Correction des travaux dirigés d'
Apprentissage par renforcement
Université Paris–Saclay

Joon Kwon

mercredi 12 novembre 2024

❧

Exercice 1. — Si on suppose $B_\pi v_* = B_* v_*$, cela signifie que v_* est un point fixe de B_π. Or v_π est l'unique point fixe de B_π. Donc $v_* = v_\pi$.

Réciproquement, supposons $v_\pi = v_*$. Alors, en utilisant le fait que v_π est point fixe de B_π et v_* est point fixe de B_*,

$$B_\pi v_* = B_\pi v_\pi = v_\pi = v_* = B_* v_*.$$

Exercice 2. —

1) On considère les ensembles $\mathscr{S} = \{s^{(1)}, s^{(2)}\}$, $\mathscr{A} = \{a^{(1)}\}$ et $\mathscr{R} = \{0\}$, et la dynamique
$$p(\cdot\,|s, a^{(1)}) = \delta_{(0,s^{(2)})}, \qquad \text{pour tout } s \in \mathscr{S}.$$
Alors, on peut facilement montrer que pour $v \neq v'$ dans $\mathbb{R}^{\mathscr{S}}$ définis par
$$v(s^{(1)}) = 1, \qquad v'(s^{(1)}) = 0 \qquad \text{et} \qquad v(s^{(2)}) = v'(s^{(2)}) = 0,$$
on a $Dv = Dv' = 0$, D n'est pas donc pas injective.

2) L'opérateur D étant affine, il est injectif si, et seulement si l'opérateur linéaire associé est injectif. On pose le vecteur

$$\mathrm{R} = \left(\sum_{(r,s') \in \mathscr{R} \times \mathscr{S}} p(r, s' | s, a) r \right)_{(s,a) \in \mathscr{S} \times \mathscr{A}} \in \mathbb{R}^{\mathscr{S} \times \mathscr{A}}$$

et la matrice

$$\mathrm{P} = \left(\sum_{r \in \mathscr{R}} p(r, s' | s, a) \right)_{((s,a),s') \in (\mathscr{S} \times \mathscr{A}) \times \mathscr{S}} \in \mathbb{R}^{(\mathscr{S} \times \mathscr{A}) \times \mathscr{S}},$$

de sorte que pour tout $v \in \mathbb{R}^{\mathscr{S} \times \mathscr{A}}$,

$$\mathrm{D}v = \mathrm{R} + \gamma \mathrm{P} v.$$

D est alors injective si, et seulement si $\operatorname{Ker} \mathrm{P} = \{0\}$.

Exercice 3. — On suppose $v_\pi \leqslant v_{\pi'}$. En utilisant la monotonie de l'opérateur D, il vient

$$q_\pi = \mathrm{D}v_\pi \leqslant \mathrm{D}v_{\pi'} = q_{\pi'}.$$

La réciproque est fausse en général, donnons un contre-exemple avec deux états $\mathscr{S} = \{s^{(1)}, s^{(2)}\}$ et deux actions $\mathscr{A} = \{a^{(1)}, a^{(2)}\}$. On considère la dynamique donnée par

$$\begin{aligned} p(\cdot | s^{(1)}, a^{(1)}) &= \delta_{(0, s^{(2)})} \\ p(\cdot | s^{(1)}, a^{(2)}) &= \delta_{(1, s^{(2)})} \\ p(\cdot | s^{(2)}, a^{(1)}) &= \delta_{(0, s^{(2)})} \\ p(\cdot | s^{(2)}, a^{(2)}) &= \delta_{(0, s^{(2)})}, \end{aligned}$$

et deux politiques stationnaires et déterministes π, π' données par

$$\pi(s^{(1)}) = \pi(s^{(2)}) = a^{(1)},$$

et

$$\pi'(s^{(1)}) = \pi'(s^{(2)}) = a^{(2)}.$$

On vérifie alors facilement que $q_\pi = q_{\pi'}$ mais que $v_{\pi'}(s^{(1)}) = 1 > 0 = v_\pi(s^{(1)})$.