CS 4649/7649
Robot Intelligence: Planning

Partially Observable MDP

Sungmoon Joo

School of Interactive Computing
College of Computing
Georgia Institute of Technology

Some slides adapted from Dr. Mike Stilman's lecture slides

Administrative

- Three lectures left
  - Nov. 25th : POMDP and Summary of Planning under Uncertainties
  - Dec. 2nd : Extension of Planning/Control: Language, Hybrid System
  - Dec. 4th : Wrap up

- Due Reminder:
  - Project report: Due Dec. 4th
  - Project report review: Due Dec. 11th
  - Project presentation & presentation evaluation: Dec. 11th

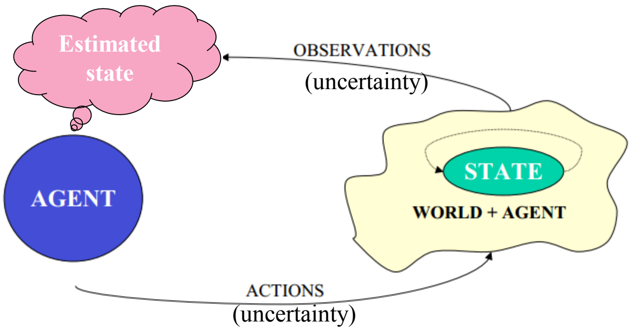
Reality

Two Sources of Error

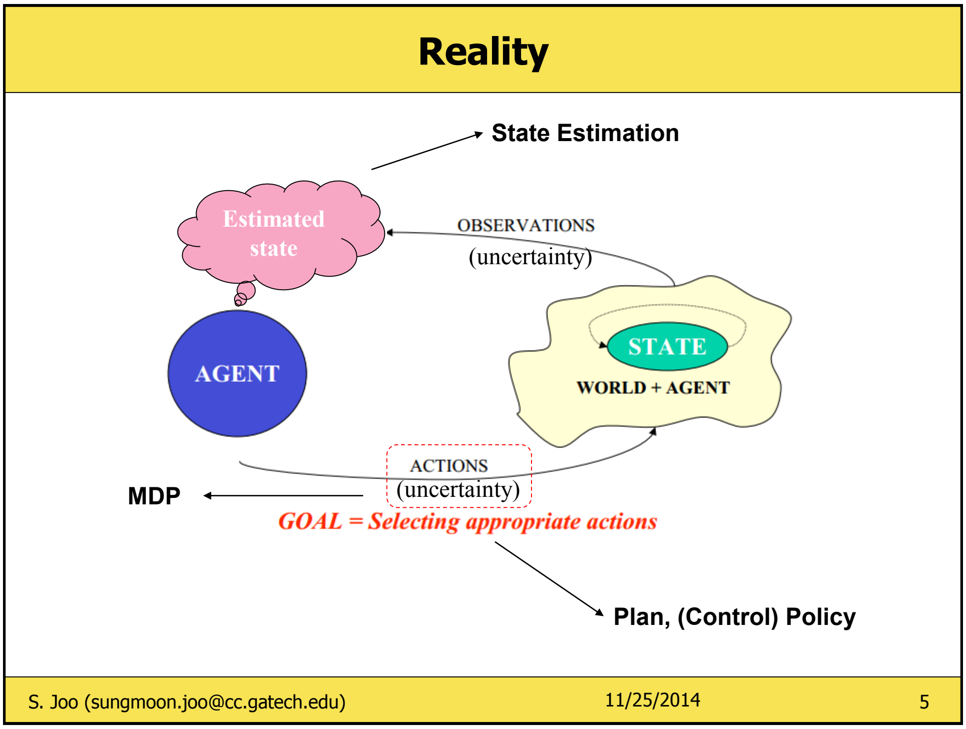
- **Sensing & State Estimation → Uncertainty**
  - Sensors have noise
  - You don't know exactly what the state is (e.g. mapping, localization,…)

- **Action Execution → Uncertainty**
  - Your actuators do not do what you tell them to
  - The system responds differently than you expect
    : Friction gears, air resistance, etc.

Reality



GOAL = Selecting appropriate actions



Reality

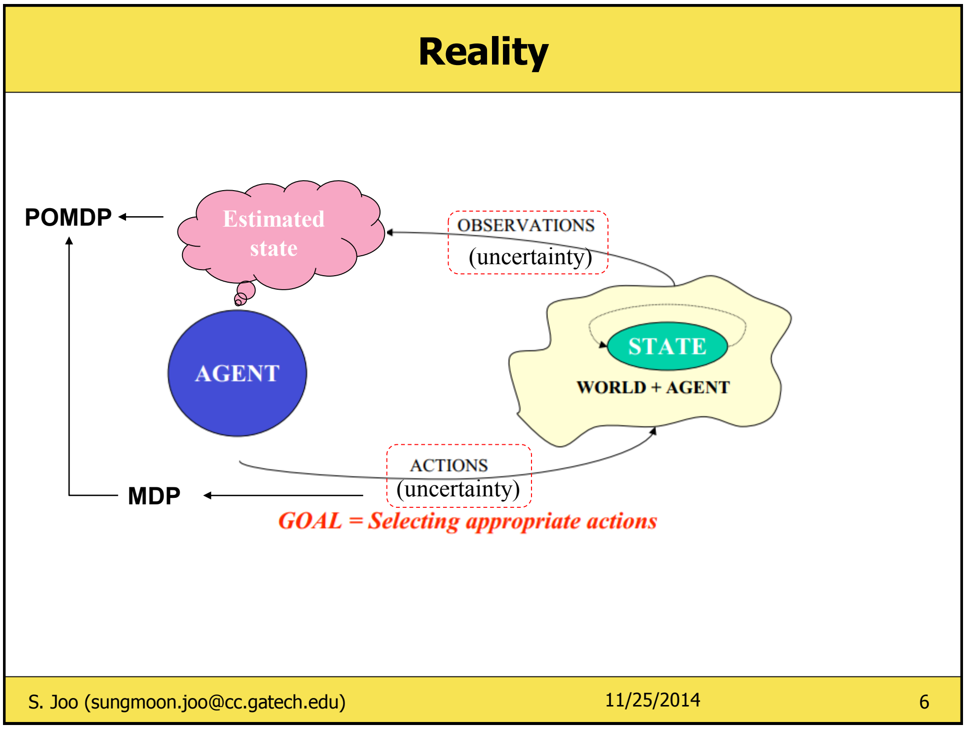
State Estimation

MDP

GOAL = Selecting appropriate actions

Plan, (Control) Policy

S. Joo (sungmoon.joo@cc.gatech.edu) 11/25/2014 5



Reality

POMDP

MDP

GOAL = Selecting appropriate actions

S. Joo (sungmoon.joo@cc.gatech.edu) 11/25/2014 6

Markov Decision Process (MDP)

- **States** $\Sigma = \{s_1, \ldots, s_n\}$ **Actions** $A = \{a_1, \ldots, a_m\}$
- **Rewards** $R = \{r(a_i, s_j)\}$
- **Transition Model**

$$\mathrm{P}(s'|a,s) : \mathrm{P}(\text{next } = s' |\text{ current } = s \text{ and action } = a)$$

Mathematical Frameworks

Action Uncertainty

Observation Uncertainty

Markov Chain **-Markov Property**	**Markov Decision Process** **(MDP)**
Hidden Markov Model (HMM)	Partially Observable MDP (POMDP)

Both

POMDP

MDP

- **States** $\Sigma = \{s_1, \ldots, s_n\}$ **Actions** $A = \{a_1, \ldots, a_m\}$
- **Rewards** $R = \{r(a_i, s_j)\}$
- **Transition Model**

Uncertainty about action outcome

$$\mathrm{P}(s'|a,s) : \mathrm{P}(\text{next } = s' |\text{ current } = s \text{ and action } = a)$$

- **Observations** $\Omega = \{o_1, \ldots, o_n\}$

Uncertainty about the state due to imperfect observation

- **Observation Function (Probability of observation "o" in state "s")**

$$\mathrm{P}(o|s)$$

Don't get to observe the state itself, instead get sensory measurements

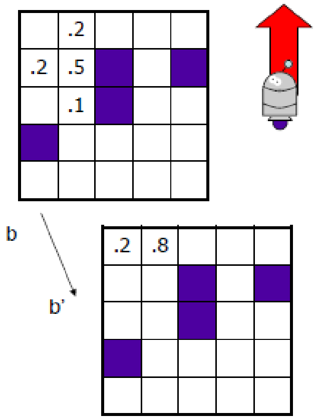
State Estimation – Belief State

- A belief state b is a probability distribution over MDP states
- Belief states have a transition model

$$b'(s) = P(s|o,a,b)$$

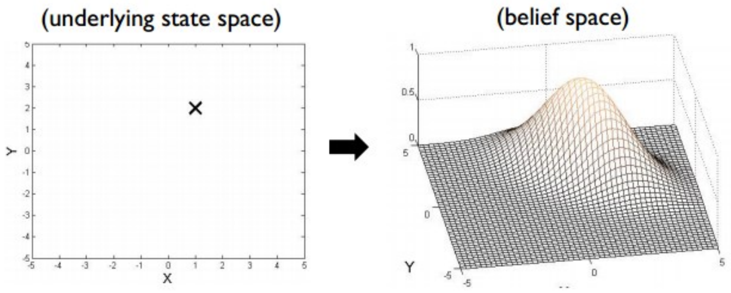
(ex. Kalman Filter)

