A mean Field Approach for Optimization in Particles System

Nicolas Gast Bruno Gaujal

Laboratoire d'informatique de Grenoble

Journées ALEA 2009, 16-20 mars 2009

Motivation, description of the problem

A Markov Decision Process

We consider:

- System of N particles evolving in a shared environment.
  - Markovian evolution.
  - Two particles are dependent only through the "mean field" (mean number of particles in state s).
- A controller
  - wants to optimize a reward (function of the mean field and environment), finite or infinite horizon.

Previous result:

Mean field limit (without control)

- When N grows, the system tends to a deterministic system (called mean field limit)

Main results

Our result

The MDP also converges to a deterministic limit.
More precisely, when N grows:

1. The optimal reward converges.
2. The optimal policy also converges.
3. The speed of convergence is $O(\sqrt{N})$ (CLT theorem).

Main results

Our result

The MDP also converges to a deterministic limit.
More precisely, when N grows:

1. The optimal reward converges.
2. The optimal policy also converges.
3. The speed of convergence is $O(\sqrt{N})$ (CLT theorem).

In this talk, we will

- Start with an example.
- Show the results and some proofs.

Model of interacting particles

- Time is discrete: $t \in \{0, 1, \ldots, \}$
- N particles with states $X_t^1 \ldots X_t^N$.
  - Finite state space $\mathcal{S} = \{1 \ldots S\}$
- Context C_t^N
  - $\mathrm{C} \in \mathbb{R}^d$ for some d.
- Population Mix: vector M_t^N: $(\mathrm{M}_t^N)_i = \sum_{n=0}^N \mathbf{1}_{X_t^n = i}$

Model of interacting particles

- Time is discrete: $t \in \{0, 1, \dots, \}$
- N particles with states $X_t^1 \dots X_t^N$.
  - Finite state space $\mathcal{S} = \{1 \dots S\}$
- Context C_t^N
  - $\mathrm{C} \in \mathbb{R}^d$ for some d.
- Population Mix: vector M_t^N: $(\mathrm{M}_t^N)_i = \sum_{n=0}^N \mathbf{1}_{X_t^n = i}$

Mean field assumption

- The evolution of C_t^N only depends on the mean field M_t^N:

$$\mathrm{C}_{t+1}^N = g(\mathrm{C}_t^N, \mathrm{M}_t^N)$$

- The evolution of a particle is independent of the others and is Markovian of kernel $K(\mathrm{C}_t^N)$.

Mean Field Limit

We are interested in the behavior of the system when N grows.

- Very popular model
  - Widely used in physics and computer science
  - Often without proof.

Theorem (Le Boudec, McDonald, Mudinger, 07)

Assume that $\mathrm{M}_0^N, \mathrm{C}_0^N \xrightarrow{a.s} m_0, c_0$ *and let us define the mean field limit:*

- $m_{t+1} = m_t K(c_t)$
- $c_{t+1} = g(c_t, m_t)$

Then for all time t: $\mathrm{M}_t^N, \mathrm{C}_t^N \xrightarrow{a.s} m_t, c_t$.

Controlling the system

We add a control variable a. The kernel matrix becomes $K(a, \mathrm{C})$

- At each time t, a controller chooses an action $a \in \mathcal{A}$.
- If the system is $\mathrm{M}_t^N, \mathrm{C}_t^N$, it gets a reward $\mathrm{r}(\mathrm{M}_t^N, \mathrm{C}_t^N)$.

Controlling the system

We add a control variable a. The kernel matrix becomes $K(a, \mathrm{C})$

- At each time t, a controller chooses an action $a \in \mathcal{A}$.
- If the system is $\mathrm{M}_t^N, \mathrm{C}_t^N$, it gets a reward $\mathrm{r}(\mathrm{M}_t^N, \mathrm{C}_t^N)$.

