Bellman Equation Derivation

$$v_\pi(s) = \mathbb{E}_\pi\left[\sum_{k\geq 0} \gamma^k R_{t+k+1} \mid S_t = s\right]$$

$$= \mathbb{E}_\pi[R_{t+1} \mid S_t = s] + \mathbb{E}_\pi\left[\sum_{k\geq 1} \gamma^k R_{t+k} \mid S_t = s\right]$$

$$= \sum_a \pi(a|s) \sum_{s',r} p(s', r \mid s, a) r + \gamma \mathbb{E}_\pi\left[\sum_{k\geq 1} \gamma^{k-1} R_{t+k+1} \mid S_t = s\right]$$

$$\sum_{s',r}$$

$$v_\pi(s) = \mathbb{E}_\pi\left[\sum_{k\geq 0} \gamma^k R_{t+k+1} \mid S_t = s\right]$$

$$= \mathbb{E}_\pi[R_{t+1} \mid S_t = s] + \mathbb{E}_\pi\left[\sum_{k\geq 1} \gamma^k R_{t+k} \mid S_t = s\right]$$

$$= \sum_a \pi(a|s) \sum_{s',r} p(s', r \mid s, a) r + \gamma \mathbb{E}_\pi\left[\sum_{k\geq 1} \gamma^{k-1} R_{t+k+1} \mid S_t = s\right]$$

$$= \sum_a \pi(a|s) \sum_{s',r} p(s', r \mid s, a) r + \gamma \mathbb{E}_\pi\left[\sum_{k\geq 0} \gamma^k R_{t+k+2} \mid S_t = s\right]$$

(Total Prob Law)

$$= \sum_a \pi(a|s) \sum_{s',r} p(s', r \mid s, a) r + \sum_a \pi(a|s) \sum_{s',r} p(s', r \mid s, a) \gamma \mathbb{E}_\pi\left[\sum_{k\geq 0} \gamma^k R_{t+2+k} \mid S_{t+1} = s', S_t = s\right]$$

(Markov Property)

$$= \sum_a \pi(a|s) \sum_{s',r} p(s', r \mid s, a) r + \sum_a \pi(a|s) \sum_{s',r} p(s', r \mid s, a) \gamma \mathbb{E}_\pi\left[\sum_{k\geq 0} \gamma^k R_{t+1+k+1} \mid S_{t+1} = s'\right]$$

$$= \sum_a \pi(a|s) \sum_{s',r} p(s', r \mid s, a)\left(r + \gamma \mathbb{E}_\pi\left[\sum_{k\geq 0} \gamma^k R_{t+1+k+1} \mid S_{t+1} = s'\right]\right)$$

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