Discrete-time Markov decision processes under risk-sensitive average cost criterion

Xian Chen
Xiamen University

The 18th Workshop on Markov Processes and Related Topics

Contents

- Risk-sensitive average optimality for discrete-time Markov decision processes, 2023, *SIAM Journal on Control and Optimization*.
- Risk-sensitive average Markov decision processes in general spaces, 2023+

Discrete-time Markov decision processes

Model $\mathcal{G} := \{X, A, \pi_n(\cdot|h_n), Q(\cdot|x,a), c(x,a)\}$.

- state space: X
- action space: A
- admissible state-action pairs: $K = \{(x,a) : x \in X, a \in A(x)\}$
- history: $H_0 := X$, $H_n = (X \times A)^n \times X$ $(n \geq 1)$
- strategy: $\pi_n(\cdot|h_n)$ $n \geq 0$ stochastic kernels on A given H_n
- transition law: $Q(\cdot|x,a)$ stochastic kernel on X given K
- cost function: $c(x,a)$

Discrete-time Markov decision processes

Model $\mathcal{G} := \{X, A, \pi_n(\cdot|h_n), Q(\cdot|x,a), c(x,a)\}$.

- state space: X
- action space: A
- admissible state-action pairs: $K = \{(x,a) : x \in X, a \in A(x)\}$
- history: $H_0 := X$, $H_n = (X \times A)^n \times X$ $(n \geq 1)$
- strategy: $\pi_n(\cdot|h_n)$ $n \geq 0$ stochastic kernels on A given H_n
- transition law: $Q(\cdot|x,a)$ stochastic kernel on X given K
- cost function: $c(x,a)$

$$x_0 \xrightarrow{\pi_0(\cdot|x_0) \to Q(\cdot|x_0,a_0)} x_1 \xrightarrow{\pi_1(\cdot|x_0,a_0,x_1) \to Q(\cdot|x_1,a_1)} x_2 \cdots$$

Discrete-time Markov decision processes

Model $\mathcal{G} := \{X, A, \pi_n(\cdot|h_n), Q(\cdot|x,a), c(x,a)\}$.

- state space: X
- action space: A
- admissible state-action pairs: $K = \{(x,a) : x \in X, a \in A(x)\}$
- history: $H_0 := X$, $H_n = (X \times A)^n \times X$ $(n \geq 1)$
- strategy: $\pi_n(\cdot|h_n)$ $n \geq 0$ stochastic kernels on A given H_n
- transition law: $Q(\cdot|x,a)$ stochastic kernel on X given K
- cost function: $c(x,a)$

$x_0 \quad {}^{\pi_0(\cdot|x_0)}$

Discrete-time Markov decision processes

Model $\mathcal{G} := \{X, A, \pi_n(\cdot|h_n), Q(\cdot|x,a), c(x,a)\}$.

- state space: X
- action space: A
- admissible state-action pairs: $K = \{(x,a) : x \in X, a \in A(x)\}$
- history: $H_0 := X$, $H_n = (X \times A)^n \times X$ $(n \geq 1)$
- strategy: $\pi_n(\cdot|h_n)$ $n \geq 0$ stochastic kernels on A given H_n
- transition law: $Q(\cdot|x,a)$ stochastic kernel on X given K
- cost function: $c(x,a)$

$$x_0 \xrightarrow{\pi_0(\cdot|x_0) \to Q(\cdot|x_0, a_0)} x_1$$

Discrete-time Markov decision processes

Model $\mathcal{G} := \{X, A, \pi_n(\cdot|h_n), Q(\cdot|x, a), c(x, a)\}$.

- state space: X
- action space: A
- admissible state-action pairs: $K = \{(x, a) : x \in X, a \in A(x)\}$
- history: $H_0 := X$, $H_n = (X \times A)^n \times X$ $(n \geq 1)$
- strategy: $\pi_n(\cdot|h_n)$ $n \geq 0$ stochastic kernels on A given H_n
- transition law: $Q(\cdot|x, a)$ stochastic kernel on X given K
- cost function: $c(x, a)$

$$x_0 \xrightarrow{\pi_0(\cdot|x_0) \to Q(\cdot|x_0, a_0)} x_1 \xrightarrow{\pi_1(\cdot|x_0, a_0, x_1) \to Q(\cdot|x_1, a_1)} x_2 \cdots$$

Discrete-time Markov decision processes

Model $\mathcal{G} := \{X, A, \pi_n(\cdot|h_n), Q(\cdot|x,a), c(x,a)\}$.

