

gz-unitree: Reinforcement learning en robotique
avec validation par moteurs de physique
multiples pour le robot *H1v2* d'Unitree

Gwenn Le Bihan <gwenn.lebihan@etu.inp-n7.fr>

6 Novembre 2025

Reinforcement Learning

Et son application à la robotique

Bases du RL

Agent

Environnement

Score

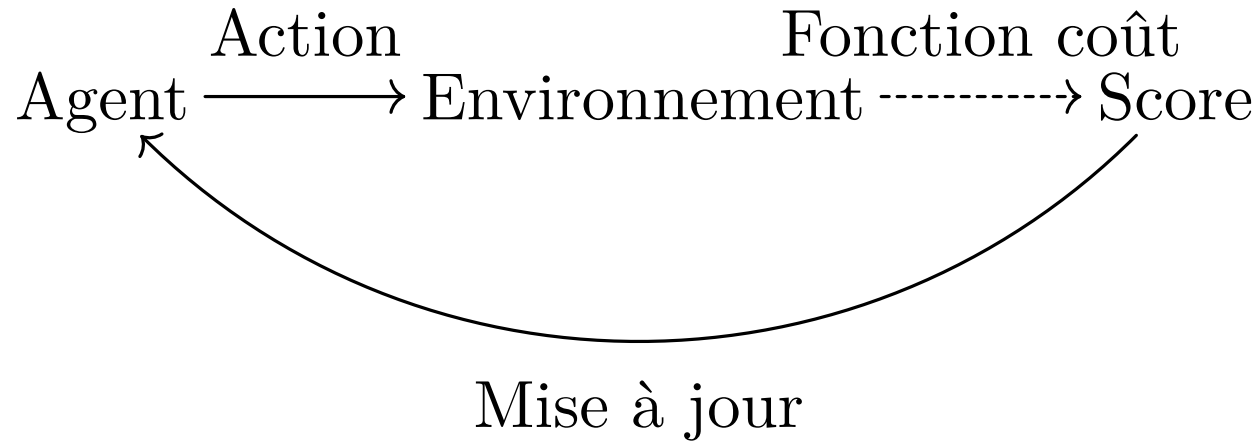
Bases du RL



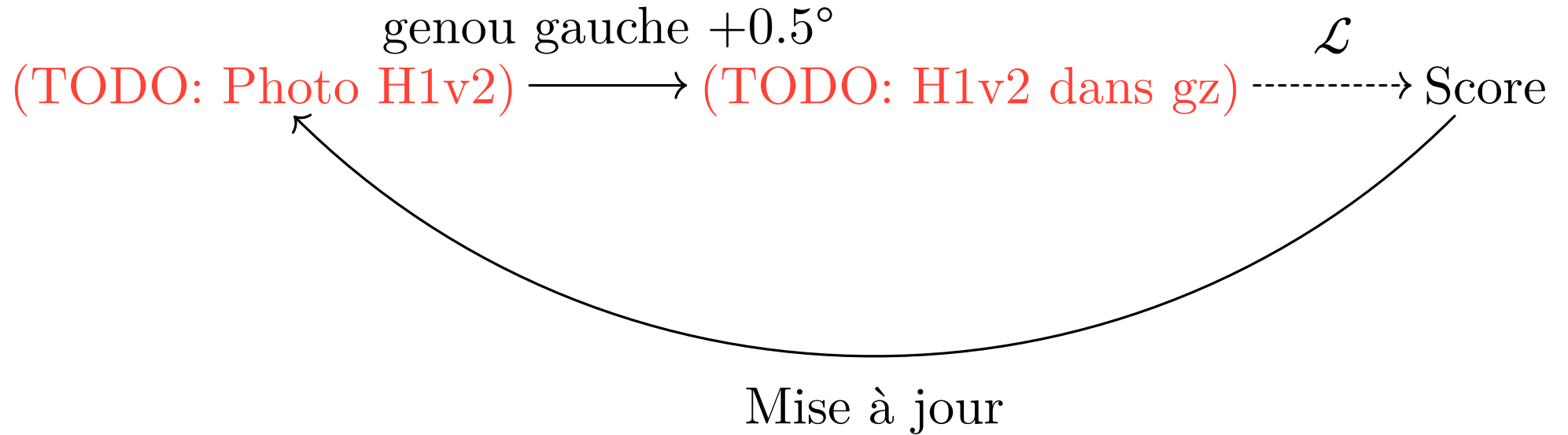
Bases du RL



Bases du RL



RL en robotique



C'est quoi \mathcal{L} ?

