On Markov decision processes with a cemetery

Yi Zhang

School of Mathematics, University of Birmingham
y.zhang.29@bham.ac.uk

1st August 2023, 18th Workshop on Markov Processes and Related Topics, Tianjin, PR China

References

This is based on joint work with Alexey Piunovskiy (University of Liverpool):

- Piunovskiy, A. and Z. (2023). Extreme occupation measures in Markov decision processes with a cemetery. arXiv:2307.03158
- Piunovskiy, A. and Z. (2023). On the structure of optimal solutions in a mathematical programming problem in a convex space. *Oper. Res. Lett.*, to appear.

MDP model

In an MDP model, one controls a Markov chain through its transition probabilities, and at each step, depending on the current state X_n and action A_n, some losses are incurred.

So an MDP model is $\{\mathbf{X}, \mathbf{A}, p, \{c_j\}_{j=0}^J\}$:

- $\mathbf{X}$: state space, assumed to be a Borel space; $\mathbf{A}$: action space, assumed to be a Borel space;
- $p(dy|x, a)$: transition probability;
- c_j, $j \in \{0, \dots, J\}$, are $[0, \infty]$-valued measurable functions on $\mathbf{X} \times \mathbf{A}$, with $J \geq 0$ being an integer.

A strategy $\pi = (\pi_n)$ is a sequence of stochastic kernels on $\mathcal{B}(\mathbf{A})$ given $x \in \mathbf{X}$. A strategy is called stationary if (π_n) are given by a common stochastic kernel, denoted by π^s. A stationary strategy is called deterministic stationary if $\pi^s(da|x) = \delta_{\varphi(x)}(da)$. For all the problems considered here, Markov strategies are sufficient [Derman, Veinott, Strauch].

Discounted MDP problem

For $\beta < 1$, the β-discounted MDP problem is

$$\text{Minimize over } \pi : \ \mathsf{E}^{\pi}_{x_0}\left[\sum_{n=0}^{\infty} \beta^n c_0(X_n, A_n)\right]$$

$$s.t. \quad \mathsf{E}^{\pi}_{x_0}\left[\sum_{n=0}^{\infty} \beta^n c_j(X_n, A_n)\right] \leq d_j, \ j \in \{1, \ldots, J\}$$

with constraint constants d_j.

For all the concerned MDP problems, assume they are consistent: there exists some strategy satisfying all the constraint inequalities.

Convex analytic approach

Suppose $0 \le \beta < 1$.

Define the occupation measure
$\mathsf{M}^\pi_\beta := \mathsf{E}^\pi_{x_0}\left[\sum_{n\ge 0}\beta^n I\{X_n \in dx, A_n \in da\}\right]$.

Define $\mathcal{D}_\beta = \{\mathsf{M}^\pi_\beta : \ \pi$ is a strategy$\}$.

Then the β-discounted MDP problem is

$$\text{Minimize over } \mathsf{M} \in \mathcal{D}_\beta\text{: } \int_{\mathbf{X}\times\mathbf{A}} c_0(x,a)\mathsf{M}(dx \times da)$$

$$s.t. \quad \int_{\mathbf{X}\times\mathbf{A}} c_j(x,a)\mathsf{M}(dx \times da) \le d_j, \ \ j \in \{1,\dots,J\}.$$

This is the convex analytic approach to discounted MDPs (This term comes from Borkar (1988, PTRF). Book treatments include Kallenberg (1983); Altman (1999); Piunovskiy (1997)). Here $\mathcal{D}_\beta$ is convex. Each extreme point is generated by a deterministic stationary strategy.

Linear programming formulation

In case $\mathbf{X}, \mathbf{A}$ are finite, and under any stationary strategy, (X_n) is positive recurrent, the last assertion can be seen as follows.

