Exam of the course *Markov decision processes : dynamic programming and applications*

Marianne Akian

ENSTA Course SOD312

Mardi 15 octobre 2024
Durée 3h

Problems 1 and 2 are independent. The solution can be written either in French or English. Documents (handwritten or typed courses and exercises notes, together with books related to the course) are allowed.

Recall that this exam is based on Lectures 1 to 5 (Tuesday Sept. 10 to Tuesday Oct. 1, 2024), and that its only purpose is to validate ENSTA Course SOD312. (Another exam will be proposed later to Master students.)

1 Problem 1

A concert will hold tonight in a concert hall on top of a private underground parking garage. There is no difficulty to find a place in the parking, the cost of which is fixed for the night to P (in euros). There is a one-way avenue which leads straight to the concert hall, on which there are N parallel parking places, numbered from 0 to $N-1$, the place number $N-1$ beeing the closest from the parking, and the place number 0 being the farthest from the parking.

For $n = 0, \dots, N-1$, let X_n be a random variable equal to 0 if the parking place number n in the avenue is not free, and equal to some measure of the difficulty to park in the parking space of place number n otherwise, which will be an integer between 1 and $\bar{x}$. X_n will be called the state of place n. We assume that the random variables $X_0, \dots, X_{N-1}$ are independent with same law on the space $\mathcal{E} := \{0, \dots, \bar{x}\}$, and we denote by q this law. We also associate a cost to park in Place number $n \leq N-1$ equal to $N - n + X_n$ (this is the time to park and to walk "converted" in euros). By convention, the underground parking garage corresponds to place number N, and we set $X_N = 1$ and assume that the fixed price P for the underground parking is > 2.

The attendee know the state X_n of place n only after having drived by the place number n. When an attendee is arriving in the avenue, he is trying to find a free parking place the closest to the concert hall and if he does not find such a place, he parks his car in the underground parking garage.

Q 1.1. Consider first an attendee who is taking the first free parking place. For each $n \in \{0, \dots, N\}$, and $x \in \mathcal{E}$, denote by $w_n(x)$ the expected cost of parking for this attendee when he is in front of Place number n and this place has state x. Write this cost in the form

$$w_n(x) = \mathbb{E}\left[r_\tau(X_\tau) \mid X_n = x\right]$$

for some stopping time $\tau \in \{n, \dots, N\}$. Precise to which filtration the stopping time is adapted and give the expression of the functions $r_n : \mathcal{E} \to \mathbb{R} \cup \{+\infty\}$, for $n \in \{0, \dots, N\}$.

Q 1.2. Determine a recurrence equation allowing to compute the functions w_n of Q. 1.1.

Q 1.3. Assume now that the attendee is trying to find a free parking place in such a way he is minimizing his expected cost. Show that this problem can be written as :

$$\min_{\tau} \mathbb{E}\left[r_\tau(X_\tau)\right]$$

where the minimization is done under all stopping times $\leq N$, with the same functions r_n as before.

Q 1.4. Define auxiliary value functions $v_n : \mathcal{E} \to \mathbb{R} \cup \{+\infty\}$ allowing to compute the value of the problem of Q. 1.3 and show that they satisfy, for $n \leq N-1$:

$$v_n(0) = \begin{cases} \sum_{i=0}^{\bar{x}} q_i v_{n+1}(i) & \text{if } n+1 \neq N \\ P & \text{if } n+1 = N \end{cases} \tag{1}$$

$$v_n(x) = \min(N - n + x, v_n(0)) \quad \text{for } x \in \mathcal{E} \setminus \{0\} \ . \tag{2}$$

Q 1.5. Show that $v_n(0)$ is a nondecreasing sequence.

Q 1.6. For all $x \in \mathcal{E} \setminus \{0\}$, let

$$n_x^* = \inf\{n \in \{0, \dots, N\} \mid N - n + x \leq v_n(0)\} \ .$$

Show n_x^* is nondecreasing with respect to x and that one cannot have $N - n + x > v_n(0)$ for $n \geq n_x^*$.

Q 1.7. Describe the optimal stopping time τ^* as a function of the numbers n_x^*, $x \in \mathcal{E} \setminus \{0\}$.

---

2 Problem 2

We consider a MDP on a state space $\mathcal{E}' = \mathcal{E} \cup \delta$, where $\mathcal{E} = \{1, \cdots, n\}$ and δ is a cemetery point (when the state equals δ at some time, it stays in δ for all following times). We assume that the action space $\mathcal{C}$ is independent of the state, and is a compact metric space (for instance a compact subset of some $\mathbb{R}^p$). We denote by $M_{xy}^{(u)}$ the transition probability of the MDP : $M_{xy}^{(u)} = \mathbb{P}\left(X_{n+1} = y \mid X_n = x, U_n = u\right)$, for $x, y \in \mathcal{E}'$ and $u \in \mathcal{C}$, and assume that it is continuous with respect to $u \in \mathcal{C}$. We consider a nonnegative cost function $c : \mathcal{E}' \times \mathcal{C} \to \mathbb{R}^+$, such that $c(x, u) = 0$ for $x = \delta$, and $c(x, u) = \exp(\gamma g(x, u))$, where $g : \mathcal{E} \times \mathcal{C} \to \mathbb{R}$ is continuous with respect to $u \in \mathcal{C}$.

