Policy Iteration: $\pi_k \leftarrow \pi_{Q^{\pi_{k-1}}}$

PD Lemma: $\forall \pi, \pi'$,

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

Lemma: $\forall f \in \mathbb{R}^{S \times A}$

$$\|V^* - V^{\pi_f}\|_\infty \leq \frac{2 \cdot \|f - Q^*\|_\infty}{1-\gamma}.$$

Proof: $V^*(s) - V^{\pi_f}(s) = \frac{1}{1-\gamma} \mathbb{E}_{\tilde{s} \sim d^{\pi_f, s}} [V^*(\tilde{s}) - Q^*(\tilde{s}, \pi_f)]$

$$= \frac{1}{1-\gamma} \mathbb{E}_{d^{\pi_f, s}} [\underbrace{Q^*(\tilde{s}, \pi^*) - Q^*(\tilde{s}, \pi_f)}].$$

$$\leq 2 \cdot \|f - Q^*\|_\infty.$$

Thm: $\|Q^* - Q^{\pi_{k+1}}\|_\infty \le \gamma \cdot \|Q^* - Q^{\pi_k}\|_\infty$.

Proof: Claim: (1) $T^{\pi_{k+1}} Q^{\pi_k} \ge T^{\pi} Q^{\pi_k}$ $\forall \pi$.

(2) $T^{\pi_{k+1}} Q^{\pi_k} \le Q^{\pi_{k+1}}$ ✓

For (1): $(T^{\pi} Q^{\pi_k})(s,a) = R(s,a) + \gamma \mathbb{E}_{s' \sim P(\cdot|s,a)}[Q^{\pi_k}(s', \pi(s'))]$.

For (2): $(T^{\pi_{k+1}} Q^{\pi_k})_{(s,a)} = \mathbb{E}\left[\sum_{t=1}^{\infty} \gamma^{t-1} r_t \,\middle|\, s_1 = s,\ a_1 = a,\ a_2 \sim \pi_{k+1},\ a_{3:\infty} \sim \pi_k\right]$.

$Q^{\pi_{k+1}}(s,a) = \mathbb{E}\left[\sum_{t=1}^{\infty} \gamma^{t-1} r_t \,\middle|\, s_1 = s,\ a_1 = a,\ a_{2:\infty} \sim \pi_{k+1}\right]$.

$$Q^* - Q^{\pi_{k+1}} = Q^* - T^{\pi_{k+1}} Q^{\pi_k} + T^{\pi_{k+1}} Q^{\pi_k} - Q^{\pi_{k+1}}$$

$$\le T^{\pi^*} Q^* - T^{\pi_{k+1}} Q^{\pi_k} \qquad (T^{\pi_{k+1}} Q^{\pi_k} - Q^{\pi_{k+1}} \le 0)$$

$$\le T^{\pi^*} Q^* - T^{\pi^*} Q^{\pi_k}$$

$$\le \gamma \cdot \|Q^* - Q^{\pi_k}\|_\infty \quad \checkmark$$

$Q^* - T^{\pi_{k+1}} Q^{\pi_k}$

$= TQ^* - TQ^{\pi_k}$.

$T^{\pi}(Tf) = Tf$. *

$T^{\pi} f \neq Tf$?

$$T^{\pi} f = R(s,a) + \gamma \mathbb{E}_{\cdots}[\, f(s', \pi_f) \,].$$

$$\pi_f = \max_{a'} f(s', a')$$

$$= Tf.$$

Linear Programming for MDPs.

Primal Form:

$$\min_{V \in \mathbb{R}^S} \; d_0^T V$$
$$\text{s.t.} \quad V \geq TV.$$

Claim: if $d_0(s) > 0 \;\; \forall s.$

then optimal sol. $V = V^*$.

Q1: why min? Q2: why linear?

Q1: $V \geq TV. \Rightarrow V \geq V^*.$ ✓

Lemma: $\forall f \geq f' \in \mathbb{R}^S,\; Tf \geq Tf'.$

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

$V \geq TV \Rightarrow TV \geq T(TV)$.

$V \geq T^2V \geq T^3V \geq \cdots \geq T^\infty V = V^*$.

$V \geq TV$: $\forall s$.

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

$\forall a$.

$V(s) \geq R(s,a) + \gamma \mathbb{E}_{\cdots} [V(s')]$.

Dual form: $\max_{d \in \mathbb{R}^{S \times A},\ d \geq 0} d^T R$.

s.t. $\forall s'$.

$\sum_{a'} d(s',a') = (1-\gamma)\alpha_0 + \gamma \sum_{s,a} P(s'|s,a) d(s,a)$.

distribution over S.

$s' \Leftrightarrow s,a \sim d$.

