Time-inconsistent mean-field stopping problems: A regularized equilibrium approach

Xiang Yu [1] **Fengyi Yuan** [1,2]

[1]Department of Applied Mathematics, The Hong Kong Polytechnic University

[2]Department of Mathematical Sciences, Tsinghua University

29th November, 2023

Contents

Discount

- For multi-period decision problems, we need to consider discount for delayed reward.
  - Agent-implied discount: the impatience. E.g.: paying you 100\$ tomorrow is indifferent to paying you 100δ\$ right now because you hate waiting for one day. $\delta \in (0, 1]$ is your one-day discount factor.
  - Market-implied discount: usually related to the interest rate.
  - Exponential discount: discount factors are constants among every delayed period. Hence, after k periods, the reward is discounted by δ^k.
- Cumulative discounted reward formulation for an MDP:

$$J(x; \pi) = \mathbb{E}^{x,\pi} \sum_{t=0}^{\infty} \delta^t r(X_t).$$

Non-exponential discount

Exponential discount: $\delta_1 = \delta_2 = \cdots = \delta_k = \cdots = \delta$, and $\delta(k) := \prod_{j=1}^{k} \delta_j = \delta^k$. Here δ_k is the one-period discount rate at period $k-1$. That is to say, each delayed period is discounted equally. But...

- There is no (practical) reason to assume that impatience is homogenous in time.
- There is no reason to anticipate that future interest rates equal to the spot rate.

Non-exponential discount

Typical examples

- Quasi-hyperbolic discount: for $K > 0$, $\beta \in (0,1)$, $\delta_1 = K\beta$, $\delta_k = \beta$, $k \geq 1$. Equivalently,

$$\delta(k) = \begin{cases} 1, & k = 0, \\ K\beta^k, & k \geq 1. \end{cases}$$

  $K < 1$: decreasing impatience (present bias); $K > 1$: increasing impatience (future bias).

- Hyperbolic discount: for $r > 0$,

$$\delta(k) = \frac{1}{1 + rk}.$$

Consistent planning

The most important consequence of considering exponential discount:

DPP

With $Q^\pi(x,a) := J(x; a \oplus_1 \pi)$, we have

$$Q^\pi(x,a) = r(x) + \delta \mathbb{E}^{x,a} Q^\pi(X_1, \pi(X_1)).$$

With $V(x) = \sup_\pi Q^\pi(x, \pi(x)) = \sup_\pi J(x;\pi)$, we have

$$V(x) = r(x) + \delta \sup_a \mathbb{E}^{x,a} V(X_1).$$

- We have an equation about V! Solving V and choosing $\pi(x) = \arg\max_a \mathbb{E}^{x,a} V(X_1)$ gives the optimal policy.
- The optimal policy is time-consistent. It is always "pure-strategy".

Consistent planning

But with non-exponential discount...

Violation of DPP

If $J(x;\pi) = \mathbb{E}^{x,\pi} \sum_{t=0}^{\infty} \delta(t) r(X_t)$ and $Q^\pi(x,a) = J(x; a \oplus_1 \pi)$, we have:

$$\begin{aligned} Q^\pi(x,a) &= r(x) + \mathbb{E}^{x,a\oplus_1\pi} \sum_{t=1}^{\infty} \delta(t) r(X_t) \\ &= r(x) + \mathbb{E}^{x,a} \mathbb{E}^{X_1,\pi} \sum_{t=1}^{\infty} \delta(t) r(X_{t-1}) \\ &= r(x) + \mathbb{E}^{x,a} \mathbb{E}^{X_1,\pi} \sum_{t=0}^{\infty} \delta(t+1) r(X_t) \\ &= r(x) + \mathbb{E}^{x,a} Q_1^\pi(X_1, \pi(X_1)). \end{aligned}$$

Want to maximize $Q^\pi \to$ need $Q_1^\pi \to$ need $Q_2^\pi \to \cdots$

Consistent planning

- Maximizing $J(x;\pi)$ still makes sense provided that future selves discount by $\delta_2, \delta_3, \cdots, \delta_{k+1}, \cdots$ However, k is the delayed time instead of the calendar time. Future selves still discount by $\delta_1, \delta_2, \cdots$
- What does $\sup_\pi J(x;\pi)$ mean at $k \geq 1$?
- Consistent planning in economics:

The equilibrium policy

Find π^*, such that for any x, a,

$$J(x; a \oplus_1 \pi^*) \leq J(x; \pi^*).$$

Stopping decision: examples

In stopping decision problems, discount is the cost of waiting.

