Planning under Uncertainty:
Partially Observable
Markov Decision Processes (POMDP)

Maxim Likhachev
Robotics Institute
Carnegie Mellon University
Graph vs. MDP vs. POMDP

• Consider a path planning example
Graph vs. MDP vs. POMDP

- Consider a path planning example
- Assume perfect action execution and full knowledge of the state (i.e., perfect localization)
Graph vs. MDP vs. POMDP

- Consider a path planning example
- Assume perfect action execution and full knowledge of the state (i.e., perfect localization)

Graph:
Implicitly defined as \(\{S, A, C\}\), where \(S\) – set of states, \(A\) – set of actions, \(C\) – costs of all \((s, a)\) pairs.
Graph vs. MDP vs. POMDP

- Consider a path planning example
- Assume perfect action execution and full knowledge of the state (i.e., perfect localization)

**Graph:**
Implicitly defined as \( \{S, A, C\} \), where \( S \) – set of states, \( A \) – set of actions, \( C \) – costs of all \((s,a)\) pairs.

Each edge is defined as: \((s, \text{succ}(s,a))\) for every \( s \) in \( S \) and every action \( a \) in \( A \) edge cost is given by \( c(s,a) \)
Consider a path planning example

Assume imperfect action execution and full knowledge of the state (i.e., perfect localization)

Let’s assume 50% chance of ending up on the left and 50% ending up on the right
Consider a path planning example

Assume **imperfect action execution** and full knowledge of the state (i.e., perfect localization)

**MDP:** Defined as \( \{S, A, T, C\} \), where \( S \) – set of states, \( A \) – set of actions, \( T(s,a,s') \) – \( \text{Prob}(s'|s, a) \). \( C \) – costs of all \( (s, a) \) pairs.
Graph vs. MDP vs. POMDP

- Consider a path planning example

- Assume imperfect action execution and full knowledge of the state (i.e., perfect localization)

**What is an optimal policy here?**

**MDP:**
*Defined as \( \{S, A, T, C\} \), where \( S \) – set of states, \( A \) – set of actions, \( T(s,a,s') \) - \( \text{Prob}(s'|s, a) \), \( C \) – costs of all \( (s,a) \) pairs*
• Consider a path planning example
• Assume **imperfect action execution** and full knowledge of the state (i.e., perfect localization)

**Graph vs. MDP vs. POMDP**

What is an optimal policy here?

**MDP:** Defined as \{S, A, T, C\}, where S – set of states, A – set of actions, \(T(s,a,s')\) - \(\text{Prob}(s' | s, a)\), C – costs of all (s,a) pairs
Graph vs. MDP vs. POMDP

- Consider a path planning example
- Assume **imperfect action execution** and full knowledge of the state (i.e., perfect localization)

**MDP (rewards version):**

Defined as \( \{ S, A, T, R \} \), where \( S \) – set of states, \( A \) – set of actions, \( T(s,a,s') \) - \( \text{Prob}(s' | s, a) \), \( R \) – rewards for all \( (s,a) \) pairs
**Graph vs. MDP vs. POMDP**

- Consider a path planning example

- Assume imperfect action execution and **partial observability of the state** (i.e., imperfect localization)

**POMDP:**

Let’s assume

*UAV initially knows it is at \( S_0 \)*

*During execution: it can only sense adjacent obstacles and being at goal*

*After taking this action, UAV doesn’t know whether it is at state \( S_1 \) or \( S_2 \)*
Consider a path planning example.

What is an optimal policy here?

Assume imperfect action execution and partial observability of the state (i.e., imperfect localization).

Graph vs. MDP vs. POMDP

Let’s assume:
UAV initially knows it is at $S_0$.
During execution: it can only sense adjacent obstacles and being at goal.

After taking this action, UAV doesn’t know whether it is at state $S_1$ or $S_2$.