- $\mathcal{D}_\beta$ can be characterized by

$$0 \le \mathsf{M} < \infty;$$
$$\mathsf{M}(x, \mathbf{A}) = \delta_{x_0}(x) + \beta \sum_{\mathbf{X}\times\mathbf{A}} p(x|y,a)\mathsf{M}(y,a).$$

  So $\mathcal{D}_\beta$ is the set of feasible solutions in a linear program in standard form: $A\mathsf{M} = b; \mathsf{M} \ge 0$, with $|X|$ equality constraints and $|X| \times |A|$ variables. Any of its extreme point (basic feasible solution) has no more than $|X|$ strictly positive components.

- For any $\mathsf{M} \in \mathcal{D}_\beta$, $\mathsf{M} = \mathsf{M}_\beta^{\pi^s}$ with $\mathsf{M}(x,a) = \mathsf{M}(x,\mathbf{A})\pi^s(a|x)$.
- For extreme point M, for each x, $\mathsf{M}(x,\mathbf{A}) = \sum_a \mathsf{M}(x,a) > 0$, and there is exactly one a such that $\mathsf{M}(x,a) > 0$.
- So for each x, there is exactly one a such that $\pi^s(a|x) = 1$, meaning π^s is deterministic stationary.

Compactness

Condition (W)

- $\mathbf{A}$ *is compact.*
- c_j, $j \in \{0, \ldots, J\}$ *are lower semicontinuous.*
- $\int_{\mathbf{X}} f(y)p(dy|x,a)$ *is continuous in* (x,a) *for each bounded continuous function* f *on* $\mathbf{X}$.

Endow $\mathcal{D}_\beta$ with the weak topology generated by bounded continuous functions on $\mathbf{X} \times \mathbf{A}$. Then under (W), $\mathcal{D}_\beta$ is compact. Moreover, there exists a $J+1$-mixed optimal strategy π^*:

$$\mathsf{M}_\beta^{\pi^*} = \sum_{k=1}^{J+1} \alpha_k \mathsf{M}_\beta^{\varphi_k}$$

for deterministic stationary strategies φ_k and $\alpha_k \geq 0$ adding to 1.

Connection to MDP with a cemetery

At each step, the process can be killed with a given positive probability $(1-\beta)$. After the process is killed, its state goes to a costless cemetery Δ, and remains there indefinitely.

Then the discounted MDP problem can be viewed as a total undiscounted problem for MDP with a cemetery.

In general, MDP model with a cemetery is

$$\{\mathbf{X} \cup \{\Delta\}, \mathbf{A}, p, \{c_j\}_{j=0}^{J}\}.$$

The cemetery Δ is costless and absorbing, uncontrolled.

For model with a cemetery, we consider

$$\text{Minimize over } \pi : \ \mathsf{E}_{x_0}^{\pi}\left[\sum_{n=0}^{\infty} c_0(X_n, A_n)\right]$$

$$s.t. \quad \mathsf{E}_{x_0}^{\pi}\left[\sum_{n=0}^{\infty} c_j(X_n, A_n)\right] \leq d_j, \ \ j \in \{1, \ldots, J\}.$$

Absorbing and uniformly absorbing model

Let T be the hitting time at Δ. Then

Definition

- Altman (1999): An MDP model with cemetery is called absorbing if $\mathsf{E}^{\pi}_{x_0}[T] < \infty$ for each π.
- Feinberg and Piunovskiy (2019, SICON): An MDP model with cemetery is called uniformly absorbing if

$$\lim_{n\to\infty} \sup_{\pi} \mathsf{E}^{\pi}_{x_0}[\sum_{t=n}^{\infty} I\{t < T\}] = 0.$$

A uniformly absorbing model is absorbing.

A discounted model can be viewed as a uniformly absorbing model.

Results for uniformly absorbing model

Occupation measure is
$\mathsf{M}^\pi(dx \times da) := \mathsf{E}^\pi_{x_0}\left[\sum_{n\geq 0} I\{X_n \in dx, A_n \in da\}\right]$, defined on $\mathcal{B}(\mathbf{X}\times\mathbf{A})$.

Denote $\mathcal{D} = \{\mathsf{M}^\pi : \ \pi \text{ is a strategy}\}$.

