Linear Programming (LP) for MDPs

$\min c^T x$
s.t. $A_1 x = b_1$
$A_2 x \le b_2$.

Δ

Primal form: $\min_{V \in \mathbb{R}^S} \; d_0^T V$.
s.t. $V \ge \mathcal{T} V$.

distribution over S.
s.t. $d_0(s) > 0$.

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

$V = V^*$?

$$V \ge \mathcal{T} V \implies V \ge V^*.$$

Monotonicity of $\mathcal{T}$: $\forall V_1, V_2 \in \mathbb{R}^S$. s.t. $V_1 \ge V_2$.
$\mathcal{T} V_1 \ge \mathcal{T} V_2$.

$$\overset{V_1}{V} \ge \overset{V_2}{\mathcal{T} V}.$$

$$\mathcal{T} V \ge \mathcal{T}(\mathcal{T} V) \ge \mathcal{T}^3 V \ge \cdots \ge \mathcal{T}^{(\infty)} V = V^*.$$

$$\forall s. \quad V(s) \ge \max_a \big(\underbrace{R(s,a) + \cdots}_{\text{Linear.}} \big)$$

$$\Updownarrow$$

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

$$V^\pi(s) = \langle d^{\pi,s}, R^\pi \rangle$$

$[R(s,a)]_{(s,a) \in S \times A}$

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

s.t. $\left[\sum_a d(s,a)\right]_{s\in S} = \gamma P^T d + (1-\gamma) d_0$

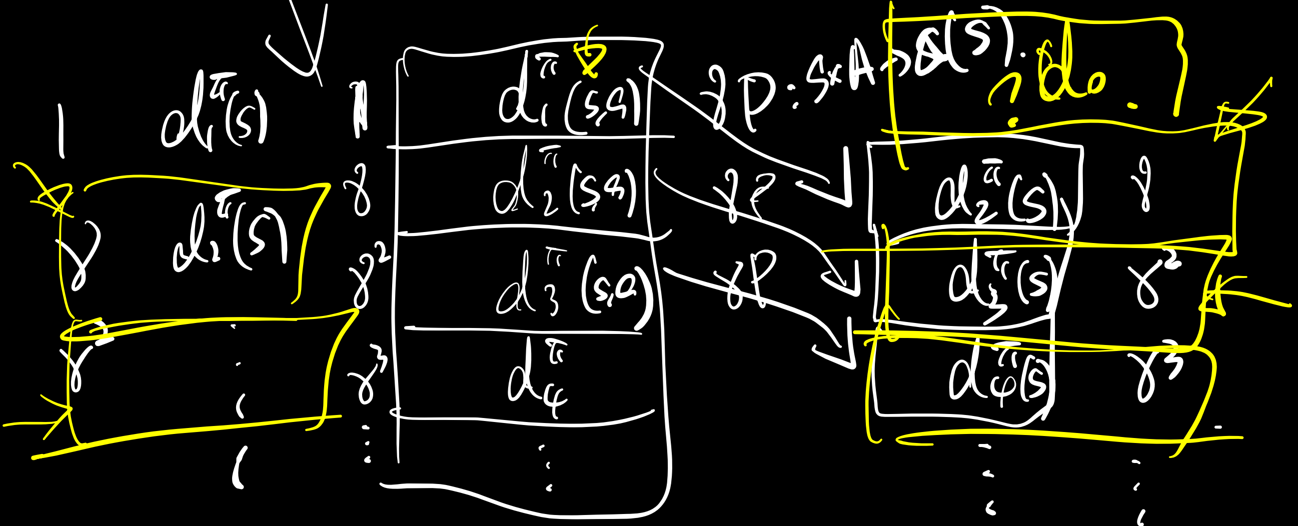
d_0 as. init. dist.

$\{ d^\pi(s,a) : \forall \pi \}$. d^{π, s_0}

$$\mathbb{E}_{s\sim d_0}[V^\pi(s)] = \frac{1}{1-\gamma} \mathbb{E}_{(s,a)\sim d^\pi}[R(s,a)].$$

$$d^\pi = (1-\gamma)\sum_{t=1}^{\infty} \gamma^{t-1} d_t^\pi$$ → marginal of (s_t, a_t) under π, d_0.

$P^T d_t^\pi$ is the $(t+1)$-th step distribution of state.



feasible d → $\pi(a|s) = \dfrac{d(s,a)}{\sum_a d(s,a)}$.