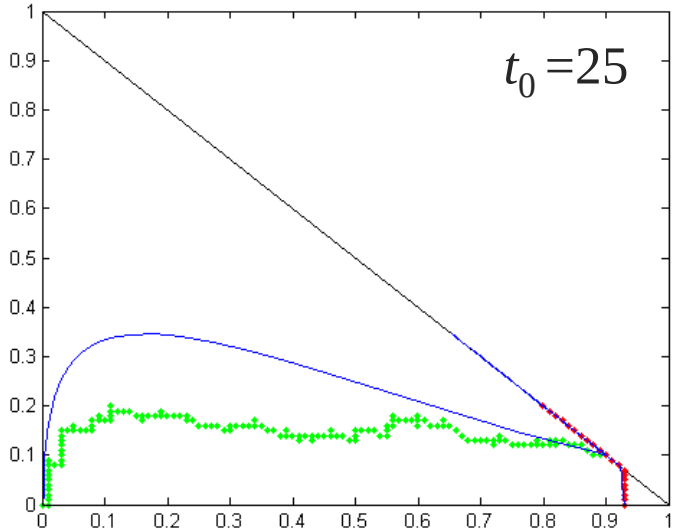
- Variants: terminal values, infinite horizon with discount

Optimal Control for Fluid Limit

- Optimal function α(t) Can be obtained with Pontryagin's maximum principle or Hamilton Jacobi Bellman equation.







3

CONVERGENCE, ASYMPTOTICALLY OPTIMAL POLICY

Convergence Theorem

- ***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

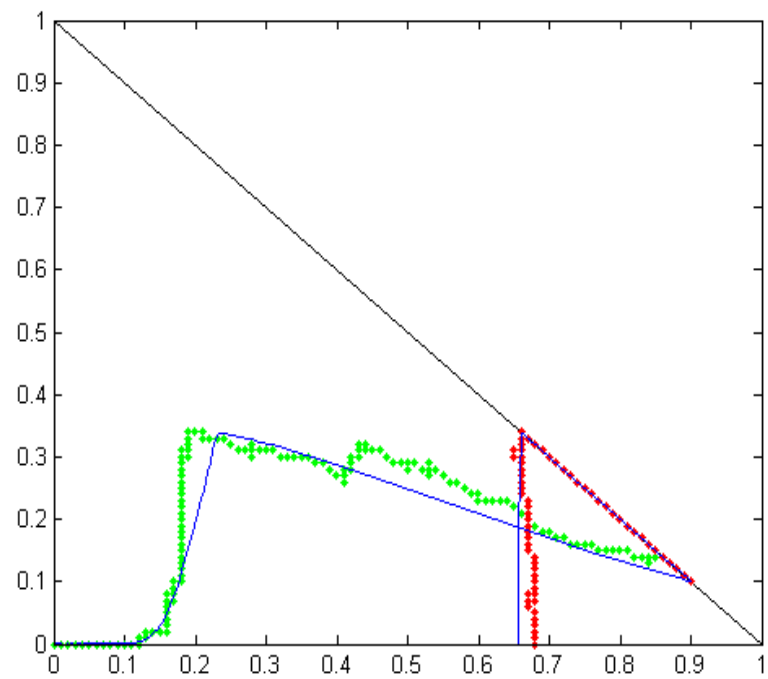
Convergence Theorem

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

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

- Does this give us an asymptotically optimal policy ?

  Optimal policy of system with N objects may not converge



Asymptotically Optimal Policy

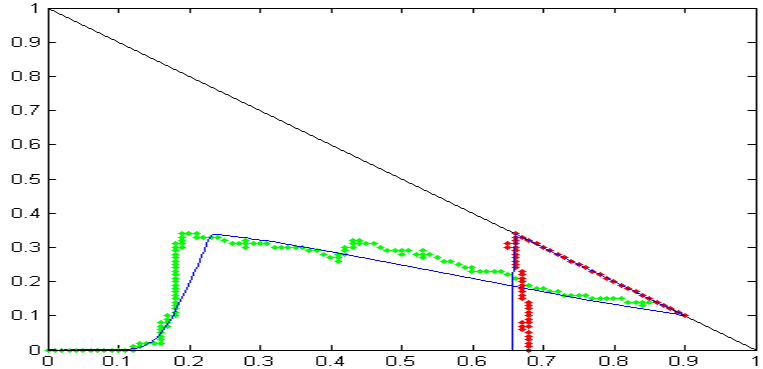
- Let α^* be an optimal policy for mean field limit
- Define the following control for the system with *N* objects
  - At time slot *k*, pick same action as optimal fluid limit would take at time *t = k I(N)*
- This defines a time dependent policy.
- Let $V^N_{\alpha^*}$ = value function when applying α^* to system with *N* objects

