

שאלה 1:

$$Q_{M'}^*(s, a) = Q_M^*(s, a) + f(s) \quad \text{נרשם שני MDP כג:}$$

המקסימום (זמור) של  $Q_{M'}^*(s)$  -  $Q_M^*(s)$  קוטר המדיניות המצטיינת

ב-  $M'$  ו-  $M$ .

כיוון של  $Q_{M'}^*(s)$  הוא אבסולוטי יותר מתק"ס:

$$\pi_{M'}^*(s) = \arg \max_a Q_{M'}^*(s, a)$$

נניח ש  $\pi$  (התוכנית) נקרא:

$$\pi_{M'}^*(s) = \arg \max_a Q_{M'}^*(s, a) = \arg \max_a Q_M(s, a) + f(s)$$

$$= \arg \max_a Q_M(s, a) = \pi_M^*(s)$$

כלומר זהו  $\pi_M^*$  -  $\pi_{M'}^*$  זהו אותו  $\pi$  -  $M$

כלומר סה"כ קיבלנו:  $\pi_{M'}^*(s) = \pi_M^*(s)$  כלומר  $\pi$  הוא המדיניות המצטיינת בשני ה-MDP.

②

$$V_{\pi^*}^{\pi^*}(s) = E^{\pi^*} \left[ \sum_{t=0}^{\infty} \gamma^t (R(s_t, a_t, s_{t+1}) + \phi(s_t) - \gamma \phi(s_{t+1})) \mid s_0 = s \right]$$

$$= E^{\pi^*} \left[ \sum_{t=0}^{\infty} \gamma^t R(s_t, a_t, s_{t+1}) \mid s_0 = s \right] + E^{\pi^*} \left[ \underbrace{\sum_{t=0}^{\infty} \gamma^t \phi(s_t) - \gamma^{t+1} \phi(s_{t+1})}_{=0} \mid s_0 = s \right]$$

$$= V_m^{\pi^*}(s) + \phi(s)$$

מלכו בליון הימני הימני שב שולח ויגז' ויגז' נין

[illegible]

M 1/28 P2 1/20/16  $\frac{2+6}{11}^*$

האיכות  
נתיב קטן  
שומר  
היינו  
מקנינו  
הקטן  
הוא

$$V_M^{\uparrow}(s) > V_M^{\pi^*}(s)$$

הערה:  $\phi(s)$  הוא פולינום

$$V_m^{\pi}(s) + \phi(s) \geq V_m^{\pi^*}(s) + \phi(s)$$

$\pi^*$        $\pi^*$        $\text{By}$      $\text{or}$      $y$      $\text{or}$      $x$

$$1. \quad V_{m_1}^{\pi^*}(s) + \phi(s) = V_{m_1'}^{\pi^*}(s)$$

$$7. \quad V_n^{\hat{\pi}}(s) + \phi(s) = V_{n'}^{\hat{\pi}}(s)$$

$$V_{m'}^{\hat{\pi}}(s) > V_{m'}^{\pi^*}(s)$$

$M' \geq -f_{\text{red}}/k \quad \pi^* \ell$   $\mu$   $\text{max}$   $\text{of}$

אשר / מנחם / ה'תשנ"ו / חסד / השבת / כנדב.