The following study is related to the problem of minimization of the possibly "positively discounted" total cost :

$$J^\pi(x) = \mathbb{E}\left[\sum_{n=0}^{\infty} \prod_{k=0}^{n} c(X_k, U_k) \mid X_0 = x\right] = \mathbb{E}\left[\sum_{n=0}^{\tau} \prod_{k=0}^{n} c(X_k, U_k) \mid X_0 = x\right] \ ,$$

among all feedback strategies $\pi = (\pi_k)_{k \geq 0}$ with $\pi_k : \mathcal{E}' \to \mathcal{C}$ and $U_k = \pi_k(X_k)$, where τ is the first arrival time of the state at point δ. Note that since c is nonnegative, this expectation exists, while it may be infinite.

We shall restrict strategies to $\mathcal{E}$ and denote by Π^0 the set of all policies $\pi : \mathcal{E} \to \mathcal{C}$. For $\pi \in \Pi^0$, we denote by $M(\pi)$ the $n \times n$ matrix with entry (x, y) equal to $M_{xy}^{(\pi(x))}$ and we set :

$$D(\pi) := \mathrm{diag}\,(c(1, \pi(1)), \cdots, c(n, \pi(n)))\,,\ A(\pi) = D(\pi)M(\pi) \in \mathbb{R}^{n\times n},\ c^{(\pi)} = D(\pi)\mathbf{1} \in \mathbb{R}^n\ ,$$

where $\mathbf{1}$ is the vector of $\mathbb{R}^n$ with all entries equal to 1.

For all functions $v : \mathcal{E} \mapsto \mathbb{R}$ identified to a vector of $\mathbb{R}^n$, we define the operators :

$$\mathcal{L}^{(\pi)}(v) = A(\pi)v + c^{(\pi)} \quad \text{and } \mathcal{L}(v) = \inf_{\pi\in\pi} \mathcal{L}^{(\pi)}(v)\ .$$

We say that a strategy $\pi = (\pi_k)_{k\geq 0}$ is γ-admissible if the following limit exists and is equal to zero :

$$\lim_{t\mapsto\infty} [A(\pi_0)A(\pi_1)\cdots A(\pi_t)]_{x,y} = 0 \quad \text{for } (x, y) \in \mathcal{E}^2\ .$$

For a stationary strategy equal to $\pi \in \pi$, this is equivalent to the condition $\rho(A(\pi)) < 1$ where $\rho(A)$ denotes the spectral radius of the matrix A.

We shall admit the following (Collatz-Wielandt) property : if $A \in \mathbb{R}^{n\times n}$ has nonnegative entries, then

$$\rho(A) < 1 \Leftrightarrow \exists\lambda \in [0, 1), w \in \mathbb{R}^n,\ \text{s.t. } w_i > 0, i = 1, \ldots, n,\ \text{and } Aw \leq \lambda w\ .$$

Q 2.1. Let $A \in \mathbb{R}^{n\times n}$ be such that $\rho(A) < 1$. Show that there exists a vectorial norm such that A is strictly contracting with respect to this norm.

Q 2.2. Show that for all $v \in \mathbb{R}^n$, there exist a policy $\pi^\sharp \in \pi$ depending on v such that $\mathcal{L}^{\pi^\sharp}(v) = \mathcal{L}(v)$.

Q 2.3. We shall say that a strategy is proper if the following limit exists and is equal to zero :

$$\lim_{t\mapsto\infty} [M(\pi_0)M(\pi_1)\cdots M(\pi_t)]_{x,y} = 0 \quad \text{for all } (x, y) \in \mathcal{E}^2\ .$$

Show that if a stationary strategy is proper then it is also a stationary γ-admissible strategy for γ small enough.

Q 2.4. Show that the operator $\mathcal{L}$ is order preserving ($v \leq w$ implies $\mathcal{L}(v) \leq \mathcal{L}(w)$).

Q 2.5. Show that for all $v \in \mathbb{R}^n$ and $\lambda > 0$, we have :

$$\mathcal{L}(v + \lambda e) \leq \mathcal{L}(v) + \overline{c}\lambda e\ ,$$

where the constant $\overline{c}$ is an upper bound of c.

Q 2.6. Deduce that $\mathcal{L}$ is Lipschitz continuous for the sup-norm $\|v\| = \max_i |v_i|$. Show that the same holds for the operators $\mathcal{L}^{(\pi)}$ with $\pi \in \pi$.

Q 2.7. Let π be a stationary γ-admissible strategy. Show that $\mathcal{L}^{(\pi)}$ has a unique fixed point $v^{(\pi)}$. How $v^{(\pi)}$ can be computed ?

Q 2.8. Show that $v^{(\pi)} > 0$ (that is $v_i^{(\pi)} > 0$, for all $i = 1, \ldots, n$).

Q 2.9. Let $\pi \in \Pi^0$ be such that there exists $v \in \mathbb{R}^n$, $v > 0$ satisfying $\mathcal{L}^{(\pi)}(v) \leq v$. Show that π is a stationary γ-admissible strategy (one can show that there exists $\lambda \in [0, 1)$ such that $A(\pi)v \leq \lambda v$).

Q 2.10. Show that $\mathcal{L}$ has at most one fixed point v in the set of vectors such that $v > 0$.

Q 2.11. Denote by $\pi_k \in \Pi^0$ the sequence of policies obtained in the policy iteration algorithm applied to the equation of $\mathcal{L}(v) = v$, starting with a stationary γ-admissible strategy $\pi_0 \in \Pi^0$. Show that the π_k is γ-admissible for all $k \geq 0$, and deduce that the algorithm is well posed.

Q 2.12. Deduce that the sequence $v^{(\pi_k)}$ of fixed points of $\mathcal{L}^{(\pi_k)}$ is nonincreasing, and that the limit is a fixed point of $\mathcal{L}$.
