Exact POMDP Solutions: α-vectors

Recap

- POMDP
- Belief Updates

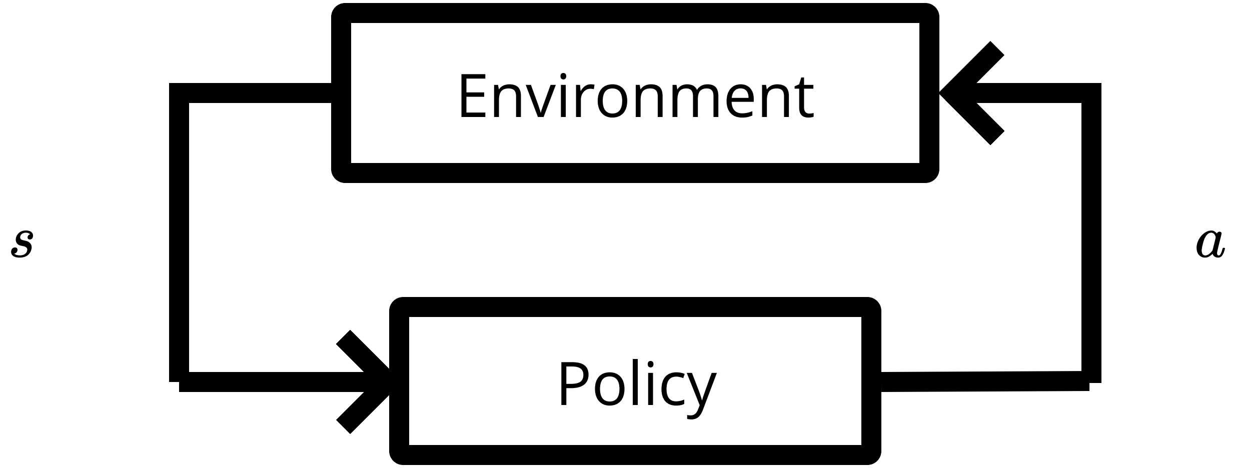
$$(S, A, O, R, T, Z, \gamma)$$

$$b_t(s) = P(s_t = s \mid h_t)$$

$$b' = \tau(b, a, o)$$

$$b'(s') \propto Z(o \mid a, s') \sum_s T(s' \mid s, a)\, b(s)$$

MDP Sense-Plan-Act Loop



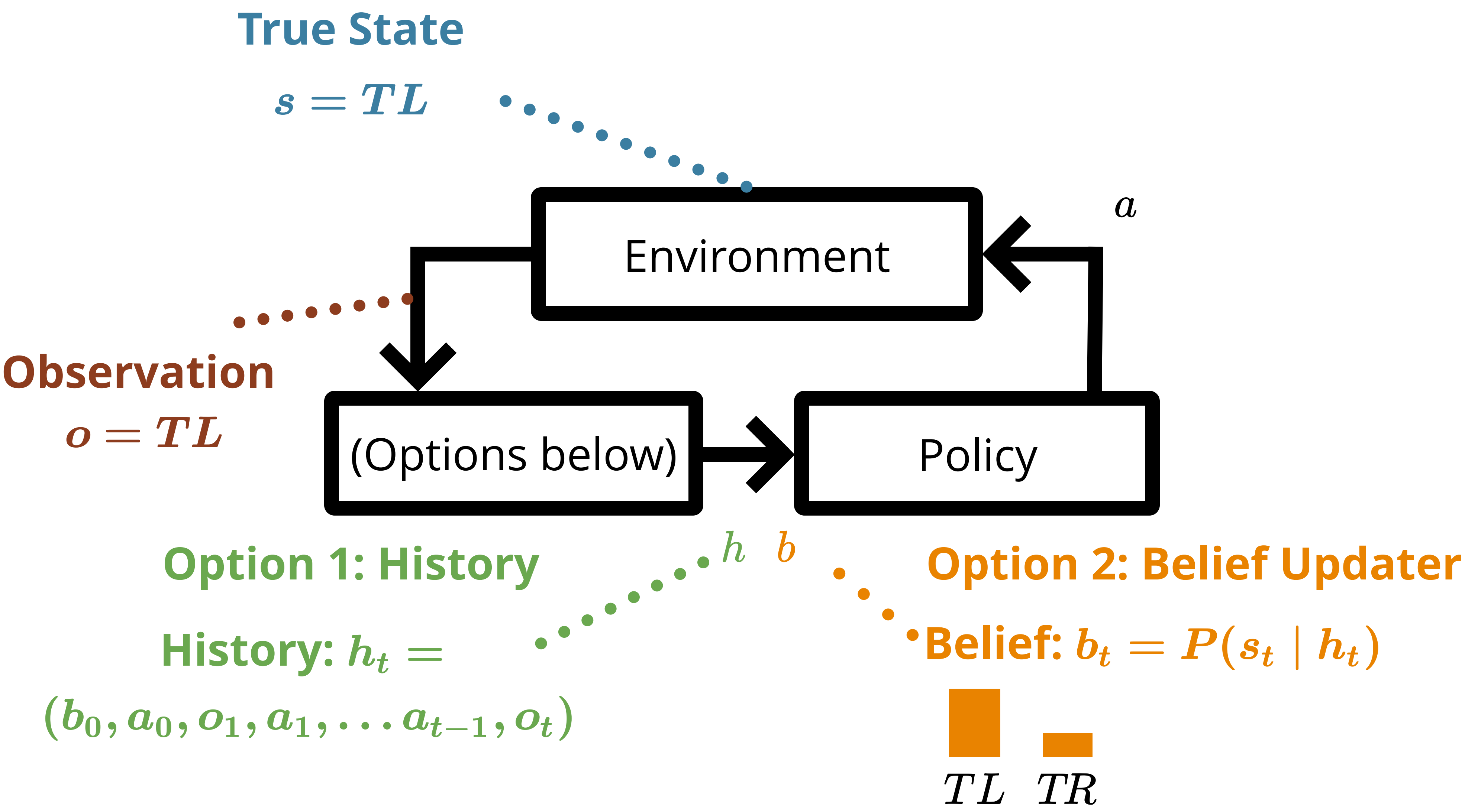
Environment

Policy

s

a

POMDP Sense-Plan-Act Loop



True State

$s = TL$

Environment

a

Observation

$o = TL$

(Options below)

Policy

h b

Option 1: History

History: $h_t = (b_0, a_0, o_1, a_1, \dots a_{t-1}, o_t)$

Option 2: Belief Updater

Belief: $b_t = P(s_t \mid h_t)$

Exercise 1: Crying Baby Belief Update

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

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

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

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

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

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

$R(s, a) = R(s) + R(a)$

$$R(s) = \begin{cases} -10 \text{ if } s = h \\ 0 \text{ otherwise} \end{cases}$$