- Planning in terms of belief states is a high-dimensional MDP where states are belief states
- POMDP is a continuous n-dimensional state space where n = # of MDP states

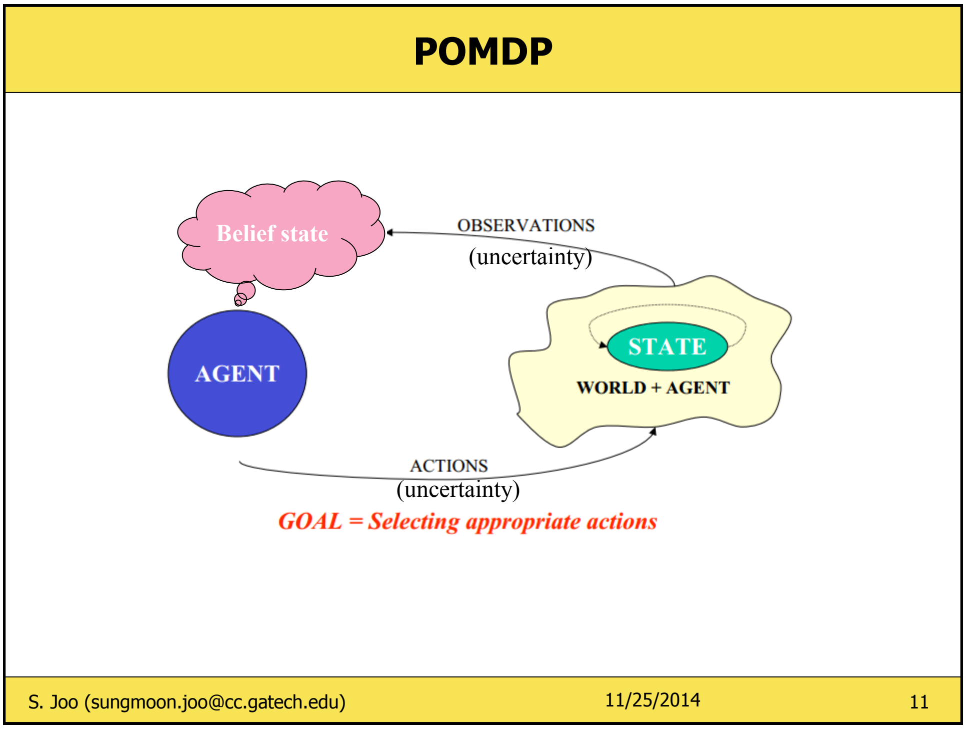


Belief States: Example

Kalman Filter: Gaussian(Mean & Covariance)

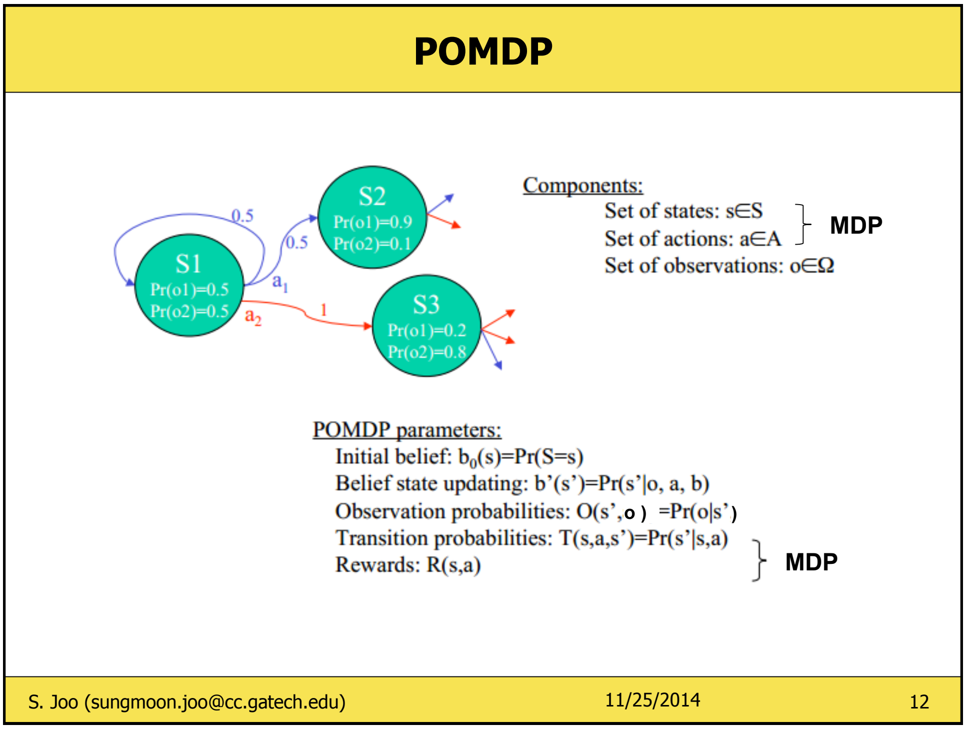


Continuous Belief States



POMDP

GOAL = Selecting appropriate actions



POMDP

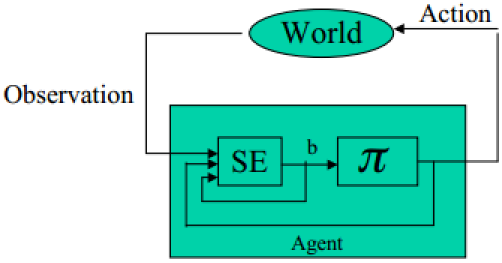
Components:

Set of states: $s \in S$ } **MDP**
Set of actions: $a \in A$
Set of observations: $o \in \Omega$

POMDP parameters:

Initial belief: $b_0(s)=\Pr(S=s)$
Belief state updating: $b'(s')=\Pr(s'|o, a, b)$
Observation probabilities: $O(s',o) = \Pr(o|s')$
Transition probabilities: $T(s,a,s')=\Pr(s'|s,a)$ } **MDP**
Rewards: $R(s,a)$

POMDP



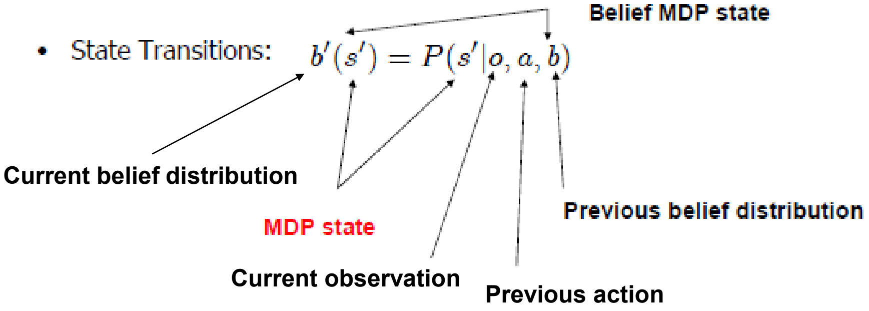
- Probability distributions over states of the underlying MDP (i.e. belief state)

- The agent keeps an internal belief state, b, that summarizes its experience(observation & control input history). The agent uses a state estimator, SE, for updating the belief state b' based on the last action a t-1, the current observation o at t, and the previous belief state b at t-1.

