

⑦

## Policy Improvement

$$\therefore \forall s \text{ BP } Q_{\pi}(s, \pi'(s)) \geq V_{\pi}(s) \quad \text{⑧}$$

$$\therefore \text{BP } V_{\pi'}(s) \geq V_{\pi}(s)$$

$$V_{\pi}(s) \stackrel{\text{BP}}{=} Q_{\pi}(s, \pi'(s)) \quad \text{⑨}$$

$$= E [R_0 + \gamma V_{\pi}(s_1) | S_0 = s, A_0 = \pi'(s)]$$

$$= E_{\pi'} [R_0 + \gamma V_{\pi}(s_1) | S_0 = s]$$

$$\stackrel{\text{BP}}{=} E_{\pi'} [R_0 + \gamma Q_{\pi}(s_1, \pi'(s_1)) | S_0 = s]$$

$$\stackrel{\text{BP}}{=} E_{\pi'} [R_0 + \gamma R_1 + \gamma^2 V_{\pi}(s_2) | S_0 = s]$$

$$\vdots$$

$$\stackrel{\text{BP}}{=} E_{\pi'} [R_0 + \gamma R_1 + \dots | S_0 = s]$$

$$= V_{\pi'}(s)$$