Sujet d'évaluation

Apprentissage par renforcement

Université Paris–Saclay

à rendre le 22 janvier 2024 au plus tard

Les deux premières parties du sujet doivent être traitées de façon intégrale. Dans la troisième et dernière partie, plusieurs pistes sont suggérées, et une grande liberté est laissée : un bon traitement d'une petite partie suffit pour obtenir une bonne note. Il doit s'agir d'un travail individuel. Il doit être envoyé par e-mail à l'adresse

`joon.kwon@inrae.fr`

et pourra comporter des fichiers sous différentes formes, par exemple : copie manuscrite scannée ou document rédigé en LaTeX, notebook Jupyter.

I. Étude théorique d'un nouvel algorithme de points fixes

Soit $d \geqslant 1$, $\mathrm{F} : \mathbb{R}^d \to \mathbb{R}^d$ une application admettant un point fixe $x_* \in \mathbb{R}^d$, $\eta > 0$, et $x_0 \in \mathbb{R}^d$. On définit l'algorithme Ada-FP comme suit :

$$x_{k+1} = x_k + \eta \frac{\mathrm{F}x_k - x_k}{\sqrt{\sum_{\ell=0}^{k} \|\mathrm{F}x_\ell - x_\ell\|_2^2}}, \qquad k \geqslant 0, \tag{Ada-FP}$$

avec la convention $0/0 = 0$. On note pour tout $k \geqslant 0$,

$$\begin{aligned} u_k &= \mathrm{F}x_k - x_k, \\ \eta_k &= \frac{\eta}{\sqrt{\sum_{\ell=0}^{k} \|\mathrm{F}x_\ell - x_\ell\|_2^2}}, \\ \mathrm{D}_k &= \max_{0 \leqslant \ell \leqslant k} \frac{1}{2} \|x_\ell - x_*\|_2^2 . \end{aligned}$$

1) Montrer que pour tout $k \geqslant 0$,

$$\|x_{k+1} - x_*\|_2^2 \leqslant \|x_k - x_*\|_2^2 + 2\eta_k u_k^\top (x_k - x_*) + \eta_k^2 \|u_k\|_2^2 .$$

2) En déduire que pour tout $k \geqslant 0$,

$$\sum_{\ell=0}^{k} u_\ell^\top (x_* - x_\ell) \leqslant \frac{\mathrm{D}_k}{\eta_k} + \sum_{\ell=0}^{k} \frac{\eta_\ell \|u_\ell\|_2^2}{2}.$$

3) a) Soit $(a_k)_{k \geqslant 0}$ une suite positive. Montrer que pour tout $k \geqslant 0$,

$$\sum_{\ell=0}^{k} \frac{a_\ell}{\sqrt{\sum_{m=0}^{\ell} a_m}} \leqslant 2\sqrt{\sum_{\ell=0}^{k} a_\ell},$$

avec la convention $0/0 = 0$.

b) En déduire que pour tout $k \geqslant 0$,

$$\sum_{\ell=0}^{k} u_\ell^\top (x_* - x_\ell) \leqslant \left(\eta + \frac{\mathrm{D}_k}{\eta}\right) \sqrt{\sum_{\ell=0}^{k} \|u_\ell\|_2^2}.$$

4) On suppose que F est γ_F-lipschitzienne pour un certain $0 \leqslant \gamma_\mathrm{F} < 1$.

a) Montrer que pour tout $k \geqslant 0$,

$$\|\mathrm{F}x_k - x_k\|_2^2 \leqslant 2(\mathrm{F}x_k - x_k)^\top (x_* - x_k).$$

b) En déduire que pour tout $k \geqslant 0$,

$$\min_{0 \leqslant \ell \leqslant k} \|\mathrm{F}x_\ell - x_\ell\|_2 \leqslant \frac{2}{\sqrt{k}} \left(\eta + \frac{\mathrm{D}_k}{\eta}\right).$$

c) Qu'en déduire sur $\min_{0 \leqslant \ell \leqslant k} \|x_\ell - x_*\|_2$?

On rappelle que dans un contexte de programmation dynamique (i.e. où la dynamique de transition p du MDP est disponible sous forme explicite), les itérations valeur (synchrones) pour l'évaluation d'une politique $\pi \in \Pi_0$ (resp. pour le contrôle) sont données par

$$v_{k+1} = \mathrm{B}_\pi v_k, \quad k \geqslant 1, \tag{$\mathrm{VI}_\pi^{(\mathrm{V})}$}$$

et

$$v_{k+1} = \mathrm{B}_* v_k, \quad k \geqslant 1, \tag{$\mathrm{VI}_*^{(\mathrm{V})}$}$$

respectivement.

5) En utilisant (Ada-FP), définir des algorithmes analogues aux intérations valeur synchrones classiques ($\mathrm{VI}_\pi^{(\mathrm{V})}$) et ($\mathrm{VI}_*^{(\mathrm{V})}$). On appellera (Ada-$\mathrm{VI}_\pi^{(\mathrm{V})}$) et (Ada-$\mathrm{VI}_*^{(\mathrm{V})}$) les algorithmes ainsi obtenus.

2. Comparaison en pratique avec les algorithmes classiques

Choisir un MDP de taille raisonnable, c'est-à-dire dont on puisse calculer les fonctions valeurs v_π, v_* ($\pi \in \Pi_0$) en un temps raisonnable. On pourra soit reprendre un MDP vu en TP, soit en trouver un dans un livre, sur internet, dans une librairie (e.g. Gymnasium) ou encore en créer un soi-même, mais pour cette partie, il est nécessaire de connaître les transitions de façon explicite.

6) Se donner une politique stationnaire $\pi \in \Pi_0$ quelconque, ainsi qu'une fonction valeur initiale $v_0 \in \mathbb{R}^{\mathcal{S}}$ tirée aléatoirement une fois pour toutes. Comparer en pratique la vitesse de convergence de (Ada-$\mathrm{VI}_\pi^{(\mathrm{V})}$) avec celle de ($\mathrm{VI}_\pi^{(\mathrm{V})}$). On pourra tracer, avec une échelle logarithmique en ordonnée, les quantités

$$\|v_k - v_\pi\|_\infty \quad \text{et} \quad \|v_k - \mathrm{B}_\pi v_k\|_\infty$$

en fonction de k. Essayer différentes valeurs pour $\eta > 0$.

7) Même question pour (Ada-$\mathrm{VI}_*^{(\mathrm{V})}$) et ($\mathrm{VI}_*^{(\mathrm{V})}$).

3. Extensions

Reprendre la démarche de la Section 2 en incorporant, par exemple, un ou plusieurs des aspects suivants :

— itérations de fonctions action-valeur,
— itérations asynchrones,
— méthode d'apprentissage par renforcement utilisant des estimateurs stochastiques des opérateurs de Bellman,
— approximation de la fonction valeur par une classe paramétrique,
— variante de (Ada-FP) définie composante par composante par :

$$x_{k+1,j} = x_{k,j} + \eta \frac{(\mathrm{F}x_k)_j - x_{k,j}}{\sqrt{\sum_{\ell=0}^{k} (\mathrm{F}(x_\ell)_j - x_{\ell,j})^2}}, \qquad 1 \leqslant j \leqslant d, \quad k \geqslant 0.$$

— tout autre aspect potentiellement pertinent.

On commencera par définir précisément les algorithmes considérés avant de les implémenter. On pourra également tenter une étude théorique des algorithmes obtenus, en gardant à l'esprit qu'il n'est pas évident de démontrer des garanties dans tous les cas.