Converting POMDP to Belief-States MDP

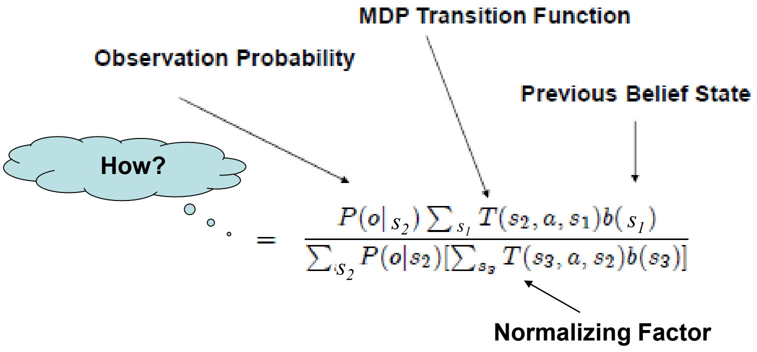
- Belief MDP State is a probability distribution over states of MDP

- State Transitions: $b'(s') = P(s'|o, a, b)$



Converting POMDP to Belief-States MDP

- Belief MDP State is a probability distribution over states of MDP
- State Transitions: $b'(s_2) = P(s_2|o,a,b)$



$$= \frac{P(o|s_2)\sum_{s_1} T(s_2,a,s_1)b(s_1)}{\sum_{s_2} P(o|s_2)[\sum_{s_3} T(s_3,a,s_2)b(s_3)]}$$

Converting POMDP to Belief-States MDP

- Belief MDP State is a probability distribution over states of MDP
- State Transitions: $b'(s_2) = P(s_2|o,a,b)$

$$= \frac{P(o|s_{2,},a,b)P(s_{21}|a,b)}{P(o|a,b)}$$

$P(s|a,b(s)) = b(s)$

$P(o|s_2,a,b) = P(o|s_2)$

Observation only Depends on state

$$= \frac{P(o|s_2)\sum_{s_{1,}} P(s_2|a,b,s_1)P(s_1|a,b)}{P(o|a,b)}$$

$P(s_2|a,b) = \sum_s P(s_2|a,b,s)P(s|a,b)$

Total Probability Theorem

Exercise: Prove it!

$$= \frac{P(o|s_2)\sum_{s_1} T(s_2,a,s_1)b(s_1)}{\sum_{s_2} P(o|s_2)[\sum_{s_3} T(s_3,a,s_2)b(s_3)]}$$

$$P(o|a,b) = \sum_s P(o|s)P(s|a,b)$$
$$= \sum_s P(o|s)\{\sum_{s'} P(s|a,b,s')P(s'|a,b(s'))\}$$

Total Probability

If {B_n: n = 1,2,3…} is a finite or countably infinite partition of a sample space, and each event B_n is measurable, then for any event A of the same probability space, the following holds

$$\Pr(A) = \sum_n \Pr(A \cap B_n) = \sum_n \Pr(A \mid B_n)\Pr(B_n) = \mathrm{E}[\Pr(A \mid B)]$$

The T.P. can also be stated for conditional probabilities. Taking the B_n as above , and assuming C is an event independent with any of the B_n

$$\Pr(A \mid C) = \sum_n \Pr(A \mid C \cap B_n)\Pr(B_n \mid C) = \sum_n \Pr(A \mid C \cap B_n)\Pr(B_n)$$

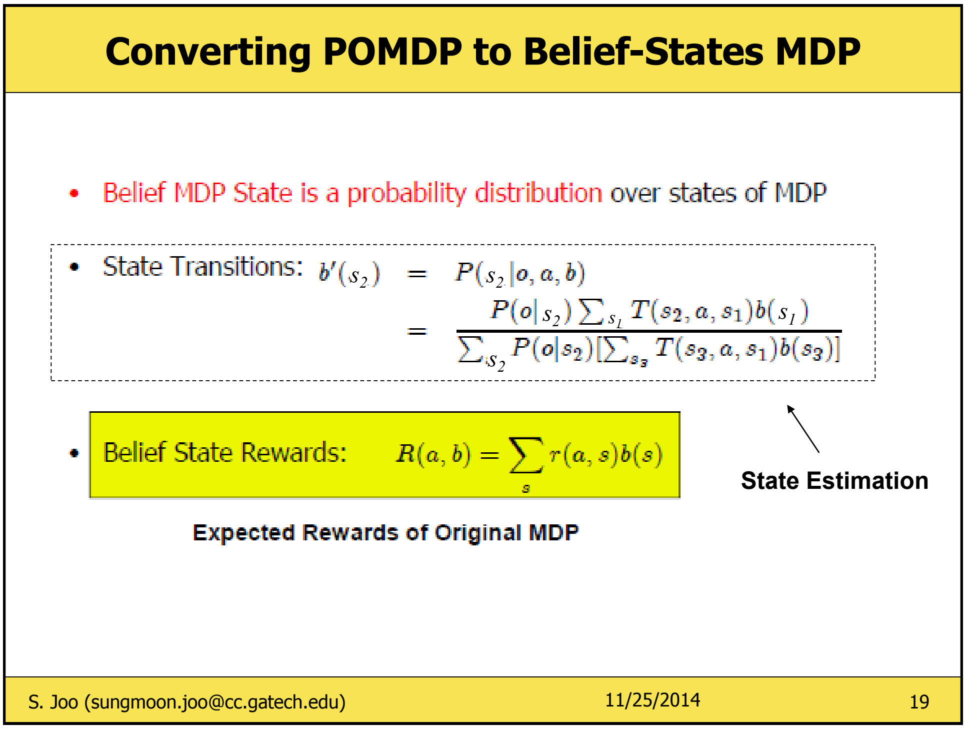
Converting POMDP to Belief-States MDP

- Belief MDP State is a probability distribution over states of MDP
- State Transitions:

$$\begin{aligned} b'(s_2) &= P(s_2 | o, a, b) \\ &= \frac{P(o|s_2, a, b)P(s_2|a, b)}{P(o|a, b)} \\ &= \frac{P(o|s_2)\sum_{s_1} P(s_2|a, b, s_1)P(s_1|a, b)}{P(o|a, b)} \\ &= \frac{P(o|s_2)\sum_{s_1} T(s_2, a, s_1)b(s_1)}{P(o|a, b)} \\ &= \frac{P(o|s_2)\sum_{s_1} T(s_2, a, s_1)b(s_1)}{\sum_{s_2} P(o|s_2)[\sum_{s_3} T(s_3, a, s_2)b(s_3)]} \end{aligned}$$

Definition of T

Definition of b



Converting POMDP to Belief-States MDP

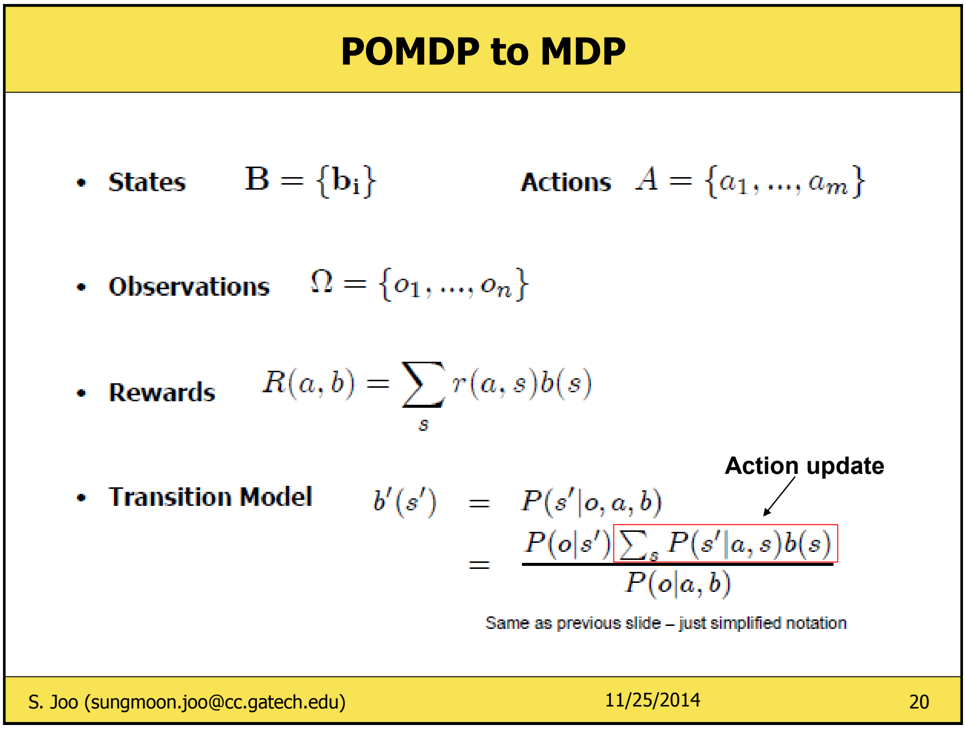
- Belief MDP State is a probability distribution over states of MDP

- State Transitions: $$b'(s_2) = P(s_2|o,a,b) = \frac{P(o|s_2)\sum_{s_1} T(s_2,a,s_1)b(s_1)}{\sum_{s_2} P(o|s_2)[\sum_{s_3} T(s_3,a,s_1)b(s_3)]}$$

- Belief State Rewards: $R(a,b) = \sum_s r(a,s)b(s)$

State Estimation

Expected Rewards of Original MDP



POMDP to MDP

- **States** $\mathrm{B} = \{\mathrm{b_i}\}$ **Actions** $A = \{a_1, \ldots, a_m\}$

- **Observations** $\Omega = \{o_1, \ldots, o_n\}$

- **Rewards** $R(a,b) = \sum_s r(a,s)b(s)$

- **Transition Model** $$b'(s') = P(s'|o,a,b) = \frac{P(o|s')\boxed{\sum_s P(s'|a,s)b(s)}}{P(o|a,b)}$$

