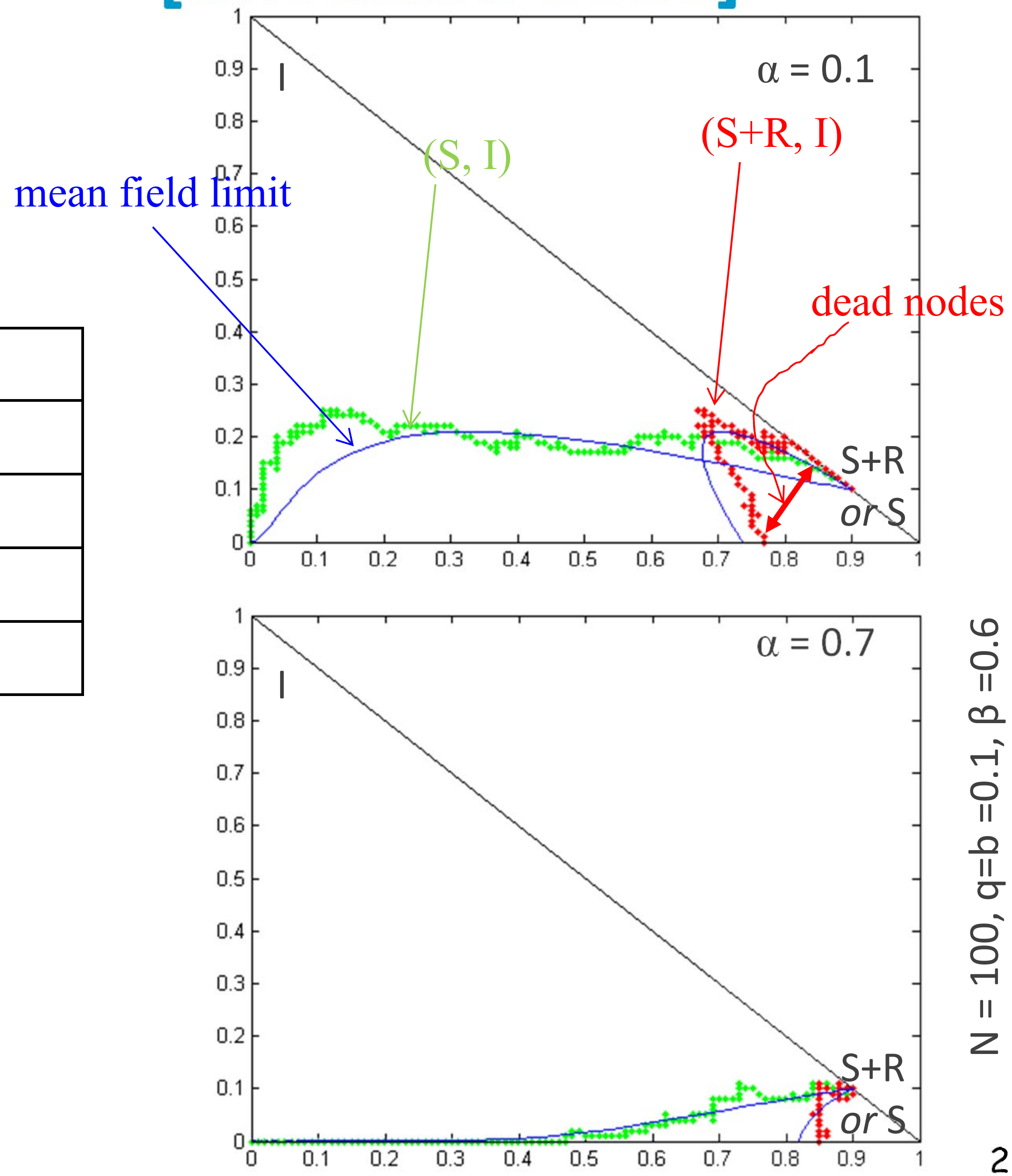
Mean Field Methods for Computer and Communication Systems
Part 3: Control

Jean-Yves Le Boudec
EPFL
ROCKS Automn School
October 2012

Virus Infection [Khouzani 2010]

transition	proba
$S \rightarrow R$	$q\ S$
$I \rightarrow R$	$b\ I$
$I \rightarrow D$	$\alpha\ I$
$I + S \rightarrow I + I$	$\beta\ IS$



