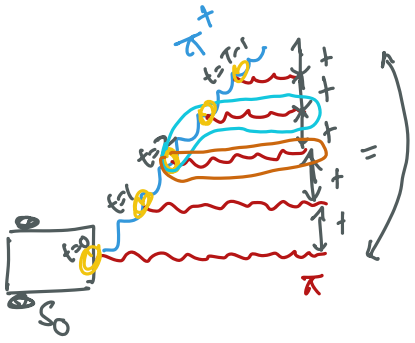
# Performance Difference Lemma (PDL)




$$V^{\pi^+}(s_0) - V^{\pi}(s_0) = \sum_{t=0}^{T-1} \mathop{E}_{s_t \sim d_t^{\pi^+}} \left[ Q^{\pi}(s_t, \pi^+(s_t)) - V^{\pi}(s_t) \right]$$

Advantage of $\pi^+$ over $\pi$

"Performance Difference is the sum over On-Policy (states $\pi^+$ visits) Advantage (of $\pi^+$ over $\pi$)" $A^{\pi}(s_t, \pi^+(s_t))$

$A < 0 \Rightarrow \pi^+$ is better than $\pi$ at that state

---

## Policy Iteration

$$\pi^+(s) = \operatorname*{argmin}_a Q^{\pi}(s,a) \quad \forall s$$

$$Q^{\pi}(s, \pi^+(s)) \leq Q^{\tilde{\pi}}(s, \pi(s)) = V^{\pi}(s) \quad \forall s$$

$$Q^{\pi}(s, \pi^+(s)) - V^{\pi}(s) \leq 0 \quad \forall s$$

$$A^{\pi}(s, \pi^+) \leq 0 \quad \forall s$$

$$V^{\pi^+}(s) - V^{\pi}(s)$$

$$V(s_0) - V(s_0)$$

$$= \sum_{t=0}^{T-1} \mathbb{E}_{s_t \sim d_t^{\pi^+}} \left[ A^{\pi}(s_t, \pi^+) \right] \leq 0$$

$$\leq 0$$

$$\Rightarrow \quad V^{\pi^+}(s_0) \leq V^{\pi}(s_0)$$

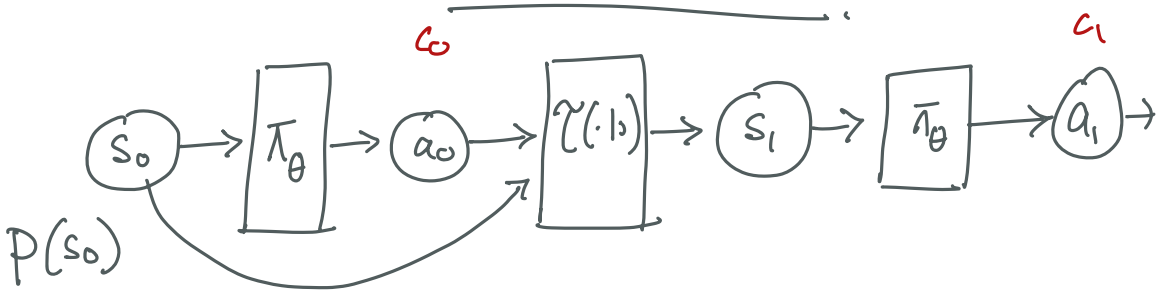
$$s_t \sim d_t^{\pi^+}$$

$$P(s_0)\, \pi(a_0|s_0)\, \underbrace{P(s_1|s_0,a_0)}_{\text{TRANSITION}}$$

$$\pi(a_2|s_2) \cdots P(s_t|s_{t-1}, a_{t-1})$$

---

# POLICY GRADIENTS




$$V^{\pi_\theta}(s_0) = \sum_{t=0}^{T-1} \gamma^t c_t$$

$$\nabla_\theta V^{\pi_\theta}(s_0) = \nabla_\theta J(\theta)$$