Action update

Same as previous slide – just simplified notation

POMDP to MDP

- **States** $\mathrm{B} = \{\mathrm{b_i}\}$ **Actions** $A = \{a_1, \ldots, a_m\}$
- **Observations** $\Omega = \{o_1, \ldots, o_n\}$
- **Rewards** $R(a,b) = \sum_s r(a,s)b(s)$
- **Transition Model**

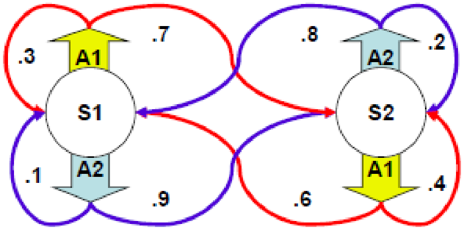
$$b'(s') = P(s'|o,a,b) = \frac{\boxed{P(o|s')}\boxed{\sum_s P(s'|a,s)b(s)}}{\boxed{P(o|a,b)}}$$

Observation update

Same as previous slide – just simplified notation

POMDP Example

$r(A1,S1) = 2$
$r(A1,S2) = 1$
$r(A2,S1) = 1$
$r(A2,S2) = 3$

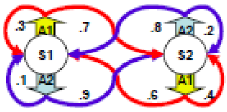


$P(O1|S1) = .9$
$P(O1|S2) = .5$
$P(O2|S1) = .1$
$P(O2|S2) = .5$

We will use a vector: [P_1 P_2] to represent a belief state:

$b(S1) = P_1$ $b(S2) = P_2 = 1 - P_1$

POMDP Example

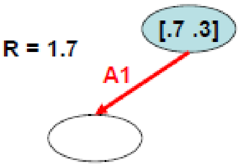


$r(A1, S1) = 2$ $P(O1|S1) = .9$

$r(A1, S2) = 1$ $P(O1|S2) = .5$

$r(A2, S1) = 1$ $P(O2|S1) = .1$

$r(A2, S2) = 3$ $P(O2|S2) = .5$

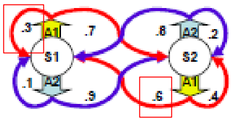


Rewards:

$$R(a, b) = \sum_s r(a, s)b(s)$$

$$\mathbf{R(A1, [.7\ \ .3]) = 2 \times .7 + 1 \times .3 = 1.7}$$

POMDP Example

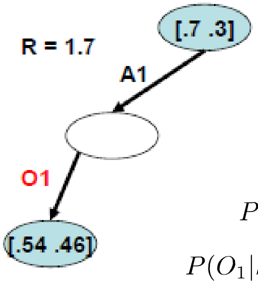


$r(A1, S1) = 2$ $P(O1|S1) = .9$

$r(A1, S2) = 1$ $P(O1|S2) = .5$

$r(A2, S1) = 1$ $P(O2|S1) = .1$

$r(A2, S2) = 3$ $P(O2|S2) = .5$



Rewards:

$$R(a, b) = \sum_s r(a, s)b(s)$$

$$\mathbf{R(A1, [.7\ \ .3]) = 2 \times .7 + 1 \times .3 = 1.7}$$

Transitions:

$$\begin{aligned} b'(s') &= P(s'|o, a, b) \\ &= \frac{P(o|s') \sum_s P(s'|a, s)b(s)}{P(o|a, b)} \end{aligned}$$

$P(S_1|A_1, S_2)b(S_2)$

$P(S_1|A_1, S_1)b(S_1)$

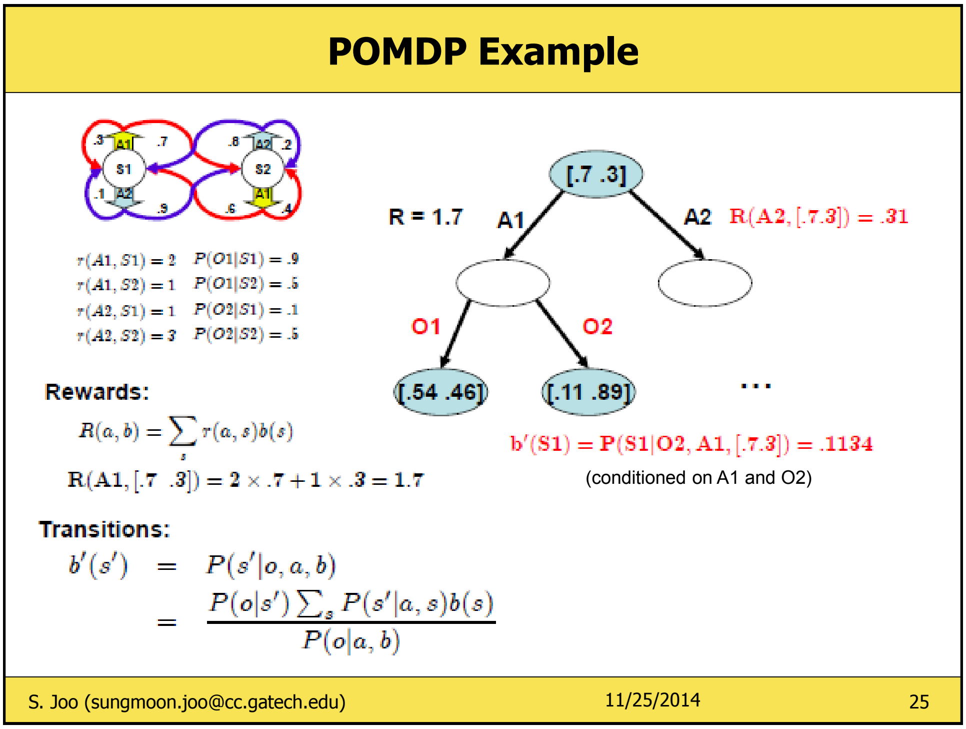
$P(O_1|S_1)$

$$b'(S1) = \frac{.9 \times (\overset{.351}{.3 \times .7 + .6 \times .3)}}{P(O1|A1, b)}$$

$$b'(S2) = \frac{.5 \times (.7 \times .7 + .4 \times .3)}{P(O1|A1, b)}\ \ .305$$

$$P(O1|A1, b) = .351 + .305 = .656$$

$$b'(S1) = .351/.656 = .54$$



POMDP Example

$r(A1, S1) = 2$ $P(O1|S1) = .9$

$r(A1, S2) = 1$ $P(O1|S2) = .5$

$r(A2, S1) = 1$ $P(O2|S1) = .1$

$r(A2, S2) = 3$ $P(O2|S2) = .5$

R = 1.7

$\mathrm{R(A2, [.7 .3]) = .31}$

Rewards:

$$R(a, b) = \sum_s r(a, s)b(s)$$

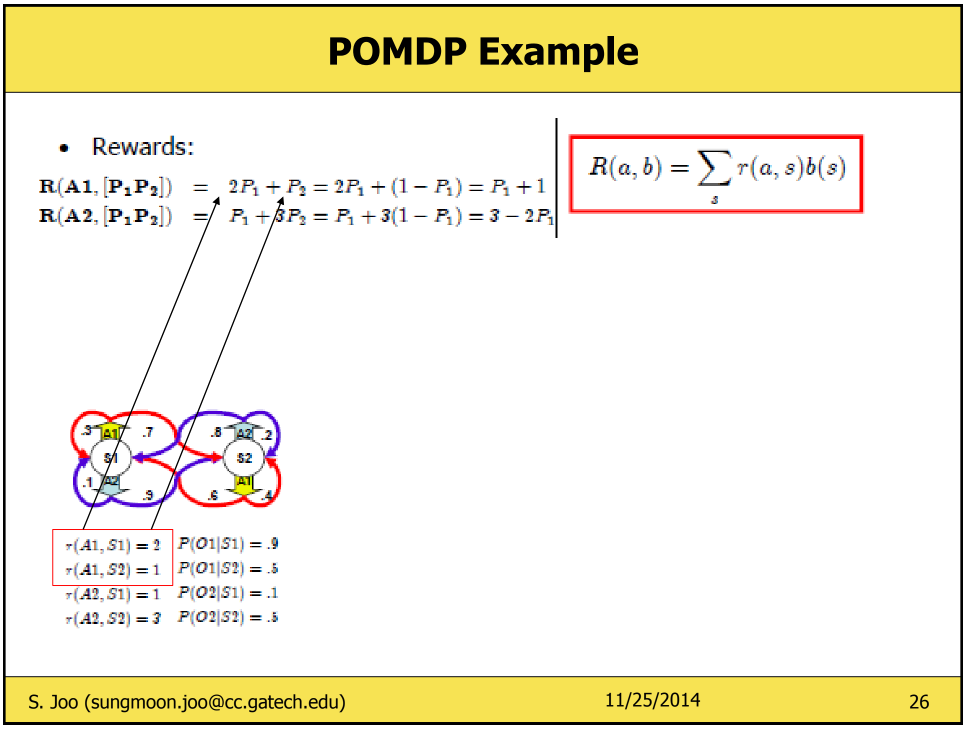
$$\mathrm{R(A1, [.7\ \ .3]) = 2 \times .7 + 1 \times .3 = 1.7}$$

