Alt. proof for monotone improvement.

PI: $\pi_{k+1} \leftarrow \pi_{Q^{\pi_k}}$.

Monotone improvement: $V^{\pi_{k+1}} \geq V^{\pi_k}$.

Performance difference Lemma. [Kakade & Langford '02]
(P-D Lemma)

$\forall \pi, \pi', s$.

$$V^{\pi'}(s) - V^{\pi}(s) = \frac{1}{1-\gamma} \mathbb{E}_{s' \sim d_s^{\pi'}} \left[Q^{\pi}(s', \pi') - V^{\pi}(s') \right].$$

$$Q^{\pi}(s', \pi') - V^{\pi}(s') = Q^{\pi}(s', \pi') - Q^{\pi}(s', \pi)$$

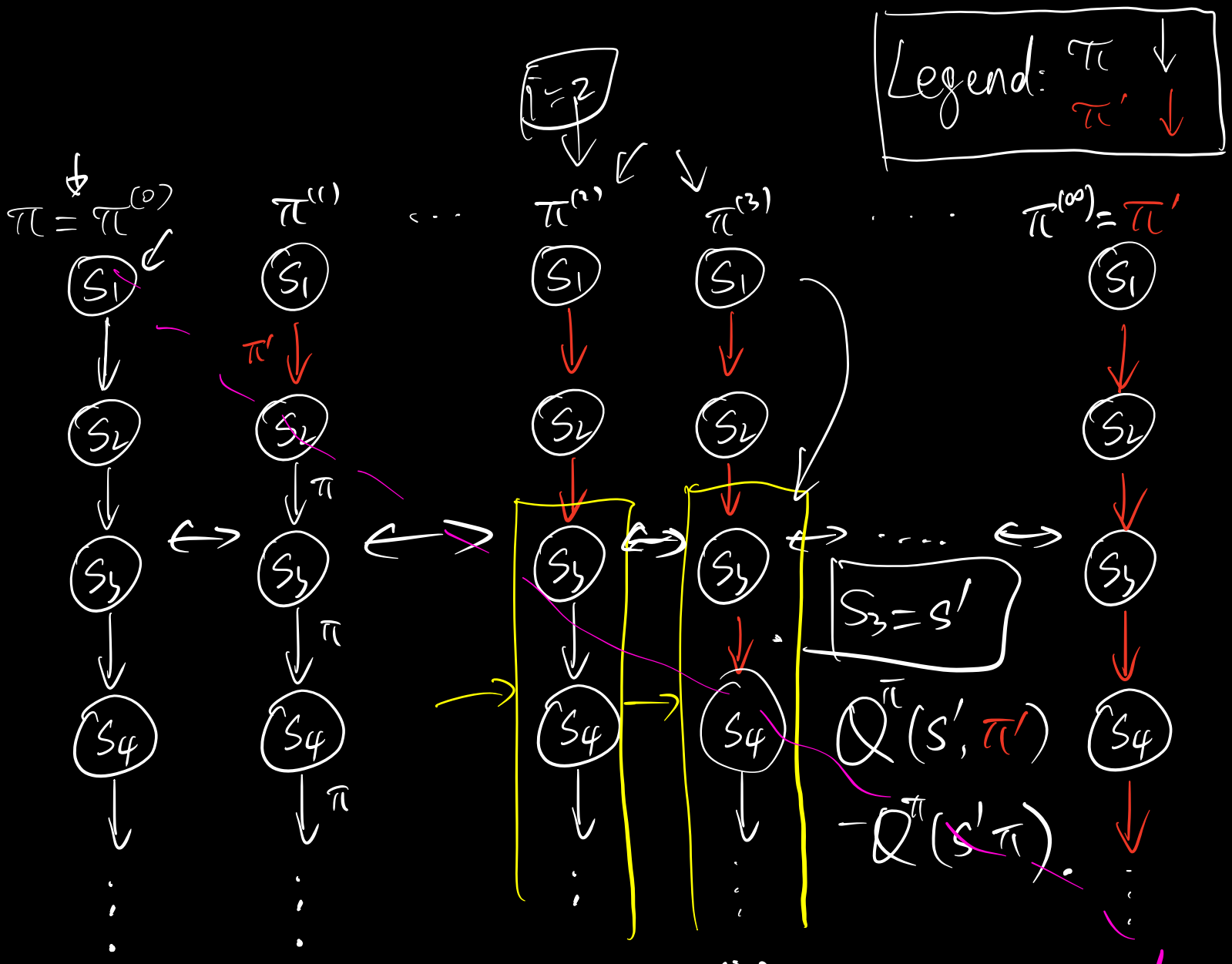
$$=: A^{\pi}(s', \pi') \quad \text{"advantage"}.$$