An example (gambling, finite states + finite time)

- Each time you roll a fair six-sided dice, you have the option to take a prize amounting to the points you get, or continue playing. You have a total of 3 chances.
- Waiting has its cost. More future rewards are needed to pursuade you to continue to play. We model this by discount. Assume $\delta_1 = 0.9$ and $\delta_2 = 0.8$.

Stopping decision: examples

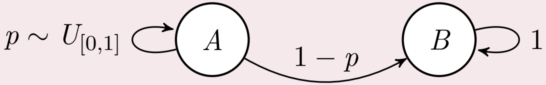
Solution (Backward Induction!)

- At Step 3, you must stop and take whatever you get.
- At Step 2, you get $3.5 \times 0.9 = 3.15$ if you wait. So stop if you get 4, 5 and 6, and wait otherwise.
- At Step 1, suppose you wait. Then w.p. 1/2 you stop at Step 2 and get 5 in expectation, and w.p. 1/2 you stop at Step 3 and get 3.5 in expectation. So in total, you get $0.9 \times \frac{1}{2} \times 5 + 0.9 \times 0.8 \times \frac{1}{2} \times 3.5 = 3.51$ if you wait at Step 1. In conclusion, you stop if you get 4,5 and 6, and wait otherwise.

If you just maximize EU of Step 1, by the method of enumeration, it is optimal to stop at Step 2 when you get 3,4,5,6.

Mean-field system with centralized stopping decision

A more complex example: continuous space + infinite time



- You lead a R&D group with many members. At each period there is a fraction p of your members completing their tasks. But you don't know how much this fraction will be. You just regard it as totally random (uniform distribution).
- You want to wait until all members complete their tasks. But there are costs for waiting (e.g., the competition or the plagiarism). So it is your decision to determine a criterion of early annoucement.

Mean-field system with centralized stopping decision

- For this particular example (with quasi-hyperbolic discount), we can prove that equilibrium policy exists and is **unique**: we have a stopping region, a continuation region and a **randomized stopping region**.
- For general infinite horizon problems:

The equilibrium policy

Find π^*, such that for any μ, a,

$$J(\mu; a \oplus_1 \pi^*) \leq J(\mu; \pi^*).$$

- Finite horizon: backward induction
- Infinite horizon: Does π^* even exist???

Contents

The transitions

- The state dynamics:

$$\mu_{k+1} = T(\mu_k, \xi^{\phi_k(\mu_k)}, Z^0),$$

where $\xi^{\phi_k(\mu_k)} \sim \mathcal{B}(\phi_k(\mu_k))$, and Z^0 is the common noise.

- To model stopping decisions:

$$T(\mu, a, z) = \begin{cases} T_0(\mu, z), & a = 0, \\ \triangle, & a = 1. \end{cases}$$

- We always denote by $\mathbb{P}^{\mu,\phi}$ (and $\mathbb{E}^{\mu,\phi}$) the probability (and its expectation) induced by initial population distribution μ and the (feed-back) policy ϕ.
- We consider a reward function r, and a general discount function δ.

The rewards

- The policy (if stationary in time) $\phi : \bar{S} \to [0,1]$ assigns to each (observed) state distribution μ a **probability** to stop. E.g., at each step you flip a biased coin and choose to stop when you get heads. The designs of such coins depend on observations (feed-back control!).
- Under the policy ϕ and observation μ, you get an expected cumulative discounted reward given by

$$J^{\phi}(\mu) := \sum_{k=0}^{\infty} \delta(k) \mathbb{E}^{\mu,\phi} r(\mu_k) \phi_k(\mu_k).$$

The rewards

Why this form of reward?