Aim of the controller

Choose the best policy to maximize:

- the finite-time expected reward between 0 and T:

$$V_\pi^N(\mathrm{M}_0^N, \mathrm{C}_0^N) = \mathbb{E}_\pi\left[\sum_{t=0}^{T} \mathrm{r}(\mathrm{M}_t^N, \mathrm{C}_t^N)\right]$$

- the expected discounted reward $(0 < \delta < 1)$:

$$V_\pi^{\delta,N}(\mathrm{M}_0^N, \mathrm{C}_0^N) = \mathbb{E}_\pi\left[\sum_{t=0}^{\infty} \delta^t \mathrm{r}(\mathrm{M}_t^N, \mathrm{C}_t^N)\right]$$

Different types of policies

Different types policies

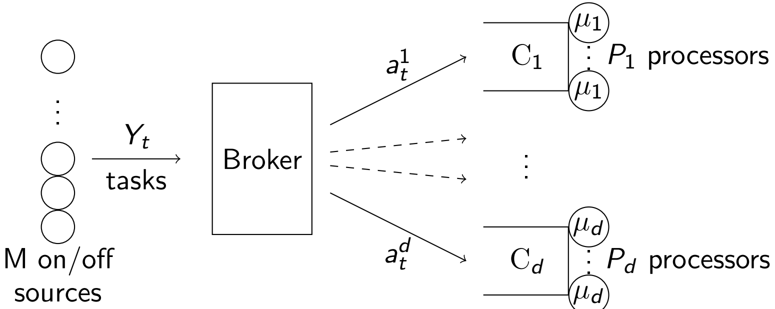
- General case – A policy is a (possibly random) function that only depends on the past: $\pi_t : (\mathrm{M}_0, \mathrm{C}_0, \dots \mathrm{M}_t, \mathrm{C}_t) \mapsto a_t$.
- Markov – Only depends on the current state: $\pi_t : (\mathrm{M}_t, \mathrm{C}_t) \mapsto a_t$
- Deterministic – π_t is not random.
- More restrictions – with limited information,...

Theorem (Optimal policies in MDP problems)

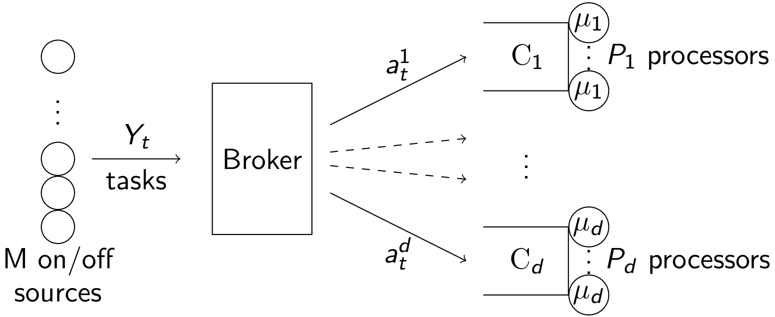
Under mild assumptions, there exists optimal policies that are Deterministic Markov. Optimal cost and policy can be computed backward:

$$V^{*N}_{t\dots T}(\mathrm{M}^N_t, \mathrm{C}^N_t) \stackrel{\text{def}}{=} \mathrm{r}_t(\mathrm{M}^N_t, \mathrm{C}^N_t) + \sup_{a \in \mathcal{A}} \mathbb{E}\left[V^{*N}_{t+1\dots T}\left(M^N_{t+1}, \mathrm{C}^N_{t+1}\right)\right]$$

Resource allocation problem



Resource allocation problem



- Size of queue i: $\mathrm{C}_{t+1}^i = \left(\mathrm{C}_t^i - \mu_i X_t^i + a_t^i Y_t\right)^+$
- Cost $=$ total waiting time $\stackrel{\text{def}}{=} \sum_i \mathrm{C}_t^i$

Computing the optimal policy

The stochastic system is hard to solve.

Policies, from optimal to approximations

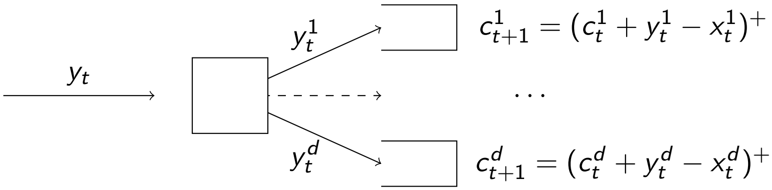
1. Brute force resolution of the MDP. But:
   - State space for particles: 2^N
   - State space for queue: $(Q_{\max})^d$
   - Expectation: 2^N states to explore.
2. Multidimensional restless bandit problem
   - Hard to solve exactly
   - Index policies
3. In practice in such systems [EGEE]
   - Use of heuristics (JSQ)

Optimal policies of the deterministic system

Solving the deterministic case

The problem becomes:

- Find $y_1^1 \dots y_T^d$ to minimize $\sum_{t=1}^{T} \sum_{i=1}^{d} c_t^i$ such that
  - $c_{t+1}^i = (c_t^i + y_t^i - x_t^i)^+$ and
  - $\sum_i y_t^i = y_t$.



