March 9

Policy Search → Policy Gradients

What to compress? → MDP

↓ Value functions!

↓ Policies

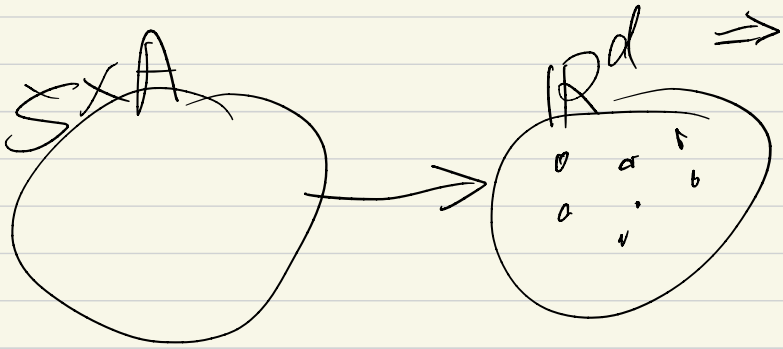
$M = (S, A, P, r)$ $\quad 0 \le \gamma < 1$

$|S|$ too big!

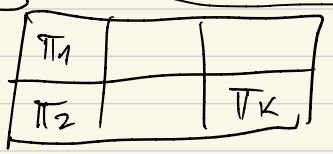
$\Pi \subsetneq ML$ ← set of memoryless policies of M

$\varphi : S \times A \to \mathbb{R}^d$

$$\Pi_\varphi = \{ \pi \in ML \mid \forall (s,a), (s',a') \in S \times A \; \varphi(s,a) = \varphi(s',a') \Rightarrow \pi(a|s) = \pi(a'|s') \}$$



$$\Pi_{\text{Boltzmann-Lin}} = \{ \pi \in ML \mid \exists w \in \mathbb{R}^d \quad \pi(a|s) \propto \exp(w^T \varphi(s,a)) \}$$

$\varphi(s,a)$ 

$\mu \in M_1(S)$ μ = row-vector

$$J(\pi) = \mu v^\pi = \sum_{s \in S} \mu(s) v^\pi(s)$$

$\operatorname{argmax}_{\pi \in \Pi} J(\pi)$	Boltzmann-Linear
"policy search"	State-aggregation

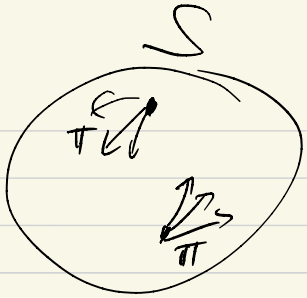
No! Vlassis Littman Barber (POMDPs)

How come?

BPS = Blind-Policy-Search

$$\varphi(s,a)_{a'} = \mathbb{I}(a = a')$$

$$\Pi \equiv \underbrace{\mathcal{M}_1([A])}_{\pi \in \mathcal{M}_1([A])}$$



$$S = A = [n]$$

$$P_a(s,a) = 1 \quad / \quad P_a(s,s') = 0, \; s' \neq a$$

$$r_a(s) = ? \qquad r_\pi(s) = \sum_a \pi(a)\, r_a(s)$$

$$\mu = \frac{1}{n} 1^T$$

$$J(\pi) = \mu v^\pi = \mu \underbrace{\left(\sum_{t=0}^{\infty} \gamma^t P_\pi^t \right)} r_\pi$$

$$P_\pi(s,s') = \sum_{a \in A} \underbrace{\pi(a|s)}_{\pi(a)} \underbrace{P_a(s,s')}_{\mathbb{I}(s'=a)} = \pi(s')$$

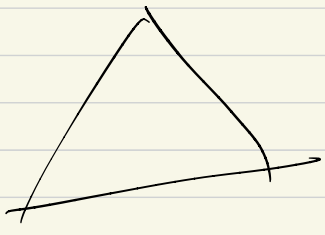
$$P_\pi = 1 \pi^T, \quad t \geq 1$$

$$P_\pi^2 = 1 \underbrace{\pi^T 1}_{1} \pi^T = 1 \pi^T$$

$$P_\pi^t = 1 \pi^T$$

$$R = (r_a(s))_{sa}$$

$$J(\pi) = \mu \left(I + \frac{\gamma}{1-\gamma} 1 \pi^T \right) R \pi$$



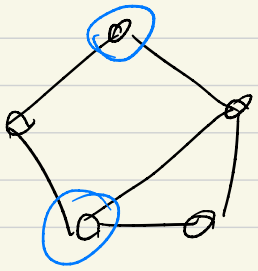
$$= \mu R\pi + \frac{\gamma}{1-\gamma} \cdot \pi^T R \pi \longrightarrow \max_{\pi \in M_1(A)}$$

Graph Theory

MAX-INDSET

$\alpha(G)$

$G = (V, E)$



indep set = neighbor-free set.

