Homework Set #2

1. **Bellman Operator**

   We define the bellman Operator and The bellman Operator assosiated with policy π as follows:

$$(TV)(s) = \max_a E\Big[R_{t+1} + \gamma V(S_{t+1}) \,|\, S_t = s, A_t = a\Big] \tag{1}$$

$$(T^\pi V)(s) = E_\pi\Big[R_{t+1} + \gamma V(S_{t+1}) \,|\, S_t = s\Big] \tag{2}$$

   (a) Prove the monotonicity of T and T^π, i.e show that:

$$V'(s) \geq V(s) \quad \forall s \implies (TV')(s) \geq (TV)(s) \quad \forall s$$

$$V'(s) \geq V(s) \quad \forall s \implies (T^\pi V')(s) \geq (T^\pi V)(s) \quad \forall s$$

   (b) Show that T and T^π are α *contraction* operators, i.e show that:

$$\|TV - TV'\|_\infty \leq \alpha \|V - V'\|_\infty$$
$$\|T^\pi V - T^\pi V'\|_\infty \leq \alpha \|V - V'\|_\infty$$

   for some $\alpha \in (0, 1)$ for any V,V'.

   (c) We define the fix point solution of an operator as the V for which:

$$(TV)(s) = V(s)$$

   prove that T and T^π have a unique fixed point solution.

   (d) Let V^π and V^* be the fixed point solutions to T^π and T , i.e:

$$T^\pi V^\pi = V^\pi$$

$$TV^* = V^*$$

   Prove that for any π:

$$TV^\pi \geqslant V^\pi$$

   and

$$V^* \geqslant V^\pi$$

2. **RL**
   Please solve question 2 in the following link: **exam-rl-questions.pdf**

3. **Easy21**
   Please solve the computer assignment in the following link: **easy21.pdf**

4. **MDP depends on the environment on via the conditional marginals**
   In this exercise we show that the MDP formulation that we saw in the class depends on the environment $p(s', r|s, a)$ only via the marginals $p(s'|s, a)$ and $p(r|s, a)$.

   (a) Write the value iteration equation explicitly.

   (b) Claim that the equation depends on $p(s', r|s, a)$ only via the marginals $p(s'|s, a)$ and $p(r|s, a)$.

   (c) Conclude that 2 environments with the same marginals $p(s'|s, a)$ and $p(r|s, a)$ results with the same optimal policy and the same value function $v^*(s)$ for all $s \in \mathcal{S}$