Optimal policy of the deterministic system

Time t	0	1	2	3	4	5	6
y_t (packets)	9	1	0	1	7	6	6
Queue 1	I						
	I						
Queue 2							
Queue 3	I						
	I						
Optimal allocation							

- Grey = "off" processors.
- I : initial packets.
- P0: packets arrived at time 0.

Optimal policy of the deterministic system

	Time t	0	1	2	3	4	5	6
	y_t (packets)	9	1	0	1	7	6	6
Queue 1		I	P0	P0				
		I	P0	P0				
		P0						
Queue 2				P0				
Queue 3		I	P0					
		I	P0					
			P0					
Optimal allocation		5						
		1						
		3						

- Grey = "off" processors.
- I : initial packets.
- P0: packets arrived at time 0.

Optimal policy of the deterministic system

Time t	0	1	2	3	4	5	6
y_t (packets)	9	1	0	1	7	6	6
Queue 1	I	P0	P0				
	I	P0	P0				
	P0						
Queue 2			P0				
Queue 3	I	P0	P1				
	I	P0					
		P0					
Optimal allocation	5	.					
	1	.					
	3	1					

- Grey = "off" processors.
- I : initial packets.
- P0: packets arrived at time 0.

Optimal policy of the deterministic system

Time t	0	1	2	3	4	5	6
y_t (packets)	9	1	0	1	7	6	6
Queue 1	I	P0	P0	P3			
	I	P0	P0				
	P0						
Queue 2			P0				
Queue 3	I	P0	P1				
	I	P0					
		P0					
Optimal allocation	5	.	.	1			
	1	.	.	.			
	3	1	.	.			

- Grey = "off" processors.
- I : initial packets.
- P0: packets arrived at time 0.

Optimal policy of the deterministic system

Time t	0	1	2	3	4	5	6
y_t (packets)	9	1	0	1	7	6	6
Queue 1	I	P0	P0	P3	P4	P4	
	I	P0	P0		P4		
	P0				P4		
					P4		
Queue 2			P0				
Queue 3	I	P0	P1		P4		
	I	P0			P4		
		P0					
Optimal allocation	5	.	.	1	5		
	1	.	.	.	.		
	3	1	.	.	2		

- Grey = "off" processors.
- I : initial packets.
- P0: packets arrived at time 0.

Optimal policy of the deterministic system

Time t	0	1	2	3	4	5	6
y_t (packets)	9	1	0	1	7	6	6
Queue 1	I	P0	P0	P3	P4	P4	
	I	P0	P0		P4	P5	
	P0				P4		
					P4		
Queue 2			P0			P5	
						P5	
Queue 3	I	P0	P1		P4	P5	
	I	P0			P4	P5	
		P0				P5	
Optimal allocation	5	.	.	1	5	1	
	1	.	.	.	.	2	
	3	1	.	.	2	3	

- Grey = "off" processors.
- I : initial packets.
- P0: packets arrived at time 0.

Optimal policy of the deterministic system

Time t	0	1	2	3	4	5	6
y_t (packets)	9	1	0	1	7	6	6
Queue 1	I	P0	P0	P3	P4	P4	P6
	I	P0	P0		P4	P5	
	P0				P4		
					P4		
Queue 2			P0			P5	P6
						P5	
Queue 3	I	P0	P1		P4	P5	P6
	I	P0			P4	P5	P6
		P0				P5	
Optimal allocation	5	.	.	1	5	1	1
	1	.	.	.	.	2	1
	3	1	.	.	2	3	2

- Grey = "off" processors.
- I : initial packets.
- P0: packets arrived at time 0.

Optimal policy of the deterministic system

Time t	0	1	2	3	4	5	6
y_t (packets)	9	1	0	1	7	6	6
Queue 1	I	P0	P0	P3	P4	P4	P6
	I	P0	P0		P4	P5	
	P0				P4		
					P4		
Queue 2			P0			P5	P6
						P5	
Queue 3	I	P0	P1		P4	P5	P6
	I	P0			P4	P5	P6
		P0				P5	
Optimal allocation	5	.	.	1	5	1	1+2
	1	.	.	.	.	2	1
	3	1	.	.	2	3	2

- Grey = "off" processors.
- I : initial packets.
- P0: packets arrived at time 0.
- 2 packets remains at the end.