POMDP:
Consider a path planning example

Assume imperfect action execution and **partial observability of the state** (i.e., imperfect localization)

**POMDP:** \(\{S, A, T, R, \Omega, O\}\), where \(S, A, T, R\) (or \(C\)) – same as in MDP, \(\Omega\) – set of all possible observation vectors \(o\), \(O(s',a,o) = \text{Prob}(o|s',a)\) – probability of seeing \(o\) after executing action \(a\) and ending up at state \(s'\)

Let’s assume
UAV initially knows it is at \(S_0\)
During execution: it can only sense adjacent obstacles and being at goal

After taking this action, UAV doesn’t know whether it is at state \(S_1\) or \(S_2\)
Graph vs. MDP vs. POMDP

- Consider a path planning example
- Assume imperfect action execution and **partial observability of the state** (i.e., imperfect localization)

**POMDP:** \(\{S, A, T, R, \Omega, O\}\), where \(S, A, T, R\) (or \(C\)) – same as in MDP, \(\Omega\) – set of all possible observation vectors \(o\), \(O(s’,a,o) – \text{Prob}(o|s’,a)\) – probability of seeing \(o\) after executing action \(a\) and ending up at state \(s’\)
Graph vs. MDP vs. POMDP

Example of POMDP problems where the robot knows its own pose perfectly (perfect localization)?

Assume imperfect action execution and partial observability of the state (i.e., imperfect localization)

POMDP: \( \{S, A, T, R, \Omega, O\} \), where \( S, A, T, R \) (or \( C \)) – same as in MDP, \( \Omega \) – set of all possible observation vectors \( o \), \( O(s',a,o) \) – \( \text{Prob}(o|s',a) \) probability of seeing \( o \) after executing action \( a \) and ending up at state \( s' \)
Belief State Space

- **Belief state** $b$: Probability distribution over the states the robot believes it is currently in

**POMDP**: $\{S, A, T, R, \Omega, O\}$, where $T(s,a,s') = P(s'|s,a)$, $R(s,a)$, $O(s',a,o) = \text{Prob}(o|s',a)$
Belief State Space

- **Belief state** $b$: Probability distribution over the states the robot believes it is currently in.

  - $b$ – a vector of size $N$ (# of states in $S$)
  - $\Sigma^N b_i = 1$, and $b_i \geq 0$ for all $i$

  - Suppose the robot knows it is initially in $s_0$.
    - Then initial $b = [1, 0, 0, 0, 0, 0, 0, 0]^T$. That is, $P(s_0) = 1$

**POMDP**: $\{S, A, T, R, O, \Omega\}$, where $T(s,a,s') = P(s'|s,a)$, $R(s,a)$, $O(s',a,o) = \text{Prob}(o|s',a)$
Belief State Space

- **Belief state** $b$: Probability distribution over the states the robot believes it is currently in

$$b = \text{a vector of size } N \text{ (# of states in } S) \text{ } \Sigma^N b_i = 1, \text{ and } b_i \geq 0 \text{ for all } i$$

Suppose the robot knows it is initially in $s_0$. Then initial $b = [1,0,0,0,0,0,0,0]^T$. That is, $P(s_0) = 1$

What is $b$ after robot takes the 1st action?

**Causal relationship**

**POMDP:** $\{S, A, T, R, \Omega, O\}$, where $T(s,a,s') = P(s'|s,a)$, $R(s,a)$, $O(s',a,o) = \text{Prob}(o|s',a)$
Belief State Space

- **Belief state** $b$: Probability distribution over the states the robot believes it is currently in.

**POMDP**: $\{S, A, T, R, \Omega, O\}$, where $T(s,a,s') = P(s'|s,a)$, $R(s,a)$, $O(s',a,o) = \text{Prob}(o|s',a)$
Belief State Space

- **Belief state** \( b \): Probability distribution over the states the robot believes it is currently in

\[
b'(s') = P(s'|b, a, o) = \frac{O(s', a, o) \sum_s T(s, a, s') \cdot b(s)}{P(o|b, a)}
\]

Here how outcome beliefs are computed

**Belief State Space**
(for \( K \) actions, \( M \) possible observations)

**POMDP:** \( \{S, A, T, R, \Omega, O\} \), where \( T(s, a, s') = P(s'|s, a) \), \( R(s, a) \), \( O(s', a, o) = \text{Prob}(o|s', a) \)
**Belief State Space**

