A non-stationary PI algorithm

Bruno Scherrer

March 26, 2022

Abstract

...

Solve the n-horizon problem

$$T_{\vec{\mu},\vec{\nu}}v = T^n v.$$

Lemma 1. *For any policy μ, there exist x, $v \leq i < i+c \leq n$, such that*

$$v_\mu(x) - v(x) \leq \frac{\gamma^{i+c}}{1-\gamma^c} \mathbb{1}_x (T^{n-i}v - T^{n-i-c}v).$$

Proof. μ_* play against $\vec{\nu}$. Once a cycle is found involving x, one loops. □

$$M_{x,c} = \{\mu \ ; \ \exists \vec{\nu}, x, c, \ \mathbb{1}_x P_{\mu^c,\vec{\nu}} = \mathbb{1}_x\}.$$

Let

$$v_{x,c} = \max_{\mu \in M_{x,c}} v_\mu(x)$$

Lemma 2. *For any policy ν, there exist x, $v \leq i < i+c \leq n$, such that*

$$v_{\vec{\mu}_1^i (\vec{\mu}_{i+1}^{i+c})^\infty, \nu^\infty}(x) - v(x) \geq \frac{\gamma^{i+c}}{1-\gamma^c} \mathbb{1}_x (T^{n-i}v - T^{n-i-c}v).$$

Proof. $\vec{\mu}$ plays against ν. Once a cycle is found, one loops. □

Algorithm: compute a policy μ' that is better than $(\vec{\mu}_{i+1}^{i+c})^\infty$ for all i and c. From any x, μ' will end up cycling in y.

Can we guarantee that one makes significant progress ?

1 The 1-player case

$$M_{x,c} = \{\pi \ ; \ \exists x, c, \ \mathbb{1}_x (P_{\pi_*})^c = \mathbb{1}_x\}$$

Let

$$T_{\vec{\pi}} v = T^n v$$

For all x, i, c such that $\mathbb{1}_x(T^{n-i}v - T^{n-i-c}v) > 0$. Then consider the next policy as $\vec{\pi}'_{x,c} = \vec{\pi}_{i+1}^{i+c}$. We have

$$v_{\vec{\pi}'}(x) - T^n v(x) \geq v_{x,c}$$

By monotonicity we get n^2 iterations.