If the model is absorbing, then $\mathcal{D} = \mathcal{D}^f := \{\mathsf{M} \in \mathcal{D} : \ \mathsf{M}(\mathbf{X}\times\mathbf{A}) < \infty\}$, because $\mathsf{M}^\pi(\mathbf{X}\times\mathbf{A}) = \mathsf{E}^\pi_{x_0}[T]$.

Theorem (Feinberg and Rothblum (2012, MOR))

Suppose the model is uniformly absorbing, then, under (W), $\mathcal{D}$ is compact in the weak topology.

Theorem (Feinberg and Rothblum (2012, MOR))

Suppose the model is uniformly absorbing, then, under (W), there exists a $J+1$-mixed optimal strategy.

The first theorem is important to the proof of the second theorem.

Our result

We want to derive the existence of $J+1$-mixed optimal strategy for models not necessarily absorbing, $\mathcal{D} \neq \mathcal{D}^f$, but we assume in particular, that strategies with infinite-valued occupation measures are not feasible.

Theorem (Piunovskiy and Z.)

Assume that, for each strategy π with $\mathsf{M}^\pi \notin \mathcal{D}^f$, there is some $\tilde{j} \in \{1, \ldots, J\}$, possibly depending on π, satisfying $\mathsf{E}_{x_0}^{\pi}[\sum_{n=0}^{\infty} c_{\tilde{j}}(X_n, A_n)] = \infty$. Then under (W), there is a $J+1$-mixed optimal strategy.

If $c_0 \equiv 0$, then $J+1$ can be replaced with J.

The theorem does not say any optimal strategy can be written as a $J+1$-optimal strategy.

These are demonstrated by the next example.

Example

Example (Adapted from Kallenberg)

$\mathbf{X} = \{0,1,2\}$, $\mathbf{A} = \{0,1\}$, $p(\{1\}|1,0) = 1$, $p(\{2\}|1,1) = 1$, $p(\{2\}|2,0) = 1$, $p(\{2\}|2,1) = p(\{\Delta\}|2,1) = \frac{1}{2}$, $p(\{1\}|0,a) = p(\{2\}|0,a) = \frac{1}{2}$ for $a \in \mathbf{A}$. The state Δ is a costless cemetery. Let $x_0 = 0$. Let $c_0(x,a) \equiv 0$, and $c_1(x,a) = 1$ for $x = 1,2$ and $c_1(0,a) \equiv 0$. Let $d_1 = 3$.

Any feasible strategy will be optimal, and any non-absorbing strategy has infinite valued objective with index 1.

The only absorbing deterministic stationary policies are specified $\varphi(1) = 1 = \varphi(2)$: the state 0 is essentially uncontrolled, and $\varphi(0)$ is immaterial.

$\mathcal{D}^f$ is a nonempty proper subset of $\mathcal{D}$.

Discussion of the example

Consider $\pi^s(\{0\}|1) = \pi^s(\{1\}|1) = \frac{1}{2}$, and $\pi^s(\{1\}|2) = 1$. Then

$$\mathsf{E}_0^{\varphi}\left[\sum_{n=0}^{\infty} c_1(X_n, A_n)\right] = \frac{1}{2}(1+2) + \frac{1}{2}2 = \frac{5}{2};$$

$$\mathsf{E}_0^{\pi^s}\left[\sum_{n=0}^{\infty} c_1(X_n, A_n)\right] = \frac{1}{2}(2+2) + \frac{1}{2}2 = 3.$$

Therefore, both φ and π^s are feasible and thus optimal.

On the other hand, $\mathsf{E}_0^{\varphi}\left[\sum_{n=0}^{\infty} c_1(X_n, A_n)\right] \neq \mathsf{E}_0^{\pi^s}\left[\sum_{n=0}^{\infty} c_1(X_n, A_n)\right]$ implies that the occupation measure of π^s cannot be represented as the convex combination of the occupation measure of φ.