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***
  $(X^N_1(t), \ldots, X^N_N(t)) \to action$
- Controller observes only object states

=> ***π*** depends on $M^N(t)$ only

$\lfloor NT \rfloor$

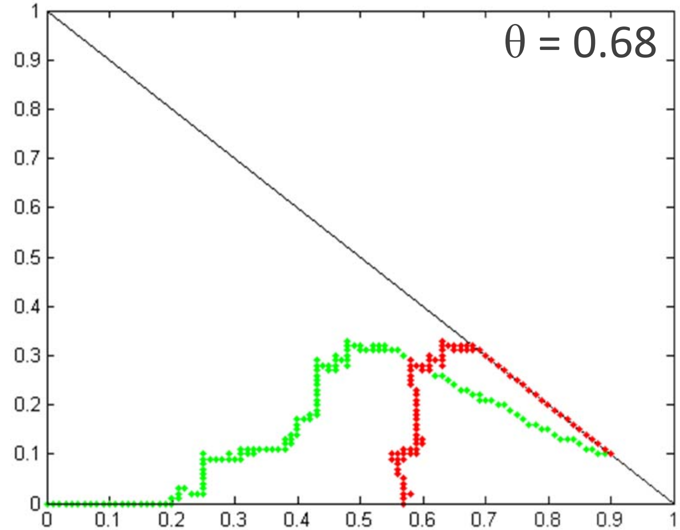
$$V_\pi^N(m) \stackrel{\text{def}}{=} \mathbb{E}\left(\sum_{k=0}^{\lfloor H^N \rfloor} r^N\left(M_\pi^N(k), \pi(M_\pi^N(k))\right) \middle| M_\pi^N(0) = m \right)$$

