

Planing, Learning and Intelligent Decision Making - Homework 4

99326 - Sebastião Carvalho, 99331 - Tiago Antunes

March 18, 2024

Contents

1 Question 1	1
1 a)	1
1 b)	1
1 c)	2

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 J doesn't depend on π , $\mathbb{E}_\pi[J] = J$, and $J^\pi = \frac{c_\pi}{1 - \gamma}$, we have that

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

So the equilibrium point is J^π , since it's the only point that satisfies the equation.

1 b)

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

Since J^π is an equilibrium point, \dot{J}^π is 0, and we have that

$$\begin{aligned}
&= (J^{(t)} - J^\pi) (\mathbb{E}_\pi[c_t + (\gamma - 1)J^{(t)}] - \dot{J}^\pi) \\
&= (J^{(t)} - J^\pi) (\mathbb{E}_\pi[c_t + (\gamma - 1)J^{(t)}] - (c_\pi + (\gamma - 1)J^\pi)) \\
&= (J^{(t)} - J^\pi) (\mathbb{E}_\pi[c_t] + (\gamma - 1)\mathbb{E}_\pi[J^{(t)}] - c_\pi - (\gamma - 1)J^\pi) \\
&= (J^{(t)} - J^\pi) (c_\pi + (\gamma - 1)J^{(t)} - c_\pi - (\gamma - 1)J^\pi) \\
&= (\gamma - 1) (J^{(t)} - J^\pi) (J^{(t)} - J^\pi) \\
&= (\gamma - 1) (J^{(t)} - J^\pi)^2
\end{aligned}$$

Since $(J^{(t)} - J^\pi)^2$ is always positive, and $\gamma - 1$ is always negative, we have that \dot{E}_t is always negative, and a negative derivative means that the function is decreasing.

Hence, the energy is always decreasing as we move forward in time.

1 c)

The result obtained in (b) suggests that the energy is always decreasing as we move forward 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 $\gamma \in [0, 1]$.