- state space: X
- action space: A
- admissible state-action pairs: $K = \{(x,a) : x \in X, a \in A(x)\}$
- history: $H_0 := X$, $H_n = (X \times A)^n \times X$ $(n \geq 1)$
- strategy: $\pi_n(\cdot|h_n)$ $n \geq 0$ stochastic kernels on A given H_n
- transition law: $Q(\cdot|x,a)$ stochastic kernel on X given K
- cost function: $c(x,a)$

$$x_0 \xrightarrow{\pi_0(\cdot|x_0) \to Q(\cdot|x_0,a_0)} x_1 \quad {}^{\pi_1(\cdot|x_0,a_0,x_1)}$$

Discrete-time Markov decision processes

Model $\mathcal{G} := \{X, A, \pi_n(\cdot|h_n), Q(\cdot|x,a), c(x,a)\}$.

- state space: X
- action space: A
- admissible state-action pairs: $K = \{(x,a) : x \in X, a \in A(x)\}$
- history: $H_0 := X$, $H_n = (X \times A)^n \times X$ $(n \geq 1)$
- strategy: $\pi_n(\cdot|h_n)$ $n \geq 0$ stochastic kernels on A given H_n
- transition law: $Q(\cdot|x,a)$ stochastic kernel on X given K
- cost function: $c(x,a)$

$$x_0 \xrightarrow{\pi_0(\cdot|x_0) \to Q(\cdot|x_0,a_0)} x_1 \xrightarrow{\pi_1(\cdot|x_0,a_0,x_1) \to Q(\cdot|x_1,a_1)} x_2 \cdots$$

Strategy

- Randomized history-dependent strategy: $\pi_n(\cdot|h_n)$ $n \geq 0$ stochastic kernels on A given H_n.
- Markov strategy: for any $n \geq 0$, if there exists a stochastic kernel ϕ_n such that $\pi_n(\cdot|h_n) = \phi_n(\cdot|x_n)$ for all $h_n \in H_n$.
- Stationary Markov strategy: if there exists a stochastic kernel ϕ such that $\pi_n(\cdot|h_n) = \phi(\cdot|x_n)$ for all $h_n \in H_n$ and $n \geq 0$.
- Deterministic stationary Markov strategy: if there exists a mapping $f : X \to A$ with $f(x) \in A(x)$ for all $x \in X$, such that $\pi_n(\cdot|h_n) = \delta_{f(x_n)}(\cdot)$ for all $h_n \in H_n$ and $n \geq 0$.

Optimality Criteria

Classical expected criteria:

- expected discounted payoff $J(x,\pi) := E_x^\pi\left[\sum_{t=0}^{\infty} \alpha^t c(x_t, a_t)\right]$
- expected finite horizon (for any fixed T) payoff
  $J(x,\pi) := E_x^\pi\left[\sum_{t=0}^{T-1} c(x_t, a_t) + g(X_T)\right]$
- expected average payoff $J(x,\pi) := \limsup_{n\to\infty} \frac{1}{n} E_x^\pi\left[\sum_{t=0}^{n-1} c(x_t, a_t)\right]$

Optimality Criteria

Classical expected criteria:

- expected discounted payoff $J(x,\pi) := E_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t c(x_t, a_t)\right]$
- expected finite horizon (for any fixed T) payoff $J(x,\pi) := E_x^\pi \left[\sum_{t=0}^{T-1} c(x_t, a_t) + g(X_T)\right]$
- expected average payoff $J(x,\pi) := \limsup_{n\to\infty} \frac{1}{n} E_x^\pi \left[\sum_{t=0}^{n-1} c(x_t, a_t)\right]$

Risk-sensitive criteria:

- risk-sensitive discounted payoff $J(x,\pi) := E_x^\pi \left[e^{\lambda \sum_{t=0}^{\infty} \alpha^t c(x_t, a_t)}\right]$
- risk-sensitive finite horizon (for any fixed T) payoff $J(x,\pi) := E_x^\pi \left[e^{\lambda(\sum_{t=0}^{T-1} c(x_t, a_t) + g(X_T))}\right]$
- risk-sensitive average payoff $J(x,\pi) := \limsup_{n\to\infty} \frac{1}{n\lambda} \ln E_x^\pi \left[e^{\lambda \sum_{t=0}^{n-1} c(x_t, a_t)}\right]$

$$\lambda : \text{risk-sensitivity coefficient} \begin{cases} \lambda > 0, \text{risk-averse} \\ \lambda < 0, \text{risk-seeking} \end{cases}$$

Optimal strategy

Definition

A strategy $\pi^* \in \Pi$ *is said to be optimal for model* $\mathcal{G}$ *if*

$$J(x, \pi^*) \leq J(x, \pi)$$

for all $x \in X$ *and* $\pi \in \Pi$.

Literature

Classical expected criteria: Puterman (1994), Hernández&Lasserre (1996, 1999), Bertsekas (2005), Bäuerle&Rieder (2011),...

Literature

Classical expected criteria: Puterman (1994), Hernández&Lasserre (1996, 1999), Bertsekas (2005), Bäuerle&Rieder (2011),...