$$R(a) = \begin{cases} -5 \text{ if } a = f \\ 0 \text{ otherwise} \end{cases}$$

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

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

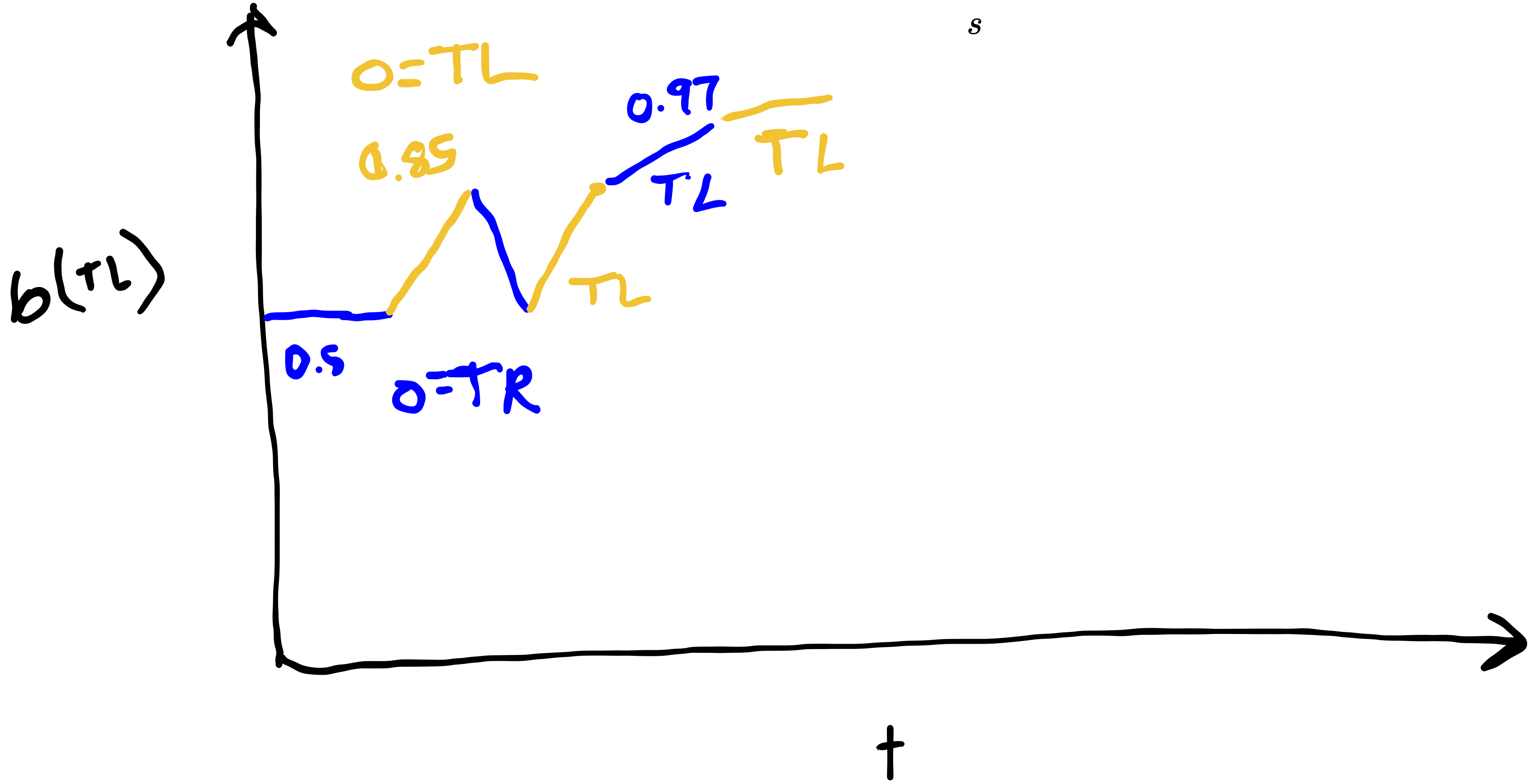
$\gamma = 0.9$

$$b'(s') \propto Z(o \mid a, s') \sum_s T(s' \mid s, a)\, b(s)$$

