# ECEN 689 Assignment 1

Diya Li ; UN 830002833

February 13, 2021

## 1

For any $V_1, V_2 \in \mathbb{R}^{|\mathcal{X}|} with V_1 >= V_2$, then for any state, we have

$$\begin{aligned} TV_2 - TV_1 &= max(R(x,a) + \gamma \sum_y P(y|x,a)V_2(y)) - max(R(x,a) + \gamma \sum_y P(y|x,a)V_1(y)) \\ &= max(\gamma \sum_y P(y|x,a)V_2(y)) - max(\gamma \sum_y P(y|x,a)V_1(y)) \\ &<= max(\gamma \sum_y P(y|x,a)V_2(y)) - max(\gamma \sum_y P(y|x,a)V_2(y)) \quad (because V_1 >= V_2) \\ &= 0 \end{aligned}$$

then $TV_2 - TV_1 <= 0$ bellman operator is monotone operator

## 2

$||f(u) - f(v)||_\infty = ||Au - Av||_\infty = ||u - v||_\infty ||A||_\infty$ , where the row sums of A is strictly less than 1, and $\sum_j |a_{ij} \leq \alpha \leq 1|$

, then $0 \leq ||u - v||_\infty ||A||_\infty = ||u - v||_\infty max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}| \leq \alpha ||u - v||_\infty \leq ||u - v||_\infty$

which means $||f(u) - f(v)||_\infty \leq \alpha ||u - v||_\infty$ for an $\alpha \in (0,1)$ and for all $u, v \in U$ the $f(\cdot)$ is a contraction mapping

## 3

Assume $v = argmax_u g_1(u)$, where $g_1 : U \longrightarrow \mathbb{R}; g_2 : U \longrightarrow \mathbb{R}$ then

$|max_u g_1(u) - max_u g_2(u)| = g_1(v) - max_u g_2(u)$ if $max_u g_1(u) \geq max_u g_2(u)$

then for the RHS, we have

$g_1(v) - max_u g_2(u) \leq g_1(v) - g_2(v)$, ( since $max_u g_2(u) \geq g_2(v)$ )

where the RHS is equal to $|g_1(v) - g_2(v)|$

then, $|max_u g_1(u) - max_u g_2(u)| \leq |g_1(v) - g_2(v)| \leq max_u |g_1(u) - g_2(u)|$

We can prove this similarly if $max_u g_1(u) \leq max_u g_2(u)$ , then

$$|max_u g_1(u) - max_u g_2(u)| \leq max_u |g_1(u) - g_2(u)|$$

# 4

## a)

we have $V^*$ as optimal value function,

$$||V^*(x)||_\infty = ||max_\pi V_\pi(x)||_\infty$$

from the linear system of $V_\pi$, the RHS

$$= ||max_\pi\{(I - \gamma P_\pi)^{-1} R_\pi\}||_\infty \leq ||(1-\gamma)^{-1} max_\pi\{R_\pi\}||_\infty = max_x\{(1-\gamma)^{-1} max_\pi\{R_\pi\}\} = R_{max}/(1-\gamma)$$

then

$$||V^*(x)||_\infty \leq R_{max}/(1-\gamma)$$

## b)

for the last inequality in b) , we have

$$k \geq \frac{log(R_{max}) - log(\varepsilon(1-\gamma))}{log 1 - log\gamma}$$

$$k \leq \frac{log(\varepsilon(1-\gamma)) - log(R_{max})}{log\gamma}$$

$$k \leq \frac{log(\frac{\varepsilon(1-\gamma)}{log(R_{max}}))}{log\gamma}$$

$$log\gamma k \leq log(\frac{\varepsilon(1-\gamma)}{R_{max}})$$

for both sides of log function, we have

$$\gamma^k \leq \frac{\varepsilon(1-\gamma)}{R_{max}}$$

then we have

$$\gamma^k R_{max} \leq \varepsilon(1-\gamma)$$

we put this inequality above for later usage, then by the contraction property of bellman operator, which is

$$||TV_1 - TV_2||_\infty \leq \gamma||V_1 - V_2||_\infty$$

we take LHS of the first inequality in b)

$$||V_k - V^*||_\infty \leq \gamma||V_{k-1} - V^*||_\infty \leq .. \leq \gamma^k||V_0 - V^*||_\infty$$

where $V_0 = 0$

$$\gamma^k||V_0 - V^*||_\infty = \gamma^k||V^*||_\infty$$

we have $||V^*(x)||_\infty \leq R_{max}/(1-\gamma)$ from a)