Lemma

Let $\tilde{\mathbb{P}}^{\mu}$ be the probability measure induced by the transition rule $\mu_{k+1} = T_0(\mu_k, Z^0)$ and the initial condition $\mu_0 = \mu$, and let $\tilde{\mathbb{E}}^{\mu}$ denote its expectation. Then for any $\phi \in \mathcal{F}$, $\mu \in \bar{S}$ and $k \in \mathbb{T}$, it holds that

$$\mathbb{E}^{\mu,\phi} r(\mu_k)\phi_k(\mu_k) = \tilde{\mathbb{E}}^{\mu} r(\mu_k)\phi_k(\mu_k)\prod_{j=0}^{k-1}(1-\phi_j(\mu_j)).$$

It is assumed by convention that $\prod_{k=0}^{-1} \equiv 1$.

- Blue part: the probability of stopping at the current step...
- Yellow part: the probability that the system has not been stopped yet.

The relaxed equilibrium

Definition

$\phi^* \in \mathcal{F}_S$ is said to be a *relaxed equilibrium* if,

$$J^{\psi \oplus_1 \phi^*}(\mu) \leq J^{\phi^*}(\mu), \forall \mu \in \bar{S}, \psi \in [0,1].$$

- The same definition as the one in Motivation part.
- If you follow some policy ϕ^* in the future, it is "optimal" to follow it now!
- Sequential game in finite horizon problem V.S. simultaneous game in infinite horizon problem.
- We do not have a "terminal" to start with when using backward induction approach.

Contents

Equilibria = fixed points!

A simple derivation from Markov property:

$$\begin{aligned} J^{\psi \oplus_1 \phi^*}(\mu) =& r(\mu)\psi + \mathbb{E}^0 \tilde{J}^{\phi^*}(T(\mu, \xi^{\psi}, Z^0)) \\ =& r(\mu)\psi + (1-\psi)\mathbb{E}^0 \tilde{J}^{\phi^*}(T_0(\mu, Z^0)), \end{aligned}$$

with (think about Q_1^{π}!)

$$\tilde{J}^{\phi^*}(\mu) = \sum_{k=0}^{\infty} \delta(1+k)\mathbb{E}^{\mu,\phi^*} r(\mu_k)\phi^*(\mu_k).$$

Lemma

ϕ^* is a relaxed equilibrium if and only if it solves the fixed point problem:

$$\phi^*(\mu) \in \arg\max_{\psi \in [0,1]} \left\{ r(\mu)\psi + (1-\psi)\mathbb{E}^0 \tilde{J}^{\phi^*}(T_0(\mu, Z^0)) \right\},$$

Equilibria = fixed points!

The optimization problem is simple so that we can solve it explicitly:

$$\phi^*(\mu) = \begin{cases} 1, & r(\mu) > f_{\phi^*}(\mu), \\ 0, & r(\mu) < f_{\phi^*}(\mu), \end{cases}$$

where $f_{\phi^*}(\mu) := \mathbb{E}^0 \tilde{J}^{\phi^*}(T_0(\mu, Z^0))$ (the reward if we choose to continue).

- The indifference principle of Game Theory: if a mixed strategy is equilibrium, pure strategies with **positive probability** are indifferent! $\longrightarrow$ only mix between indifferent strategies.
- Definition in Huang and Zhou (2019):

$$\phi^*(\mu) = \begin{cases} 1, & r(\mu) \geq f_{\phi^*}(\mu), \\ 0, & r(\mu) < f_{\phi^*}(\mu), \end{cases}$$

Equilibria = fixed points!

The next task: how do we solve the fixed point of

$$\phi^*(\mu) = \begin{cases} 1, & r(\mu) > f_{\phi^*}(\mu), \\ 0, & r(\mu) < f_{\phi^*}(\mu). \end{cases}$$

- Even proving the existence is not straightforward. **Kakutani–Glicksberg–Fan** theorem must be used (if possible).
- We choose to use the method of regularization, which produces a Lipschitz approximation to (possibly discontinuous) ϕ^*.
- Existence of the relaxed equilibrium is obtained by the vanishing of regularization.