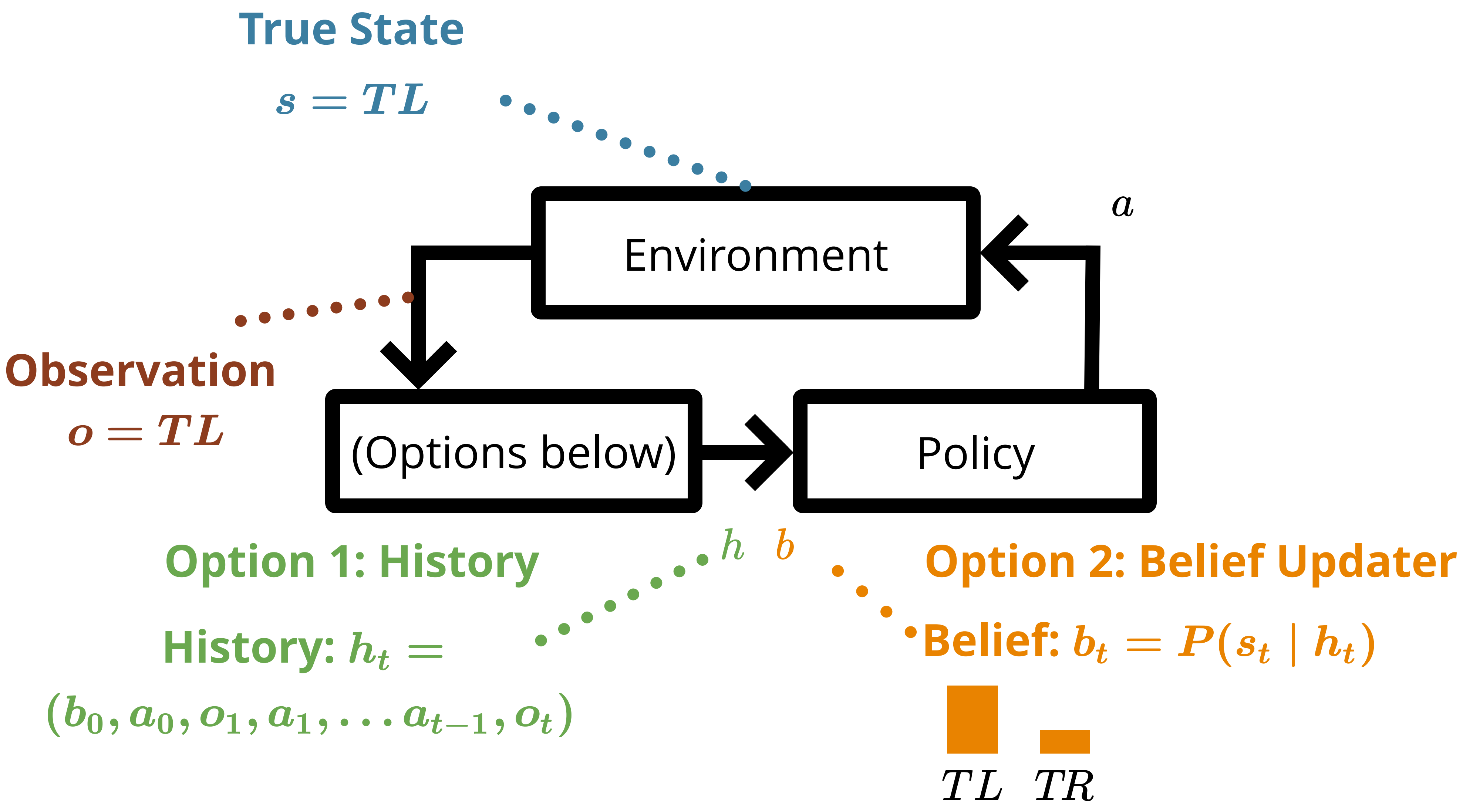
Starting at a $b(h) = 0$, calculate b' with $a = \neg f$ and $o = c$.

Belief Dynamics

$$b'(s') \propto Z(o \mid a, s') \sum_{s} T(s' \mid s, a)\, b(s)$$



POMDP Sense-Plan-Act Loop



True State

$s = TL$

Environment

a

Observation

$o = TL$

(Options below)