$\mathrm{b'(S1) = P(S1|O2, A1, [.7 .3]) = .1134}$

(conditioned on A1 and O2)

Transitions:

$$\begin{aligned} b'(s') &= P(s'|o, a, b) \\ &= \frac{P(o|s') \sum_s P(s'|a, s)b(s)}{P(o|a, b)} \end{aligned}$$



POMDP Example

- Rewards:

$$\begin{aligned} \mathbf{R(A1, [P_1 P_2])} &= 2P_1 + P_2 = 2P_1 + (1 - P_1) = P_1 + 1 \\ \mathbf{R(A2, [P_1 P_2])} &= P_1 + 3P_2 = P_1 + 3(1 - P_1) = 3 - 2P_1 \end{aligned}$$

$$\boxed{R(a, b) = \sum_s r(a, s)b(s)}$$

$r(A1, S1) = 2$ $P(O1|S1) = .9$

$r(A1, S2) = 1$ $P(O1|S2) = .5$

$r(A2, S1) = 1$ $P(O2|S1) = .1$

$r(A2, S2) = 3$ $P(O2|S2) = .5$

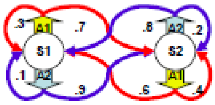
POMDP Example

- Rewards:

$$\begin{aligned}\mathbf{R}(\mathbf{A1},[\mathbf{P_1P_2}]) &= 2P_1 + P_2 = 2P_1 + (1-P_1) = P_1 + 1\\ \mathbf{R}(\mathbf{A2},[\mathbf{P_1P_2}]) &= P_1 + 3P_2 = P_1 + 3(1-P_1) = 3 - 2P_1\end{aligned}$$

- Transitions:

$$\begin{aligned}b'(s') &= P(s'|o,a,b(s))\\ &= \frac{P(o|s')\sum_s P(s'|a,s)b(s)}{P(o|a,b(s))}\\ &= \frac{P(o|s')P(s'|a,b(s))}{P(o|a,b(s))}\end{aligned}$$



$r(A1,S1) = 2$ $P(O1|S1) = .9$
$r(A1,S2) = 1$ $P(O1|S2) = .5$
$r(A2,S1) = 1$ $P(O2|S1) = .1$
$r(A2,S2) = 3$ $P(O2|S2) = .5$

POMDP Example

- Rewards:

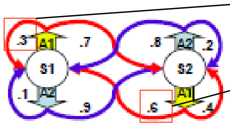
$$\begin{aligned}\mathbf{R}(\mathbf{A1},[\mathbf{P_1P_2}]) &= 2P_1 + P_2 = 2P_1 + (1-P_1) = P_1 + 1\\ \mathbf{R}(\mathbf{A2},[\mathbf{P_1P_2}]) &= P_1 + 3P_2 = P_1 + 3(1-P_1) = 3 - 2P_1\end{aligned}$$

- Transitions:

$$\begin{aligned}b'(s') &= P(s'|o,a,b(s))\\ &= \frac{P(o|s')\sum_s P(s'|a,s)b(s)}{P(o|a,b(s))}\\ &= \frac{P(o|s')\boxed{P(s'|a,b(s))}}{P(o|a,b(s))}\end{aligned}$$

$$\begin{aligned}\mathbf{P}(\mathbf{S1}|\mathbf{A1},[\mathbf{P_1P_2}]) &= P(S1|A1,S1)P_1 + P(S1|A1,S2)P_2\\ &= .3P_1 + .6P_2 = .3P_1 + .6(1-P_1)\\ &= .6 - .3P_1\\ \mathbf{P}(\mathbf{S2}|\mathbf{A1},[\mathbf{P_1P_2}]) &= .7P_1 + .4P_2 = .4 + .3P_1\\ \mathbf{P}(\mathbf{S1}|\mathbf{A2},[\mathbf{P_1P_2}]) &= .1P_1 + .8P_2 = .8 - .7P_1\\ \mathbf{P}(\mathbf{S2}|\mathbf{A2},[\mathbf{P_1P_2}]) &= .9P_1 + .2P_2 = .2 + .7P_1\end{aligned}$$

T.P.



$r(A1,S1) = 2$ $P(O1|S1) = .9$
$r(A1,S2) = 1$ $P(O1|S2) = .5$
$r(A2,S1) = 1$ $P(O2|S1) = .1$
$r(A2,S2) = 3$ $P(O2|S2) = .5$

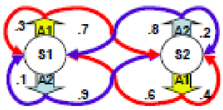
POMDP Example

- Rewards:

$$\begin{aligned}\mathbf{R}(\mathbf{A1},[\mathbf{P_1P_2}]) &= 2P_1+P_2=2P_1+(1-P_1)=P_1+1\\ \mathbf{R}(\mathbf{A2},[\mathbf{P_1P_2}]) &= P_1+3P_2=P_1+3(1-P_1)=3-2P_1\end{aligned}$$

$$\begin{aligned}b'(s') &= P(s'|o,a,b(s))\\ &= \frac{P(o|s')\sum_s P(s'|a,s)b(s)}{P(o|a,b(s))}\\ &= \frac{P(o|s')P(s'|a,b(s))}{\boxed{P(o|a,b(s))}}\end{aligned}$$

- Transitions:



$$\begin{aligned}r(A1,S1)&=2 & P(O1|S1)&=\boxed{.9}\\ r(A1,S2)&=1 & P(O1|S2)&=\boxed{.5}\\ r(A2,S1)&=1 & P(O2|S1)&=.1\\ r(A2,S2)&=3 & P(O2|S2)&=.5\end{aligned}$$

$$\begin{aligned}\mathbf{P}(\mathbf{S1}|\mathbf{A1},[\mathbf{P_1P_2}]) &= P(S1|A1,S1)P_1+P(S1|A1,S2)P_2\\ &= .3P_1+.6P_2=.3P_1+.6(1-P_1)\\ &= .6-.3P_1\\ \mathbf{P}(\mathbf{S2}|\mathbf{A1},[\mathbf{P_1P_2}]) &= .7P_1+.4P_2=.4+.3P_1\\ \mathbf{P}(\mathbf{S1}|\mathbf{A2},[\mathbf{P_1P_2}]) &= .1P_1+.8P_2=.8-.7P_1\\ \mathbf{P}(\mathbf{S2}|\mathbf{A2},[\mathbf{P_1P_2}]) &= .9P_1+.2P_2=.2+.7P_1\end{aligned}$$

$$\begin{aligned}\mathbf{P}(\mathbf{O1}|\mathbf{A1},[\mathbf{P_1P_2}]) &= P(O1|S1)\boxed{P(S1|A1,[P_1])}+P(O1|S2)\boxed{P(S2|A1,[P_1])}\\ &= .9(.6-.3P_1)+.5(.4+.3P_1)\\ &= .74-.12P_1\\ \mathbf{P}(\mathbf{O2}|\mathbf{A1},[\mathbf{P_1P_2}]) &= .26+.12P_1\\ \mathbf{P}(\mathbf{O1}|\mathbf{A2},[\mathbf{P_1P_2}]) &= .82-.28P_1\\ \mathbf{P}(\mathbf{O2}|\mathbf{A2},[\mathbf{P_1P_2}]) &= .18+.28P_1\end{aligned}$$

POMDP Example

- Rewards:

$$\begin{aligned}\mathbf{R}(\mathbf{A1},[\mathbf{P_1P_2}]) &= 2P_1+P_2=2P_1+(1-P_1)=P_1+1\\ \mathbf{R}(\mathbf{A2},[\mathbf{P_1P_2}]) &= P_1+3P_2=P_1+3(1-P_1)=3-2P_1\end{aligned}$$

$$\begin{aligned}b'(s') &= P(s'|o,a,b(s))\\ &= \frac{P(o|s')\sum_s P(s'|a,s)b(s)}{P(o|a,b(s))}\\ &= \frac{P(o|s')P(s'|a,b(s))}{P(o|a,b(s))}\end{aligned}$$