Numerical example

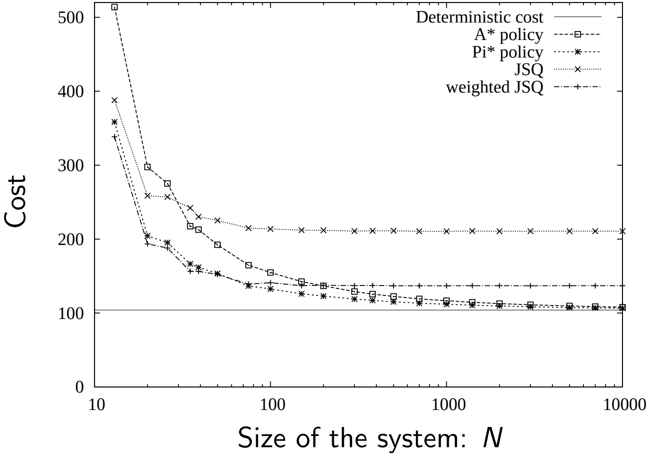
This gives us two policies

- π^* : at t, we apply $\pi^*_t(\mathrm{M}^N_t, \mathrm{C}^N_t)$
- a^* : we apply $a^*_t \stackrel{\text{def}}{=} \pi^*_t(m_t, c_t)$

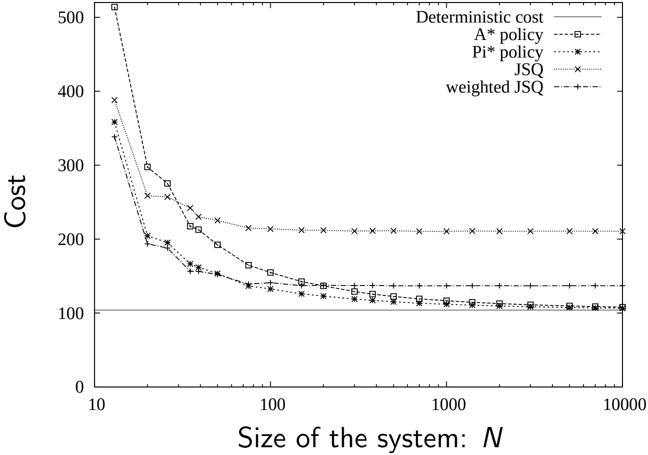
We want to compare

- V^{*N} – optimal cost for the system of size N
- $V^N_{a^*}$ – cost when applying a^*
- $V^N_{\pi^*}$ – cost when applying π^*
- V^N_{JSQ} – cost of Join Shortest Queue.
- v^* – cost of the deterministic system.

Cost convergence



Cost convergence



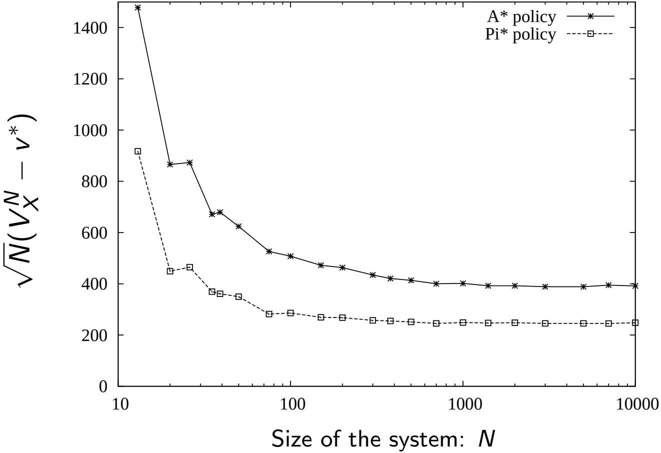
Theorem

$$\lim V_{\pi^*}^N(\mathrm{M},\mathrm{C}) = \lim V_{a^*}^N(\mathrm{M},\mathrm{C}) = v^*(m,c) \;\left(= \lim V^{*N}(\mathrm{M},\mathrm{C})\right)$$