C'est très simple:

$$\mathcal{L}_r(\pi', \pi) := \mathbb{E}_{(s_t, a_t)_{t \in \mathbb{N}} \in \mathcal{C}} \sum_{t=0}^{\infty} \frac{Q_{\pi}(s_t, a_t)}{Q_{\pi'}(s_t, a_t)} A_{\pi, r}(s_t, a_t)$$

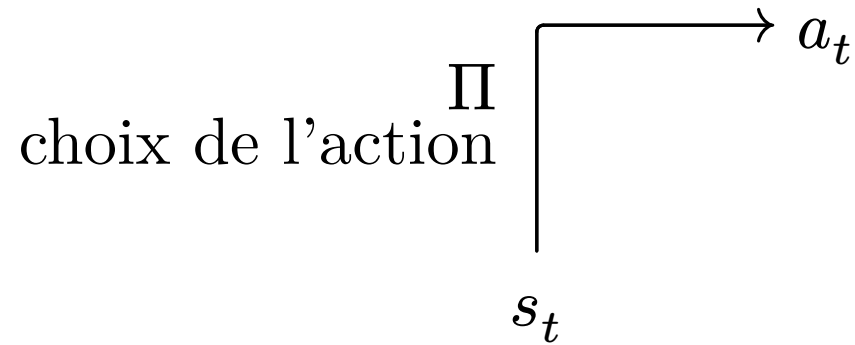
Comparaison des politiques

