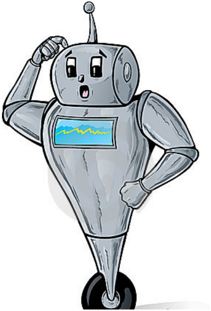
QUANTUM POMDPS

Jenny Barry

6.845 Final Project Presentation
December 12, 2012

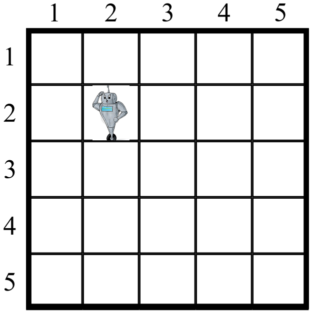
POMDPs



Robots...

- Don't know where they are.
- Don't know what they are doing.
- Don't understand what they are seeing.

POMDPs

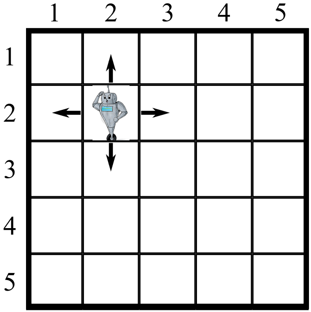


States: (i, j)

Partially Observable Markov Decision Process (POMDP)

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

POMDPs

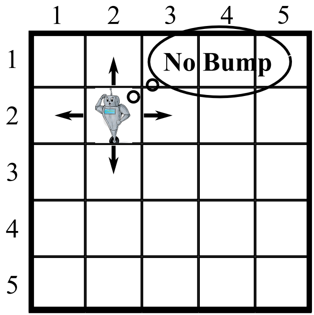


States: (i, j)
Actions: (L, R, U, D, S)

Partially Observable Markov Decision Process (POMDP)

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

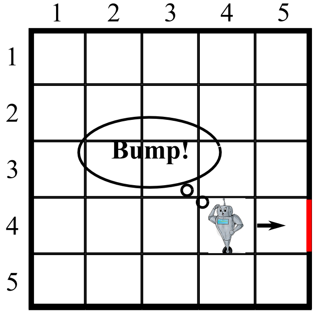
POMDPs



States: (i, j)
Actions: (L, R, U, D, S)
Observations: No Bump, Bump

Partially Observable Markov Decision Process (POMDP)

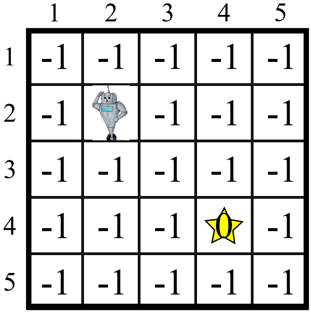
- S, A, Ω: Possible states, actions, observations



States: (i, j)
Actions: (L, R, U, D, S)
Observations: No Bump, Bump

Partially Observable Markov Decision Process (POMDP)

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

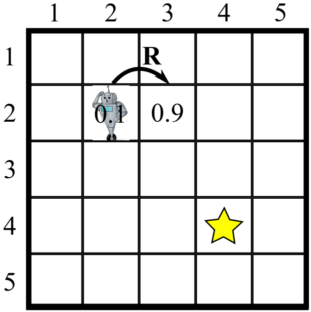


States: (i, j)
Actions: (L, R, U, D, S)
Observations: No Bump, Bump
Rewards: 0 at ☆, -1 else

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

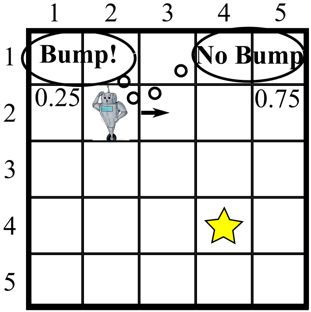
POMDPs



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

Partially Observable Markov Decision Process (POMDP)

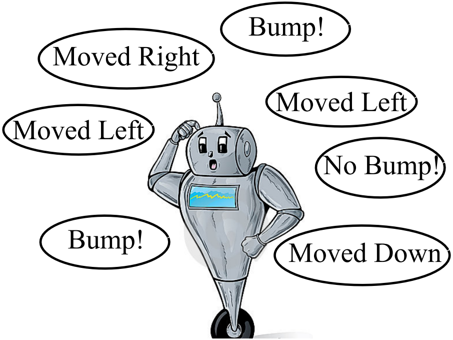
- S, A, Ω: Possible states, actions, observations
- $R(s_i, a_j)$: Reward for taking action a_j in state s_i
- $T(s_i|a_j, s_k)$: Probability of transitioning to s_i starting in s_j taking action a_j



