Problem Set 3

(1) $M = \langle S, A, R, T, \gamma \rangle$, $\gamma \in [0,1)$, $(|S \times A| < \infty)$ finite

$\pi : S \to \Delta(A)$; $d^\pi(s) = (1-\gamma)\sum_{t=0}^{\infty} \gamma^t p^\pi(s_t = s)$

$\beta \in \Delta(S)$ = initial state distribution.

$p^\pi(s_0 = s) = \beta(s)$ $\forall \pi$ & $s \in S$.

$$d^\pi(s') = (1-\gamma)\beta(s') + \gamma \sum_{s\in S}\sum_{a \in A} T(s'|s,a)\pi(a|s) d^\pi(s).$$

$$\sum d^\pi(s) = \sum (1-\gamma)\left[\sum \gamma^t p^\pi(s_t = s)\right]$$

$$= (1-\gamma)\sum_{1}^{\infty} \gamma^t \sum_{s} p^\pi(s_t = s) = (1-\gamma)\sum \gamma^t = 1-\gamma \frac{1}{1-\gamma} = 1.$$

$\Rightarrow \sum d^\pi(s) = 1$; $\pi : S \to \Delta(A)$; $\beta \in \Delta(S)$

$$d^\pi(s') = (1-\gamma)\beta(s') + \gamma \sum\sum_{(s,a)} T(s'|s,a)\pi(a|s) d^\pi(s).$$

$$d^\pi(s') = (1-\gamma)\left[\sum_{t=1}^{\infty} \gamma^t p^\pi(s_t = s) + p^\pi(s_0 = s')\right]$$

$$= (1-\gamma)\beta(s') + \underbrace{(1-\gamma)\sum_{t=1}^{\infty} \gamma^t p^\pi(s_t' = s)}$$

$$= (1-\gamma)\sum_{t=1}^{\infty} \gamma^t \left[\sum_{a} p^\pi(s_{t-1} = s') \begin{bmatrix} P(s'|s,a) \\ \pi(a|s) \end{bmatrix}\right]$$

$$= (1-\gamma)\gamma \sum_{t=0}^{\infty} \gamma^t \left[\sum_{a} p^\pi(s_t = s') P(s'|s,a)\pi(a|s)\right]$$

$$= (1-\gamma)\gamma \frac{1}{1-\gamma} \sum_{(s,a)} \left(p^\pi(s_t = s') P(s'|s,a)\pi(a|s)\right)$$

$$= \gamma \sum_{s,a} p^\pi(s_t = s') P(s'|s,a)\pi(a,s).$$

② π; π'; we have :

$$\|d^{\pi} - d^{\pi'}\|_1 \le \frac{2\gamma}{1-\gamma} E_{s \sim d^{\pi}}[D_{TV}(\pi(\cdot|s) \| \pi'(\cdot|s))].$$

$D_{TV}(\cdot) = \frac{1}{2} \sum_{a \in A} |\pi(a|s) - \pi'(a|s)|$ = total variance distance between π, π' at s.

$$\frac{2\gamma}{1-\gamma} E[\tfrac{1}{2} \sum |\pi(a|s) - \pi'(a|s)|]$$

$$= \frac{\gamma}{1-\gamma} E[\sum_{a \in A} |\pi(a|s) - \pi'(a|s)|].$$

$$= \frac{\gamma}{1-\gamma} E[\sum |\pi(a|s) - \pi'(a|s)|]$$

$$= \gamma \sum_{t=0}^{\infty} \gamma^t E[\,] = \sum_{t=1}^{\infty} \gamma^t E[\,] = E[\sum_{t=1}^{\infty} \gamma^t \sum_{a \in A} |\pi(a|s) - \pi'(a|s)|].$$

$$d^{\pi} = (1-\gamma) \sum_{t=0}^{\infty} \gamma^t P^{\pi}(s_t = s).$$ $\boxed{\text{check solution}}$

③ $x^{\pi} \in \Delta(S \times A)$: $x^{\pi}(s,a) = d^{\pi}(s)\pi(a|s)$.

$d^{\pi}(s)\pi(a|s)$

$$\|x^{\pi} - x^{\pi'}\| = \sum_{s,a} |x^{\pi}(s,a) - x^{\pi'}(s,a)| = \sum | d^{\pi}(s,a)\pi(a|s) - d^{\pi'}(s,a')\pi'(a|s)|$$

$|ab - cd| \le k|a-b|$ ~~assume~~.

$$\Rightarrow |ab - cd| \le k(a-b). \quad = \sum | d^{\pi}(s,a)\pi(a|s) - d^{\pi}(s,a)\pi'(a|s) + d^{\pi}(s,a)\pi'(a|s) - d^{\pi'}(s,a)\pi'(a|s)|$$

$$= \underbrace{\sum d^{\pi}(s,a)|\pi(a|s) - \pi'(a|s)|}_{2E[D_{TV}]} + \sum |d^{\pi} - d^{\pi'}| \underbrace{\pi'(a|s)}_{=1 \text{ over } \sum s}$$

$$2E[D_{TV}]. \quad + \quad \le \frac{2\gamma}{1-\gamma} E[D_{TV}].$$

$$\left(2 + \frac{2\gamma}{1-\gamma}\right) E[D_{TV}] = \frac{2}{1-\gamma} E[D_{TV}].$$

CS CamScanner

$$R_{MAX} = \max_{(s,a) \in S \times A} |R(s,a)|$$

$$E_{s_0 \sim \beta}\left[V^{\pi}(s_0) - V^{\pi'}(s_0)\right] \leq \frac{2R_{MAX}}{1-\gamma} E[D_{TV}].$$

$$E[V^{\pi}(s_0) - V^{\pi'}(s_0)] = E[V^{\pi}(s_0)] - E[V^{\pi'}(s_0)].$$

$$= R^T x^{\pi} - R^T x^{\pi'}$$

$$= R^T (x^{\pi} - x^{\pi'})$$

$$\leq |R^T(x^{\pi} - x^{\pi'})| \leq \underset{MAX}{|R|_{\infty}} \|x^{\pi} - x^{\pi'}\|$$

$$\leq \frac{2R_{MAX}}{1-\gamma} E[D_{TV}]$$