Last Time

POMDP Policies
- History or Belief-Based

Approximate
- Particle Filter

This Time

POMDP policy structure
- Alpha Vectors
- POMDP Value Iteration

$$\pi(b) = \begin{cases} \text{OL} & \text{if } b(TR) > 0.85 \\ \text{OR} & \text{if } b(TL) > 0.85 \\ \text{L} & \text{otherwise} \end{cases}$$

(DMU)

Crying Baby Problem

$S = \{h, \neg h\}$

$A = \{f, \neg f\}$

$O = \{c, \neg c\}$

$R(s,a) = -5\,\mathbb{1}(a=f) - 10\,\mathbb{1}(s=h)$

$T(h \mid \neg h, \neg f) = 0.1$

$T(\neg h \mid \cdot, f) = 1.0$

$Z(c \mid \cdot, \neg h) = 0.1$

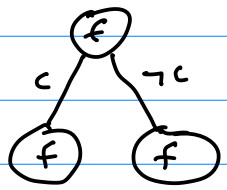
$Z(c \mid \cdot, h) = 0.8$

Conditional Plan: Fixed-Depth History Based

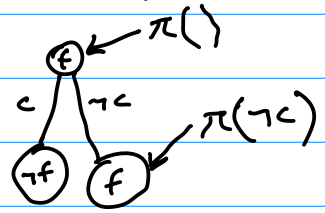
Fixed Initial Action

1-step: (f) (¬f)

2 step:



Policy



$2^3 = 8$

# of possible conditional plans $|A|^{(|O|^h - 1)/(|O|-1)}$

Value of a conditional plan π at state s

$V^\pi(s)$

1 step

$V^\pi(s) = R(s, \pi(1))$

$\pi =$ (h) $V^\pi(h) = -15$ $V^\pi(\neg h) = -5$

$\pi =$ ($\neg$h) $V^\pi(h) = -10$ $V^\pi(\neg h) = 0$

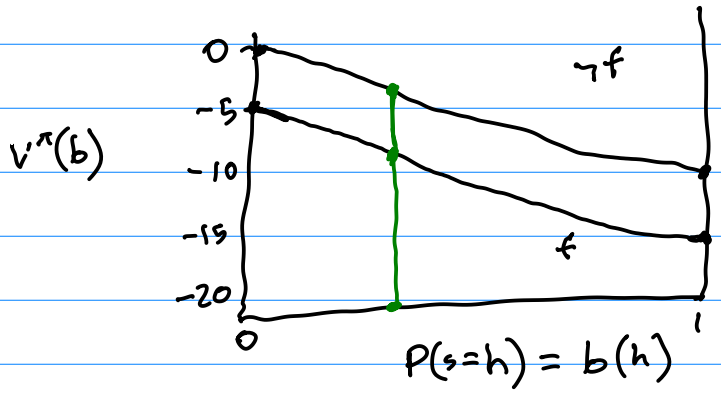
belief b

$$V^\pi(b) = \sum_s V^\pi(s)\, b(s) = \underset{s\sim b}{E}\left[V^\pi(s)\right]$$

Alpha Vector

$\alpha_\pi[s] = V^\pi(s)$

$V^\pi(b) = \alpha_\pi^T b$



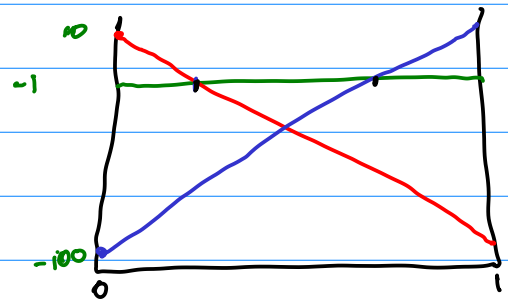
b(h) = 0.35

Tiger POMDP

$S = \{TL, TR\}$

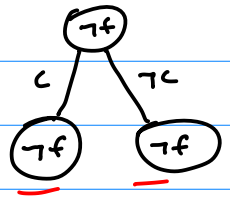
$A = \{OL, OR, L\}$

$R = \begin{cases} +10 & \text{if open non-tiger door} \\ -100 & \text{if open tiger door} \\ -1 & \text{if listen} \end{cases}$

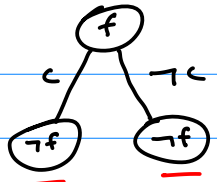


(OL) (OR) (L)

$\pi_1 =$



$\pi_2 =$



$$V^{\pi}(s) = R(s, \pi(1)) + \gamma \left[\sum_{s'} T(s' \mid s, \pi(1)) \sum_{o} Z(o \mid \pi(1), s') V^{\pi(o)}(s') \right]$$