- Transitions:

$$\begin{aligned}\mathbf{P}(\mathbf{S1}|\mathbf{A1},\mathbf{O1},[\mathbf{P_1P_2}]) &= P(O1|S1)P(S1|A1,[P_1P_2])/P(O1|A1,[P_1P_2])\\ &= .90(.60-.30P_1)/(.74-.12P_1)\\ &= (.54-.27P_1)/(.74-.12P_1)\\ \mathbf{P}(\mathbf{S2}|\mathbf{A1},\mathbf{O1},[\mathbf{P_1P_2}]) &= (.20+.15P_1)/(.74-.12P_1)\\ \mathbf{P}(\mathbf{S1}|\mathbf{A2},\mathbf{O1},[\mathbf{P_1P_2}]) &= (.72-.63P_1)/(.82-.28P_1)\\ \mathbf{P}(\mathbf{S2}|\mathbf{A2},\mathbf{O1},[\mathbf{P_1P_2}]) &= (.10+.35P_1)/(.82-.28P_1)\\ \mathbf{P}(\mathbf{S1}|\mathbf{A1},\mathbf{O2},[\mathbf{P_1P_2}]) &= (.06-.03P_1)/(.26+.12P_1)\\ \mathbf{P}(\mathbf{S2}|\mathbf{A1},\mathbf{O2},[\mathbf{P_1P_2}]) &= (.20+.15P_1)/(.26+.12P_1)\\ \mathbf{P}(\mathbf{S1}|\mathbf{A2},\mathbf{O2},[\mathbf{P_1P_2}]) &= (.08-.07P_1)/(.18+.28P_1)\\ \mathbf{P}(\mathbf{S2}|\mathbf{A2},\mathbf{O2},[\mathbf{P_1P_2}]) &= (.10+.35P_1)/(.18+.28P_1)\end{aligned}$$

$$\begin{aligned}&\boxed{\mathbf{P}(\mathbf{S1}|\mathbf{A1},[\mathbf{P_1P_2}]) = .6-.3P_1}\\ &\mathbf{P}(\mathbf{S2}|\mathbf{A1},[\mathbf{P_1P_2}]) = .4+.3P_1\\ &\mathbf{P}(\mathbf{S1}|\mathbf{A2},[\mathbf{P_1P_2}]) = .8-.7P_1\\ &\mathbf{P}(\mathbf{S2}|\mathbf{A2},[\mathbf{P_1P_2}]) = .2+.7P_1\end{aligned}$$

$$\begin{aligned}&\boxed{\mathbf{P}(\mathbf{O1}|\mathbf{A1},[\mathbf{P_1P_2}]) = .74-.12P_1}\\ &\mathbf{P}(\mathbf{O2}|\mathbf{A1},[\mathbf{P_1P_2}]) = .26+.12P_1\\ &\mathbf{P}(\mathbf{O1}|\mathbf{A2},[\mathbf{P_1P_2}]) = .82-.28P_1\\ &\mathbf{P}(\mathbf{O2}|\mathbf{A2},[\mathbf{P_1P_2}]) = .18+.28P_1\end{aligned}$$

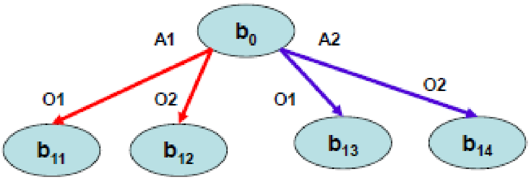
How to Solve Belief-State MDP?

- **States** $\mathrm{B} = \{\mathrm{b_i}\}$ **Actions** $A = \{a_1, ..., a_m\}$
- **Observations** $\Omega = \{o_1, ..., o_n\}$
- **Rewards** $R(a, b) = \sum_s r(a, s) b(s)$
- **Transition Model** $$\begin{aligned} b'(s') &= P(s'|o, a, b) \\ &= \frac{P(o|s') \sum_s P(s'|a, s) b(s)}{P(o|a, b)} \end{aligned}$$

Same as previous slide – just simplified notation

Solving a POMDP

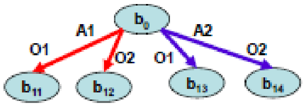
- Convert to MDP and then use Value Iteration
- **How to use Value Iteration in a Continuous State Space?**



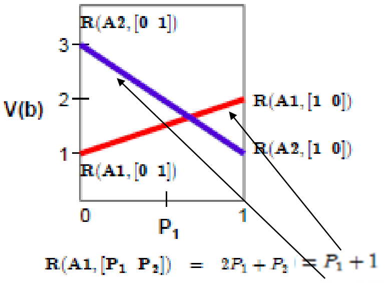
$$V(b) = \max_a [R(a, b) + \gamma \sum_{b'} P(b'|a, b) V(b')]$$

Solving a POMDP

- Convert to MDP and then use Value Iteration
- How to use Value Iteration in a Continuous State Space?



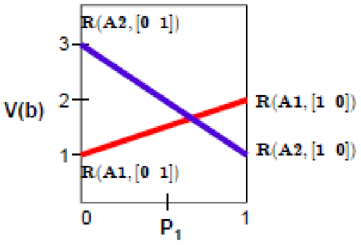
$$V(b) = \max_a [R(a,b) + \gamma \sum_{b'} P(b'|a,b) V(b')]$$



$$\mathbf{R}(\mathbf{A1},[\mathbf{P_1}\ \ \mathbf{P_2}]) = 2P_1 + P_2 = P_1 + 1$$
$$\mathbf{R}(\mathbf{A2},[\mathbf{P_1}\ \ \mathbf{P_2}]) = P_1 + 3P_2 = 3 - 2P_1$$

- **Property: Value Function is Piece-wise Linear & Convex**

Solving a POMDP: Step1



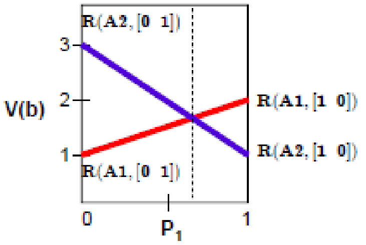
$$V(b) = \max_a [R(a,b) + \gamma \sum_{b'} P(b'|a,b) V(b')]$$

$$\mathbf{R}(\mathbf{A1},[\mathbf{P_1}\ \ \mathbf{P_2}]) = 2P_1 + P_2$$
$$\mathbf{R}(\mathbf{A2},[\mathbf{P_1}\ \ \mathbf{P_2}]) = P_1 + 3P_2$$

$$\mathbf{V_1}(\mathbf{A1},\mathbf{b}) = R(A1,b) + \gamma \cdot 0 = 2P_1 + P_2$$
$$\mathbf{V_1}(\mathbf{A2},\mathbf{b}) = R(A2,b) + \gamma \cdot 0 = P_1 + 3P_2$$

- Values are Piece-wise Linear & Convex (Lines or Hyperplanes)

Solving a POMDP: Step1



$$V(b) = \max_a [R(a,b) + \gamma \sum_{b'} P(b'|a,b)V(b')]$$

$$\begin{aligned}\mathbf{R}(\mathbf{A1}, [\mathbf{P_1}\ \ \mathbf{P_2}]) &= 2P_1 + P_2\\ \mathbf{R}(\mathbf{A2}, [\mathbf{P_1}\ \ \mathbf{P_2}]) &= P_1 + 3P_2\end{aligned}$$

$$\begin{aligned}\mathbf{V_1}(\mathbf{A1}, \mathbf{b}) &= R(A1, b) + \gamma \cdot 0 = 2P_1 + P_2\\ \mathbf{V_1}(\mathbf{A2}, \mathbf{b}) &= R(A2, b) + \gamma \cdot 0 = P_1 + 3P_2\end{aligned}$$

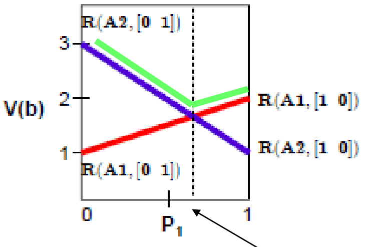
- Values are Piece-wise Linear & Convex (Lines or Hyperplanes)
- Another Interpretation of Value Lines = Vectors of Coefficients on Belief

$$\Psi_1 = \{[2\ \ 1], [1\ \ 3]\}$$

$$\begin{aligned}\mathbf{V_1}(\mathbf{A1}, \mathbf{b}) &= [2\ \ 1] \cdot [P_1\ \ P_2]\\ \mathbf{V_1}(\mathbf{A2}, \mathbf{b}) &= [1\ \ 3] \cdot [P_1\ \ P_2]\end{aligned}$$

Crossover?