Policy

h b

Option 1: History

History: $h_t =$
$(b_0, a_0, o_1, a_1, \ldots a_{t-1}, o_t)$

Option 2: Belief Updater

Belief: $b_t = P(s_t \mid h_t)$

Guiding Quesiton

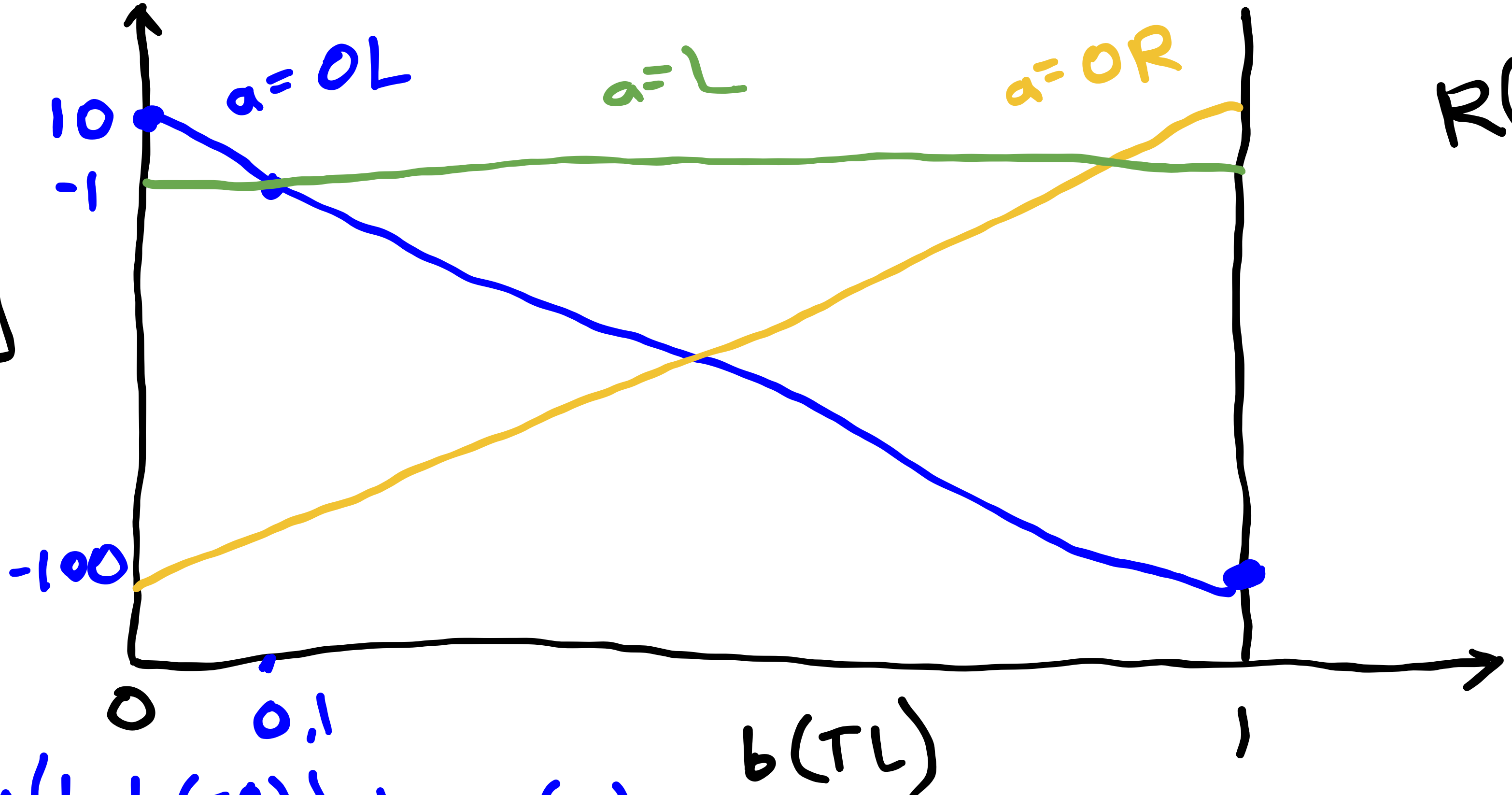
How do we calculate the optimal action in a POMDP?

One-step utility

One-step utility

Reward: +10 empty door
-1 Listen
-100 Tiger

$R(b,a)$
$= E_{s\sim b}[R(s,a)]$



$R(b,a) = \overline{r}_a \cdot b$

↑

α-vector

$R(b,a) = 10\cdot(1-b(TL)) - 100\, b(TL) = -110\, b + 10$

Exercise 2: Crying Baby 1-Step Utility

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

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