The regularization

We shall consider the following regularization to the original problem:

$$J_\lambda^{\phi}(\mu) := \sum_{k=0}^{\infty} \delta_\lambda(k)\mathbb{E}^{\mu,\phi}\left[r(\mu_k)\phi_k(\mu_k) + \lambda\mathcal{E}(\phi_k(\mu_k))\right],$$

$$\tilde{J}_\lambda^{\phi}(\mu) := \sum_{k=0}^{\infty} \delta_\lambda(k+1)\mathbb{E}^{\mu,\phi}\left[r(\mu_k)\phi_k(\mu_k) + \lambda\mathcal{E}(\phi_k(\mu_k))\right],$$

where $\mathcal{E}(\phi) := -\phi\log\phi - (1-\phi)\log(1-\phi)$, and $\delta_\lambda(k) := \delta(k)\left(\frac{1}{1+\lambda}\right)^{k^2}$.

- The entropy regularization is to encourage exploration (so the resulted equilibria ϕ_λ are inherently of **mixed strategy**).
- The choice of δ_λ is purely technical, and the exponent k^2 is not special (subject to certain technical constraints).

Regularized equilibria

- Regularized equilibria are defined in the same way as relaxed equilibria, with J replaced by J_λ, and $\tilde{J}$ replaced by $\tilde{J}_\lambda$.
- Another simple derivation from Markov property:

$$\begin{aligned} J_\lambda^{\psi \oplus_1 \phi^*}(\mu) =& r(\mu)\psi - \lambda\psi\log\psi - \lambda(1-\psi)\log(1-\psi) \\ & + \sum_{k=1}^{\infty} \delta(k)\mathbb{E}^{\mu,\psi}\mathbb{E}^{\mu_1,\phi^*}[r(\mu_k)\phi(\mu_k) - \lambda\mathcal{E}(\phi(\mu_k))] \\ =& r(\mu)\psi - \lambda\psi\log\psi - \lambda(1-\psi)\log(1-\psi) \\ & + (1-\psi)\mathbb{E}^0\tilde{J}_\lambda^{\phi^*}(T_0(\mu, Z^0)). \end{aligned}$$

Regularized equilibria

- ϕ_λ is a regularized equilibrium if and only if it solves

$$\phi_\lambda(\mu) \in \arg\max_{\psi\in[0,1]} \{ r(\mu)\psi + (1-\psi)\mathbb{E}^0\mathcal{T}_2^\lambda(\phi_\lambda)(T_0(\mu, Z^0)) - \lambda\psi\log\psi - \lambda(1-\psi)\log(1-\psi)\},$$

with

$$\mathcal{T}_2^\lambda(\phi)(\mu) := \tilde{J}_\lambda^\phi(\mu).$$

- The optimization problem can still be solved explicitly:

$$\phi_\lambda(\mu) = \frac{1}{1+\exp\left(\frac{1}{\lambda}[\mathbb{E}^0\mathcal{T}_2^\lambda(\phi_\lambda)(T_0(\mu, Z^0)) - r(\mu)]\right)} =: \mathcal{T}_1^\lambda \circ \mathcal{T}_2^\lambda(\phi_\lambda)(\mu).$$

- ϕ_λ is a regularized equilibrium if and only it is a fixed point (**Equilibria=Fixed points!**) of $\mathcal{T}_1^\lambda \circ \mathcal{T}_2^\lambda$.

Regularized equilibria

Notations:

$$f^{\lambda}_{\phi_\lambda}(\mu) = \mathcal{T}^{\lambda}_2(\phi_\lambda)(\mu) = \mathbb{E}^0 \tilde{J}^{\phi^*}_{\lambda}(T_0(\mu, Z^0)).$$

Then, ϕ_λ is a regularized equilibrium if and only it solves

$$\phi_\lambda(\mu) = \frac{\exp\left(\frac{1}{\lambda} r(\mu)\right)}{\exp\left(\frac{1}{\lambda} r(\mu)\right) + \exp\left(\frac{1}{\lambda} f^{\lambda}_{\phi_\lambda}(\mu)\right)}.$$

- The original problem: compare r and f_{ϕ^*}. Choose the one that strictly dominates, mix between two if they are indifferent. Discontinuous policy, bang-bang type (not exactly because mixture exists).
- The regularized problem: choose the **soft-max** between r and $f^{\lambda}_{\phi_\lambda}$. Continuous policy, inherently mixed strategy ($\phi_\lambda \in (0, 1)$).

