Travaux dirigés d'
Apprentissage par renforcement

Université Paris–Saclay

Joon Kwon

mardi 5 décembre 2023

Exercice 1 (*Bornes d'erreur et critère d'arrêt*). — Soit π une politique stationnaire, $v \in \mathbb{R}^{\mathcal{S}}$ et $q \in \mathbb{R}^{\mathcal{S}\times\mathcal{A}}$.

1) Montrer que :

$$\text{(i)}\ \|v - v_\pi\|_\infty \leqslant \frac{\|v - \mathrm{B}_\pi v\|_\infty}{1-\gamma}, \qquad \text{(iii)}\ \|v - v_*\|_\infty \leqslant \frac{\|v - \mathrm{B}_* v\|_\infty}{1-\gamma},$$
$$\text{(ii)}\ \|q - q_\pi\|_\infty \leqslant \frac{\|q - \mathrm{B}_\pi q\|_\infty}{1-\gamma}, \qquad \text{(iv)}\ \|q - q_*\|_\infty \leqslant \frac{\|q - \mathrm{B}_* q\|_\infty}{1-\gamma}.$$

2) Soit $\varepsilon > 0$. En déduire un critère d'arrêt pour les itération valeur qui nous assure d'avoir une solution approchée à moins d'ε (en norme $\|\cdot\|_\infty$).

3) Montrer qu'une itération valeur donne nécessairement une politique optimale au bout d'un nombre fini d'itérations.

Exercice 2 (*Convergence en temps fini de l'itération de politiques*). — Montrer qu'une itération de politiques donne une politique optimale au bout d'un nombre fini d'itérations.

Exercice 3 (*Amélioration gloutonne par rapport à plusieurs politiques*). — Soit $\pi_1, \dots, \pi_M$ des politiques stationnaires, et

$$\pi \in \Pi_g \left[\max_{1 \leqslant m \leqslant M} v_{\pi_m} \right].$$

Montrer que

$$v_\pi \geqslant \max_{1 \leqslant m \leqslant M} v_{\pi_m}, \qquad \text{et} \qquad q_\pi \geqslant \max_{1 \leqslant m \leqslant M} q_{\pi_m}.$$