

TRAVAUX DIRIGÉS D'
APPRENTISSAGE
PAR RENFORCEMENT
UNIVERSITÉ PARIS–SACLAY

Joon Kwon

lundi 25 novembre 2024



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 :

$$\begin{aligned} \text{(i)} \quad \|v - v_{\pi}\|_{\infty} &\leq \frac{\|v - B_{\pi}v\|_{\infty}}{1 - \gamma}, & \text{(iii)} \quad \|v - v_{*}\|_{\infty} &\leq \frac{\|v - B_{*}v\|_{\infty}}{1 - \gamma}, \\ \text{(ii)} \quad \|q - q_{\pi}\|_{\infty} &\leq \frac{\|q - B_{\pi}q\|_{\infty}}{1 - \gamma}, & \text{(iv)} \quad \|q - q_{*}\|_{\infty} &\leq \frac{\|q - B_{*}q\|_{\infty}}{1 - \gamma}. \end{aligned}$$

- 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 π_1, \dots, π_M des politiques stationnaires, et

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

Montrer que

$$v_\pi \geq \max_{1 \leq m \leq M} v_{\pi_m}, \quad \text{et} \quad q_\pi \geq \max_{1 \leq m \leq M} q_{\pi_m}.$$