Solving a POMDP: Step1



$$V(b) = \max_a [R(a,b) + \gamma \sum_{b'} P(b'|a,b)V(b')]$$

$$\begin{aligned}\mathbf{R}(\mathbf{A1}, [\mathbf{P_1}\ \ \mathbf{P_2}]) &= 2P_1 + P_2\\ \mathbf{R}(\mathbf{A2}, [\mathbf{P_1}\ \ \mathbf{P_2}]) &= P_1 + 3P_2\end{aligned}$$

$$\begin{aligned}\mathbf{V_1}(\mathbf{A1}, \mathbf{b}) &= R(A1, b) + \gamma \cdot 0 = 2P_1 + P_2\\ \mathbf{V_1}(\mathbf{A2}, \mathbf{b}) &= R(A2, b) + \gamma \cdot 0 = P_1 + 3P_2\end{aligned}$$

- Values are Piece-wise Linear & Convex (Lines or Hyperplanes)
- Another Interpretation of Value Lines = Vectors of Coefficients on Belief

$$\Psi_1 = \{[2\ \ 1], [1\ \ 3]\}$$

$$\begin{aligned}\mathbf{V_1}(\mathbf{A1}, \mathbf{b}) &= [2\ \ 1] \cdot [P_1\ \ P_2]\\ \mathbf{V_1}(\mathbf{A2}, \mathbf{b}) &= [1\ \ 3] \cdot [P_1\ \ P_2]\end{aligned}$$

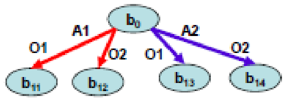
Crossover:

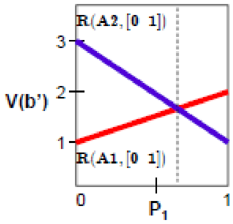
$$\begin{aligned}[2\ \ 1] \cdot [P_1\ \ P_2] &= [1\ \ 3] \cdot [P_1\ \ P_2]\\ 2P_1 + (1 - P_1) &= P_1 + 3(1 - P_1)\\ P_1 &= \boxed{2/3}\end{aligned}$$

Solving a POMDP: Step2

$$V(b) = \max_a [R(a,b) + \gamma \sum_{b'} P(b'|a,b)\boxed{V(b')}]$$

$$\boxed{\mathbf{V}(\mathbf{b}') = ?}$$





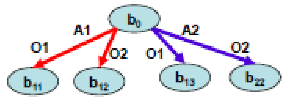
$$\Psi_1 = \{[2\ \ 1], [1\ \ 3]\}$$

$$\begin{aligned}\mathbf{V}_1(\mathbf{A1}, \mathbf{b}) &= [2\ \ 1] \cdot [P_1\ \ P_2] \\ \mathbf{V}_1(\mathbf{A2}, \mathbf{b}) &= [1\ \ 3] \cdot [P_1\ \ P_2]\end{aligned}$$

Solving a POMDP: Step2

$$V(b) = \max_a [R(a,b) + \gamma \sum_{b'} P(b'|a,b)V(b')]$$

$$\mathbf{V}(\mathbf{b}') = \mathbf{max}_{\mathbf{v}\in\Psi}(\mathbf{v} \cdot \mathbf{b}')$$



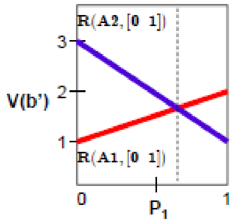
A bit of math trickery:

$$P(b'|a,b) = P(o|a,b)$$

$$\begin{aligned} b' &= [P(S1|a,o,b)\ \ P(S2|a,o,b)] \\ b' &= [\frac{P(o|S1)P(S1|a,b)}{P(o|a,b)}\ \ \frac{P(o|S2)P(S2|a,b)}{P(o|a,b)}] \\ \underline{b}' &= [P(o|S1)P(S1|a,b)\ \ P(o|S2)P(S2|a,b)] \end{aligned}$$

$$V(b) = \max_a [R(a,b) + \gamma \sum_{b'} \cancel{P(o|a,b)} \max_{v\in\Psi} \frac{\mathbf{v} \cdot \underline{\mathbf{b}}'}{\cancel{P(o|a,b)}}]$$

$$\boxed{V(b) = \max_a [R(a,b) + \gamma \sum_{b'} \max_{v\in\Psi}(\mathbf{v} \cdot \underline{\mathbf{b}}')]}$$



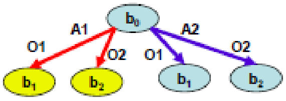
$$\Psi_1 = \{[2\ \ 1], [1\ \ 3]\}$$

$$\begin{aligned}\mathbf{V}_1(\mathbf{A1}, \mathbf{b}) &= [2\ \ 1] \cdot [P_1\ \ P_2] \\ \mathbf{V}_1(\mathbf{A2}, \mathbf{b}) &= [1\ \ 3] \cdot [P_1\ \ P_2]\end{aligned}$$

Solving a POMDP: Step2

$$V(b) = \max_a [R(a,b) + \gamma \sum_{b'} P(b'|a,b)V(b')]$$

$$V(b) = \max_a \boxed{R(a,b) + \gamma \sum_{b'} \max_{v \in \Psi}(\mathbf{v} \cdot \mathbf{b}')}$$

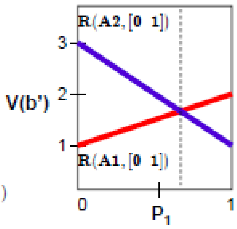


Computing the Values for Action 1

$$V_2(A1,b) = R(A1,b) + \gamma(\max_{v \in \Psi}(\mathbf{v} \cdot \underline{\mathbf{b}}_1') + \max_{v \in \Psi}(\mathbf{v} \cdot \underline{\mathbf{b}}_2'))$$

$$\begin{aligned} \underline{b}_1' &= [P(O1|S1)P(S1|A1,b) \;\; P(O1|S2)P(S2|A1,b)] \\ &= [.054 - .27P_1 \;\; .20 + .15P_1] \\ \underline{b}_2' &= [.06 - .03P_1 \;\; .20 + .15P_1] \end{aligned}$$

$$\begin{aligned} V_2(A1,b) &= (P_1+1) + \gamma([2\;\;1] \cdot \underline{b}_1' + [2\;\;1] \cdot \underline{b}_2') = (P_1+1) + \gamma(1.6 - .3P_1) \\ &\;\; or \\ V_2(A1,b) &= (P_1+1) + \gamma([2\;\;1] \cdot \underline{b}_1' + [1\;\;3] \cdot \underline{b}_2') = (P_1+1) + \gamma(1.94 - .03P_1) \\ V_2(A1,b) &= (P_1+1) + \gamma([1\;\;3] \cdot \underline{b}_1' + [2\;\;1] \cdot \underline{b}_2') = (P_1+1) + \gamma(1.46 - .27P_1) \\ V_2(A1,b) &= (P_1+1) + \gamma([1\;\;3] \cdot \underline{b}_1' + [1\;\;3] \cdot \underline{b}_2') = (P_1+1) + \gamma(1.86 - .6P_1) \end{aligned}$$



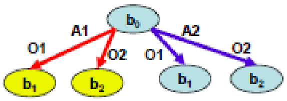
$$\Psi_1 = \{[2\;\;1], [1\;\;3]\}$$

$$\begin{aligned} \mathbf{V}_1(\mathbf{A1}, \mathbf{b}) &= [2\;\;1] \cdot [P_1\;\;P_2] \\ \mathbf{V}_1(\mathbf{A2}, \mathbf{b}) &= [1\;\;3] \cdot [P_1\;\;P_2] \end{aligned}$$

Solving a POMDP: Step2

$$V(b) = \max_a [R(a,b) + \gamma \sum_{b'} P(b'|a,b)V(b')]$$

$$V(b) = \max_a \boxed{R(a,b) + \gamma \sum_{b'} \max_{v \in \Psi}(\mathbf{v} \cdot \mathbf{b}')}$$

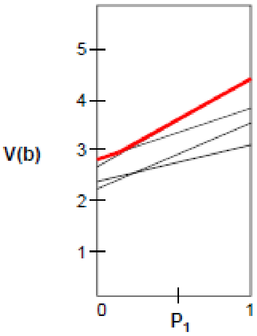


Computing the Values for Action 1

$$V_2(A1,b) = R(A1,b) + \gamma(\max_{v \in \Psi}(\mathbf{v} \cdot \underline{\mathbf{b}}_1') + \max_{v \in \Psi}(\mathbf{v} \cdot \underline{\mathbf{b}}_2'))$$

$$\begin{aligned} \underline{b}_1' &= [P(O1|S1)P(S1|A1,b) \;\; P(O1|S2)P(S2|A1,b)] \\ &= [.054 - .27P_1 \;\; .20 + .15P_1] \\ \underline{b}_2' &= [.06 - .03P_1 \;\; .20 + .15P_1] \end{aligned}$$

$$\begin{aligned} V_2(A1,b) &= 2.44 + .73P_1 \equiv [3.17\;\;2.44] \\ &\;\; or \\ V_2(A1,b) &= 2.75 + 1.0P_1 \equiv [3.77\;\;2.75] \\ V_2(A1,b) &= 2.31 + 1.2P_1 \equiv [3.56\;\;2.31] \\ V_2(A1,b) &= 2.62 + 1.5P_1 \equiv [4.16\;\;2.62] \end{aligned}$$



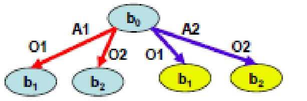
Solving a POMDP: Step2

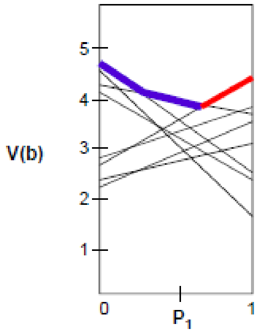
$$V(b) = \max_a [R(a,b) + \gamma \sum_{b'} P(b'|a,b)V(b')]$$

$$V(b) = \max_a \boxed{[R(a,b) + \gamma \sum_{b'} \max_{v\in\Psi}(\mathrm{v}\cdot\underline{\mathbf{b}}')}]$$