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



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

	1	2	3	4	5
1	0.04	0.04	0.04	0.04	0.04
2	0.04	0.04	0.04	0.04	0.04
3	0.04	0.04	0.04	0.04	0.04
4	0.04	0.04	0.04	0.04	0.04
5	0.04	0.04	0.04	0.04	0.04

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

	1	2	3	4	5
1	0.04	0.04	0.04	0.04	0.04
2	0.04	0.04	0.04	0.04	0.04
3	0.04	0.04	0.04	0.04	0.04
4	0.04	0.04	0.04	0.04	0.04
5	0.04	0.04	0.04	0.04	0.04

Move Right

See No Bump

	1	2	3	4	5
1	0	0.07	0.07	0.07	0
2	0	0.07	0.07	0.07	0
3	0	0.07	0.07	0.07	0
4	0	0.07	0.07	0.07	0
5	0	0.07	0.07	0.07	0

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 $B \subset \mathbb{R}^{|S|}$:

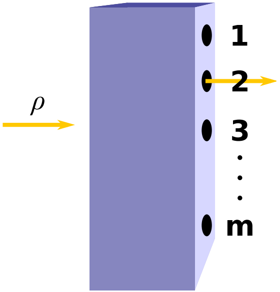
- $\vec{b}_i = \Pr(s_i)$
- $\sum_i \vec{b}_i = |\vec{b}|_1 = 1$

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

QOMDPs



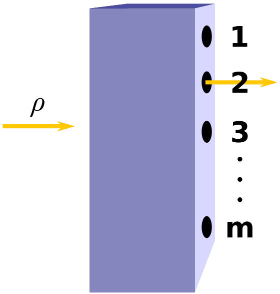
Definition: Superoperator

$\mathbf{S} = \{K_1, ..., K_m\}$

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

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

QOMDPs



Definition: Superoperator

$\mathbf{S} = \{K_1, \ldots, K_m\}$

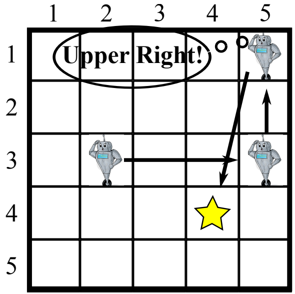
- $\sum_{i=1}^m K_i^\dagger K_i = \mathbb{I}$
- $\Pr[\text{Observation } i] = \mathrm{Tr}(K_i \rho K_i^\dagger)$

$$\rho \to \frac{K_i \rho K_i^\dagger}{\mathrm{Tr}(K_i \rho K_i^\dagger)}$$

Quantum Observable Markov Decision Process (QOMDP)

- S: Hilbert space
- Ω: Set of observations
- $\mathscr{A}$: Set of quantum superoperators
- R: Reward function
- ρ_0: Starting state

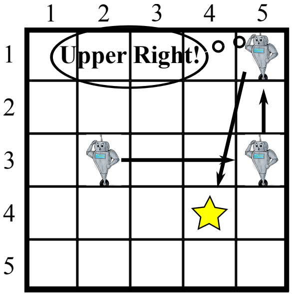
POMDPs are Hard...



Strategy:

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

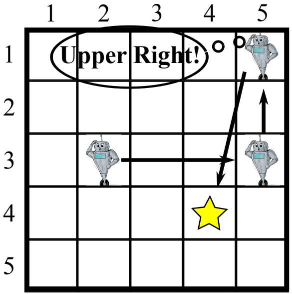
POMDPs are Hard...



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

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

POMDPs are Hard...



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

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

Policy Existence Problem (PEP)

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

- If $h =$poly(S), PEP is in PSPACE and **PSPACE-Complete**.
- If $h = \infty$, PEP is **undecidable**.

...BUT QOMDPS ARE HARDER

POMDPs $\subseteq$ QOMDPs

- PEP with $h = $poly$(d)$ is at least PSPACE-Complete
- ✓ PEP with $h = \infty$ is **UNDECIDABLE**

...but QOMDPs are Harder

POMDPs $\subseteq$ 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.

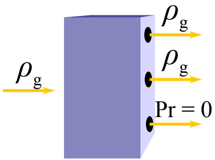
Proof Sketch: There are only $O((|A||\Omega|)^h)$ policies. Try them all.

POMDPs $\subseteq$ 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.



