QUANTUM POMDPs

Jenny Barry

6.845 Final Project Presentation
December 12, 2012
**Robots...**

- Don’t know where they are.
- Don’t know what they are doing.
- Don’t understand what they are seeing.
**Partially Observable Markov Decision Process (POMDP)**

- $S$, $A$, $\Omega$: Possible states, actions, observations
**PARTIALLY OBSERVABLE MARKOV DECISION PROCESS (POMDP)**

- **S, A, Ω**: Possible states, actions, observations
**POMDPs**

**States:** $(i, j)$

**Actions:** $(L, R, U, D, S)$

**Observations:** No Bump, Bump

---

**PARTIALLY OBSERVABLE MARKOV DECISION PROCESS (POMDP)**

- $S, A, \Omega$: Possible states, actions, observations
POMDPs

**States:** \((i, j)\)

**Actions:** \((L, R, U, D, S)\)

**Observations:** No Bump, Bump

**Partially Observable Markov Decision Process (POMDP)**

- \(S, A, \Omega\): Possible states, actions, observations
POMDPs

States: (i, j)
Actions: (L, R, U, D, S)
Observations: No Bump, Bump
Rewards: 0 at ✺, -1 else

Partially Observable Markov Decision Process (POMDP)
- $S, A, \Omega$: Possible states, actions, observations
- $R(s_i, a_j)$: Reward for taking action $a_j$ in state $s_i$
POMDPs

Partialy Observable Markov Decision Process (POMDP)

- $S, A, \Omega$: Possible states, actions, observations
- $R(s_i, a_j)$: Reward for taking action $a_j$ in state $s_i$
- $T(s_i | a_j, s_k)$: Probability of transitioning to $s_i$ starting in $s_j$ taking action $a_j$

States: $(i, j)$
Actions: $(L, R, U, D, S)$
Observations: No Bump, Bump
Rewards: 0 at ⭐, -1 else
**POMDPs**

**States:** (i, j)

**Actions:** (L, R, U, D, S)

**Observations:** No Bump, Bump

**Rewards:** 0 at ★, -1 else

---

**Partially Observable Markov Decision Process (POMDP)**

- **S, A, Ω:** Possible states, actions, observations
- **R(s_i, a_j):** Reward for taking action \(a_j\) in state \(s_i\)
- **T(s_i|a_j, s_k):** Probability of transitioning to \(s_i\) starting in \(s_j\) taking action \(a_j\)
- **O(o_i|a_j, s_k):** Probability of observing \(o_i\) given that action \(a_j\) ended in \(s_k\)
**Belief States**

**Definition: Belief State**

POMDP $P = \langle S, A, \Omega, R, T, O \rangle \Rightarrow$ Belief space $B \subset \mathbb{R}^{|S|}$:

- $\vec{b}_i = \Pr(s_i)$
- $\sum_i \vec{b}_i = |\vec{b}|_1 = 1$
**Belief States**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>2</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>3</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>4</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>5</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
</tr>
</tbody>
</table>

**Definition: Belief State**

POMDP $P = \langle S, A, \Omega, R, T, O \rangle \Rightarrow$ Belief space $B \subset \mathbb{R}^{|S|}$:

- $\vec{b}_i = \text{Pr}(s_i)$
- $\sum_i b_i = |\vec{b}|_1 = 1$
**Belief States**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>2</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>3</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>4</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>5</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
</tr>
</tbody>
</table>

**Move Right**

**See No Bump**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0</td>
</tr>
</tbody>
</table>

**Definition: Belief State**

POMDP $P = \langle S, A, \Omega, R, T, O \rangle \Rightarrow$ Belief space $B \subset \mathbb{R}^{|S|}$:

- $\vec{b}_i = \Pr(s_i)$
- $\sum_i \vec{b}_i = |\vec{b}|_1 = 1$
**Belief States**

**Definition: Belief State**

POMDP $P = \langle S, A, \Omega, R, T, O \rangle \Rightarrow$ Belief space $B \subset \mathbb{R}^{|S|}$:

- $\vec{b}_i = \text{Pr}(s_i)$
- $\sum_i \vec{b}_i = |\vec{b}|_1 = 1$

**Belief Markov Decision Process**

- $B$: Belief space (continuous)
- $A$: Robot’s actions
- $\tau(\vec{b}'|a_i, \vec{b})$: Probability of $\vec{b}'$ after taking action $a_i$ in state $\vec{b}$.
- $\rho(\vec{b}, a_i) = \sum_i \vec{b}_i R(s_i, a_i)$: Reward for taking action $a_i$ in state $\vec{b}$
- $b_0$: Starting belief state