Risk-sensitive average criterion:

- Howard&Matheson (1972 Management Sci.)
- Hernández-Hernández&Marcus (1996 SCL), Borkar&Meyn (2002 MOR), Cavazos-Cadena (2018 MOR), Biswas&Pradhan (2022 ESAIM), Saucedo-Zul et al. (2020 JOTA)
- Di Masi&Stettner (1999, 2007 SICON; 2000 SCL), Jaśkiewicz (2007 AAP; 2007 SCL), Stettner (2020 SICON, 2021 AMO)
- Biswas&Borkar (2023+)

Literature

Classical expected criteria: Puterman (1994), Hernández&Lasserre (1996, 1999), Bertsekas (2005), Bäuerle&Rieder (2011),...

Risk-sensitive average criterion:

- Howard&Matheson (1972 Management Sci.)
- Hernández-Hernández&Marcus (1996 SCL), Borkar&Meyn (2002 MOR), Cavazos-Cadena (2018 MOR), Biswas&Pradhan (2022 ESAIM), Saucedo-Zul et al. (2020 JOTA)
- Di Masi&Stettner (1999, 2007 SICON; 2000 SCL), Jaśkiewicz (2007 AAP; 2007 SCL), Stettner (2020 SICON, 2021 AMO)
- Biswas&Borkar (2023+)

Risk-sensitive analysis: Shen&Stannat&Obermayer (2013 SICON), Bäuerle&Rieder (2014 MOR), Bäuerle&Glauner (2021 EJOR)

Example

Example 1

The control model is given as follows: $X = \{0, 1, 2, \ldots\}$, $A = \{0\}$, $Q(1|0,0) = p$, $Q(0|0,0) = 1 - p$, $Q(i+1|i,0) = p$, $Q(i-1|i,0) = 1 - p$ for all $i \geq 1$, $c(0,0) = \varpi$ and $c(i,0) = 0$ for all $i \geq 1$, where the constants $p \in (0, \frac{1}{2})$ and $\varpi < \ln \frac{1}{2\sqrt{p(1-p)}}$. Take the risk-sensitivity parameter $\lambda = 1$.

Example

Example 2

The controlled linear Gaussian system is given by $x_{t+1} = Ux_t + Wa_t + \xi_t$ for all $t \geq 0$, where the state $x_t \in \mathbb{R}^n$, the action $a_t \in \mathbb{R}^m$, the matrices $U \in \mathbb{R}^{n\times n}$, $W \in \mathbb{R}^{n\times m}$ and the Gaussian white noise ξ_t is i.i.d. with $\xi_t \sim N(0, \Sigma)$. We assume that the rank of matrices U and Σ equals n and $\gamma := \|U\|^2 < 1$. Let $l(x) := \frac{(1-\gamma)\delta}{2} x^T U^T Ux + 1$ for all $x \in X := \mathbb{R}^n$ for some δ satisfying some condition.

(E1) For each $x \in X$, $A(x)$ is compact and A is compact.

(E2) $l(\cdot) - \sup_{a\in A(\cdot)} \lambda c(\cdot, a)$ is coercive.

(E3) For any compact set $C \subset X$, $\sup_{x\in C, a\in A} c(x, a) < \infty$.

(E4) For each $x \in X$, $c(x, a)$ is lower semi-continuous in $a \in A(x)$.

Model

- state space: X countable set/Borel space
- action space: A Borel space
- admissible state-action pairs: $K = \{(x, a) : x \in X, a \in A(x)\}$
- history: $H_0 := X$, $H_n = (X \times A)^n \times X$ $(n \geq 1)$
- strategy: $\pi_n(\cdot|h_n)$ $n \geq 0$ stochastic kernels on A given H_n.
- transition law: $Q(\cdot|x, a)$ stochastic kernel on X
- cost function: $c(x, a)$

Model $\mathcal{G} := \{X, A, \pi_n(\cdot|h_n), c(x, a), Q(\cdot|x, a)\}$.

Risk-sensitive average payoff $J(x, \pi) := \limsup_{n\to\infty} \frac{1}{n\lambda} \ln E_x^\pi \left[e^{\lambda \sum_{t=0}^{n-1} c(x_t, a_t)}\right]$.