Existence of regularized equilibria

Theorem

Under certain technical assumptions (of T_0, r and Z^0), there exist a regularized equilibrium $\phi_\lambda \in \mathcal{F}_S^{\mathrm{Lip}}$ for any regularization parameter $\lambda > 0$.

Proof ideas:

- Prove that $\mathcal{T}_1^\lambda \circ \mathcal{T}_2^\lambda$ admits a fixed point, using Schauder's theorem.
- Obtain compactness from Arzela-Ascoli. We use Lipschitz continuity with respect to μ, which is guaranteed by the regularization.
- Almost all estimates blow up when $\lambda \to 0$!

Regularized equilibria as ε-equilirbria

Theorem

Under certain technical assumptions (of T_0, r and Z^0), for any $\varepsilon > 0$, ϕ_λ is an ε-equilibrium of the original problem, i.e., for every $\mu \in \bar{S}$,

$$J^{\psi \oplus_1 \phi_\lambda}(\mu) \leq J^{\phi_\lambda}(\mu) + \varepsilon, \forall \psi \in [0,1].$$

provided that λ is sufficiently small.

Proof idea: From definitions of the regularized equilibrium and the total reward without λ, we may write

$$J^{\psi \oplus_1 \phi_\lambda}(\mu) \leq J^{\phi_\lambda}(\mu) + \delta J^\lambda + \delta \mathcal{E}^\lambda.$$

δJ^λ comes from the regularization of discount function, and $\delta \mathcal{E}^\lambda$ comes from the entropy.

Existence of relaxed equilibria

Theorem

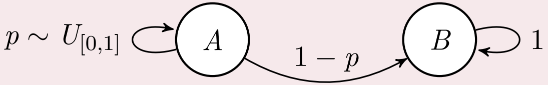
Under certain technical assumptions (of T_0, r and Z^0), there exist a relaxed equilibrium $\phi_0 \in \mathcal{F}_S$. Moreover, for any convergent subsequence of $\{\phi_\lambda\}_{\lambda>0}$ (in the sense of weak- convergence), it converges to ϕ_0.*

Proof ideas:

- Use the Banach-Alaoglu theorem to obtain a candidate relaxed equilibrium.
- Prove the candidate relaxed equilibrium is indeed relaxed equilibrium. Key step: softmax→max. But the limit in the indifference region is not clear! This gives mixed strategy.
- Switch between $\mathbb{P}^{\mu,\phi}$ and $\tilde{\mathbb{P}}^{\mu}$ as appropriate.

An example

A more complex example: continuous space + infinite time



- Take (the quasi-hyperbolic discount)

$$\delta(k) = \begin{cases} 1, & k = 0, \\ K\beta^k, & k \geq 1. \end{cases}$$

- The most popular in discrete time setting due to its tractability (usually admits explicit solution). (Control problem with this discount function: Jaśkiewicz and Nowak (2021).

An example

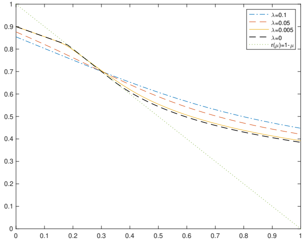
Proposition

Suppose $K \in \left(\frac{2}{(3-\beta)\beta}, \frac{1}{\beta}\right)$. *Denote* $a := \frac{1-K\beta}{1-K\beta/2}$, $b := \frac{1-\beta}{2-\beta}$. *Then, there exists a unique (!) relaxed equilibrium of this example, which is given by*

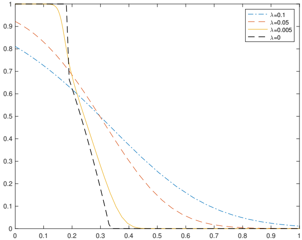
$$\phi_0(\mu) = \begin{cases} 1, & 0 \le \mu \le a, \\ \dfrac{1-\beta-(2-\beta)\mu}{\beta(K-1)(1-\mu)}, & a < \mu \le b, \\ 0, & b < \mu \le 1. \end{cases}$$

Announce the project if the finishing rate is high enough, wait if it is too low, and give your group a chance if it is intermediate.

