

7. MDPs AND REINFORCEMENT LEARNING: (14 points)

Recall that the Bellman equation for the utility of a state in an MDP is

$$U(s) = R(s) + \gamma \max_a \sum_{s'} T(s, a, s') U(s')$$

(a) (6) Sometimes, MDPs are formulated with a reward function $R(s, a, s')$ where the reward depends on the action taken and the resulting outcome state. Rewrite the Bellman equation for this formulation.

(e) (8) Any finite search problem can be translated into a reinforcement learning problem such that the optimal solution to one is also a solution to the other. Explain precisely *how* to go from a traditional state-based search problem to a reinforcement learning problem, AND how to translate the solution back. *Hint: you may find the $R(s, a, s')$ formulation to be useful.*