Multiplicative ergodic theory

Theorem (Balaji&Meyn 2000 SPA)

Suppose that $\{X_k\}$ is an irreducible and aperiodic Markov chain with countable state space X, and that the sublevel set $\{x : F(x) \leq n\}$ is finite for each n. Suppose that there exists $V : X \to [1, \infty)$, a finite set C and a constant $b < \infty$, satisfying

$$\sum_{y \in X} V(y) P(x, y) \leq e^{-F(x)} V(x) + b I_C(x).$$

Then there exist a function $\check{F} : X \to R$ and a constant $\Lambda > 0$ such that
(i) $\Lambda = \lim_{n \to \infty} \frac{1}{n} \ln E_x \left[e^{\sum_{t=0}^{n-1} F(X_t)} \right]$
(ii) $\check{F}(x) = \lim_{n \to \infty} E_x \left[e^{\sum_{t=0}^{n-1} (F(X_t) - \Lambda)} \right]$
(iii) $(\check{F}, \Lambda)$ *solves the multiplicative Poisson equation*

$$e^F P \check{F} = e^\Lambda \check{F}.$$

Multiplicative ergodic theory

Theorem (Balaji&Meyn 2000 SPA)

Suppose that $\{X_k\}$ is an irreducible and aperiodic Markov chain with countable state space X, and that the sublevel set $\{x : F(x) \leq n\}$ is finite for each n. Suppose that there exists $V : X \to [1, \infty)$, a finite set C and a constant $b < \infty$, satisfying

$$\sum_{y \in X} V(y)P(x,y) \leq e^{-F(x)}V(x) + bI_C(x).$$

Then there exist a function $\check{F} : X \to R$ and a constant $\Lambda > 0$ such that
(i) $\Lambda = \lim_{n\to\infty} \frac{1}{n} \ln E_x \left[e^{\sum_{t=0}^{n-1} F(X_t)} \right]$
(ii) $\check{F}(x) = \lim_{n\to\infty} E_x \left[e^{\sum_{t=0}^{n-1} (F(X_t) - \Lambda)} \right]$
(iii) $(\check{F}, \Lambda)$ *solves the multiplicative Poisson equation*

$$e^F P \check{F} = e^\Lambda \check{F}.$$

(iv) For any fixed $z \in X$, Λ is the unique solution to

$$E_z \left[e^{\sum_{t=0}^{\tau_z - 1} (F(X_t) - \Lambda)} \right] = 1,$$

and $\check{F}(x) = E_x \left[e^{\sum_{t=0}^{\tau_z - 1} (F(X_t) - \Lambda)} \right]$.

Assumptions

Assumption

(i) *For any* $f \in F$*, the Markov chain associated with the transition law* $Q(\cdot|\cdot, f(\cdot))$ *is aperiodic and irreducible.*

(ii) *For each* $i \in X$*, the set* $A(i)$ *is compact. Moreover,* $c(i,\cdot)$ *and* $Q(j|i,\cdot)$ *are lower semi-continuous on* $A(i)$ *for all* $i, j \in X$*.*

(iii) *There exist a real-valued function* $w \geq 1$ *on* X*, a norm-like function* $l \geq 0$ *on* X*, a constant* $d > 0$ *and a finite set* $C \subseteq X$ *such that*

$$\sum_{j \in S} w(j) Q(j|i,a) \leq e^{-l(i)} w(i) + d I_C(i)$$

for all $(i,a) \in K$*. Moreover,* $l(\cdot) - \sup_{a \in A(\cdot)} \lambda c(\cdot, a)$ *is norm-like.*

Auxiliary functions

Fix $z \in C$. For any $i \in X$, $f \in F$, and $\rho \in \mathbb{R}^+ := [0, \infty)$, the risk-sensitive first passage function is defined by

$$v(i, f, \rho) := E_i^f \left[e^{\lambda \sum_{t=0}^{\tau_z - 1} (c(x_t, f(x_t)) - \rho)} \right].$$

Auxiliary functions

Fix $z \in C$. For any $i \in X$, $f \in F$, and $\rho \in \mathbb{R}^+ := [0, \infty)$, the risk-sensitive first passage function is defined by

$$v(i, f, \rho) := E_i^f \left[e^{\lambda \sum_{t=0}^{\tau_z - 1} (c(x_t, f(x_t)) - \rho)} \right].$$

For each $\rho \in \mathbb{R}^+$ and $i \in X$, set

$$v^*(i, \rho) := \inf_{f \in F} v(i, f, \rho),$$

which is referred to as the risk-sensitive first passage optimal value function.

Auxiliary functions

Fix $z \in C$. For any $i \in X$, $f \in F$, and $\rho \in \mathbb{R}^+ := [0, \infty)$, the risk-sensitive first passage function is defined by

$$v(i, f, \rho) := E_i^f \left[e^{\lambda \sum_{t=0}^{\tau_z - 1} (c(x_t, f(x_t)) - \rho)} \right].$$

For each $\rho \in \mathbb{R}^+$ and $i \in X$, set

$$v^*(i, \rho) := \inf_{f \in F} v(i, f, \rho),$$

which is referred to as the risk-sensitive first passage optimal value function. Moreover, define

$$\mathbb{G} := \{\rho \in \mathbb{R}^+ : v^*(z, \rho) \leq 1\}, \ \rho^* := \inf \mathbb{G}.$$

Main Result

Main Theorem

Under the above Assumptions, the following statements are true.