An example: the convergence as $\lambda \to 0$



(a) Graphs of functions $\tilde{V}_\lambda$ and r.



(b) Graphs of equilibrium stopping strategies ϕ_λ.

Contents

The N-agent problem (N-MDP)

Consider the N-agent problem:

$$X_{k+1}^{i,N,\phi} = T^r\left(X_k^{i,N,\phi}, \frac{1}{N}\sum_{j\in[N]}\delta_{X_k^{i,N,\phi}}, \mathbb{1}_{\{U_{k+1}\le\phi_N(\vec{X}_k^{N,\phi})\}}, Z_{k+1}^i, Z_{k+1}^0\right),$$
$$X_0^{i,N,\phi} = \xi^i,$$

with

$$T^r(x,\mu,a,z',z) = \begin{cases} T_0^r(x,\mu,z',z), & a=0,\\ \triangle_S, & a=1.\end{cases}$$

- "r" stands for **r**epresentative agent.
- $\{Z_k^i\}_{i\in[N],k\in\mathbb{T}}$: the idiosyncratic noises. $\{Z_k^0\}_{k\in\mathbb{T}}$: the common noise. $\{U_k\}_{k\in\mathbb{T}}$: the random device (the coin) **for the social planner** to determine whether to stop or not .

The limit problem (Limit-MDP)

- The total reward of the N-agent problem

$$J^{i,N,\phi}(\xi^i) = \sum_{k=0}^{\infty} \delta(k)\mathbb{E}\left[f\left(X_k^{i,N,\phi}, \frac{1}{N}\sum_{j\in[N]} \delta_{X_k^{i,N,\phi}} \right) \phi_N(\vec{X}_k^{N,\phi}) \right].$$

- Consider the limit as $N \to \infty$, we get the trantition

$$X_{k+1}^{i,\phi} = T^r\left(X_k^{i,\phi}, \mathbb{P}^0_{X_k^{i,\phi}}, \mathbb{1}_{\{U_{k+1}\le\phi(\mathbb{P}^0_{X_k^{i,\phi}})\}}, Z_{k+1}^i, Z_{k+1}^0 \right),$$
$$X_0^{i,\phi} = \xi^i,$$

and the total reward

$$J^{i,\phi}(\xi^i) = \sum_{k=0}^{\infty} \delta(k)\mathbb{E}\left[f\left(X_k^{i,\phi}, \mathbb{P}^0_{X_k^{i,\phi}} \right) \phi(\mathbb{P}^0_{X_k^{i,\phi}}) \right].$$

The convergence result of (N-MDP) $\to$ (Limit-MDP)

Theorem

Under certain technical assumptions (of T_0^r, f and Z^i), for any given $\phi \in \mathcal{F}_S^{\mathrm{Lip}}$, $k \in \mathbb{T}$, $i \in [N]$ and $\lambda > 0$, we have that

$$\mathbb{E} d(X_k^{i,N,\phi}, X_k^{i,\phi}) \leq C_1(k, L_5, \|\phi\|_{\mathrm{Lip}}) M_N,$$
$$\mathbb{E} |J_\lambda^{i,N,\phi}(\xi^i) - J_\lambda^{i,\phi}(\xi^i)| \leq C_2(\lambda, L_5, L_6, L_7, \|\phi\|_{\mathrm{Lip}}) M_N,$$

- $J_\lambda^{i,N,\phi}$ and $J_\lambda^{i,\phi}$ are defined in the same way as $J^{i,N,\phi}$ and $J^{i,\phi}$, with the discount function replaced by $\delta_\lambda := \delta(k)\left(\frac{1}{1+\lambda}\right)^{k^2}$ (only regularize the discount function).
- M_N is the (non-asymptotic) approximation upper bound of empirical measures under Wasserstein metric.

