

RL and Planning Under Uncertainty (ANU, Sem2, 2008)
Final Tutorial: Review Questions
Tutorial Instructor: Scott Sanner

Questions:

1. Prove that optimal LP solution does yield V^* :

Variables: V^*

Minimize: $\|V^*\|_1$

Subject to: $0 \geq R_a + \gamma T_a V^* - V^*, \forall a \in A$

2. Assume you have an algorithm to solve any well-defined one-shot partially observed stochastic game. Describe how to solve any well-defined finite horizon POSG using this one-shot algorithm (assume that after each round of multiagent actions is taken, a single observation is emitted). What is a caveat of using such an approach in practice?
3. Justify the eligibility trace update for function approximation. Recall that eligibilities are defined on parameters θ as opposed to states in this case:

$$\vec{e}^{t+1} = \gamma \lambda \vec{e}^t + \nabla_{\vec{\theta}^t} V_{\vec{\theta}^t}(s^t)$$

Hint: when function approximation is exact, want to show that can recover the original eligibility trace formula for state-based eligibilities. What was the formula for state-based eligibilities?

4. Formalize an optimal solution to the following preference elicitation problem:
- User has an (unknown) linear utility function over fixed set of integer- or real-valued attributes, e.g. in an airfare setting might have attributes:
(cost, hours in air, number layovers)
 - System has to take one action on behalf of user (e.g., select a ticket to purchase): action leads to fixed distribution over attribute values for which user receives appropriate reward according to their utility function.
 - System wants to maximize user's expected utility.
 - For a fixed cost, system can query user on linear refinements of their utility function: "on which side of hyperplane is your utility function?"
 - Or system can execute final action based on current beliefs at which point user receives their utility for this selection and problem terminates.