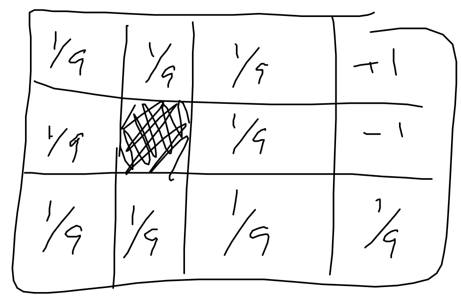
Partially-Observable MDPs

Monday, November 7, 2016 1:33 PM

Partially Observable MDPs

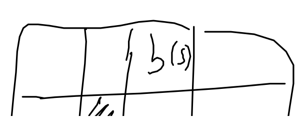
- What if your sensors are unreliable?
- You don't know what state you are in and thus don't know which action to perform, $\pi(s)$.
- Belief state: probability distribution over the states
  - b = <1/9, 1/9, 1/9, 1/9, …>
  - b(s) is the probability of being in state s
- Sensor model: $P(e \mid s)$
  - Probability that you receive observation e given you are in state s



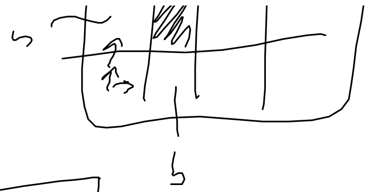
- **Agent update:** $b'(s') = \alpha P(e \mid s') \sum_s P(s' \mid s, a) b(s)$
  - Shorthand: b' = FORWARD(b, a, e)
- **If you had a policy:**
  - Given b, execute a = $\pi(b)$
  - Receive percept e
  - Set b' = FORWARD(b, a, e)
- **Problem:** MDP creates $\pi(s)$, but POMDP requires $\pi(b)$

- **Convert POMPD to a MDP**

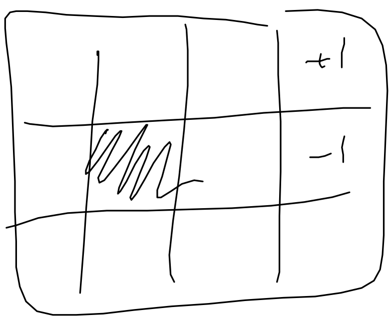
- Transition function over states $T(s, a, s') = P(s' \mid s, a)$
- State reward function: $R(s)$



- Belief reward function: $\rho(b) = \sum_s b(s)\, R(s)$



- Belief transition function: $T(b, a, b') = P(b' \mid b, a)$
  - $P(b' \mid b, a) = P(b' \mid a, b) = \sum_e P(b' \mid e, a, b)\, \boxed{P(e \mid a, b)}$
    $= \sum_e P(b' \mid e, a, b) \sum_{s'} P(e \mid s') \sum_s P(s' \mid s, a)\, b(s)$
  - $P(b' \mid e, a, b) = 1$ if $b' = \text{FORWARD}(b, a, e)$
    or 0 otherwise



9 states

s

Approximation Technique