$$\gamma^k||V^*||_\infty \leq \gamma^k(R_{max}/(1-\gamma)) \leq \varepsilon$$

then

$$||TV_1 - TV_2||_\infty \leq \varepsilon$$

**c)**

$$||(V_{m+1} - V_m) + (V^* - V_{m+1})||_\infty \leq ||V_{m+1} - V_m||_\infty + ||V^* - V_{m+1}||_\infty \leq ||V_{m+1} - V_m||_\infty + \gamma||V_m - V^*||_\infty$$

then,

$$(1-\gamma)||V_m - V^*||_\infty \leq ||V_{m+1} - V_m||_\infty$$

$$||V_m - V^*||_\infty \leq (||V_{m+1} - V_m||_\infty)/(1-\gamma)$$

from the inequality of $||V_{m+1} - V_m||_\infty$ in c)

$$||V_{m+1} - V_m||_\infty/(1-\gamma) \leq \varepsilon/2\gamma$$

$$\gamma||V_m - V^*||_\infty \leq \varepsilon/2$$

from the contraction property

$$||V_{m+1} - V^*||_\infty \leq \gamma||V_{m+1} - V^*||_\infty \leq \varepsilon/2$$

Then for the value iteration that reach the maximum of the

# 5

From the definition of infinity norm, we have

$$||V_1^* - V_2^*||_\infty = max_x|V_1^* - V_2^*|$$

$$= max_x|max_a(R(x,a) + \gamma\sum_y P_1(y|x,a)V_1^*(y)) - max_a(R(x,a) + \gamma\sum_y P_2(y|x,a)V_2^*(y))|$$

From the definition of question 3), we have

$$RHS \leq max_x(max_a|R(x,a) + \gamma \sum_y P_1(y|x,a)V_1^*(y) - R(x,a) - \gamma \sum_y P_2(y|x,a)V_2^*(y)|)$$

$$\leq \gamma max_x(max_a| \sum_y P_1(y|x,a)V_1^*(y) - \sum_y P_2(y|x,a)V_2^*(y)|)$$

$$= \gamma max_x(max_a| \sum_y P_1(y|x,a)V_1^*(y) - \sum_y P_2(y|x,a)V_1^*(y) - (\sum_y P_2(y|x,a)V_2^*(y) - \sum_y P_2(y|x,a)V_1^*(y))|)$$

from the contraction property

$$\leq max_x(max_a|(\sum_y P_1(y|x,a) - \sum_y P_2(y|x,a)|) \cdot ||V_1^*(y)||_\infty + \gamma||V_2^* - V_1^*||_\infty$$

$$\leq \gamma \varepsilon R_{max} + \gamma||V_2^* - V_1^*||_\infty$$

then we have,

$$||V_2^* - V_1^*||_\infty \leq (\gamma \varepsilon R_{max})/(1-\gamma)^2$$

## 6

From the definition of infty norm

$$||V^* - V_{\bar{\pi}}||_\infty = max_x|V^*(x) - V_{\bar{\pi}}(x)|$$

by definition of optimal q value function, $V^*(x) = max_a Q^*(x,a)$, and the definition of q value function of a policy $\pi$, we have

$$||\bar{V} - V^*||_\infty = max_x|max_a \bar{Q}(x,a) - max_a Q^*(x,a)|$$

$$\leq max_x max_a|\bar{Q}(x,a) - Q^*(x,a)|$$

$$= ||\bar{Q} - Q^*|| \leq \varepsilon$$

also by triangle inequality, we have

$$||V^* - V_{\bar{\pi}}||_\infty \leq ||V^* - \bar{V}||_\infty + ||\bar{V} - V_{\bar{\pi}}||_\infty$$

we conclude $V_{\bar{\pi}} = R(x,\bar{\pi}(x)) + \gamma \sum_y P(y|x,\bar{\pi}(x)) = max_a(R(x,a) + \gamma \sum_y P(y|x,a)V^* - \gamma\varepsilon \sum_y P(y|x,a)) + \gamma \sum_y P(y|x,\bar{\pi}(x))V_{\bar{\pi}}(y) - \gamma \sum_y P(y|x,\bar{\pi}(x))V^*(y) - \gamma\varepsilon \sum_y P(y|x,\bar{\pi}(x)) = V^* + \gamma \sum_y P(y|x,\bar{\pi})V_{\bar{\pi}}(y) - \gamma \sum_y P(y|x,\bar{\pi})V^*(y) - 2\gamma\varepsilon$ (the middle equation is too long to put in one line)

We have $V^* \leq V_{\bar{\pi}}$, then

$$V^* - V_{\bar{\pi}} \leq \gamma||V^* - V_{\bar{\pi}}||_\infty + 2\varepsilon$$

$$||V^* - V_{\bar{\pi}}||_\infty \leq 2\varepsilon/(1-\gamma)$$

# 7

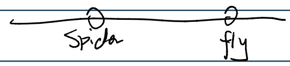
consider a MDP for this question $M = (X, A, P, R, \gamma)$, where X defined as the states of distance between the spider and fly, A defined as action of the spider, and P defined the transition probability function from one state to another state, R is a reward function and $\gamma$ is the discount factor.