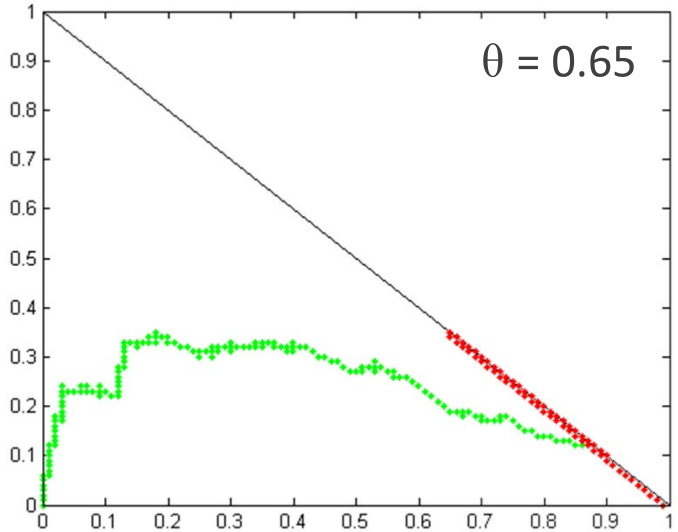
Example

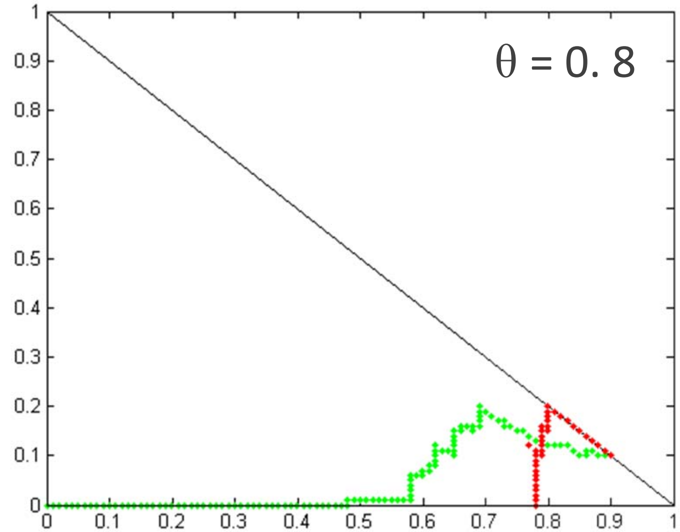
Policy π: set α=1 when R+S < θ

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

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

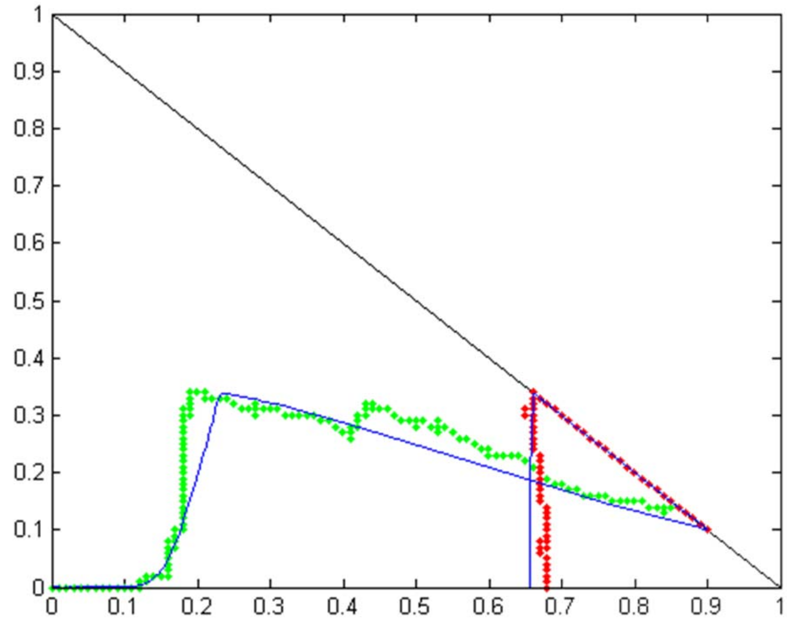
m is the initial condition of occupancy measure

- 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

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

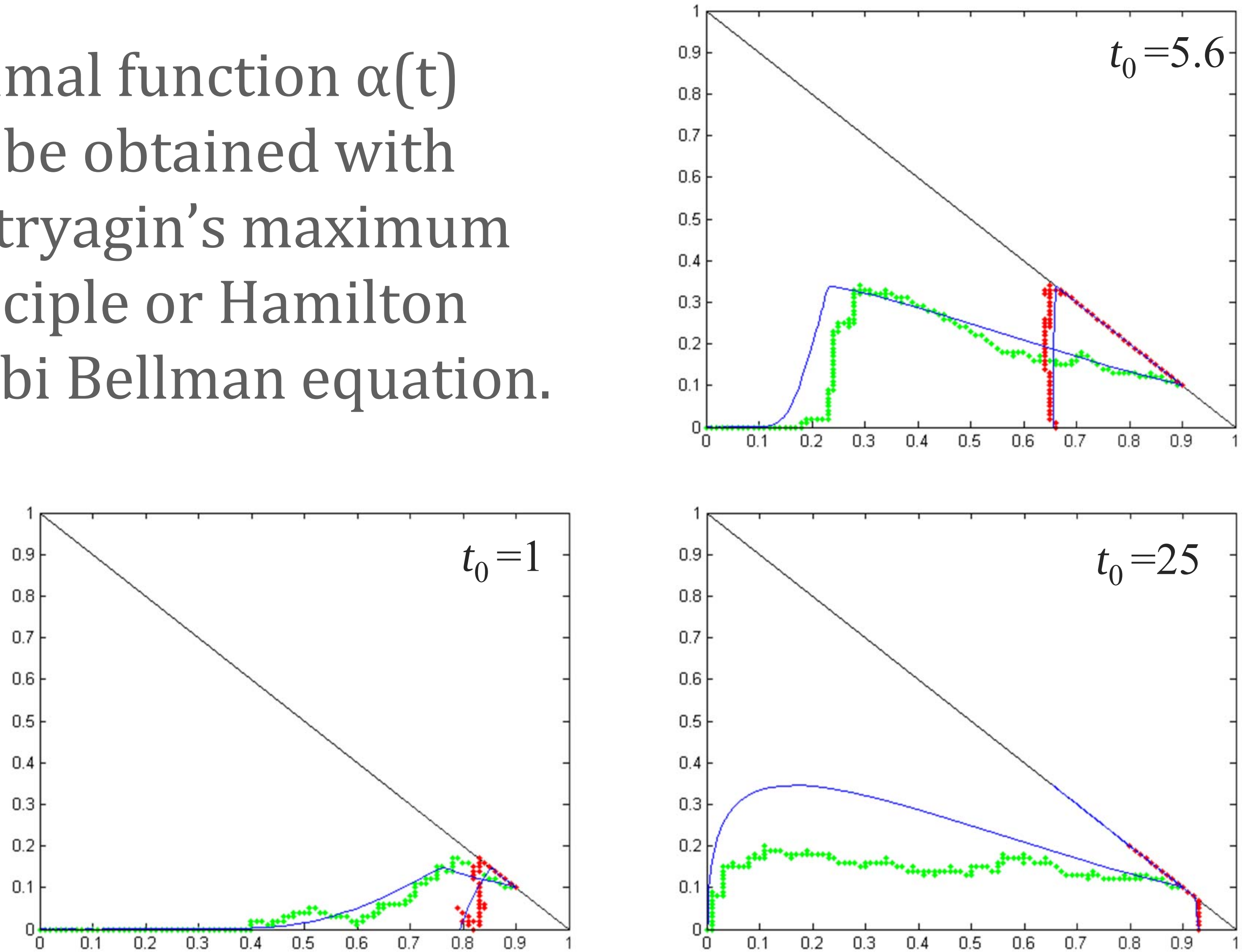
$$v_*(m_0) = \sup_\alpha v_\alpha(m_0).$$