(a) *There exists a unique positive function* u^* *on* X *with* $u^*(z) = 1$ *such that*

$$u^*(i) = \inf_{a \in A(i)} \left\{ e^{\lambda(c(i,a)-\rho^*)} \sum_{j \in S} u^*(j) Q(j|i,a) \right\} \tag{1}$$

for all $i \in X$. *Moreover, we have* $u^*(i) = v^*(i, \rho^*)$ *for all* $i \in X$.

(b) *There exists* $f^* \in F$ *with* $f^*(i) \in A(i)$ *attaining the infimum in (1) and* $\rho^* = J(i, f^*) = \inf_{\pi \in \Pi} J(i, \pi)$ *for all* $i \in X$.

(c) *A stationary policy* $f \in F$ *is optimal if and only if*

$$e^{\lambda(c(i,f(i))-\rho^*)} \sum_{j \in S} v^*(j, \rho^*) Q(j|i, f(i)) = \inf_{a \in A(i)} \left\{ e^{\lambda(c(i,a)-\rho^*)} \sum_{j \in S} v^*(j, \rho^*) Q(j|i,a) \right\}$$

for all $i \in X$.

Proposition

For any $f \in F$, $(\rho^f, v(i, f, \rho^f))$ solves the multiplicative Poisson equation

$$v(i, f, \rho^f) = e^{\lambda(c(i,f(i)) - \rho^f)} \sum_{j \in X} v(j, f, \rho^f) Q(j|i, f(i)).$$

For $n \geq 2$, define

$$c_n(i, a) := c(i, a) + \frac{1}{n} \left[l(i) - \max_{a \in A(i)} \lambda c(i, a) \right] \vee 0 \text{ for all } (i, a) \in K.$$

We need to introduce the new transition law as follows: for any $n \geq 2$ and $i, j \in X$,

$$\begin{aligned}&\widetilde{Q}_n^f(j|i)\\&:= \frac{1}{v_n(i, f, \rho_n^f)} e^{\lambda(c_n(i,f(i))-\rho_n^f)} Q(j|i, f(i)) v_n(j, f, \rho_n^f)\end{aligned}$$

Key Lemma

There exist a subsequence of $\{n\}$ (denoted by the same sequence) and a constant $R > 1$ such that $\sup_{n\geq 1} \widetilde{E}_z^{f,n}[R^{\tau_z}] < \infty$.

Multiplicative ergodic theory

Theorem (Balaji&Meyn 2000 SPA)

Suppose that $\{X_k\}$ is an irreducible and aperiodic Markov chain with countable state space X, and that the sublevel set $\{x : F(x) \leq n\}$ is finite for each n. Suppose that there exists $V : X \to [1, \infty)$, a finite set C and a constant $b < \infty$, satisfying

$$\sum_{y \in X} V(y)P(x,y) \leq e^{-F(x)}V(x) + bI_C(x).$$

Then there exist a function $\check{F} : X \to R$ and a constant $\Lambda > 0$ such that
(i) $\Lambda = \lim_{n\to\infty} \frac{1}{n} \ln E_x \left[e^{\sum_{t=0}^{n-1} F(X_t)} \right]$
(ii) $\check{F}(x) = \lim_{n\to\infty} E_x \left[e^{\sum_{t=0}^{n-1} (F(X_t)-\Lambda)} \right]$
(iii) $(\check{F}, \Lambda)$ solves the multiplicative Poisson equation

$$e^F P\check{F} = e^\Lambda \check{F}.$$

(iv) For any fixed $z \in X$, Λ is the unique solution to

$$E_z \left[e^{\sum_{t=0}^{\tau_z-1} (F(X_t)-\Lambda)} \right] = 1,$$

and $\check{F}(x) = E_x \left[e^{\sum_{t=0}^{\tau_z-1} (F(X_t)-\Lambda)} \right]$.

Multiplicative ergodic theory

Theorem (Balaji&Meyn 2000 SPA)

Suppose that $\{X_k\}$ is an irreducible and aperiodic Markov chain with countable state space X, and that the sublevel set $\{x : F(x) \leq n\}$ is finite for each n. Suppose that there exists $V : X \to [1, \infty)$, a finite set C and a constant $b < \infty$, satisfying

$$\sum_{y \in X} V(y)P(x, y) \leq e^{-F(x)}V(x) + bI_C(x).$$

Then there exist a function $\check{F} : X \to R$ and a constant $\Lambda > 0$ such that
(i) $\Lambda = \lim_{n\to\infty} \frac{1}{n} \ln E_x \left[e^{\sum_{t=0}^{n-1} F(X_t)} \right]$
(ii) $\check{F}(x) = \lim_{n\to\infty} E_x \left[e^{\sum_{t=0}^{n-1} (F(X_t)-\Lambda)} \right]$
(iii) $(\check{F}, \Lambda)$ *solves the multiplicative Poisson equation*

$$e^F P\check{F} = e^\Lambda \check{F}.$$

(iv) For any fixed $z \in X$, Λ is the unique solution to

$$E_z \left[e^{\sum_{t=0}^{\tau_z -1} (F(X_t)-\Lambda)} \right] = 1,$$

and $\check{F}(x) = E_x \left[e^{\sum_{t=0}^{\tau_z -1} (F(X_t)-\Lambda)} \right]$.