- **Belief state** $b$: Probability distribution over the states the robot believes it is currently in

  \[ b': P(s'|b,a,o) \text{ for every } s' \text{ in } S; \]

  \[ b'(s') = P(s'|b,a,o) = \frac{O(s',a,o) \sum_s [T(s,a,s') * b(s)]}{P(o|b,a)} \]

**Derivation:**

\[ P(s'|b,a,o) = \frac{P(o|b,a,s') P(s'|b,a)}{P(o|b,a)} = \frac{P(o|s',a) \sum_s [P(s'|s,a) * P(s)]}{P(o|b,a)} \]

**POMDP:** $\{S, A, T, R, O, O\}$, where $T(s,a,s') = P(s'|s,a)$, $R(s,a)$, $O(s',a,o) = \text{Prob}(o|s',a)$
Belief State Space

- **Belief state** $b$: Probability distribution over the states the robot believes it is currently in.

**What is Belief State Space?**

It is MDP!
We just need to compute transition probabilities $\tau(b,a,b') = P(b'|b,a)$ and reward function $\rho(b,a)$.

**Belief State Space**
(for K actions, M possible observations)

**POMDP:** $\{S, A, T, R, \Omega, O\}$, where $T(s,a,s') = P(s'|s,a)$, $R(s,a)$, $O(s',a,o) = \text{Prob}(o|s',a)$.
**Belief State Space**

- **Belief state** \( b \): Probability distribution over the states the robot believes it is currently in

\[
\tau(b,a,b') = P(b'|b,a) = \sum_{o \text{ leading to } b'} P(o|b,a) = \sum_{s} P(o|s',a) \sum_{s'} P(s'|s,a) b(s)
\]

**POMDP:** \( \{S, A, T, R, \Omega, O\} \), where \( T(s,a,s') = P(s'|s,a) \), \( R(s,a) \), \( O(s',a,o) = \text{Prob}(o|s',a) \)
Belief State Space

- **Belief state** $b$: Probability distribution over the states the robot believes it is currently in

$$\tau(b, a, b') = P(b'|b, a) = \sum_o \text{leading to } b, P(o|b, a) \sum_s \text{leading to } b, \sum_s P(o'|s, a) \sum_s P(s'|s, a) b(s)$$

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

**POMDP**: $\{S, A, T, R, \Omega, O\}$, where $T(s, a, s') = P(s'|s, a)$, $R(s, a)$, $O(s', a, o) = \text{Prob}(o|s', a)$
Belief State Space

- **Belief state** $b$: Probability distribution over the states the robot believes it is currently in

$$\tau(b,a,b') = P(b'|b,a) = \sum_o \text{leading to } b, P(o|b,a) = \sum_o \text{leading to } b, \sum_s P(o|s', a) \sum_s P(s'|s, a) b(s)$$

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

*Belief State Space*
(for K actions, M possible observations)

The size of Belief MDP is infinite 😥

We can even use Value Iteration you studied, can’t we?

So, finding an optimal policy for POMDP = finding an optimal policy for Belief MDP 😊

POMDP

---

Carnegie Mellon University
Belief State Space

- **Belief state** $b$: Probability distribution over the states the robot believes it is currently in
- Popular techniques for solving POMDPs
  - by discretizing belief statespace into a finite # of states [Lovejoy, ‘91]
  - by taking advantage of the geometric nature of value function [Kaelbing, Littman & Cassandra, ‘98]
  - by sampling-based approximations [Pineau, Gordon & Thrun, ‘03]

**Belief State Space**
(for $K$ actions, $M$ possible observations)

**POMDP**: $\{S, A, T, R, \Omega, O\}$, where $T(s,a,s’) = P(s’|s,a)$, $R(s,a)$, $O(s’,a,o) = \text{Prob}(o|s’,a)$
What You Should Know…

- What problems should be modeled as planning on Graphs vs. MDPs vs. POMDPs
- How POMDPs can be transformed into a Belief MDP
- How to plan in Belief MDP