En Reinforcement Learning

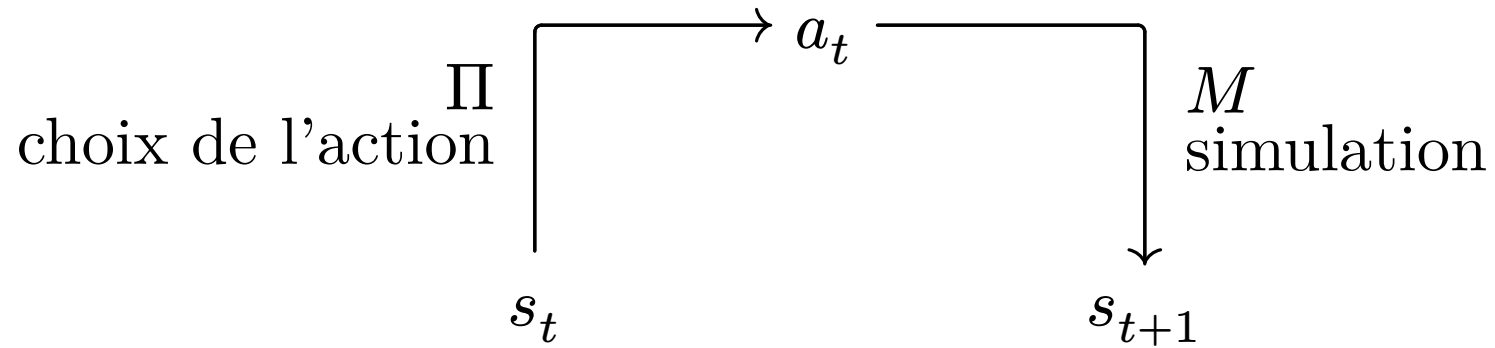
Comparaison des politiques

$$s_t$$

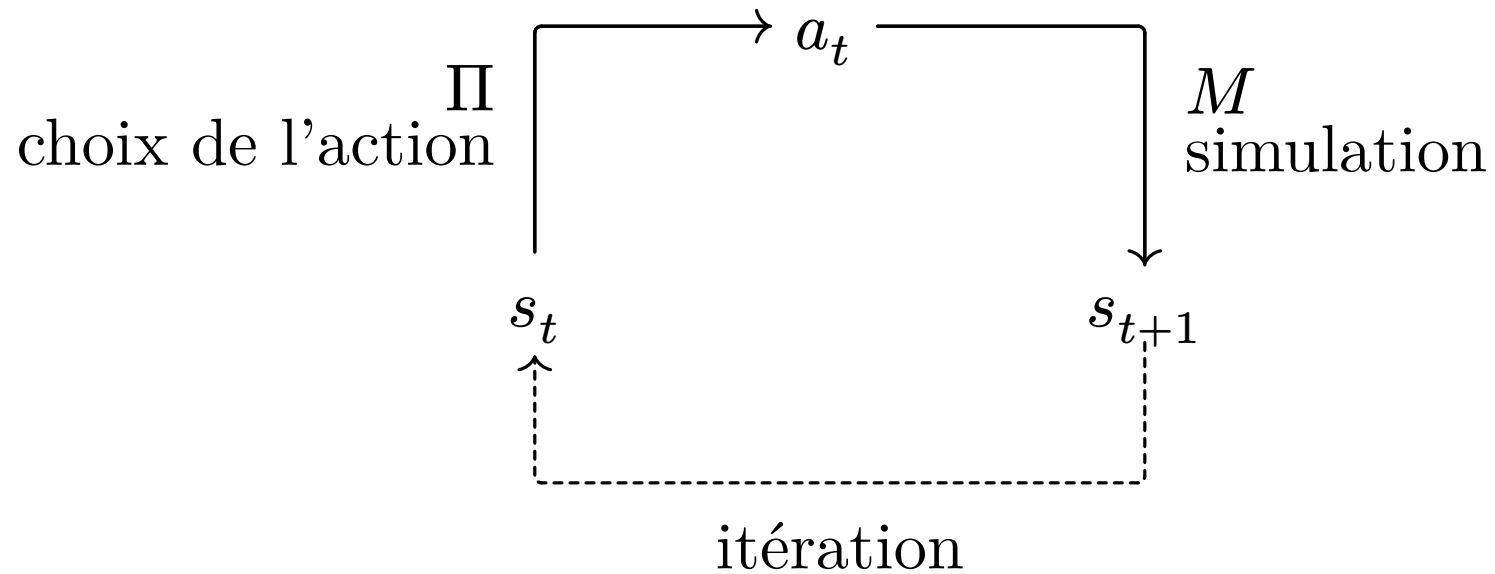
Comparaison des politiques



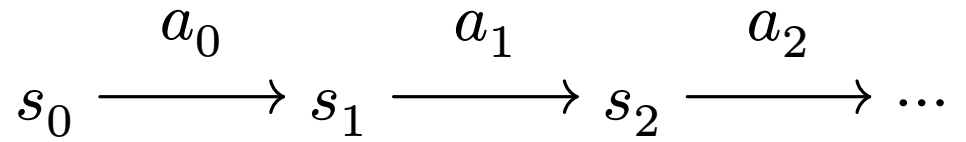
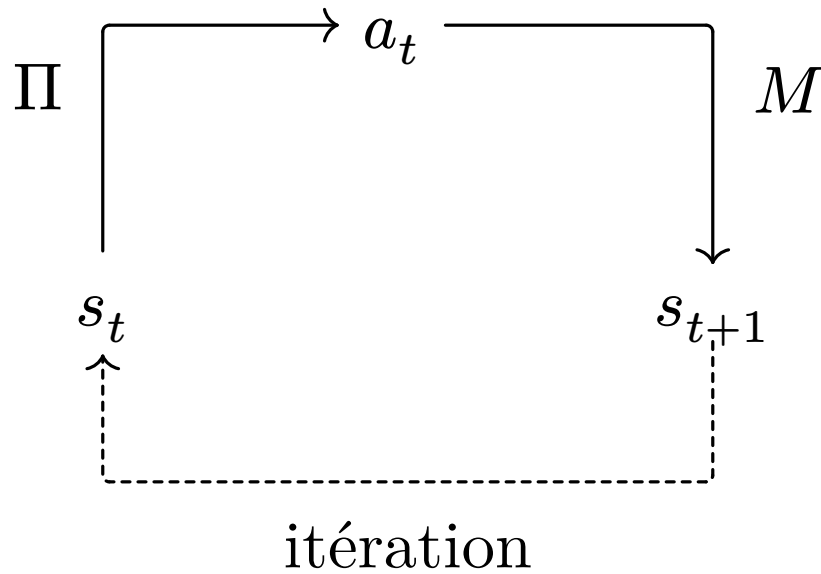
Comparaison des politiques