From the definition of probability that spider's action for each state, for the transition probability $p_x = P(y|x, a), x, y \in X, a \in A$ with certain states.

I put the rest of the possible solution by handwriting work in the next pages, cuz it will save some time, please check

# Question 7 ecen 689

Friday, February 12, 2021 10:55 PM




( I hand write the question after got some hint from class, but run out of time, so I can not put it in latex )

consider one unit away from the spider and fly will move to same direction which is $x_1 \to x_1$

$P\{\text{the spider will move}\} = 2p$, since the spider will move left or right, that is $p+p$

$P\{\text{the spider will stay}\} = 1-2p$

Consider spider will catch the fly, which is $x_1 \to x_0$

$P\{\text{spider will move}\} = 1-2p$

$P\{\text{spider will stay}\} = p$

Consider if the spider and fly will move to two unit, which is spider will stay

$P\{\text{spider will stay}\} = p$

So the bellman optimation function is

$$V^*(x) = 1 + pV^*(x) + (1-2p)V^*(x-1) + pV^*(x-2), \text{ for } x \geq 2$$

but for the special cases $V^*(1)$, consider the above statement

$$V^*(1) = 1 + \min\{2pV^*(1),\ pV^*(2) + (1-2p)V^*(1)\}$$

$$V^*(2) = 1 + pV^*(2) + (1-2p)V^*(1)$$

$\Rightarrow$

$$V^*(1) = 1 + \min\{2p \cdot V^*(1),\ [p + p(1-2p)V^*(1) + (1-p)(1-2p)V^*(1)]/(1-p)\}$$

Consider the one unit away bellman function of $V^*(1)$

$$\begin{cases} V^*(1) = 1 + 2pV^*(1) & \text{if } 2pV^*(1) < pV^*(2) + (1-2p)V^*(1) \\ \quad \Rightarrow 2pV^*(1) < [p + (1-2p)V^*(1)]/(1-p) & \\ V^*(1) = pV^*(2) + (1-2p)V^*(1) & \text{if } 2pV^*(1) > pV^*(2) + (1-2p)V^*(1) \\ \quad \Rightarrow 2pV^*(1) \geq p/(1-p) + (1-2p)V^*(1)/(1-p) & \end{cases}$$

$\Rightarrow V^*(1) = 1/(1-2p)$

$\Rightarrow 2p(1-p) \leq (1-p)(1-2p)$

$\Rightarrow 2p - 2p^2 \leq 1-2p-p-2p^2 \Rightarrow p \leq 1/3$

$\Rightarrow V^*(1) = 1/p$

$\Rightarrow (1-2p)V^*(1) \leq 2(1-p)$

$\Rightarrow p \geq 1/3$

$$\Rightarrow \begin{cases} V^*(1) = 1/(1-2p) & \text{if } p \leq \frac{1}{3} \\ V^*(1) = \frac{1}{p} & \text{if } p \geq \frac{1}{3} \end{cases}$$

then for the two unit away from the spider to catche the fly

we have $V^*(2) = \frac{1}{1-p}(1 + V^*(1) + 2pV^*(1))$

$= \frac{1}{(\ldots)}(1 + \min[2p \cdot V^*(1),\ p + p(1-2p)V^*(1) + (1-p)(1-2p)V^*(1)/(1-p)]$

we have $V^*(2) = \frac{1}{1-p}(1+V^*(1) + 2pV'(1))$

$= \frac{1}{1-p}(1+(1-2p)(1+\min\{2p\cdot V^*(1),\ p+p(1-2p)V^*(1)+(1-p)(1-2p)V^*(1)\}/(1-p)]$

similar to the closed form of $V^*(1)$

$$V^*(2) = \begin{cases} \frac{1}{p} & \text{if } p \geq \frac{2}{3} \\ 2/(1-p) & \text{if } p \leq \frac{1}{3} \end{cases}$$