Kontoyiannis&Meyn 2003 AAP, 2005 EJP; Wu 2004 PTRF; Hennion 2007 PTRF

Let $\mathcal{B}$ be an abstract Banach space, $\mathcal{L}(\mathcal{B})$ is the Banach algebra of bounded operators on $\mathcal{B}$, and $Q \in \mathcal{L}(\mathcal{B})$. We denote by $r(Q)$ the spectral radius of Q, and by $Q|_G$ its restriction to a Q-invariant subspace G.

Definition

(i) The essential spectral radius of $Q \in \mathcal{L}(\mathcal{B})$, denoted by $r_e(Q)$, is the infimum of $r(Q)$ and of the real number $\rho \geq 0$ such that we have

$$\mathcal{B} = F_\rho \oplus H_\rho,$$

where F_ρ and H_ρ are Q-invariant subspaces such that H_ρ is closed and $r(Q_{H_\rho}) < \rho$, $dim F_\rho < \infty$ and the eigenvalues of Q_{F_ρ} have a modulus $\geq \rho$.
(ii) When $r_e(Q) < r(Q)$, the operator Q is said to be quasi-compact.

Gelfand's formula

Let $\mathcal{K}(\mathcal{B})$ be the ideal of compact operators on $\mathcal{B}$. For any $Q \in \mathcal{L}(\mathcal{B})$, we have

$$r_e(Q) = \lim_{n\to\infty} \left(\inf\{||Q^n - V|| : V \in \mathcal{K}(\mathcal{B})\}\right)^{1/n}.$$

Let Q be a bounded positive kernel on $(X, \mathcal{X})$. For any positive measurable g and $x \in X$, define $Qg(x) := \int_X g(y)Q(x, dy)$. Then the kernel Q defines a positive bounded operator on the Banach space of bounded measurable complex valued functions on $(X, \mathcal{X})$ equipped with the supremum norm.

Theorem (Hennion 2007 PTRF)

Assume that there exist a probability measure ν and a positive measurable function α on $(X \times X, \mathcal{X} \otimes \mathcal{X})$, such that the functions $\alpha(x, \cdot)$, $x \in X$, are uniformly ν-integrable. Define the bounded positive kernel $T_{\nu,\alpha}$ as

$$T_{\nu,\alpha}(x, A) := \int_A \alpha(x, y)\nu(dy), \ (x, A) \in X \times \mathcal{X}.$$

If $S = Q - T_{\nu,\alpha} \geq 0$ and $r(S) < r(Q)$, then the operator Q is quasi-compact.

Assumption

(i) *For any* $f \in F$*, the Markov chain associated with the transition law* $Q(\cdot|\cdot, f(\cdot))$ *is aperiodic and irreducible.*

(ii) *For each* $x \in X$*,* $A(x)$ *is compact,* $c(x, a)$ *is lower semi-continuous in* $a \in A(x)$ *and* $\int_X u(y)Q(dy|x, a)$ *is continuous in* $a \in A(x)$ *for all* $u \in B_b(X)$*.*

(iii) *There exist a real-valued measurable function* $w \geq 1$ *on* X*, a norm-like function* $l \geq 0$ *on* X*, a constant* $d > 0$ *and a set* $C \subseteq X$ *such that*

$$\int_X w(y)Q(dy|x, a) \leq e^{-l(x)}w(x) + dI_C(x) \text{ for all } (x, a) \in K.$$

Moreover, $l(\cdot) - \sup_{a \in A(\cdot)} \lambda c(\cdot, a)$ *is coercive.*

Assumption

(iv) *There exist a probability measure* ν_1 *on* $\mathcal{B}(X)$ *and a nonnegative real-valued measurable function* q *on* $K \times X$ *such that* $Q(dy|x,a) = q(x,a,y)\nu_1(dy)$ *for all* $(x,a) \in K$*. For each* $f \in F$, $\{q(x, f(x), \cdot), x \in C\}$ *is uniformly integrable with respect to the measure* ν_1*.*

(v) *There exist a probability measure* ν_2 *on* $\mathcal{B}(X)$*, a positive integer* n_0 *and a constant* $\beta \in (0,1)$ *such that* $Q^{n_0}(\cdot|x,f) \geq \beta I_C(x)\nu_2(\cdot)$ *for all* $x \in X$ *and* $f \in F$*.*

Main Result

Main Theorem

Under the above Assumptions, the following statements are true.