Speed of convergence – central limit theorem

We plot

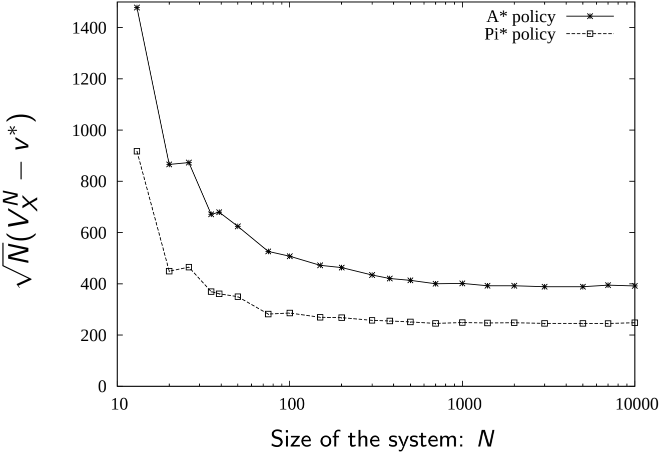
- $\sqrt{N}(V_{\pi^*}^N(\mathrm{M},\mathrm{C}) - v^*(m,c))$
- $\sqrt{N}(V_{a^*}^N(\mathrm{M},\mathrm{C}) - v^*(m,c))$



Speed of convergence – central limit theorem

We plot

- $\sqrt{N}(V_{\pi^*}^N(\mathrm{M},\mathrm{C}) - v^*(m,c)) \longrightarrow \gamma$
- $\sqrt{N}(V_{a^*}^N(\mathrm{M},\mathrm{C}) - v^*(m,c)) \longrightarrow \gamma'$



Cost convergence

Summary of the assumptions

(A1) **Independence of the users, Markov system** – If environment is C, action is a, the evolution of particles is independent and Markovian with kernel $K(a, \mathrm{C})$.

(A2) **Compact action set** – The action set $\mathcal{A}$ is compact.

(A3) **Continuity of K, g, r** – the mappings $K(a, \mathrm{C})$, $g(\mathrm{C}, \mathrm{M}, a)$ and $\mathrm{r}_t(\mathrm{M}, \mathrm{C})$ are continuous deterministic functions, uniformly continuous in a.

(A4) **Almost sure initial state** – $\mathrm{M}_0^N, \mathrm{C}_0^N \xrightarrow{a.s} m_0, c_0$. There exists B such that almost surely $\mathrm{C}_0^N \leq B$.

Optimal reward convergence

Theorem (Convergence of the optimal reward)

Under assumptions (A1,A2,A3,A4), almost surely:

$$\lim_{N\to\infty} V_T^{*N}(M_0^N, C_0^N) = \lim_{N\to\infty} V_{a_0^* \dots a_{T-1}^*}^N(M_0^N, C_0^N) = v_T^*(m_0, c_0)$$

In particular, this shows that:

- Optimal cost converges
- Fixed policy (a^*) is asymptotically optimal

Remark that π^* is also asymptotically optimal but not asymptotically better.

Optimal reward convergence : sketch of proof

Lemma

Under assumptions (A1,A3,A4) and for any fixed actions $a_0 \ldots a_{T-1}$:

$$\lim_{N\to\infty} V^N_{a_0\ldots a_{T-1}}(\mathrm{M}^N_0, \mathrm{C}^N_0) = v_{a_0\ldots a_{T-1}}(m_0, c_0).$$

Optimal reward convergence : sketch of proof

Lemma

Under assumptions (A1,A3,A4) and for any fixed actions $a_0 \dots a_{T-1}$:

$$\lim_{N\to\infty} V^N_{a_0\dots a_{T-1}}(\mathrm{M}^N_0, \mathrm{C}^N_0) = v_{a_0\dots a_{T-1}}(m_0, c_0).$$

Proof of theorem.

- $V^{*N} \geq V^N_{a^*} \xrightarrow{a.s} v^*$
- Use backward computation to compare:
  1. $V^{*N}_{t\dots T}(\mathrm{M},\mathrm{C}) \stackrel{\mathrm{def}}{=} \mathrm{r}_t(\mathrm{M},\mathrm{C}) + \sup_{a\in\mathcal{A}} \mathbb{E}\left[V^{*N}_{t+1\dots T}\left(\Phi^N_a(\mathrm{M},\mathrm{C})\right)\right]$
  2. $v^*_{t\dots T}(m,c) \stackrel{\mathrm{def}}{=} \mathrm{r}_t(m,c) + \sup_{a\in\mathcal{A}} \left[v^*_{t+1\dots T}\left(\Phi_a(m,c)\right)\right]$
- Proof that any actions maximizing 1 will also maximize 2.