$R(s', \neg f)$

$$V^{\pi}(s) = R(s, \pi(1)) + \gamma \sum_{s'} T(s' \mid s, \pi(1)) R(s', \neg f)$$

$V^{\pi_1}(h) = -10 + 0.9(1.0 \cdot -10) = -19$

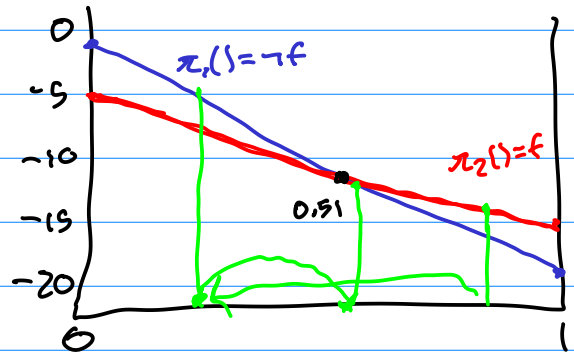
$V^{\pi_1}(\neg h) = 0 + 0.9(0.9 \cdot 0 + 0.1 \cdot -10) = -0.9$

$V^{\pi_2}(h) = -15 + 0.9(1.0 \cdot 0) = -15$

$V^{\pi_2}(\neg h) = -5 + 0.9(1.0 \cdot 0) = -5$

$\alpha_{\pi_1} = [-0.9, -19]$

$\alpha_{\pi_2} = [-5, -15]$



$$\pi(b) = \left[\underset{\pi_i}{\operatorname{argmax}} \, \alpha_i^T b \right]()$$

In general

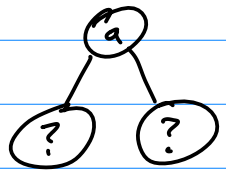
$$V^*(b) = \max_{\alpha \in \Gamma^*} \alpha^T b$$

↑ set of α vectors

Convex, Piecewise Linear Value Functions

POMDP Value Iteration

Starting with 1 step plans
Creating all Possible 2-step Plans
Remove Suboptimal Plans
Create 3 step plans
⋮



loop

$$\Gamma^1 = \bigcup_{a \in A} \Gamma^a$$

where $\Gamma^a = \bigoplus_{o \in O} \Gamma^{a,o}$

where $\Gamma^{a,o} = \left\{ \frac{1}{|O|} r_a + \gamma \alpha^{a,o} : \alpha \in \Gamma \right\}$

where $\alpha^{a,o}[s] = \sum_{s'} Z(o|a,s') T(s'|s,a) \alpha[s']$

MDP Value Iteration

loop

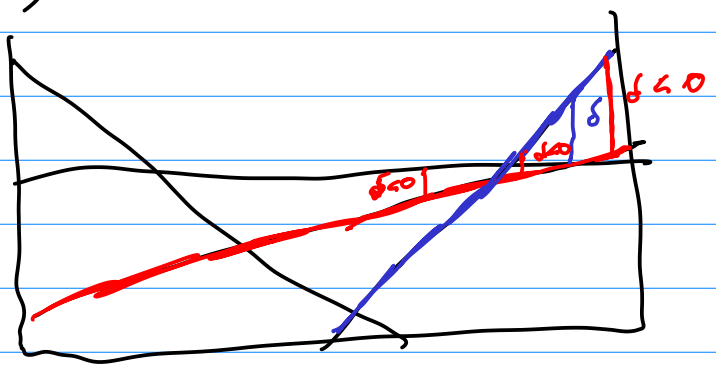
$V' \leftarrow B[V]$

$$B[V](s) = \max_a R(s,a) + \gamma E_{s' \sim T(s,a)}[V(s')]$$

$$\Gamma_1 \oplus \Gamma_2 = \{ \alpha_1 + \alpha_2 : \alpha_1 \in \Gamma_1, \alpha_2 \in \Gamma_2 \}$$

$r_a[s] = R(s,a)$

Pruning



$$\begin{aligned} \underset{b,\delta}{\text{maximize}} \quad & \delta \\ \text{subject to} \quad & b \geq 0 \\ & 1^T b = 1 \\ & \alpha^T b \geq \alpha'^T b + \delta \quad \forall \alpha' \in \Gamma \end{aligned}$$

if $\delta > 0$, then α is not dominated

if $\delta \leq 0$, α is dominated, we can prune

b that maximizes δ is called "witness belief"

- WITNESS
- Incremental Pruning