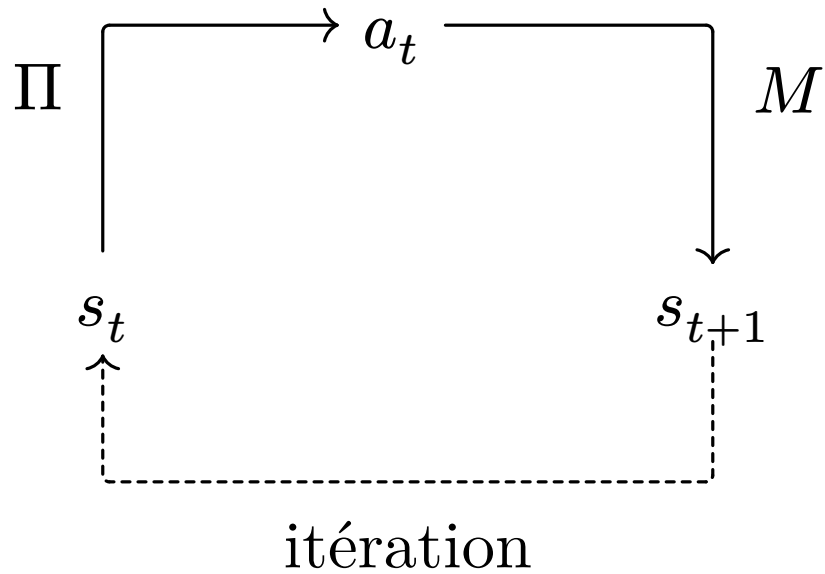
Comparaison des politiques



Comparaison des politiques



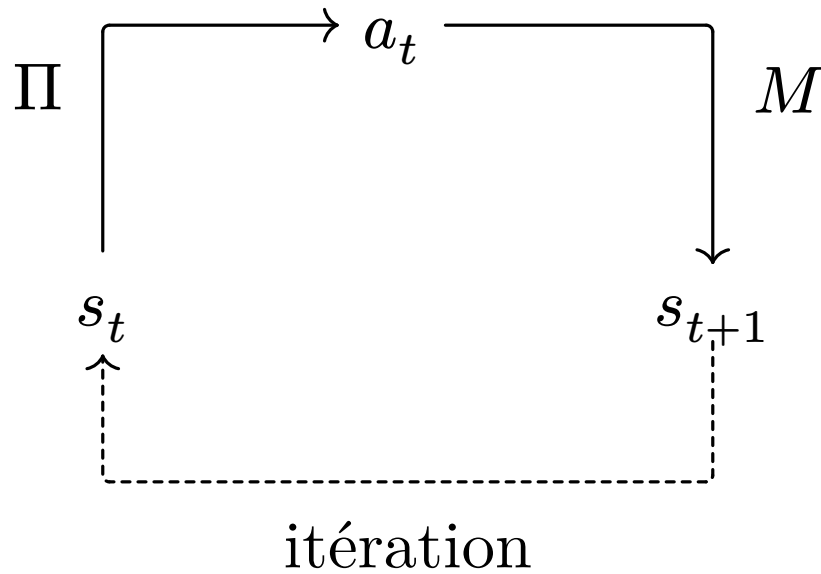
Comparaison des politiques



$$s_0 \xrightarrow{a_0} s_1 \xrightarrow{a_1} s_2 \xrightarrow{a_2} \dots$$

$$((s_0, a_0), (s_1, a_1), (s_2, a_2), \dots)$$

Comparaison des politiques



$$s_0 \xrightarrow{a_0} s_1 \xrightarrow{a_1} s_2 \xrightarrow{a_2} \dots$$

$$((s_0, a_0), (s_1, a_1), (s_2, a_2), \dots) \in \mathcal{C}$$

Comparaison des politiques

A := actions possibles

S := états possibles

$$\mathcal{C} := \left\{ \left\{ \begin{array}{ll} c_0 &= (s_0, a_0) \\ \forall t \in \mathbb{N} & c_{t+1} = (M(c_t), a_t) \end{array} \right. \middle| (s_0, a) \in S \times A^{\mathbb{N}} \right\}$$