$s \sim P(\cdot|s,a)$

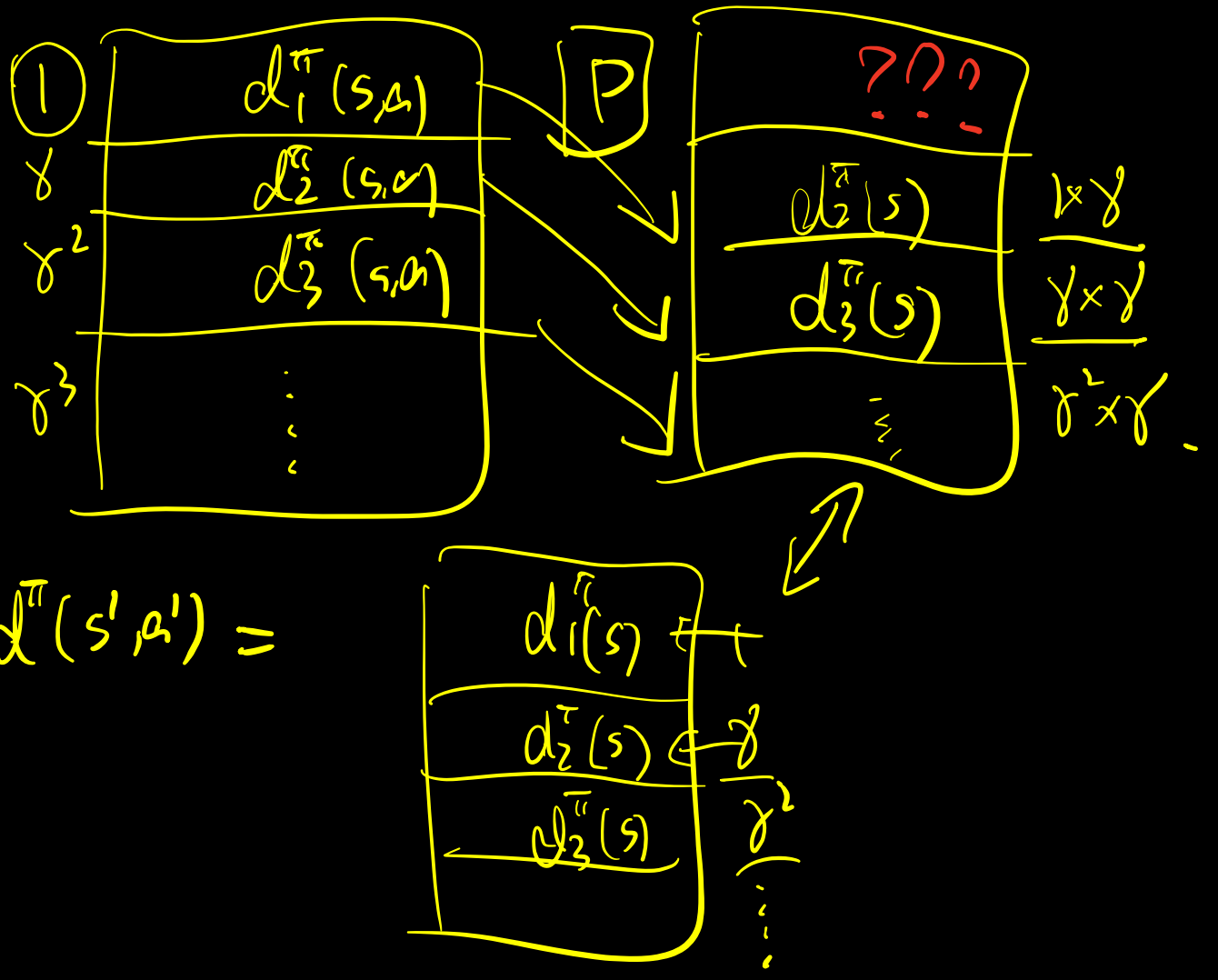
Claim: feasible space.

$= \{ d^\pi(s,a) : \forall \pi \}$.

$d^{\pi,s}$. d^{π}.

$$d^{\pi}(s,a) = (1-\gamma)\sum_{t=1}^{\infty} \gamma^{t-1} d_t^{\pi}$$

→ distribution of (s_t, a_t).



$$\sum_{a'} d^{\pi}(s',a') =$$

$$Q^{\pi} = T^{\pi} Q^{\pi}$$

$$Q^{*} = T^{*} Q^{*}$$

$$Q^{*} = Q^{\pi^{*}}.$$

$u \in \mathbb{R}^n, \quad v \in \mathbb{R}^n.$

$\langle u, v \rangle$

$$|u^T v| \leq \|u\|_1 \, \|v\|_\infty.$$

$\mathbb{E}_{a \sim \pi}[Q(s,a)]$

$Q(s,\pi) - Q(s,\pi').$

$= \langle Q(s,\cdot), \pi(\cdot|s) \rangle$

$- \langle Q(s,\cdot), \pi'(\cdot|s) \rangle.$