NP-hard

$\alpha(G) = \max\{|V'| \mid V' \subseteq V \text{ indep. in } G\}$

cubic (3-regular)

Theorem (Motzkin-Straus '65):

$|V| = n$

$$\frac{1}{\alpha(G)} = \min_{y \in M_1(n)} y^T (G + I) y$$

$\{0,1\}^{n \times n} \ni G$ vertex adj. mx. of $G = (V, E)$

$$G_{ij} = 1 \iff (i,j) \in E, \quad i \neq j$$

Pick G 3-regular.

$$R := -(I + G)$$

$$J(\pi) = -\mu(I+G)\pi - \frac{\gamma}{1-\gamma} \pi^T (I+G)\pi$$

$$= \frac{-1}{n} \mathbb{1}^T (I + G)\pi - \frac{\gamma}{1-\gamma} \pi^T (I + G)\pi$$

$$= -\frac{4}{n} - \frac{\gamma}{1-\gamma} \dots$$

$$(\mathbb{1}^T G)_j = \sum_i G_{ij} = 3 \quad , \quad \mathbb{1}^T G = 3\,\mathbb{1}^T$$

$$\max_{\pi \in \Pi} J(\pi) = -\frac{4}{n} - \frac{\gamma}{1-\gamma} \frac{1}{\alpha(G)} \geq -\frac{4}{n} - \frac{\gamma}{1-\gamma} \frac{1}{m}$$

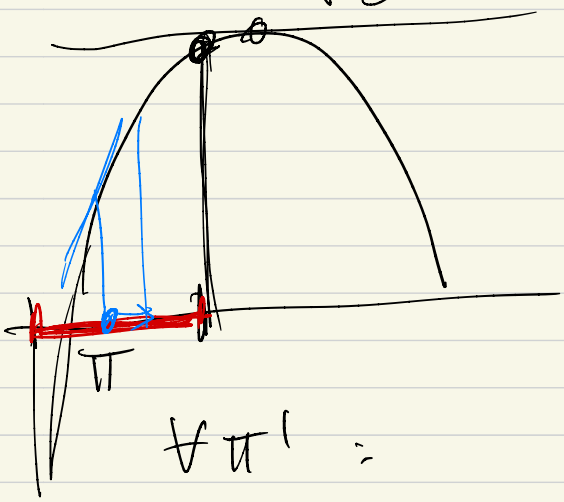
$$\Leftrightarrow \quad \alpha(G) \geq m$$

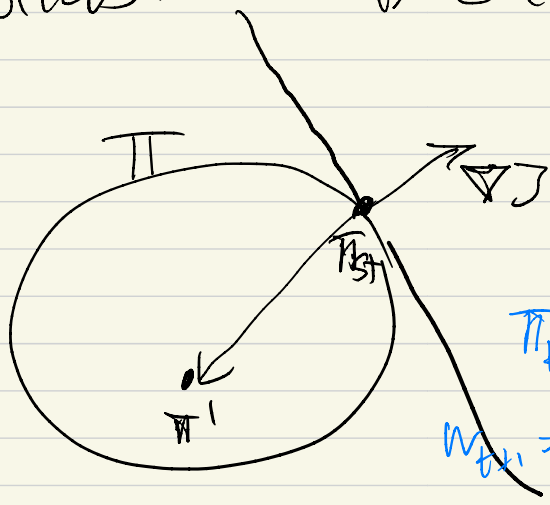
Conclusion: Policy search can be hard!

Relax the problem!

Stationary points! $\nabla J(\pi) = 0$

$\nabla J = 0$





$$\pi_{t+1} = \pi_t + \eta \nabla J(\pi_t)$$

$$w_{t+1} = w_t + \eta \underbrace{\nabla_\pi J(\pi_{w_t}) \nabla_w \pi_{w_t}}_{\nabla_w J(\pi_{w_t})}$$

$$\forall \pi' : \quad \langle \nabla J(\pi_{st}), \pi' - \pi_{st} \rangle \leq 0$$

1. Computation (nice & benign)

2. Approx. guarantees

π_{st} stat. point

$\Rightarrow \mu \cdot v^{\pi_{st}} \geq \mu v^{\pi^*} - ?$

$\varepsilon(M, \Pi)$

$\varepsilon(M, \varphi)$

gradient ascent

1.1 $\nabla_w J(\pi_w) = ?$ simulator?

1.2 Which algo? → Agnostic to Π

Nope!

?

change vanilla PG method

Natural-PG

$\nabla_w J(\pi_w) = \nabla_\pi J(\pi_w) \nabla_w \pi_w$