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

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

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

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

$$R(s, a) = R(s) + R(a)$$

$$R(s) = \begin{cases} -10 \text{ if } s = h \\ 0 \text{ otherwise} \end{cases}$$

$$R(a) = \begin{cases} -5 \text{ if } a = f \\ 0 \text{ otherwise} \end{cases}$$

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

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

$$\gamma = 0.9$$

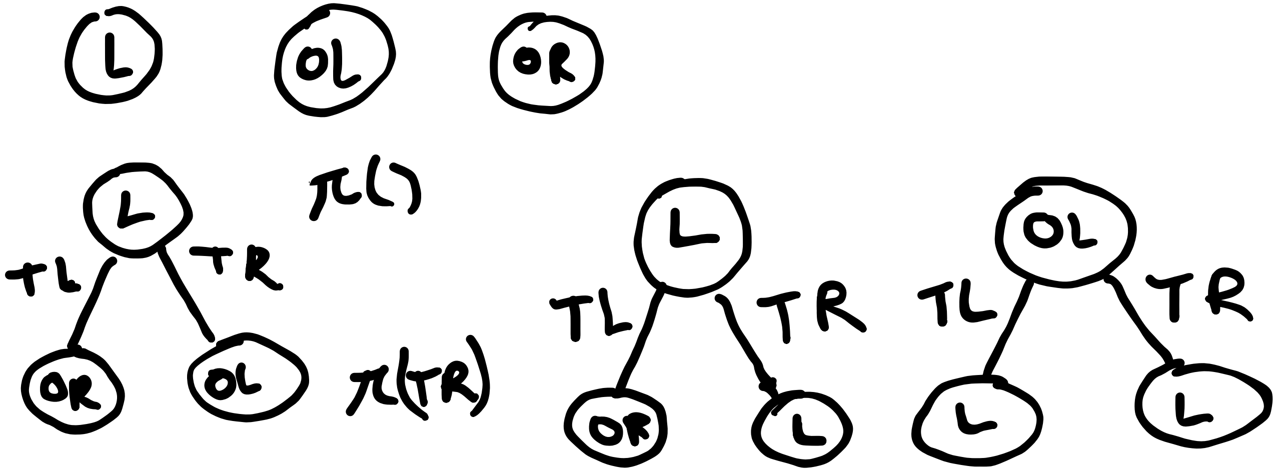
Draw the 1-step utility α-vectors for the Crying Baby problem.

