



Linköping University

TDDC17 ARTIFICIAL INTELLIGENCE
Lab 5: Reinforcement Learning

Robin Andersson (roban591)
Lawrence Thanakumar Rajappa (lawra776)

October 10, 2019

1 Part 2

1. In the report, a) describe your choices of state and reward functions, and b) describe in your own words the purpose of the different components in the Q-learning update that you implemented. In particular, what are the Q-values?

a) The state function is split into ten discrete pieces from angle $-\pi$ to π . This is done to cover all possible angles that the controller can have. The reward function is

$$\frac{\pi}{|\phi|},$$

where ϕ is the current angle of the controller. The reason we chose this formula is to make the angles around π give as low reward as possible and to make angles gain a higher reward the closer they are to zero.

b) When we update the Qtable we use the following formula

$$Q(s, a) = Q(s, a) + \alpha(R(s) + \gamma \max_{a'} Q(s', a') - Q(s, a)),$$

where

- $Q(s, a)$ represents the Q-value of the previous state and action.
- α is calculated using the N-value of the previous state and action.
- $R(s)$ represents the reward from the previous action.
- γ is a constant.
- $Q(s', a')$ represents the Q-value of the new state and a new action where a' is chosen as the action that gives the highest Q-value.