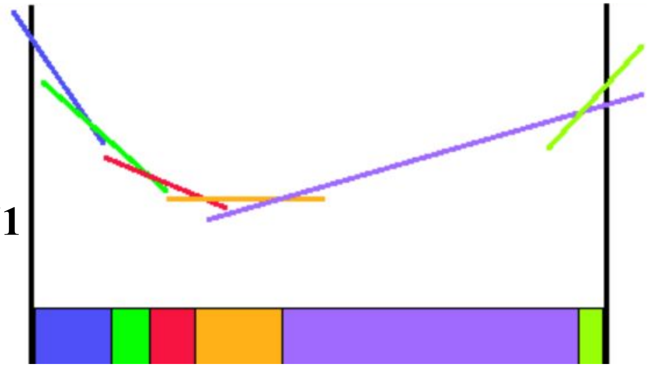
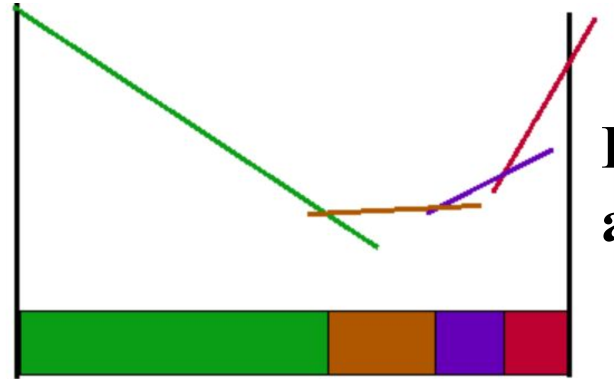
16-782
Planning & Decision-making in Robotics
Planning under Uncertainty: Partially Observable Markov Decision Processes (POMDP) (cont.)

Alex LaGrassa
Robotics Institute
Carnegie Mellon University

Value Iteration (Horizon = 3)

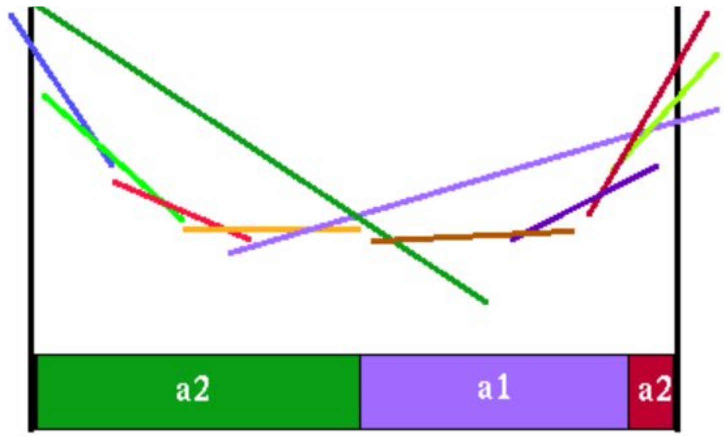
First action: a_1



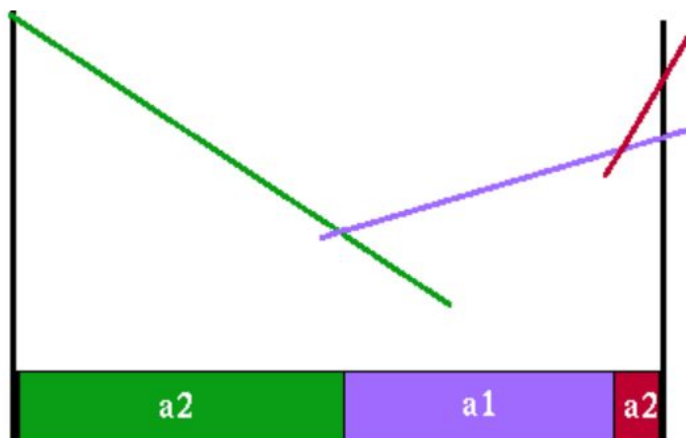


First action: a_2

a_1 and a_2 before pruning



After pruning



Source: Cassandra 1999

Algorithm sketch

Initialize list of plans and α's

while true:

  Compute all strategies

  Update each $\alpha_p(s) = \Sigma_{s'} P(s'|s,a)[R(s,a,s') + \gamma\Sigma_o P(o|s'a)\alpha_{p.o}(s')$

  Remove dominated plans

  If the maximum difference between $V_t(b)$ and $V_{t-1}(b) < \epsilon(\gamma)$:

    break

Return V

Exact POMDP value iteration

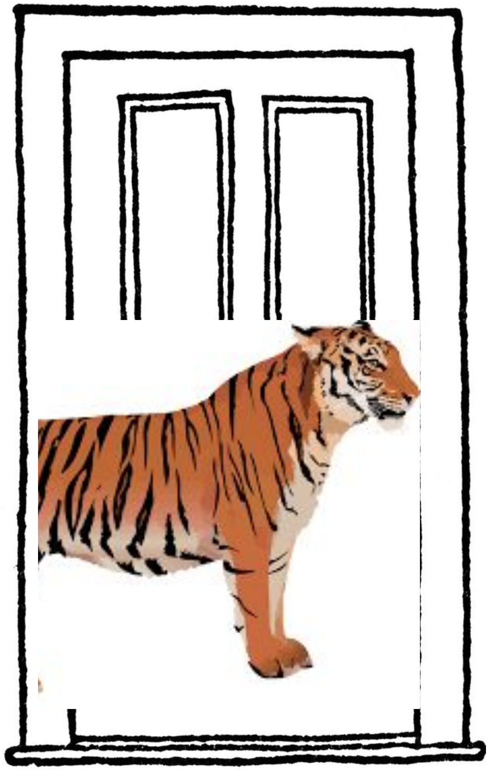
- Value functions remain PWLC
- Value functions over longer horizons do **not** necessarily become more complex
- Can still be quite expensive
  - Generation
  - Pruning

Other methods for solving POMDPs

- Point-based Value Iteration - approximation
- Sampling points from reachable belief space (SARSOP)
- Maintain sparse representation of belief tree online (DESPOT)
- Monte Carlo sampling of states and histories (POMCP)

Generally difficult to do long-horizon planning with POMDPs

Tiger problem



-100

+10

States:

s_l, s_r

Actions:

left

right

listen

Transition model:

Either left or right results in reset

s_l:0.5 s_r:0.5

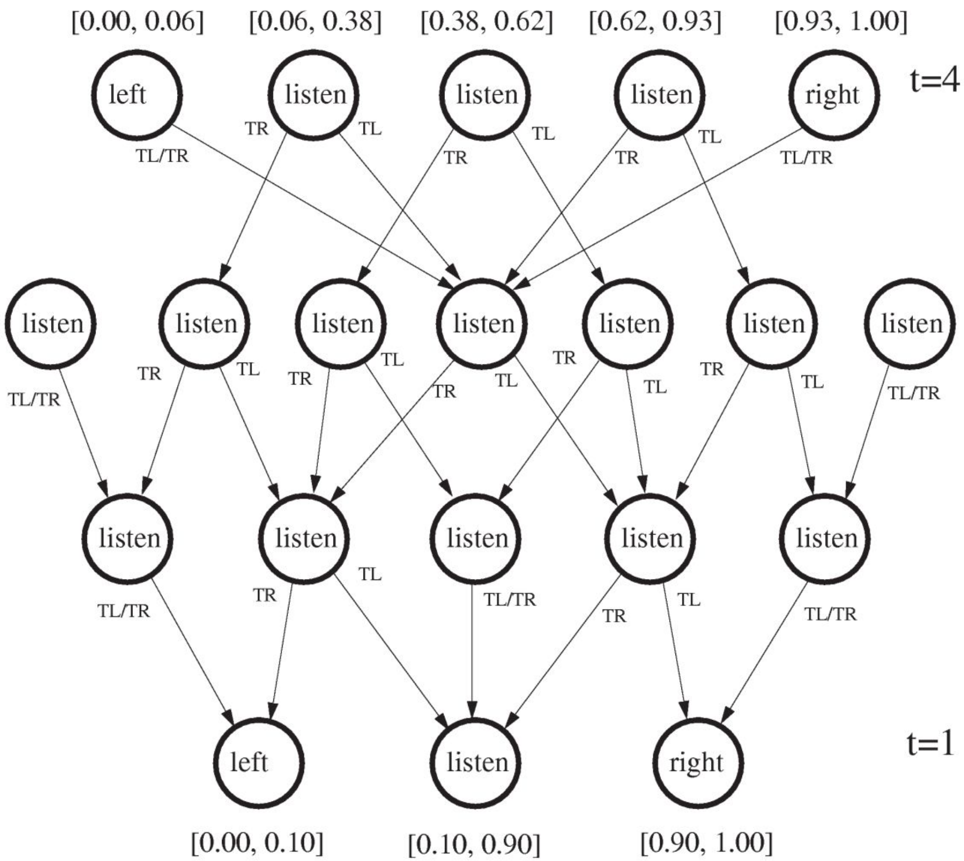
Observations:

TL, TR

$P(TL \mid s_l) = 0.85$

converse for s_r

Tiger problem: policy structure for horizon=3

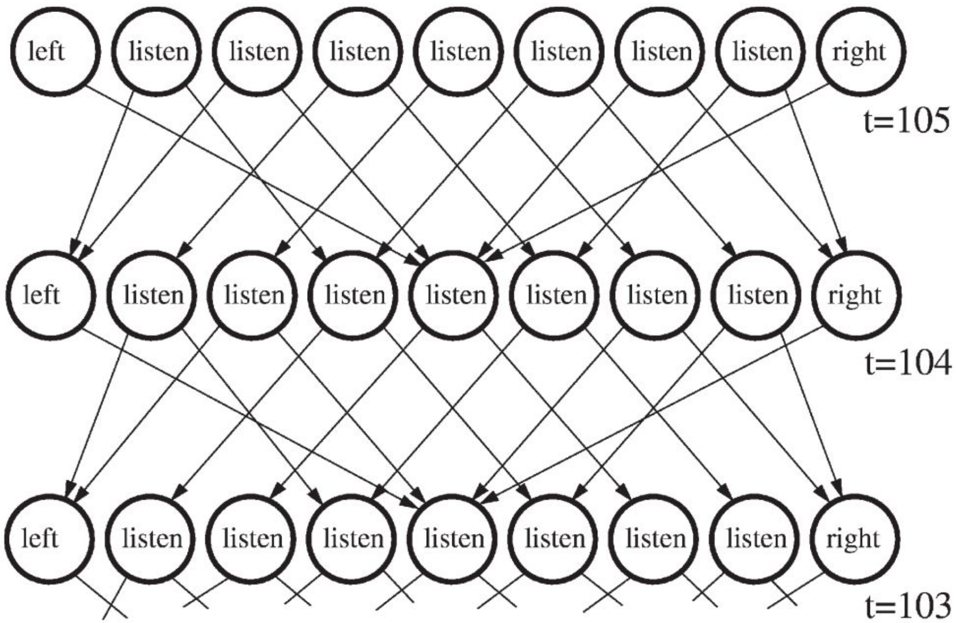


- Open door if fairly certain
- Q: no arrows into 2 nodes at t=3 Why?

- Most sets of observations end in opening a door for the optimal policy

L.P. Kaelbling et al. *Planning and acting in partially observable stochastic domains*. 1998

Tiger problem: policy structure for long horizon



- For $0< \gamma <1$ future rewards are less important
- What is the policy?
- Optimal policy is stationary

L.P. Kaelbling et al. *Planning and acting in partially observable stochastic domains*. 1998

Summary

- The finite-horizon value function is PWLC
- POMDPs can be solved exactly in some cases
  - Finite horizon
  - Not too many actions/observations
- Problem structure can be exploited