

# CSE 250A Quiz 9

Thursday December 6, 2012

*Instructions.* You should do this quiz in partnership with exactly one other student. Write both your names at the top of this page. Discuss the answer to the question with each other, and then write your joint answer below the question. It is ok if you overhear other students' discussions, because you still need to decide if they are right or wrong. You have seven minutes.

Let  $\pi$  be a deterministic, stationary policy. The function  $Q^\pi : S \times A \rightarrow \mathbb{R}$  gives the expected total reward achieved by starting in state  $s \in S$ , doing action  $a \in A$ , and then acting according to the policy  $\pi$  afterwards. "Afterwards" means in every state after  $s$ , not including  $s$ . Write down carefully the linear Bellman equation satisfied by  $Q^\pi(s, a)$ .

*Answer.*

$$Q^\pi(s, a) = r(s, a) + \gamma \sum_{s'} p(s'|s, a) Q^\pi(s', \pi(s')).$$

*Additional comments.* The unknowns of this equation are the values  $Q^\pi(s, a)$  for every  $(s, a)$  pair. This equation is linear because on the righthand side, the coefficient of each unknown  $Q^\pi(s', \pi(s'))$  is  $p(s'|s, a)$  which is a known constant.