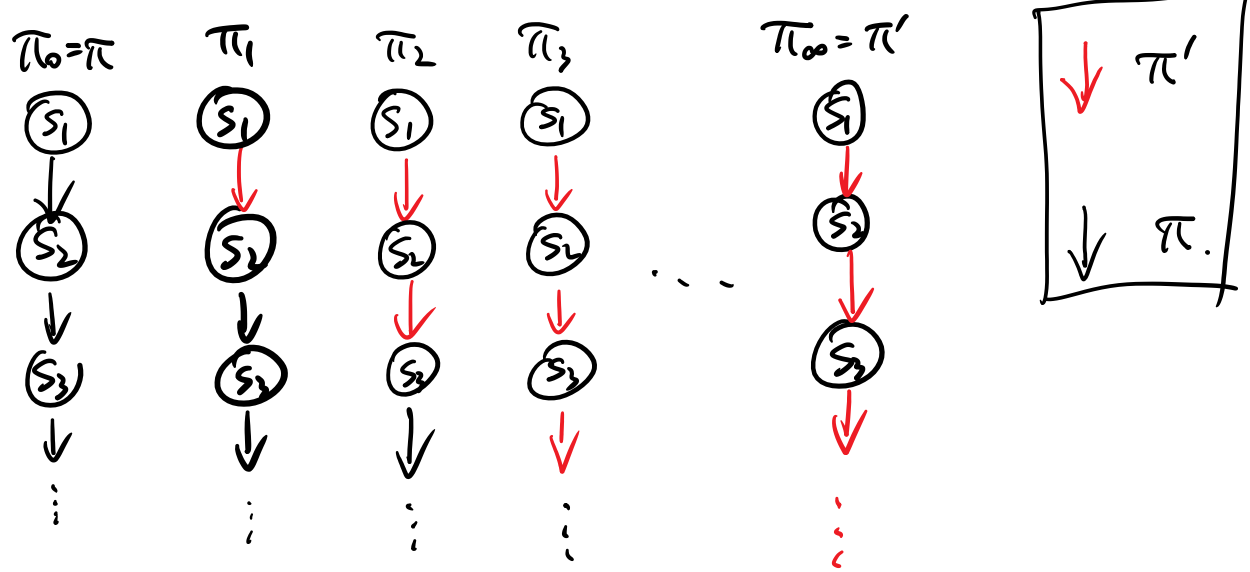
Proof

Saturday, August 25, 2018 4:15 PM

Proof: Define π_i as an non-stationary policy, where the first i steps π' is followed, and π is followed for the remaining steps.

Note: $s_1 = s$



$$V^{\pi'}(s) - V^{\pi}(s) = \sum_{i=0}^{\infty} \left(V^{\pi_{i+1}}(s) - V^{\pi_i}(s) \right)$$

$$= \sum_{i=0}^{\infty} \gamma^i \sum_{s' \in S} \mathbb{P}[s_{i+1} = s' \mid s_1 = s, \pi'] \left(Q^{\pi}(s', \pi'(s')) - Q^{\pi}(s', \pi(s')) \right)$$

$$= \sum_{s' \in S} \sum_{i=0}^{\infty} \gamma^i \mathbb{P}[s_{i+1} = s' \mid s_1 = s, \pi'] \; A^{\pi}(s', \pi')$$

$$= \frac{1}{1-\gamma} \sum_{s' \in S} \eta_s^{\pi'}(s') \; A^{\pi}(s', \pi') = \frac{1}{1-\gamma} \mathbb{E}_{s' \sim \eta_s^{\pi'}} \left[A^{\pi}(s', \pi') \right]$$