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.



Convergence Theorem

- ***Theorem*** [Gast 2012]
  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

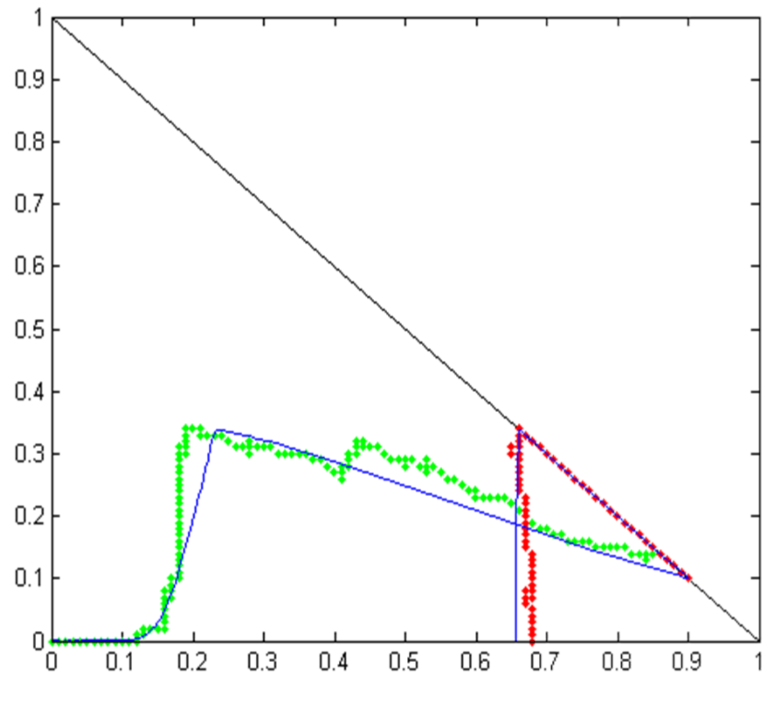
Convergence Theorem

- ***Theorem*** [Gast 2012]
  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.

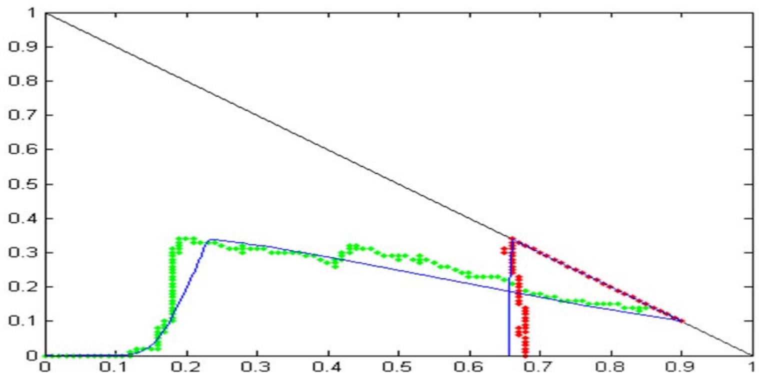
- Let $V_{\alpha^*}^N$ = value function when applying α^* to system with *N* objects