Use P-D Lemma to establish monotone impr.

$$V^{\pi_{k+1}}(s) - V^{\pi_k}(s) = \frac{1}{1-\gamma} \mathbb{E}_{s' \sim d_s^{\pi_{k+1}}} \left[Q^{\pi_k}(s', \pi_{k+1}) - Q^{\pi_k}(s', \pi_k) \right]$$

Recall: $\pi_{k+1} \leftarrow \pi_{Q^{\pi_k}}$.

i.e. $\forall s, \quad \pi_{k+1}(s) = \operatorname{argmax}_a Q^{\pi_k}(s, a)$.



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

$$= \sum_{i=0}^{\infty} \gamma^i \mathbb{E}_{s' \sim d_{s,i+1}^{\pi'}} \left[Q^{\pi}(s', \pi') - Q^{\pi}(s', \pi) \right]$$

$s_1 = s$

$$= \frac{1}{1-\gamma} \mathbb{E}_{s' \sim d_s^{\pi'}} \left[Q^{\pi}(s', \pi') - Q^{\pi}(s', \pi) \right]$$

Linear Programming

$$\max \; x_1 + x_2 \quad \text{s.t.} \quad x_1 - 2x_2 \le 1.$$

Solve for V^* (primal).

Choose any $\mu \in \mathbb{R}^S$ s.t. $\forall s,\ \mu(s) > 0$, $\sum_s \mu(s) = 1$.

$$\min_{V \in \mathbb{R}^S} \mu^T V \quad \text{s.t.} \quad TV \le V$$

① why is opt. V^*?

② why linear?

① : $TV \le V \Rightarrow T(TV) \le TV \le V$.

$$V^* = T^\infty V \le V.$$

∴ any feasible V must $V \ge V^*$.

& V^* is feasible.

② Constraint : $\forall s$.

$$\max_a \left(R(s,a) + \gamma \mathbb{E}_{s' \sim P(\cdot|s,a)}[V(s')] \right) \le V(s)$$

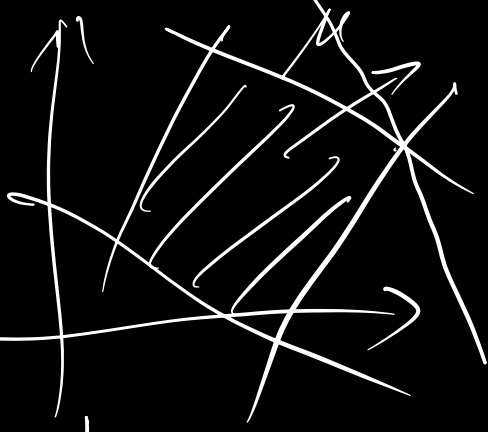
$\forall s, a$

$$R(s,a) + \gamma \mathbb{E}_{s' \sim P(\cdot|s,a)}[V(s')] \le V(s).$$

Dual form. $\mathbb{R}^{S\times A}$.