$\square$

Speed of convergence: a central limit theorem

Additional assumptions

(A4-bis) **Initial gaussian variable** – (A4) + there exists G_0 gaussian such that $\sqrt{N}((\mathrm{M}_0^N, \mathrm{C}_0^N) - (m_0, c_0)) \xrightarrow{\mathcal{L}} G_0$.

(A5) **Continuous differentiability** – g, K_{ij} and r_t are C^1.

or

(A5-bis) **Differentiability in $a_1 \dots a_T$** – g, K_{ij} and r_t are differentiable in each (m_t, c_t) for all t.

or

(A5-ter) **Continuous lipschitz** – g, K_{ij} and r_t are lipschitz on all compact set of their set of definition.

- (A4) : natural, *e.g* if initial states are chosen independently.
- (A5/bis/ter): give different results.

Speed of convergence: a central limit theorem

Theorem (CLT for the evolution of particles)

Under assumptions (A1,A2,A3,A4bis,A5-bis), if the actions taken by the controller are $a_0 \dots a_{T-1}$, there exist gaussian vectors of mean 0, $G_1 \dots G_{T-1}$ such that for every t:

$$\sqrt{N}\left((\mathrm{M}_0^N, \mathrm{C}_0^N) - (m_0, c_0), \dots, (\mathrm{M}_t^N, \mathrm{C}_t^N) - (m_t, c_t)\right) \xrightarrow{\mathcal{L}} G_0, \dots, G_t$$

The covariance of G_t can be effectively computed.

Speed of convergence: a central limit theorem

Theorem (CLT for the evolution of particles)

Under assumptions (A1,A2,A3,A4bis,A5-bis), if the actions taken by the controller are $a_0 \dots a_{T-1}$, there exist gaussian vectors of mean 0, $G_1 \dots G_{T-1}$ such that for every t:

$$\sqrt{N}\left((\mathrm{M}_0^N, \mathrm{C}_0^N) - (m_0, c_0), \dots, (\mathrm{M}_t^N, \mathrm{C}_t^N) - (m_t, c_t)\right) \xrightarrow{\mathcal{L}} G_0, \dots, G_t$$

The covariance of G_t can be effectively computed.

Theorem (CLT for reward)

Under assumptions (A1,A2,A3,A4bis,A5-ter), $\exists \beta, \gamma, \beta', \gamma'$ such that when N does to infinity:

- $\sqrt{N}\left| V_T^{*N}(\mathrm{M}_0^N, \mathrm{C}_0^N) - V_{a^*}^N(\mathrm{M}_0^N, \mathrm{C}_0^N) \right| \leq_{\mathrm{st}} \beta + \gamma \|G_0\|_\infty$
- $\sqrt{N}\left| V_T^{*N}(\mathrm{M}_0^N, \mathrm{C}_0^N) - v_{0\dots T}^*(m_0, c_0) \right| \leq_{\mathrm{st}} \beta' + \gamma' \|G_0\|_\infty$

Speed of convergence: a central limit theorem

Sketch of the proof (evolution).

- at time t, we are in state $(m_t, c_t) + \frac{1}{\sqrt{N}} G_t + o(\frac{1}{\sqrt{N}})$.
  - New alea A_t created by dispersion.
  - $K(\mathrm{C}_t) = K(c_t) + \frac{1}{\sqrt{N}} \mathrm{D}KG_t + o(\frac{1}{\sqrt{N}})$.
- thus

$$\begin{aligned} \mathrm{M}^N_{t+1} &\approx (m_t + \tfrac{1}{\sqrt{N}} G_t)(K(c_t) + \tfrac{1}{\sqrt{N}} \mathrm{D}KG_t) + \tfrac{1}{\sqrt{N}} A_t \\ &\approx m_t K(c_t) + \tfrac{1}{\sqrt{N}}(G_t K + m_t \mathrm{D}KG_t + A_t) \end{aligned}$$

- Same for C^N_t.

$\square$

Discounted case

Assumptions

(A6) **Homogeneity in time** – reward and probability kernel K_t do not depend on t.