Alpha Vectors for Conditional Plans

Conditional Plans: fixed-depth history-based policies

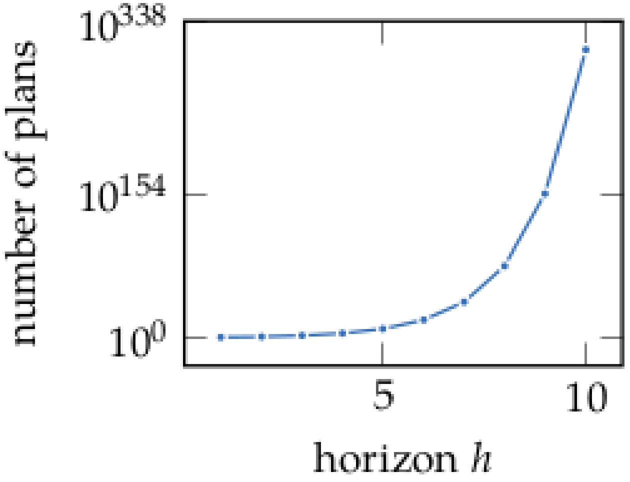
1 Step:

2 Step:



$|A|^{\frac{(|O|^{h}-1)}{(|O|-1)}}$

27 two step plans!



Alpha Vectors for Conditional Plans

For 1-step: $U^{\pi}(s) = R(s, \pi())$

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

POMDP Value Functions

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

Exercise: 2-Step Crying Baby α Vectors

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

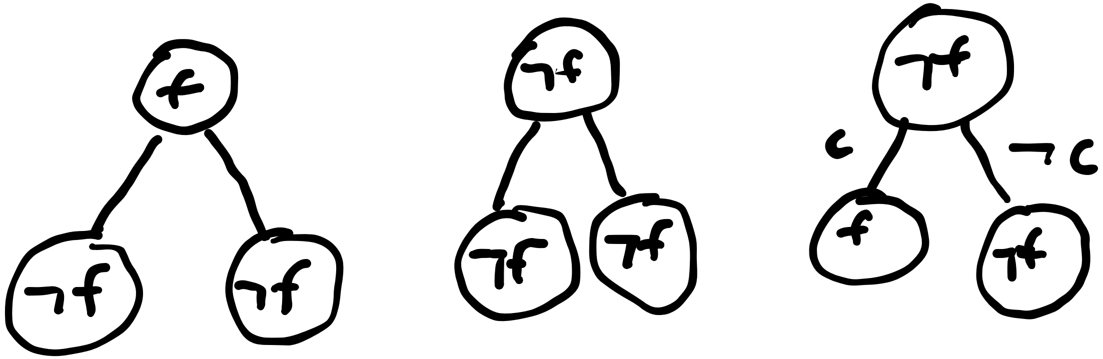
$A = \{f, \neg f\}$

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

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

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

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



$$U^\pi(s) = R(s, \pi()) + \gamma \left[\sum_{s'} T(s' \mid s, \pi()) \sum_o O(o \mid \pi(), s') U^{\pi(o)}(s') \right]$$