$\max\ d^T R$. $\quad (d_\mu^\pi)^T R = (1-\gamma) V^\pi(\mu)$

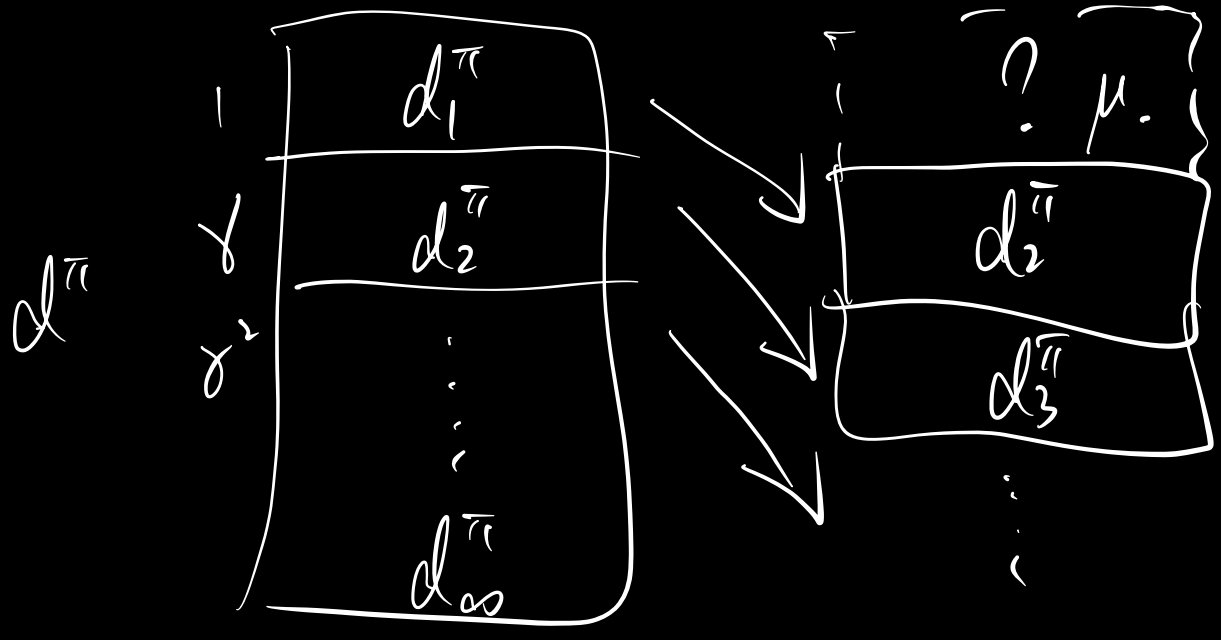
$d \geq 0$



s.t. $\forall s'$.

$$\sum_{a'} d(s',a') = \boxed{\mu(s')} + \gamma \sum_{s,a} d(s,a) P(s'|s,a).$$

any feasible $d = d_\mu^\pi$. $\forall \pi$.



$$V^{\pi^{(3)}}(s_1) - V^{\pi^{(2)}}(s_1).$$

$$= \mathbb{E}\left[\sum_{t=1}^{\infty} \gamma^{t-1} r_t \,\middle|\, s_1, \pi^{(3)}\right]$$

$$- \mathbb{E}\left[\sum_{t=1}^{\infty} \gamma^{t-1} r_t \,\middle|\, s_1, \pi^{(2)}\right].$$

$$= \mathbb{E}\left[\sum_{t=3}^{\infty} \gamma^{t-1} r_t \,\middle|\, s_1, \pi^{(3)}\right]$$

$$- \mathbb{E}\left[\sum_{t=3}^{\infty} \cdots \gamma^2 \cdot Q^{\pi}(s', \pi)\right]$$

$$= \mathbb{E}\left[\mathbb{E}\left[\sum_{t=3}^{\infty} \gamma^{t-1} r_t \,\middle|\, s_1, s_3 = s', \pi^{(3)}\right]\right]$$

s_3

$d_3^{\pi'}$

$$- \mathbb{E}\left[\quad\quad s_3 = s', \pi^{(2)}\right]$$

$Q^{\pi}(s', \pi)$.