(a) *There exist a constant* $\eta^* \geq 1$, *a positive measurable function* $u^* \in B_w(X) := \{u : \sup_{x\in X} \frac{|u(x)|}{w(x)} < \infty\}$ *and* $f^* \in F$ *such that for all* $x \in X$,

$$\eta^* u^*(x) = \inf_{a\in A(x)} \left\{ e^{\lambda c(x,a)} \int_X u^*(y) Q(dy|x,a) \right\} = e^{\lambda c(x,f^*)} \int_X u^*(y) Q(dy|x,f^*).$$

(b) *The policy* $f^* \in F$ *in part (a) is optimal and* $\frac{1}{\lambda} \ln \eta^* = J(x, f^*) = \inf_{\pi\in\Pi} J(x,\pi)$ *for all* $x \in X$.

Define $\widehat{s}(x) := \frac{\beta I_C(x)}{2\sup_{x\in C} w(x)}$, $\widetilde{Q}_f^w(x, dy) := \frac{1}{w(x)} e^{\lambda c(x,f)} w(y) Q(dy|x, f)$, $\widehat{Q}_f(dy|x) := \widetilde{Q}_f^{w,n_0}(dy|x) - \widehat{s}(x)\nu_2(dy)$ and $v_f(x) := w(x) \sum_{k=0}^{\infty} \eta_f^{-(k+1)n_0} \widehat{Q}_f^k \widehat{s}(x)$ for any $x \in X$. Denote by η^f the spectral radius of $\widetilde{Q}_f^w$.

Proposition

For any $f \in F$, $(\eta^f, v_f(x))$ solves the multiplicative Poisson equation

$$\eta^f v_f(x) = e^{\lambda c(x,f(x))} \int_{x\in X} v_f(y) Q(dy|x, f).$$

Define $\widehat{s}(x) := \frac{\beta I_C(x)}{2\sup_{x\in C} w(x)}$, $\widetilde{Q}_f^w(x,dy) := \frac{1}{w(x)} e^{\lambda c(x,f)} w(y) Q(dy|x,f)$, $\widehat{Q}_f(dy|x) := \widetilde{Q}_f^{w,n_0}(dy|x) - \widehat{s}(x)\nu_2(dy)$ and $v_f(x) := w(x)\sum_{k=0}^{\infty} \eta_f^{-(k+1)n_0} \widehat{Q}_f^k \widehat{s}(x)$ for any $x \in X$. Denote by η^f the spectral radius of $\widetilde{Q}_f^w$.

Proposition

For any $f \in F$, $(\eta^f, v_f(x))$ solves the multiplicative Poisson equation

$$\eta^f v_f(x) = e^{\lambda c(x,f(x))} \int_{x\in X} v_f(y) Q(dy|x,f).$$

Main idea:
Prove that $\widetilde{Q}_f^w$ is quasi-compact on the Banach space of all bounded measurable functions on $(X,\mathcal{X})$.

Example

Example 1

The control model is given as follows: $X = \{0, 1, 2, \ldots\}$, $A = \{0\}$, $Q(1|0, 0) = p$, $Q(0|0, 0) = 1 - p$, $Q(i + 1|i, 0) = p$, $Q(i - 1|i, 0) = 1 - p$ for all $i \geq 1$, $c(0, 0) = \varpi$ and $c(i, 0) = 0$ for all $i \geq 1$, where the constants $p \in (0, \frac{1}{2})$ and $\varpi < \ln \frac{1}{2\sqrt{p(1-p)}}$. Take the risk-sensitivity parameter $\lambda = 1$.

This example is given to illustrate the facts:
(1) the key assumption in Jaśkiewicz (2007 AAP) fails to hold;
(2) the near-monotone condition in Hernández-Hernández&Marcus (1999 AMO) fails to hold;
(3) the set of states in which the limit point of the discount relative function is finite in Cavazos-Cadena&Salem-Silva (2010 AMO) is empty.

Example

Example 2

The controlled linear Gaussian system is given by $x_{t+1} = Ux_t + Wa_t + \xi_t$ for all $t \geq 0$, where the state $x_t \in \mathbb{R}^n$, the action $a_t \in \mathbb{R}^m$, the matrices $U \in \mathbb{R}^{n\times n}$, $W \in \mathbb{R}^{n\times m}$ and the Gaussian white noise ξ_t is i.i.d. with $\xi_t \sim N(0, \Sigma)$. We assume that the rank of matrices U and Σ equals n and $\gamma := \|U\|^2 < 1$. Let $l(x) := \frac{(1-\gamma)\delta}{2} x^T U^T Ux + 1$ for all $x \in X := \mathbb{R}^n$ for some δ satisfying some condition.

(E1) For each $x \in X$, $A(x)$ is compact and A is compact.

(E2) $l(\cdot) - \sup_{a\in A(\cdot)} \lambda c(\cdot, a)$ is coercive.

(E3) For any compact set $C \subset X$, $\sup_{x\in C, a\in A} c(x, a) < \infty$.