$R(s, a) = R(s) + R(a)$

$$R(s) = \begin{cases} -10 \text{ if } s = h \\ 0 \text{ otherwise} \end{cases}$$

$$R(a) = \begin{cases} -5 \text{ if } a = f \\ 0 \text{ otherwise} \end{cases}$$

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

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

$\gamma = 0.9$

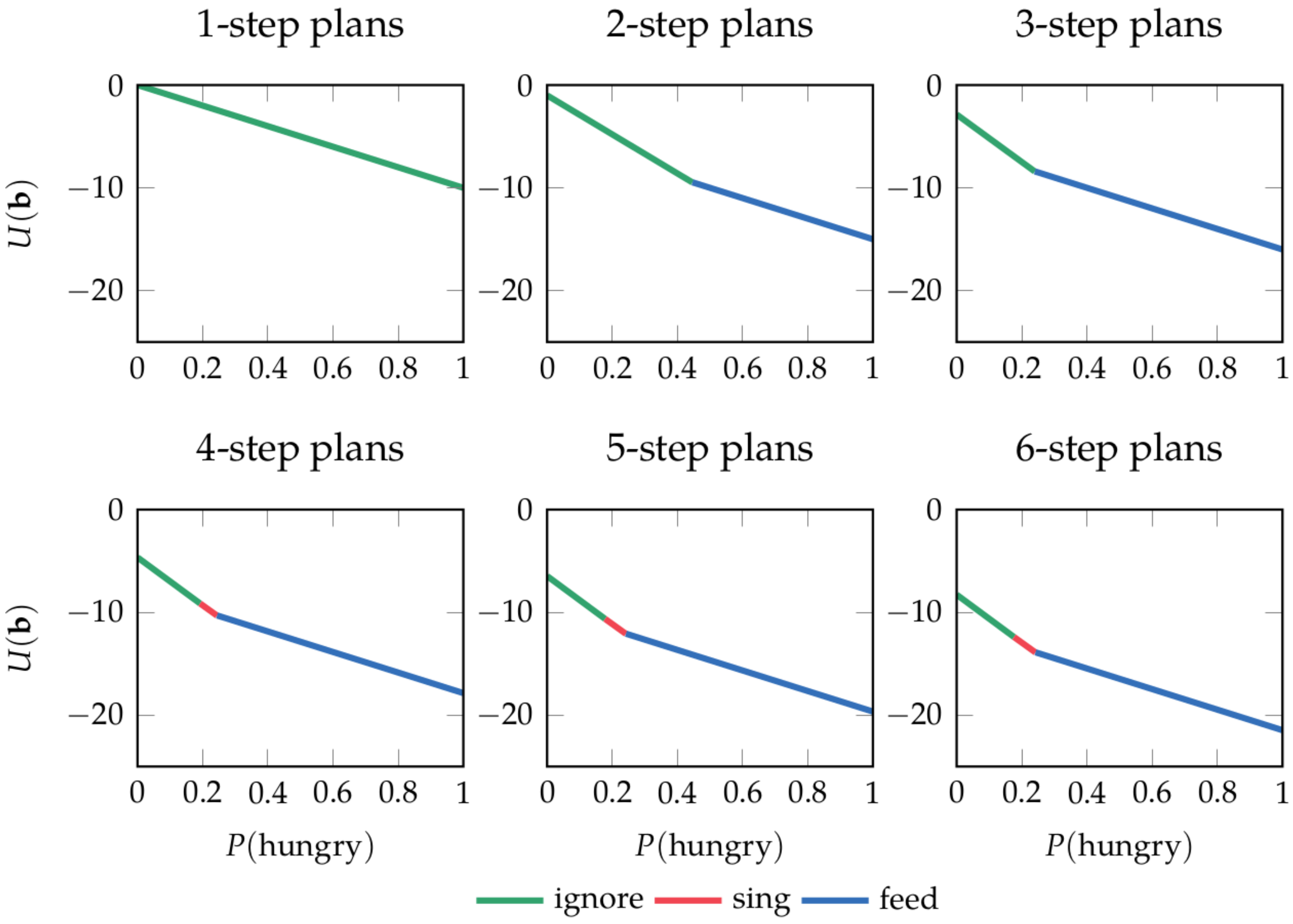
α-Vector Pruning

Alpha Vector Expansion

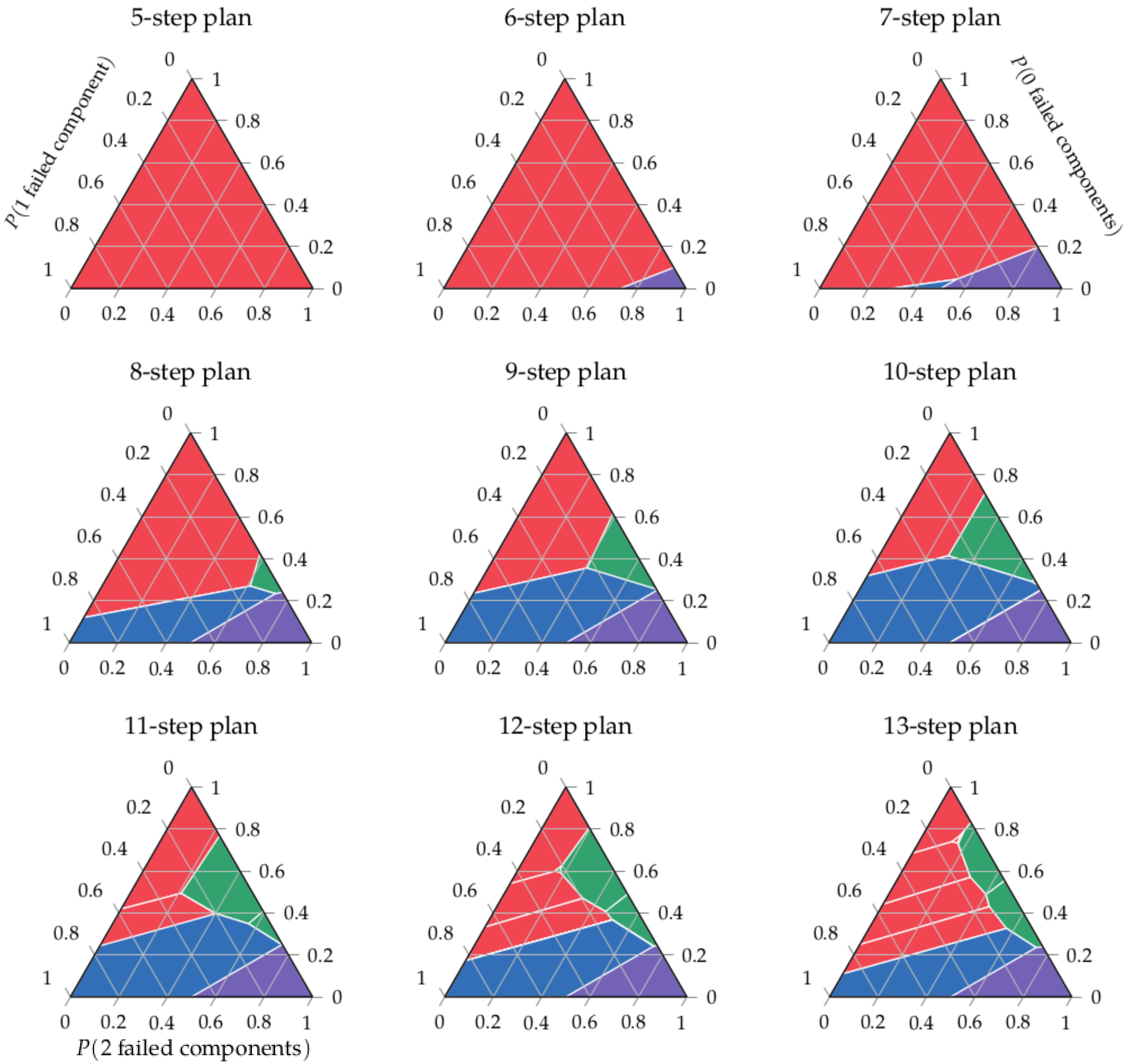
POMDP Value Iteration (horizon d)

```
$\Gamma^0 \leftarrow \emptyset$
for $n \in 1 \ldots d$
    Construct $\Gamma^n$ by expanding with $\Gamma^{n-1}$
    Prune $\Gamma^n$
```

Finite Horizon POMDP Value Iteration



Finite Horizon POMDP Value Iteration



🟥 manufacture
🟩 examine
🟦 interrupt
🟪 replace

Recap

- A POMDP is an MDP on the belief space
- The value function of a discrete POMDP can be represented by a set of α-vectors
- Each α vector corresponds to a conditional plan