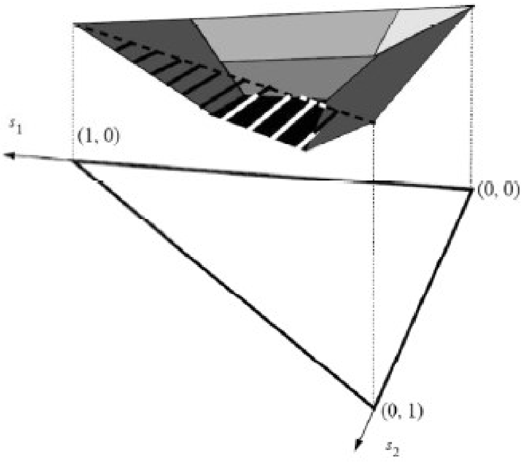
Complete Maximized Value Function: V2!

$$V_2(A2,b) = R(A2,b) + \gamma(\max_{v\in\Psi}(\mathrm{v}\cdot\underline{\mathbf{b}}_1') + \max_{v\in\Psi}(\mathrm{v}\cdot\underline{\mathbf{b}}_2'))$$





POMDP in Higher Dimensions: Hyperplanes

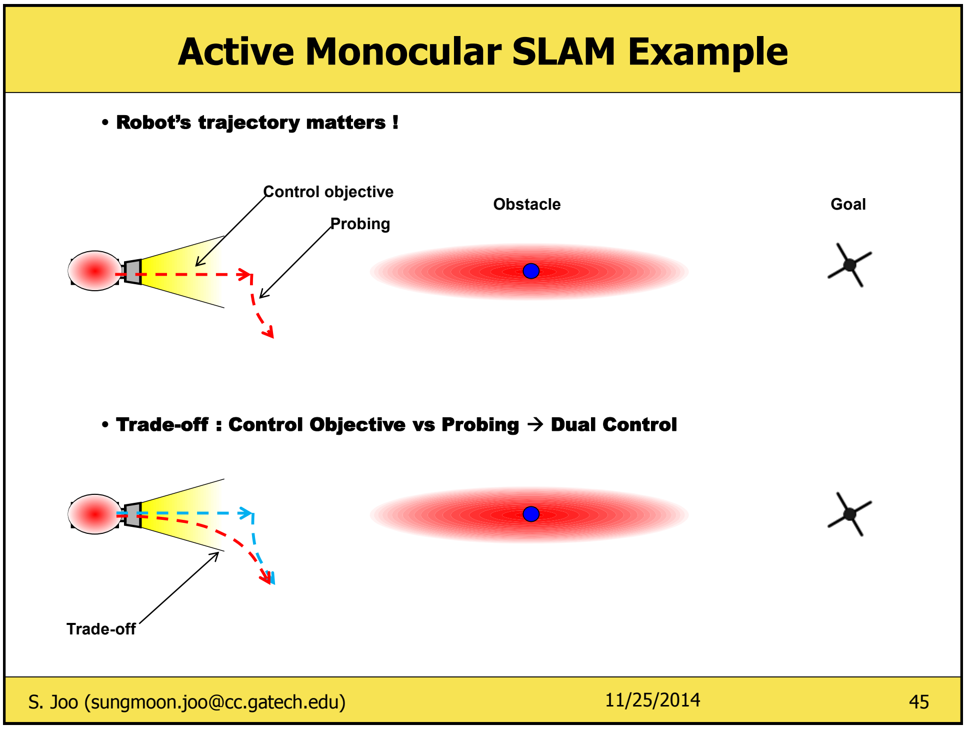


POMDP Summary

- Complex but Powerful technique
  - State explodes upon conversion to MDP
  - State becomes difficult to understand upon conversion to MDP
  - Unique cohesive method that trades off:
    : Value of ascertaining state
    : Value of pursuing a goal
- Exist more efficient algorithms:
  - Witness Algorithm (Littman '94)
  - Policy Iteration (Sondik, Hansen '97)
- Typically complexity is still prohibitive for large problems

POMDP Summary

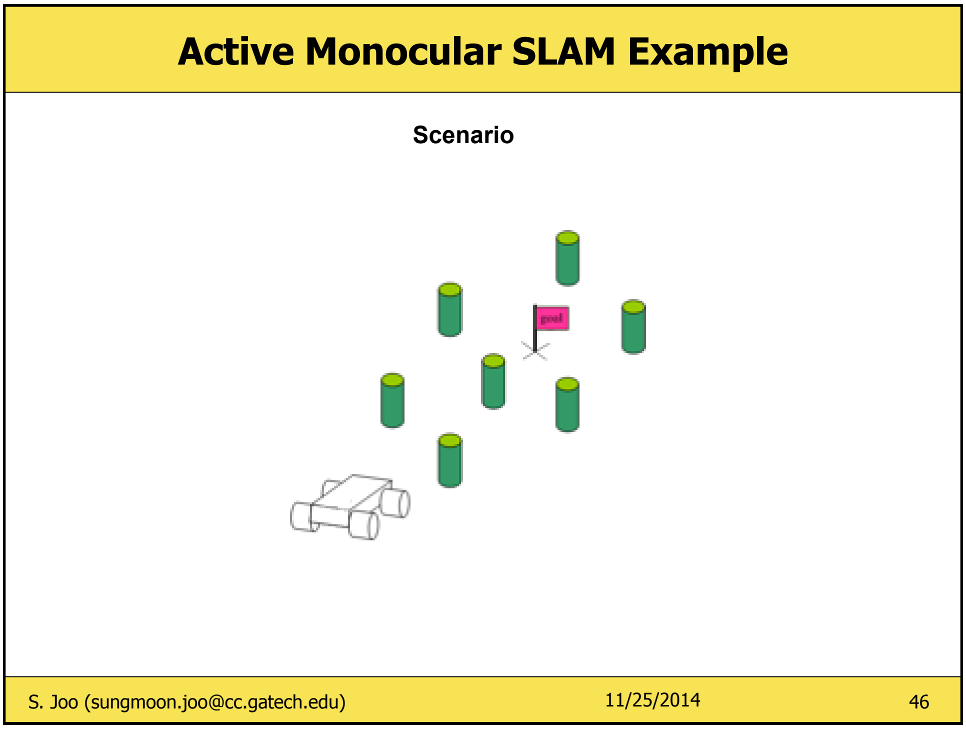
- Canonical solution method 1 – Covered today
  - Run value iteration, but now the state space is the space of probability distributions
    :value and optimal action for every possible probability distribution
    :will automatically trade off information gathering actions versus actions that affect the underlying state
- Canonical solution method 2 – Finite-horizon/MPC-style
  - Search over sequences of actions with limited look-ahead
  - Branching over actions and observations
- Canonical solution method 3 – LQG-style
  - Plan in the MDP
  - Run probabilistic inference (filtering) to track probability distribution
  - Choose optimal action for MDP for what is currently the most likely state



Active Monocular SLAM Example

- **Robot's trajectory matters !**

- **Trade-off : Control Objective vs Probing → Dual Control**



Active Monocular SLAM Example

Scenario

Active Monocular SLAM Example

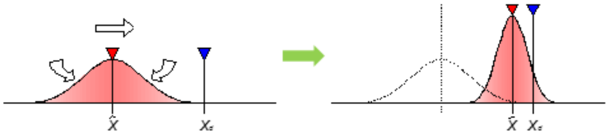
Stochastic Performance Index – Dual Effect

Conventional Control Objective

$$\begin{aligned}\delta J = E[\,&(\hat{\mathbf{x}}_{robot} - \mathbf{x}^{d}_{robot})^{T} \mathcal{Q}_{robot} (\hat{\mathbf{x}}_{robot} - \mathbf{x}^{d}_{robot}) + \mathbf{u}^{T} \mathcal{R}_{c} \mathbf{u} \\ &+ (\hat{\mathbf{x}}_{robot} - \mathbf{x}_{robot})^{T} \mathcal{Q}_{robot} (\hat{\mathbf{x}}_{robot} - \mathbf{x}_{robot}) \\ &+ (\hat{\mathbf{x}}_{obstacle} - \mathbf{x}_{obstacle})^{T} \mathcal{Q}_{obstacle} (\hat{\mathbf{x}}_{obstacle} - \mathbf{x}_{obstacle}) | \mathcal{I}\,] \delta t\end{aligned}$$

Localization Performance

Mapping Performance



$$J = \int_{0}^{\infty} (\hat{\mathrm{x}}_{robot} - \mathrm{x}^{d}_{robot})^{T} \mathcal{Q}_{robot} (\hat{\mathrm{x}}_{robot} - \mathrm{x}^{d}_{robot}) + \mathrm{u}^{T} \mathcal{R}_{C} \mathrm{u} + \mathrm{Tr}[\,\mathcal{Q}_{SLAM}\, \mathrm{P}_{SLAM}\,] dt$$

Sungmoon Joo, “SLAM-based nonlinear optimal control approach to robot navigation with limited resources”

Active Monocular SLAM Example