(E4) For each $x \in X$, $c(x, a)$ is lower semi-continuous in $a \in A(x)$.

Policy Iteration Algorithm

1. (Initialization) Set $k = 0$ and select any stationary policy $f_0 \in F$.
2. (Policy evaluation) For the policy f_k, the function v_k on X and the constant $e^{\lambda \rho_k}$ are the unique solution to the multiplicative Poisson equation satisfying

$$v_k(z) = 1 \text{ and } e^{\lambda \rho_k} v_k(i) = \sum_{j \in X} v_k(j) e^{\lambda c(i, f_k(i))} Q(j|i, f_k(i)) \text{ for all } i \in X.$$

3. (Policy improvement) Choose f_{k+1} to satisfy

$$f_{k+1}(i) \in \operatorname{argmin}_{a \in A(i)} \left\{ e^{\lambda c(i,a)} \sum_{j \in X} v_k(j) Q(j|i, a) \right\} \text{ for all } i \in X,$$

setting $f_{k+1} = f_k$ if possible.
4. If $f_{k+1} = f_k$, stop and set $f^* = f_k$. Otherwise, let $k \leftarrow k + 1$ and return to step 2.

Policy Iteration Algorithm

1. (Initialization) Set $k = 0$ and select any stationary policy $f_0 \in F$.
2. (Policy evaluation) For the policy f_k, the function v_k on X and the constant $e^{\lambda \rho_k}$ are the unique solution to the multiplicative Poisson equation satisfying

$$v_k(z) = 1 \text{ and } e^{\lambda \rho_k} v_k(i) = \sum_{j \in X} v_k(j) e^{\lambda c(i, f_k(i))} Q(j|i, f_k(i)) \text{ for all } i \in X.$$

3. (Policy improvement) Choose f_{k+1} to satisfy

$$f_{k+1}(i) \in \operatorname{argmin}_{a \in A(i)} \left\{ e^{\lambda c(i,a)} \sum_{j \in X} v_k(j) Q(j|i, a) \right\} \text{ for all } i \in X,$$

setting $f_{k+1} = f_k$ if possible.
4. If $f_{k+1} = f_k$, stop and set $f^* = f_k$. Otherwise, let $k \leftarrow k + 1$ and return to step 2.

Theorem

$\lim_{k \to \infty} \rho_k = \rho^*$ *and* $\lim_{k \to \infty} v_k(i) = v^*(i, \rho^*)$ *for all* $i \in X$.

For any $i, j \in X$ and $k \geq 0$, define

$$p_{i,j}(f_k) := \frac{e^{\lambda(c(i,f_k(i))-\rho_k)}Q(j|i, f_k(i))v_k(j)}{v_k(i)} \text{ and } \mu_i(f_k) := \frac{w(i)}{v_k(i)}.$$

Denote by $\widehat{P}_i^k$ the probability measure associated with the transition law $p_{\cdot,\cdot}(f_k)$ for any initial state $i \in X$.

Key Lemma

There exist constants $\alpha \in (0,1)$ and $L^* > 0$ such that $\sum_{j \in S} \mu_j(f_k) \left| \widehat{P}_i^k(i_t = j) - \nu_k(j) \right| \leq L^* \alpha^t \mu_i(f_k)$ for all $i \in X$, $k \geq 1$ and $t \geq 1$.

Remark

(a) We obtain optimality equation without compact support condition on the cost and simultaneous Doeblin condition in Cavazos-Cadena (2018 MOR), without the boundedness condition on the cost in Masi&Stettner (1999, 2007 SICON; 2000 SCL) and Jaśkiewicz (2007 SCL) and without the requirement that there exists a state $i' \in X$ such that $Q(j|i', a) > 0$ for all $j \in X \setminus \{i'\}$ and $a \in A(i')$ in Biswas&Pradhan (2022 ESAIM).

(b) Our approach is different from the technique of using the nonlinear version of Krein-Rutman theorem in Biswas&Pradhan (2022 ESAIM).

(c) We prove the convergence of policy iteration algorithm under conditions different from Biswas&Pradhan (2022 ESAIM) and Borkar&Meyn (2002 MOR).
(c1) We do not require the following conditions in Biswas&Pradhan (2022 ESAIM): (1) there exists a state $i' \in X$ such that $Q(j|i', a) > 0$ for all $j \in X \setminus \{i'\}$ and $a \in A(i')$; (2) there exists a constant $\zeta \in (0, 1)$ such that $\max_{i \in A(i)} c(i, a) \leq \zeta l(i)$ for all $i \in X$; (3) there exists a state $j' \in X$ such that $inf_{a \in A(i)} Q(j'|i, a) > 0$ for all $i \in X$.
(c2) We do not require the norm-like condition on the cost and some less easily verifiable conditions in Borkar&Meyn (2002 MOR).

Thank you !