- ***Theorem*** [Gast 2012]

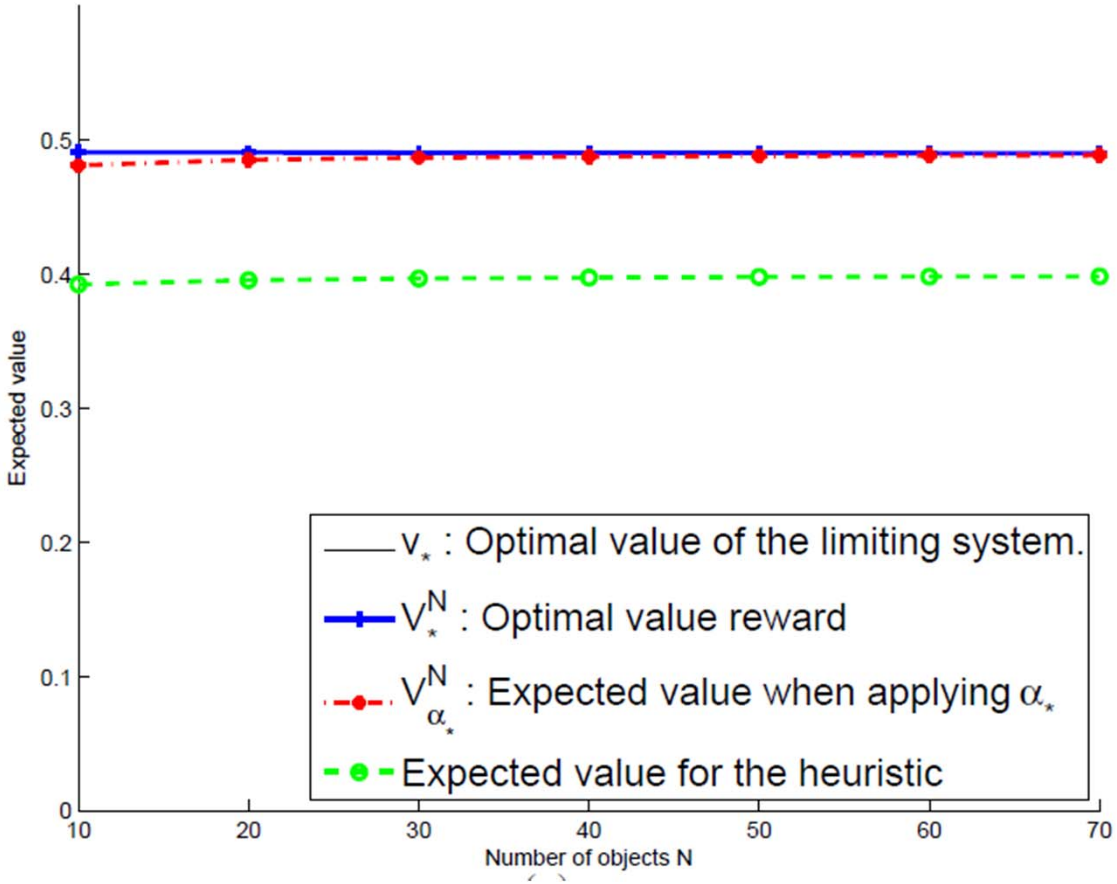
$$\lim_{N\to\infty} \left| V_{\alpha^*}^N - V_*^N \right| = 0$$

Optimal value for system with N objects (MDP)

Value of this policy



1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1



Expected value
0.5
0.4
0.3
0.2
0.1
0
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 2012]
  N. G. Gast, B. Gaujal and J.-Y. Le Boudec. *Mean Field for Markov Decision Processes: from Discrete to Continuous Optimization,* in IEEE Transactions on Automatic Control, vol. 57, num. 8, 2012.
- [Khouzani 2010]
  M.H.R. Khouzani, S. Sarkar and E. Altman. *Maximum Damage Walware Attack in Mobile Wireless Networks.* Infocom 2010