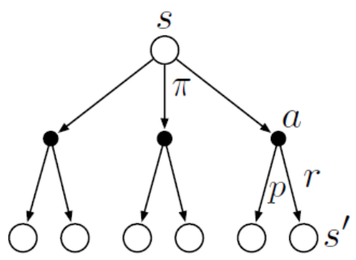
Bellman Backup Diagrams + Corresponding Equations



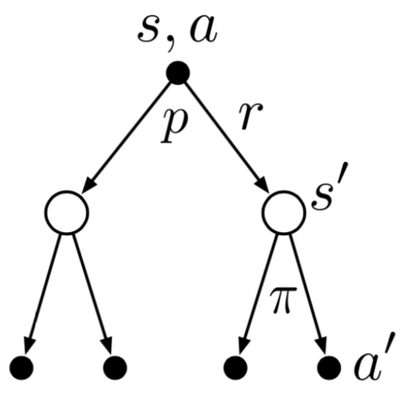
Backup diagram for v_π

$$v_\pi(s) \doteq \mathbb{E}[G_t|S_t = s]$$

$$= \mathbb{E}[R_{t+1} + \gamma G_{t+1}|S_t = s]$$

$$= \sum_a \pi(a|s) \sum_{s',r} p(s', r|s, a) \left[r + \gamma \mathbb{E}[G_{t+1}|S_{t+1} = s']\right]$$

$$\boxed{= \sum_a \pi(a|s) \sum_{s',r} p(s', r|s, a) \left[r + \gamma v_\pi(s')\right]}$$



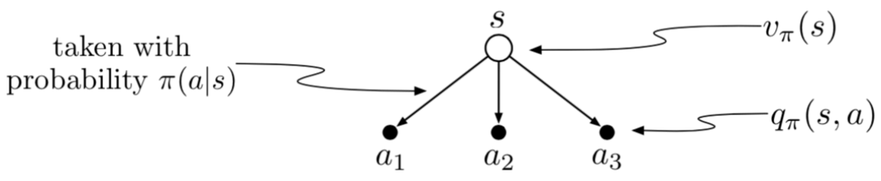
$$q_\pi(s, a) = \mathbb{E}_\pi \left[G_t|S_t = s, A_t = a\right]$$

$$= \mathbb{E}_\pi \left[R_{t+1} + \gamma G_{t+1}|S_t = s, A_t = a\right]$$

$$= \sum_{s',r} p(s', r|s, a) \left[\mathbb{E}_\pi[R_{t+1}|S_t = s', A_t = a'] + \gamma \mathbb{E}[G_{t+1}|S_t = s', A_t = a']\right]$$

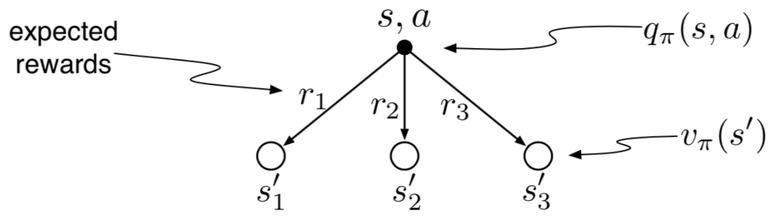
$$= \sum_{s',r} p(s', r|s, a) \left[r + \gamma \sum_{a'} \mathbb{E}[G_{t+1}|S_t = s']\right]$$

$$\boxed{q_\pi(s, a) = \sum_{s',r} p(s', r|s, a) \left[r + \gamma \sum_{a'} \pi(a'|s') q_\pi(s', a')\right]}$$



$$v_\pi(s) = \mathbb{E}_\pi \left[G_t|S_t = s\right]$$

$$\boxed{v_\pi(s) = \sum_{a \in A} \pi(a|s) q_\pi(s, a)}$$



$$q_\pi^1(s, a) = p(s_1, r_1|s, a) \left[r_1 + \gamma v_\pi(s_1)\right]$$

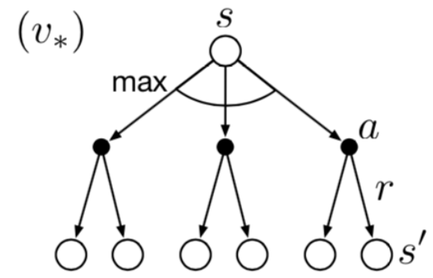
$$q_\pi^2(s, a) = p(s_2, r_2|s, a) \left[r_2 + \gamma v_\pi(s_2)\right]$$

$$q_\pi^3(s, a) = p(s_3, r_3|s, a) \left[r_3 + \gamma v_\pi(s_3)\right]$$

$$q_\pi(s, a) = q_\pi^1(s, a) + q_\pi^2(s, a) + q_\pi^3(s, a)$$

$$\boxed{= \sum_{s',r} p(s', r|s, a) \left[r + \gamma v_\pi(s')\right]}$$

Bellman Optimality Diagrams + Corresponding Equations



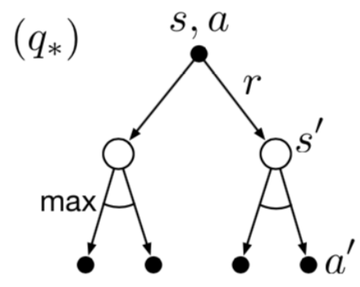
$$v_*(s) = \max_{a \in A(s)} q_{\pi_*}(s, a)$$

$$= \max_a \mathbb{E}_{\pi_*} \left[G_t|S_t = s, A_t = a\right]$$

$$= \max_a \mathbb{E}_{\pi_*} \left[R_{t+1} + \gamma G_{t+1}|S_t = s, A_t = a\right]$$

$$= \max_a \mathbb{E}_{\pi_*} \left[R_{t+1} + \gamma v_*(S_{t+1})|S_t = s, A_t = a\right]$$

$$\boxed{= \max_a \sum_{s',r} p(s', r|s, a) \left[r + \gamma v_*(s')\right]}$$



$$q_*(s, a) = \mathbb{E}\left[R_{t+1} + \gamma \max_{a'} q_*(S_{t+1}, a')|S_t = s, A_t = a\right]$$

$$\boxed{= \sum_{s',r} p(s', r|s, a) \left[r + \gamma \max_{a'} q_*(s', a')\right]}$$