Conclusion: the optimal strategy φ is an 1-mixed optimal strategy, the optimal strategy π^s is not a mixed strategy.

Difference between absorbing and uniformly absorbing models

That $\mathcal{D}$ is convex compact in a locally convex Hausdorff space is used in [Feinberg and Rothblum (2012, MOR)] when treating uniformly absorbing models.

For absorbing but not uniformly absorbing models, under (W), it can still happen that $\mathcal{D}$ is not compact in the weak topology. (This contradicts some claims made in the literature.)

Let $\mathcal{P} := \{\mathsf{P}^{\pi}_{x_0} : \pi \text{ is a strategy}\}$.

Endow $\mathcal{D}$ with the strongest topology such that $O : \mathcal{P} \to \mathcal{D}$ defined by

$$O(\mathsf{P}) := \sum_{n=0}^{\infty} \mathsf{P}(X_n \in dx, A_n \in da)$$

is continuous.

Under (W), $\mathcal{P}$, endowed with the weak topology, is compact [Schal (1975, SPA)]. Now $\mathcal{D}$ is compact under (W).

Some formalities

The MDP problem can still be written as

$$\text{Minimize over } \mathsf{M} \in \mathcal{D}\text{:} \int_{\mathbf{X}\times\mathbf{A}} c_0(x,a)\mathsf{M}(dx \times da)$$

$$s.t. \quad \int_{\mathbf{X}\times\mathbf{A}} c_j(x,a)\mathsf{M}(dx \times da) \leq d_j, \; j \in \{1,\ldots,J\}.$$

If the model is not absorbing, then $\mathcal{D}$ contains some infinite measures.

Consequently, $\mathcal{D}$ is not a subset of a vector space.

In standard texts, cone or convex sets are defined as certain subsets of a vector space. Formally speaking the above problem is not a convex program.

So we replace a vector space with an abstract cone.

Abstract cone

Definition

Consider a nonempty set X with an element $\mathbf{0} \in X$ and the operations of addition, denoted as $x + y \in X$ for $x, y \in X$, and scalar multiplication by non-negative real numbers, denoted as $ax \in X$ for all $x \in X$ and $a \in [0, \infty)$. The set X is called a cone (over $[0, \infty)$) if it satisfies the following conditions:

- $a(x + y) = ax + ay$ for all $a \in [0, \infty)$ and $x, y \in X$;
- $a(bx) = (ab)x$ for all $a, b \in [0, \infty)$ and $x \in X$;
- $1x = x$ and $0x = \mathbf{0}$ for all $x \in X$;
- $x + y = y + x$ for all $x, y \in X$;
- $x + (y + z) = (x + y) + z$ for all $x, y, z \in X$;
- $(a + b)x = ax + bx$ for all $a, b \in [0, \infty)$ and $x \in X$;
- $x + \mathbf{0} = x$ for all $x \in X$.

Convex set and affine function

Definition (Convex set in a cone)

A subset E of a cone X is called a convex set (in X) if $\alpha x + (1-\alpha)y \in E$ for all $x, y \in E$ and $\alpha \in [0,1]$.

Definition (Affine function)

A $(-\infty, +\infty]$-valued function $f(\cdot)$ defined on a convex set E in a cone is called affine if

$$f(\alpha x_1 + (1-\alpha)x_2) = \alpha f(x_1) + (1-\alpha) f(x_2)$$

for all $x_1, x_2 \in E$ and all $\alpha \in [0,1]$.

Example

Example

Let $E' \subseteq \mathbb{R}^2$ be a closed disk of radius 0.5 and centered at $(1, 1.5)$, Then this set E is convex in the cone $\mathbb{R}^2$, and $\vec{u} = (1, 1)$ is Pareto optimal in E'. It is also an extreme point of E. Thus, the minimal face containing $\vec{u}$ in E' is $G_{E'}(\vec{u}) = \{\vec{u}\}$.