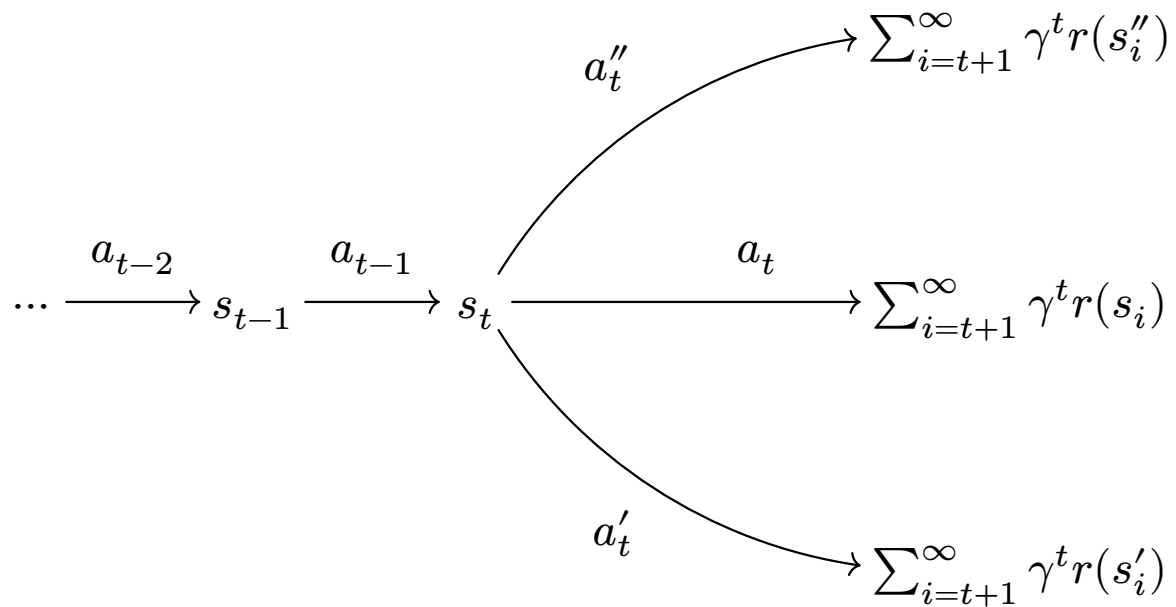
Comparaison des politiques: *Avantage A*

À quel point est-il mieux de choisir a_t plutôt qu'une autre action?

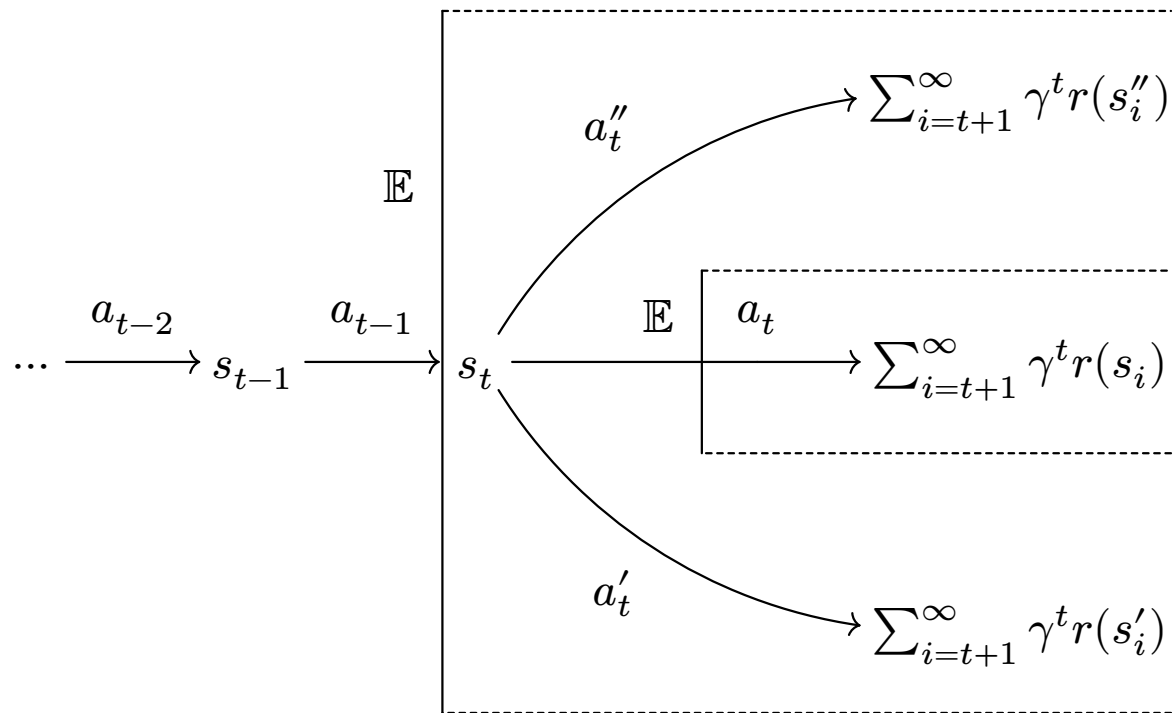
Comparaison des politiques: *Avantage A*

