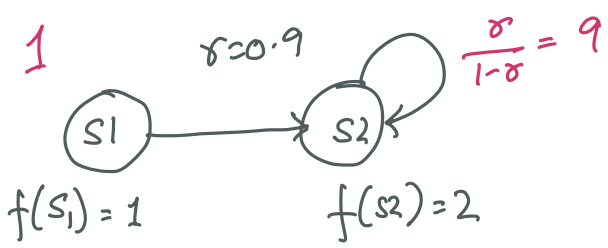


Init $\omega = 1$

$V(s) = \omega f(s)$

$V^{new}(S_1) = 0 + \gamma \cdot V(S_2) = 2 = 1.8$

$V^{new}(S_2) = 0 + \gamma \cdot V(S_2) = 2 = 1.8$

Init $\omega = 1$.

$V(S_1) = \omega \cdot 1$

$V(S_2) = \omega \cdot 2$

LEAST SQUARES

$$\text{ERROR} = 1 \cdot \underbrace{(\omega \cdot 1}_{V(S_1)} - 1.8)^2 + 9 \underbrace{(\omega \cdot 2}_{V(S_2)} - 1.8)^2$$

$$2(\omega - 1.8) + 2 \times 9 (2\omega - 1.8) = 0$$

$$\omega - 1.8 + 18(2\omega - 1.8) = 0$$

$$\omega = \frac{1.8(1+18)}{37} \approx 0.92$$

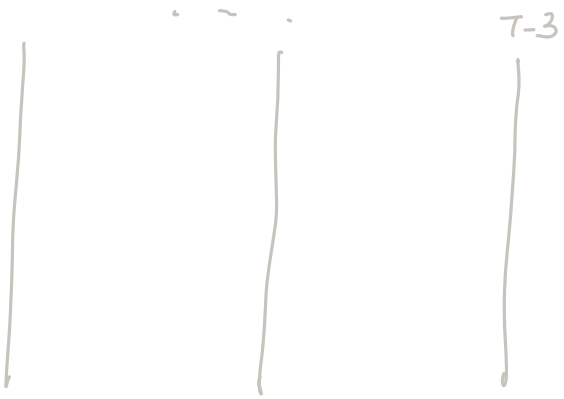
$$2(\omega - 1.8) + 2 \times 2 (2\omega - 1.8) = 0$$

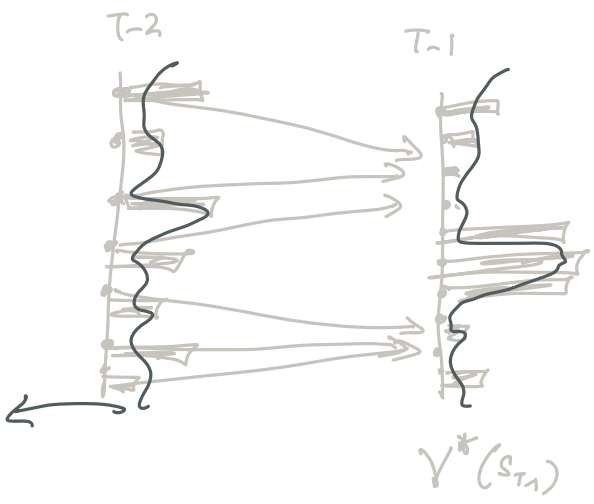
$$\omega - 1.8 + 4\omega - 3.6 = 0$$

$$\omega = \frac{1.8 + 3.6}{5} = 1.08$$

DYNAMIC PROGRAMMING

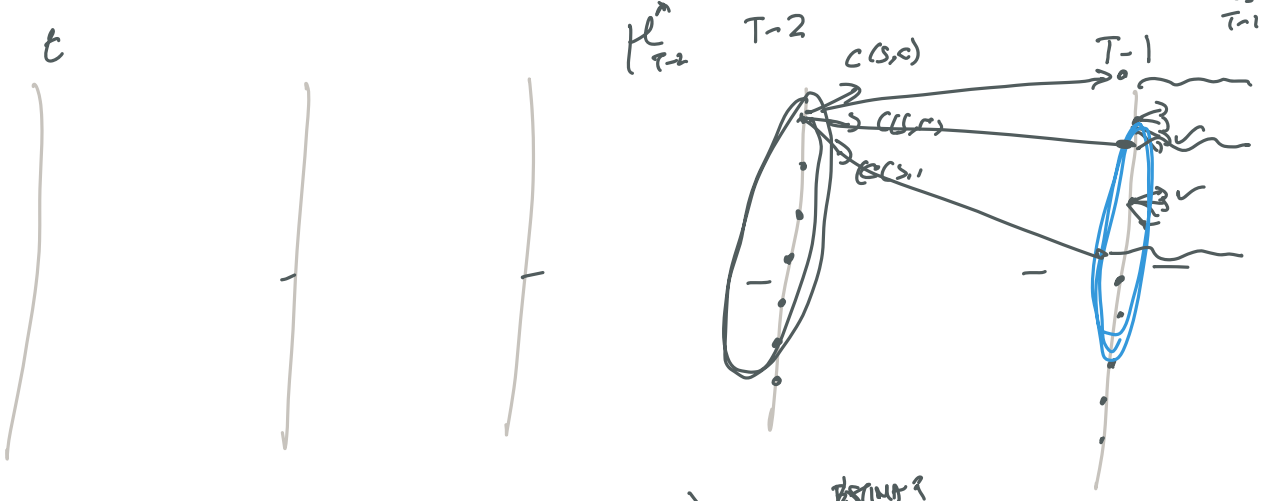
FITTED DYNAMIC PROGRAMMING





Policy Search via Dynamic Programming

First $\pi_{\theta, T-1} \approx \arg\min_a C(s,a)$



$$\pi^*_{T-2} = \arg\min_{a_{t-2}} C(s_{t-2}, a_{t-2}) + \cancel{V^*(s_{t-1})}$$

ESTIMATE?

ROLLOUT π^*_{T-1}

$$\pi^*_{T-1}(s) = \arg\min_a C(s,a)$$

RESTART DISTRIBUTION $\mu^*_{T-1}(s)$