Planing, Learning and Intelligent Decision Making - Homework 4

99326 - Sebastião Carvalho, 99331 - Tiago Antunes

March 17, 2024

Contents

1 Question 1

1 a)

To find the equilibrium points for the o.d.e (2), we need to solve the following equation:

$\dot{J} = 0 \Leftrightarrow \mathbb{E}_\pi[c_t + (\gamma - 1)J] = 0 \Leftrightarrow \mathbb{E}_\pi[c_t] + (\gamma - 1)\,\mathbb{E}_\pi[J] = 0$

Since $\mathbb{E}_\pi[c_t] = c_\pi$,

$c_\pi = (1-\gamma)\,\mathbb{E}_\pi[J] \Leftrightarrow \mathbb{E}_\pi[J] = \frac{c_\pi}{1-\gamma}$

Since $\mathbb{E}_\pi[\mathrm{J}^\pi] = \mathrm{J}^\pi$, and $\mathrm{J}^\pi = \frac{c_\pi}{1-\gamma}$, we have that $\mathbb{E}_\pi[J] = \mathbb{E}_\pi[\mathrm{J}^\pi] = \mathrm{J}^\pi$, so the equilibrium point is J^π.

1 b)

$$
\begin{aligned}
\dot{E}_t &= \frac{d}{dt}\frac{1}{2}\left(J^{(t)} - \mathrm{J}^\pi\right)^2 \\
&= \frac{1}{2}\frac{d}{dt}\left(J^{(t)} - \mathrm{J}^\pi\right)^2 \\
&= \frac{1}{2}2\left(J^{(t)} - \mathrm{J}^\pi\right)\frac{d}{dt}\left(J^{(t)} - \mathrm{J}^\pi\right) \\
&= \left(J^{(t)} - \mathrm{J}^\pi\right)\dot{J} \\
&= \left(J^{(t)} - \mathrm{J}^\pi\right)\mathbb{E}_\pi[c_t + (\gamma - 1)J] \\
&= \left(J^{(t)} - \mathrm{J}^\pi\right)(\mathbb{E}_\pi[c_t + (\gamma - 1)J] - \dot{\mathrm{J}^\pi}) \\
&= \left(J^{(t)} - \mathrm{J}^\pi\right)(\mathbb{E}_\pi[c_t + (\gamma - 1)J] - (c_\pi + (\gamma - 1)\,\mathrm{J}^\pi)) \\
&= \left(J^{(t)} - \mathrm{J}^\pi\right)(\mathbb{E}_\pi[c_t] + (\gamma - 1)\,\mathbb{E}_\pi[J] - c_\pi - (\gamma - 1)\,\mathrm{J}^\pi)
\end{aligned}
$$

$$= \left(J^{(t)} - \mathrm{J}^{\pi}\right)\left(c_{\pi} + (\gamma - 1)\,\mathbb{E}_{\pi}[J] - c_{\pi} - (\gamma - 1)\,\mathrm{J}^{\pi}\right)$$
$$= (\gamma - 1)\left(J^{(t)} - \mathrm{J}^{\pi}\right)\left(\mathbb{E}_{\pi}[J] - \mathrm{J}^{\pi}\right)$$

1 c)

The result obtained in (b) suggests that the energy is always decreasing as we move foward in time, unless $J^{(t)} = J^{\pi}$, at which point the derivative is 0, meaning that the value will not change. Given its equation and that the energy is only 0 if $J^{(t)} = J^{\pi}$, we conclude that $J^{(t)}$ converges to J^{π} until they're equal, and then it will permanently stay equal. Knowing that in this case, the state $\mathcal{X} = \{x\}$, we can conclude that if the states are visited infinitely often, the TD(0) algorithm converges to J^{π}, for any $\gamma \in [0, 1]$.