It is known that there are nonnegative $b_1, b_2 \geq 0$ with $b_1 + b_2 > 0$ and a constant β such that $b_1 \cdot 1 + b_2 \cdot 1 = \beta$ and $b_1 v_1 + b_2 v_2 \geq \beta$ for all $\vec{v} = (v_1, v_2) \in E'$.

Now Consider $E = E' \cup \{(\infty, v_2) : v_2 \in [0, 2]\}$. Then this set E is convex in the cone $(-\infty, \infty]^2$, and $\vec{u} = (1, 1)$ is still Pareto optimal in E. It is also an extreme point of E. Thus, $G_E(\vec{u}) = \{\vec{u}\}$.

Discussion of the example

If the blue assertion with E' being replaced by E held, then there would be nonnegative $b_1, b_2 \geq 0$ with $b_1 + b_2 > 0$ and a constant β such that $b_1 \cdot 1 + b_2 \cdot 1 = \beta$ and $b_1 v_1 + b_2 v_2 \geq \beta$ for all $\vec{v} = (v_1, v_2) \in E$.

The requirement of $b_1 \cdot 1 + b_2 \cdot 1 = \beta$ and $b_1 v_1 + b_2 v_2 \geq \beta$ for all $\vec{v} = (v_1, v_2) \in E' \subset E$ imply that $b_1 = 0$, and $b_2 = \beta > 0$, as given by the unique supporting hyperplane in $\mathbb{R}^2$ of the disk E' at $\vec{u} = (1, 1)$.

On the other hand, since $0 \cdot \infty := 0$, for $(\infty, 0) \in E$, $b_1 \cdot \infty + b_2 \cdot 0 = 0 < \beta$, yielding a contradiction.

Optimization problem

- E is a nonempty convex set in a cone.
- E is endowed with a topology: E is compact.
- W_j $j = 0, \ldots, J$ are lower semicontinuous affine functions, bounded from below.
- We consider the following optimization problem

$$\begin{aligned} &\text{Minimize over } x \in E\text{: } W_0(x) \\ &\text{subject to } W_j(x) \leq d_j, \; j = 1, 2, \ldots, J, \end{aligned} \tag{1}$$

Mixed optimal solution

Theorem (Piunovskiy and Z.)

Suppose that the nonempty convex set E (in a cone) is compact, problem (1) is consistent, and $\hat{\mathcal{C}}(E)$ is total. Then there exists a $J+1$-mixed optimal solution for problem (1).

- $\hat{\mathcal{C}}(E)$ is the space of bounded below, lower semicontinuous, affine $(-\infty, \infty]$-valued functions on E.
- An optimal solution x^* is called a $J+1$-mixed optimal solution if

$$x^* = \sum_{k=1}^{J+1} \alpha_k x_k,$$

where $\alpha_k \in [0,1]$, $\sum_{k=1}^{J+1} \alpha_k = 1$, and x_k is extreme in E for each $k = 1, 2, \ldots, J+1$.

Proof for the absorbing model

MDP problem

$$\text{Minimize over } \mathsf{M} \in \mathcal{D}\text{:} \int_{\mathbf{X}\times\mathbf{A}} c_0(x,a)\mathsf{M}(dx \times da)$$

$$s.t. \quad \int_{\mathbf{X}\times\mathbf{A}} c_j(x,a)\mathsf{M}(dx \times da) \le d_j, \ \ j \in \{1,\dots,J\}.$$

can be written in form of problem (1) by setting $E = \mathcal{D}$, $x = \mathsf{M}$, $W_j(x) = \int c_j(x,a)\mathsf{M}(dx \times da)$.

Assume the model is absorbing. Then every extreme point of $\mathcal{D}$ is generated by a deterministic stationary strategy, and under (W),

$$\left\{ \int_{\mathbf{X}\times\mathbf{A}} c(x,a)\mathsf{M}(dx \times da) : c \text{ is nonnegative lsc} \right\}$$

is a total family of nonnegative lsc functions on $\mathcal{D}$.

Now the theorem in the previous slide can be directly applied to absorbing MDP models.

Thank you

Thank you very much.