## Advanced AI Techniques

Prof. Dr. Burgard, Prof. Dr. Nebel, Dr. Kersting
M. Ragni, A. Rottmann
University of Freiburg
Department of Computer Science
WS 2006/2007

# Exercise Sheet 6

**Due: Tuesday, 12. December 2006**

### Exercise 6.1

How would the policy iteration algorithm of slide 77 be defined for action values? Give a complete algorithm for computing $Q^*$, analogous to the algorithm of slide 77.

### Exercise 6.2

You plan to go skiing in a skiing region with three slopes: A (red), B (red) and C (blue), and three lifts X (to the top of slope A), Y (to slope B), Z (to slope C). Going down a red slope, you get twice as much pleasure as going down a blue one. From A, you can either run down to lift Y or to lift Z. From B, you can either go to lift X or Y. From C, you can go down to lift X or Z. There is always the option of waiting on top of a slope (without getting any pleasure from it). Assume that you are a bad skier: The transition probabilities are 0.6 for slope B, and 0.75 for slopes A and C. In all other cases, you end in hospital for the rest of the time (each time step in hospital rewards the negative of the pleasure of running down the blue slope).

(a) Draw the MDP diagram of the skiing trip.

(b) What are the state-value Bellmann equations for the random policy that selects either one of the three actions in each state with probability $\frac{1}{3}$ ? (Choose $\gamma = \frac{2}{3}$)

(c) Calculate the state-value functions for each of the states for the random policy. (You do not have to derive the result, you can use an equation solver.)

(d) For one of the states (being at the top of A, or of B, or of C), verify by inserting the corresponding values from the solution (c) that the Bellmann equation is solved.