(A7) **Bounded reward** – $\sup_{\mathrm{M},\mathrm{C}} |\mathrm{r}(\mathrm{M},\mathrm{C})| < \infty$.

$$V_\pi^{\delta N}(\mathrm{M}_0^N, \mathrm{C}_0^N) = \mathbb{E}_\pi\left[\sum_{t=0}^{\infty} \delta^t \mathrm{r}(\mathrm{M}_t^N, \mathrm{C}_t^N)\right].$$

Discounted case

Assumptions

(A6) **Homogeneity in time** – reward and probability kernel K_t do not depend on t.

(A7) **Bounded reward** – $\sup_{\mathrm{M},\mathrm{C}} |\mathrm{r}(\mathrm{M},\mathrm{C})| < \infty$.

$$V_\pi^{\delta N}(\mathrm{M}_0^N, \mathrm{C}_0^N) = \mathbb{E}_\pi\left[\sum_{t=0}^{\infty} \delta^t \mathrm{r}(\mathrm{M}_t^N, \mathrm{C}_t^N)\right].$$

Theorem

Under assumptions (A1,A2,A3,A4,A6,A7),

$$\lim_{N\to\infty} V_*^{\delta N}(\mathrm{M}^N, \mathrm{C}^N) = v_*^\delta(m, c) \text{ a.s.}$$

Discounted case – sketch of the proof

Sketch of the proof.

Thanks to A7 (bounded reward), finite time convergence suffices. For T big enough:

$$\sum_{t=0}^{\infty} \delta^t \mathrm{r}(\mathrm{M}_t^N, \mathrm{C}_t^N) \approx \sum_{t=0}^{\mathbf{T}} \delta^t \mathrm{r}(\mathrm{M}_t^N, \mathrm{C}_t^N)$$

$\square$

Discounted case – sketch of the proof

Sketch of the proof.

Thanks to A7 (bounded reward), finite time convergence suffices. For T big enough:

$$\sum_{t=0}^{\infty} \delta^t \mathrm{r}(\mathrm{M}_t^N, \mathrm{C}_t^N) \approx \sum_{t=0}^{\mathbf{T}} \delta^t \mathrm{r}(\mathrm{M}_t^N, \mathrm{C}_t^N)$$

$\square$

Problem for other asymptotic limits

- Average infinite reward: $\lim_{T\to\infty} \frac{1}{T} \mathbb{E}_\pi \sum_{t=0}^{T} \mathrm{r}(\mathrm{M}_t, \mathrm{C}_t)$
- Problem with exchanging limits $N \to \infty$ and $T \to \infty$:
    - ex: $K_{i1}(\mathrm{C}) = f(\mathrm{M}_1)$ with $f(x) = x$ or $f(x) = 8(\frac{1}{2} - x)^3$.
    - Convergence to a set such that $A = f(A)$.

Conclusion

1. We presented a model for optimization for interacting particles with mean field interraction.
2. In general, the stochastic case is impossible to solve and we restrict the problem to different types of policies:
   - With limited information
   - Based on heuristics
   - Index
   - ...

   We showed that asymptotically this distinction collapses.
3. In particular, we showed that:
   - a^*, π^* are asymptotically optimal.
   - Give bounds for the speed of convergence in $\sqrt{N}$.

Computational issues

How to apply this in practice?

We distinguish 3 cases:

1. Solve the deterministic system:
   - apply a^* or π^*.
2. Design an approximation algorithm for the deterministic system
   - also an approximation (asymptotically) for stochastic problem.
3. Use brute force computation:
   - $v^*_{t\dots T}(m,c) = \mathrm{r}(m,c) + \sup_a v^*_{t+1\dots T}(\phi_a(m,c))$
   - Still costly but reduce computational time by removing the expectation $\mathbb{E}_{\mathrm{M,C}}$.
   - We can compute forward.

References

- Paper corresponding to this talk:
  - Gast N., Gaujal B. – *A Mean Field Approach for Optimization in Particles Systems and Applications* – RR 6877, `http://mescal.imag.fr/membres/nicolas.gast/`
- Mean field models:
  - Le Boudec, McDonald, Mudinger – *A Generic Mean Field Convergence Result for Systems of Interacting Objects* – QEST 2007
  - J.Y. Le Boudec, Benaïm, M. – *A Class Of Mean Field Interaction Models for Computer and Communication Systems* – Performance Evaluation
- Infinite horizon results:
  - Borkar – *Stochastic Approximation: A Dynamical Systems Viewpoint* – Cambridge University Press 2008