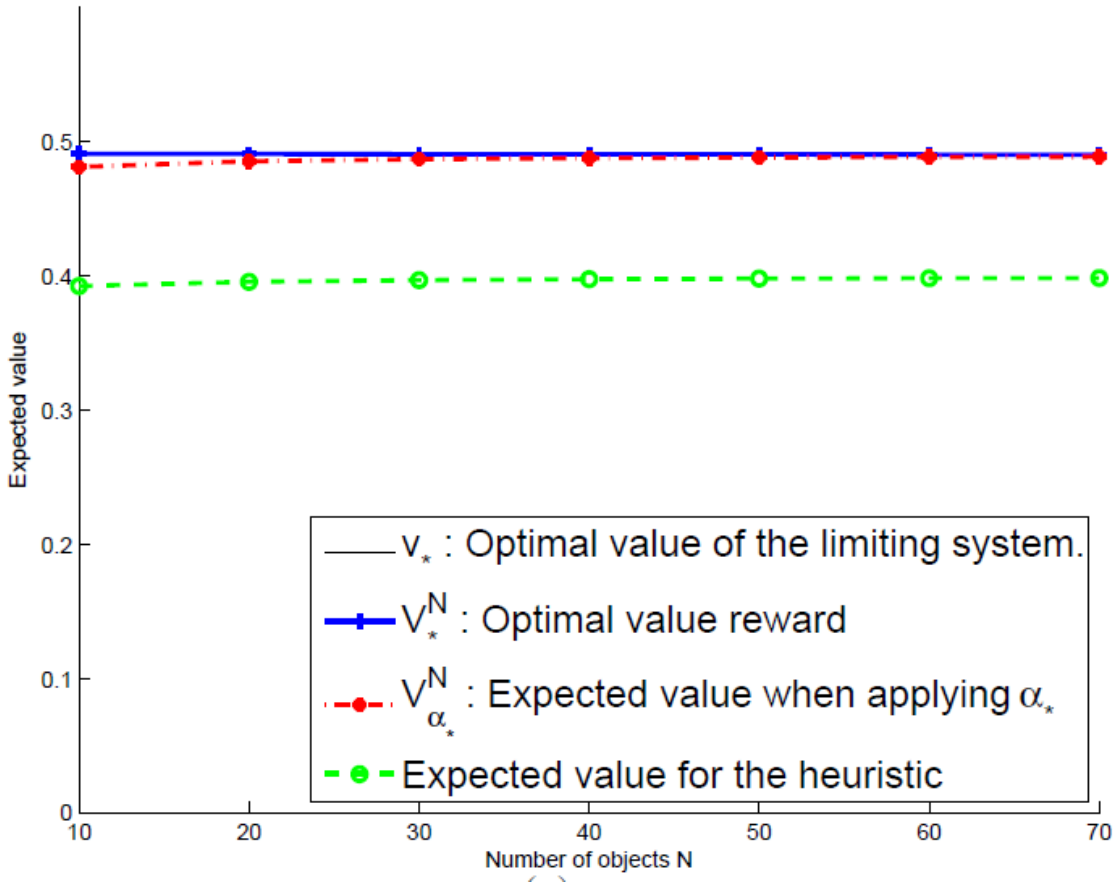
- ***Theorem*** [Gast 2011]

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

Optimal value for system with N objects (MDP)

Value of this policy


0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


Expected value
0
0.1
0.2
0.3
0.4
0.5
10
20
30
40
50
60
70
Number of objects N
v_* : Optimal value of the limiting system.
V_*^N : Optimal value reward
$V_{\alpha_*}^N$: Expected value when applying α_*
Expected value for the heuristic

Conclusions

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

Questions ?

[Gast et al.(2010)Gast, Gaujal, and Le Boudec] Nicolas Gast, Bruno Gaujal, and Jean-Yves Le Boudec. Mean field for Markov Decision Processes: from Discrete to Continuous Optimization. Technical Report arXiv:1004.2342v2, 2010.

[Benaim and Le Boudec(2008)] M. Benaim and J.Y. Le Boudec. A class of mean field interaction models for computer and communication systems. *Performance Evaluation*, 65(11-12):823–838, 2008.

[Bordenave et al.(2007)Bordenave, McDonald, and Proutiere] C. Bordenave, D. McDonald, and A. Proutiere. A particle system in interaction with a rapidly varying environment: Mean field limits and applications. *Arxiv preprint math/0701363*, 2007.

[Ethier and Kurtz(2005)] Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes, Characterization and Convergence*. Wiley, 2005.

[Khouzani 2010]

M.H.R. Khouzani, Saswati Sarkar, and Eitan Altman. Maximum damage malware attack in mobile wireless networks. In *IEEE Infocom*, San Diego, 2010.