*I know that I know nothing.* - Socrates
**Definition: Superoperator**

\[ S = \{K_1, \ldots, K_m\} \]

- \[ \sum_{i=1}^{m} K_i^\dagger K_i = \mathbb{I} \]
- \[ \text{Pr}[\text{Observation } i] = \text{Tr}(K_i \rho K_i^\dagger) \]

\[ \rho \rightarrow \frac{K_i \rho K_i^\dagger}{\text{Tr}(K_i \rho K_i^\dagger)} \]
**QOMDPs**

**Definition: Superoperator**

\[ S = \{K_1, \ldots, K_m\} \]

\[ \sum_{i=1}^{m} K_i^\dagger K_i = \mathbb{I} \]

\[ \Pr[\text{Observation } i] = \text{Tr}(K_i \rho K_i^\dagger) \]

\[ \rho \rightarrow \frac{K_i \rho K_i^\dagger}{\text{Tr}(K_i \rho K_i^\dagger)} \]

---

**Quantum Observable Markov Decision Process (QOMDP)**

- **S**: Hilbert space
- **Ω**: Set of observations
- **A**: Set of quantum superoperators
- **R**: Reward function
- **ρ₀**: Starting state
POMDPs are hard...

Strategy:

1. Localize: go right until wall, then up
2. Go to goal
POMDPs are Hard...

\[ \pi(\vec{b}) = \begin{cases} 
R & \text{if } \sum_{s \in \text{Right Wall}} \vec{b}(s) < 1 \\
U & \text{if } \sum_{s \in \text{Right Wall}} \vec{b}(s) = 1 \\
& \text{and } \vec{b}(\text{Upper Right}) < 1 \\
\text{Go to goal} & \text{if } ||\vec{b}||^2 = 1 
\end{cases} \]

**Policy**: \( \pi(\vec{b}, t) = a \) specifies action to take in belief \( \vec{b} \) at time \( t \)
POMDPs are Hard...

**Policy**: \( \pi(\vec{b}, t) = a \) specifies action to take in belief \( \vec{b} \) at time \( t \)

**Policy Existence Problem (PEP)**

Given POMDP \( P = \langle S, A, \Omega, R, T, O \rangle \), decide if there is some policy \( \pi \) that has expected future reward at least \( V \) over the next \( h \) timesteps.

- If \( h = \text{poly}(S) \), PEP is in PSPACE and **PSPACE-Complete**.
- If \( h = \infty \), PEP is **Undecidable**.
...but QOMDPs are harder

\[ \text{POMDPs} \subseteq \text{QOMDPs} \]

- PEP with \( h = \text{poly}(d) \) is at least PSPACE-Complete
- \( \checkmark \) PEP with \( h = \infty \) is UNDECIDABLE
...but QOMDPs are Harder

POMDPs ⊆ QOMDPs

✓ PEP with $h = \text{poly}(d)$ is **PSPACE-Complete**

✓ PEP with $h = \infty$ is **UNDECIDABLE**

**Theorem**

PEP for QOMDPs with $h = \text{poly}(d)$ is in PSPACE.

**Proof Sketch:** There are only $O((|A||\Omega|)^h)$ policies. Try them all.
...but QOMDPs are harder

POMDPs ⊆ QOMDPs

✓ PEP with $h = \text{poly}(d)$ is PSPACE-COMPLETE
✓ PEP with $h = \infty$ is UNDECIDABLE

**Goal-State Reachability Problem (GRP)**

Assume the Q(P)OMDP has an absorbing goal state. Decide if there is a policy that reaches this goal state with probability 1.
...BUT QOMDPs ARE HARDER

POMDPs ⊆ QOMDPs

✓ PEP with $h = \text{poly}(d)$ is **PSPACE-COMPLETE**
✓ PEP with $h = \infty$ is **UNDECIDABLE**

- GRP is **DECIDABLE** for POMDPs
- GRP is **UNDECIDABLE** for QOMDPs

**GOAL-STATE REACHABILITY PROBLEM (GRP)**

Assume the Q(P)OMDP has an absorbing goal state. Decide if there is a policy that reaches this goal state with probability 1.
Given a superoperator $S = \{K_1, \ldots, K_m\}$ and starting state $\rho_0$, decide if there is some finite sequence of measurements that can never be observed if $\rho_0$ is continually fed back into $S$.