$$\dots \xrightarrow{a_{t-2}} s_{t-1} \xrightarrow{a_{t-1}} \dots$$

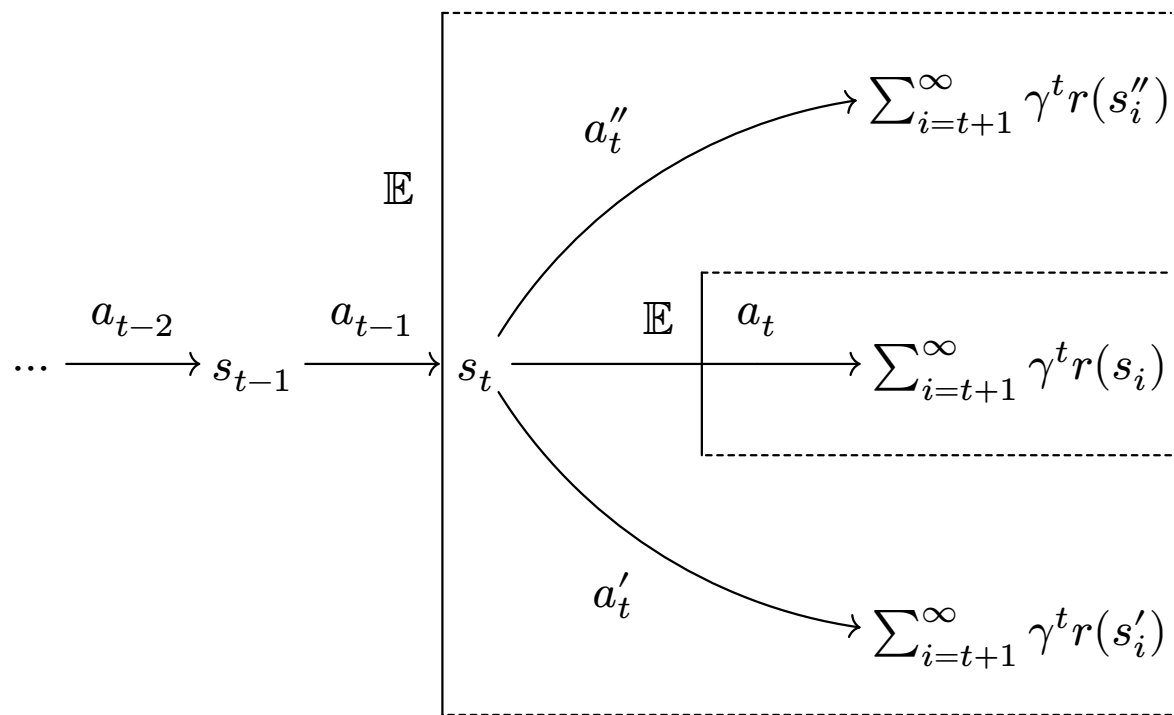
Comparaison des politiques: Avantage A



Comparaison des politiques: Avantage A



Comparaison des politiques: Avantage A



$$A_{\pi,r}(s, a) := \mathbb{E}(\text{avec } a_t) - \mathbb{E}(\text{\`a } t - 1)$$

C'est quoi \mathcal{L} ?

$$\mathcal{L}_r(\pi', \pi) :=$$

C'est quoi \mathcal{L} ?

$$\mathcal{L}_r(\pi', \pi) := \mathbb{E}_{(s_t, a_t)_{t \in \mathbb{N}} \in \mathcal{C}}$$

C'est quoi \mathcal{L} ?

$$\mathcal{L}_r(\pi', \pi) := \mathbb{E}_{(s_t, a_t)_{t \in \mathbb{N}} \in \mathcal{C}} \sum_{t=0}^{\infty}$$

C'est quoi \mathcal{L} ?

$$\mathcal{L}_r(\pi', \pi) := \mathbb{E}_{(s_t, a_t)_{t \in \mathbb{N}} \in \mathcal{C}} \sum_{t=0}^{\infty} \frac{Q_{\pi}(s_t, a_t)}{Q_{\pi'}(s_t, a_t)}$$

C'est quoi \mathcal{L} ?

$$\mathcal{L}_r(\pi', \pi) := \mathbb{E}_{(s_t, a_t)_{t \in \mathbb{N}} \in \mathcal{C}} \sum_{t=0}^{\infty} \frac{Q_{\pi}(s_t, a_t)}{Q_{\pi'}(s_t, a_t)} A_{\pi, r}(s_t, a_t)$$

Optimisation de Π

Mise à jour de la politique RL