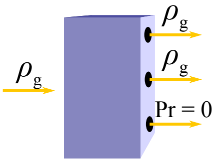
...BUT QOMDPS ARE HARDER

POMDPs $\subseteq$ QOMDPs

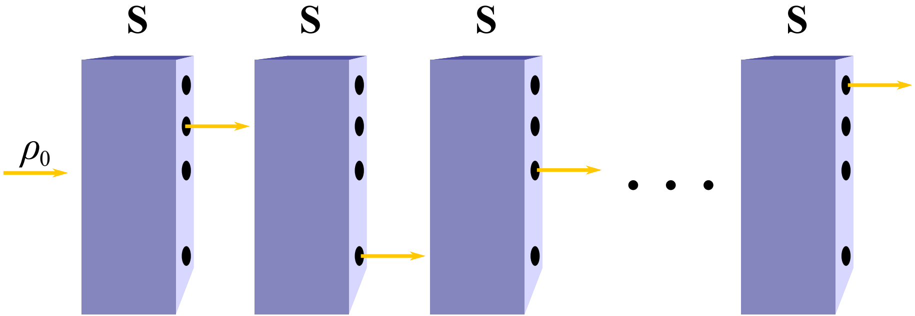
- ✓ PEP with $h = \text{poly}(d)$ is **PSPACE-COMPLETE**
- ✓ PEP with $h = \infty$ is **UNDECIDABLE**
- GRP is **DECIDABLE** for POMDPs
- GRP is **UNDECIDABLE** for QOMDPs

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.



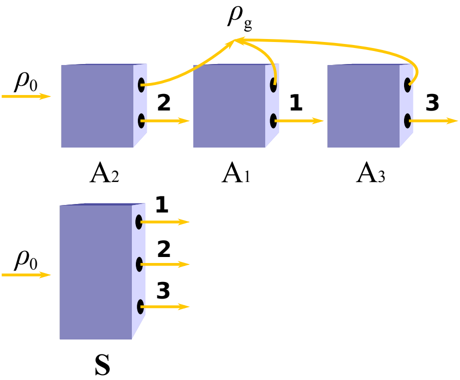
GRP for QOMDPs: QMOP



Quantum Measurement Occurrence Problem (QMOP)

Given a superoperator $\mathbf{S} = \{K_1, ..., K_m\}$ and starting state ρ_0, decide if there is some finite sequence of measurements that can never be observed if ρ_0 is continually fed back into $\mathbf{S}$.

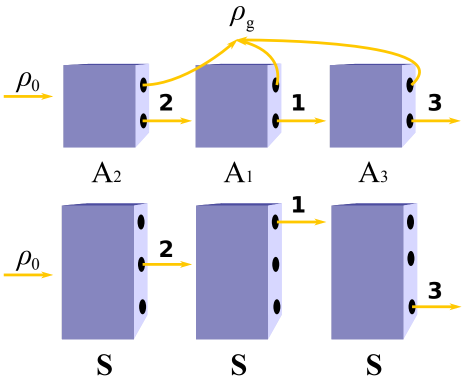
QMOP is **UNDECIDABLE** [Eisert12]



Given QMOP $\mathbf{S} = \{K_1, \ldots, K_m\}$:

- 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

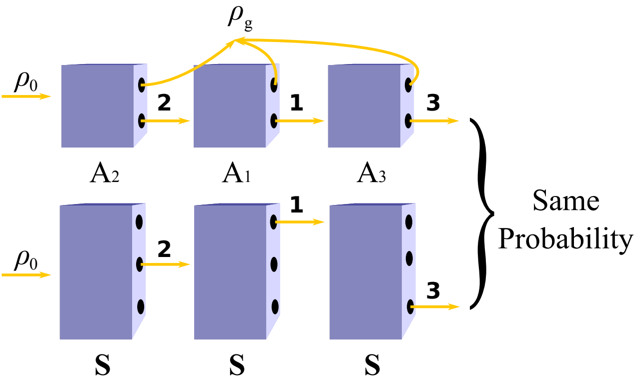
GRP for QOMDPs: Reduction from QMOP



Given QMOP $\mathbf{S} = \{K_1, \ldots, K_m\}$:

- 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

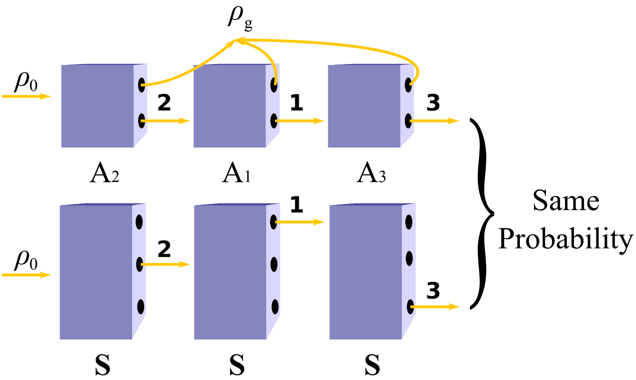
GRP for QOMDPs: Reduction from QMOP



Given QMOP $\mathbf{S} = \{K_1, \ldots, K_m\}$:

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

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



Theorem

GRP for QOMDPs is undecidable.

CONVERSION TO PLUS/ZERO LAND

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

Conversion to Plus/Zero Land

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

- 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.

Conversion to Plus/Zero Land

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

- 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.

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