**QMOP is UNDECIDABLE** [Eisert12]
Given QMOP $S = \{K_1, \ldots, K_m\}$:

- $m$ actions. Action $i$ either:
  - Transitions according to $K_i$
  - Transitions to goal state

- $m + 1$ observations:
  - At-Goal
  - Observation $i$ from QMOP
Given QMOP $S = \{K_1, ..., K_m\}$:

- $m$ actions. Action $i$ either:
  - Transitions according to $K_i$
  - Transitions to goal state
- $m + 1$ observations:
  - 1 At-Goal
  - 2 Observation $i$ from QMOP
Given QMOP $S = \{K_1, \ldots, K_m\}$:

- $m$ actions. Action $i$ either:
  - Transitions according to $K_i$
  - Transitions to goal state

- $m + 1$ observations:
  - At-Goal
  - Observation $i$ from QMOP

$$\Pr(\rho_n \neq \text{goal} \mid \text{actions } j_1, \ldots, j_n) = \Pr(\text{Observing sequence } j_1, \ldots, j_n).$$

$\Rightarrow$ Path to goal of probability 1 if and only some sequence unobservable.
**Theorem**

GRP for QOMDPs is undecidable.
Goal-State Reachability for POMDPs

Conversion to Plus/Zero Land

\[
\begin{bmatrix}
0.2 & 0 & 0.8 \\
0.3 & 0.1 & 0.6 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
1 \\
0 \\
0
\end{bmatrix}
\rightarrow
\begin{bmatrix}
+ & 0 & + \\
+ & + & + \\
0 & 0 & +
\end{bmatrix}
\begin{bmatrix}
+ \\
0 \\
0
\end{bmatrix}
= \begin{bmatrix}
+ \\
0 \\
0
\end{bmatrix}
\]
**GOAL-STATE REACHABILITY FOR POMDPs**

## Conversion to Plus/Zero Land

\[
\begin{bmatrix}
0.2 & 0 & 0.8 \\
0.3 & 0.1 & 0.6 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
1 \\
0 \\
0
\end{bmatrix}
\rightarrow
\begin{bmatrix}
+ & 0 & + \\
+ & + & + \\
0 & 0 & +
\end{bmatrix}
\begin{bmatrix}
+ \\
0 \\
0
\end{bmatrix}
= \begin{bmatrix}
+ \\
0 \\
0
\end{bmatrix}
\]

- Convert POMDP probabilities to plus/zero
- Finitely many \((2^{|S|} - 1)\) states
- Finitely many policies

\(\Rightarrow\) We find the goal state or repeat a previously seen state in finite time.
**Goal-State Reachability for POMDPs**

**Conversion to Plus/Zero Land**

\[
\begin{bmatrix}
0.2 & 0 & 0.8 \\
0.3 & 0.1 & 0.6 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
1 \\
0 \\
0
\end{bmatrix}
\rightarrow
\begin{bmatrix}
+ & 0 & + \\
+ & + & + \\
0 & 0 & +
\end{bmatrix}
\begin{bmatrix}
+ \\
0 \\
0
\end{bmatrix}
=\begin{bmatrix}
+ \\
0 \\
0
\end{bmatrix}
\]

- Convert POMDP probabilities to plus/zero
- Finitely many \((2^{|S|} - 1)\) states
- Finitely many policies

⇒ We find the goal state or repeat a previously seen state in finite time.

**Theorem**

GRP for POMDPs is decidable.
**FUTURE WORK**

**COMPLEXITY PROBLEMS**

- Complexity separations using non-negative properties of POMDPs
- Complexity separations using value function structure of POMDPs
- What if we don’t know the starting state in a QOMDP?
FUTURE WORK

COMPLEXITY PROBLEMS
- Complexity separations using non-negative properties of POMDPs
- Complexity separations using value function structure of POMDPs
- What if we don’t know the starting state in a QOMDP?

ALGORITHMS
- Algorithms for solving QOMDPs
- Algorithms for approximating QOMDPs
FUTURE WORK

COMPLEXITY PROBLEMS

- Complexity separations using non-negative properties of POMDPs
- Complexity separations using value function structure of POMDPs
- What if we don’t know the starting state in a QOMDP?

ALGORITHMS

- Algorithms for solving QOMDPs
- Algorithms for approximating QOMDPs

APPLICATIONS

- Reward structure for QOMDPs
- Practical applications of QOMDPs