Several remarks on (N-MDP)→ (Limit-MDP)

- We obtain convergence results under fixed Lipschitz policy, which is sufficient due to the regularization (of (MF-MDP)).
- Similar results for open-loop control problems are obtained in Motte and Pham (2022). Because we consider (feed-back) policies, the Lipschitz continuity of ϕ seems indispensable. We achieve such a continuity via regularization.
- The constants before C_1 and C_2 depends on λ and $\|\phi\|_{\text{Lip}}$, both of which blow up when $\lambda \to 0$.
- By introducing δ_λ, we get in exchange an improved convergence rate from M_N^γ ($\gamma \leq 1$) to M_N, comparing to Motte and Pham (2022).

Constructing (MF-MDP) from (Limit-MDP)

We call our original MDP (the one with T_0, r and states μ, e.t.c.) by (MF-MDP).

Proposition

Take $T_0(\mu, z) := T_0^r(\cdot, \mu, \cdot, z)_{\#}(\mu \times \mathcal{L}(\mathcal{Z})')$, $\triangle := \delta_{\triangle_S}$, *and* $r(\mu) := \int_S f(x, \mu)\mu(\mathrm{d}x)$. *Then,* **(Limit-MDP)** *becomes* **(MF-MDP)**.

A remark: If S, the state space of **(N-MDP) or (Limit-MDP)**, is finite, then all technical assumptions are satisfied naturally. But the state space of **(MF-MDP)** is always continuous.

Regularized equilibria of (MF-MDP) as ε-equilibria of (N-MDP)

Theorem

For any $\varepsilon > 0$, ϕ_λ is an ε-equilibrium for **(N-MDP)** *with N agents, provided that λ is sufficiently small and N is sufficiently large.*

Regularized equilibria of (MF-MDP) as ε-equilibria of (N-MDP)

Proof ideas:

- Regularized equilibrium of (MF-MDP):

$$J_\lambda^{\psi\oplus_1\phi_\lambda}(\nu_0) \leq J_\lambda^{\phi_\lambda}(\nu_0).$$

- Regularization error of (N-MDP):

$$\frac{1}{N}\sum_{i\in[N]} |J^{i,N,\psi\oplus_1\phi_\lambda}(\xi^i) - J_\lambda^{i,N,\psi\oplus_1\phi_\lambda}(\xi^i)| < \varepsilon,$$

$$\text{and} \quad \frac{1}{N}\sum_{i\in[N]} |J^{i,N,\phi_\lambda}(\xi^i) - J_\lambda^{i,N,\phi_\lambda}(\xi^i)| < \varepsilon.$$

- Approximation error of (N-MDP) to (Limit-MDP):

$$\frac{1}{N}\sum_{i\in[N]} |J_\lambda^{i,N,\psi\oplus\phi_\lambda}(\xi^i) - J_\lambda^{i,\psi\oplus\phi_\lambda}(\xi^i)| \leq C_3 M_N.$$

A big picture about three MDPs

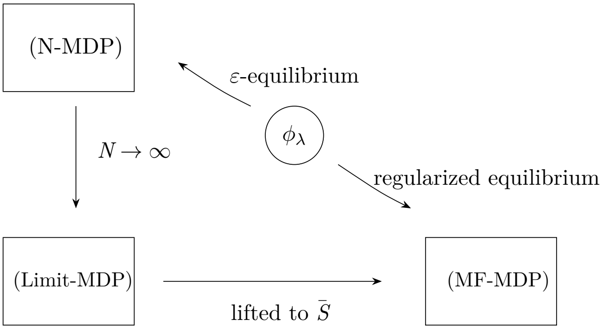


Figure: Relation among different MDP models.

References

Y.-J. Huang and Z. Zhou (2019): The Optimal equilibrium for time-inconsistent stopping problems—the discrete-time case. *SIAM Journal on Control and Optimization*, 57(1), 590-609.

A. Jaśkiewicz and A. S. Nowak (2021): Markov decision processes with quasi-hyperbolic discounting. *Finance and Stochastics*, 25(2), 189-229.

M. Motte and H. Pham (2022): Mean-field Markov decision processes with common noise and open-loop controls. *The Annals of Applied Probability*. 32(20), 1421-1458.

Thank you!

The paper is available at https://arxiv.org/abs/2311.00381.

Contact me at
yfy19@mails.tsinghua.edu.cn or fengyi.yuan@polyu.edu.hk