# Advanced AI Techniques (WS04)

Exercise sheet 13
Deadline: Thursday, 10 Feb 05

**Exercise 1 (4 points)**

*Consider the Hidden Markov Model from last week's exercise no. 12 (which also was the example of the lecture).*

*What is the the optimal belief based policy for the time horizon $T = 2$? By performing the necessary calculations en detail, verify the claimed results for $V_1(b|z_2)$, $\bar{V}_1(b|u_3), and \bar{V}_2(b)$ from the lecture (which were not derived in detail on the lecture slides) and give reference to the formulas you use.*

**Exercise 2 (6 points)**

*Imagine an agent standing in front of two closed doors. Behind one of the doors is a tiger and behind the other is a large reward. If the agent opens the door with the tiger, then a large penalty is received (presumably in the form of some amount of bodily injury). Instead of opening one of the two doors, the agent can listen, in order to gain some information about the location of the tiger. Unfortunately, listening is not free; in addition, it is also not entirely accurate. There is a chance that the agent will hear a tiger behind the left-hand door when the tiger is really behind the right-hand door, and vice versa. Immediately after the agent opens a door and receives a reward or penalty, the problem resets, randomly relocating the tiger behind one of the two doors.*

*The transition and observation models can be described in detail as follows. We refer to the state of the world when the tiger is on the left as sl and when it is on the right as sr . The actions are LEFT, RIGHT, and LISTEN. There are only two possible observations which are given after the LISTEN action: to hear the tiger on the left (TL) or to hear the tiger on the right (TR). The LISTEN action does not change the state of the world. When the world is in state sl, the LISTEN action results in observation TL with probability 0.85 and the observation TR with probability 0.15; conversely for world state sr. The LEFT and RIGHT actions cause a transition to world state sl with probability 0.5 and to state sr with probability 0.5 (essentially resetting the problem). No matter what state the world is in, the LEFT and RIGHT actions result in either observation with probability 0.5. The reward for opening the correct door is +10 and*

*the penalty for choosing the door with the tiger behind it is -100. The cost of listening is -1.*

*Derive the optimal undiscounted finite-horizon policies for the tiger problem.*

*a. Begin with the situation-action mapping for the time horizon $T = 1$ at time step $t = 1$ when the agent only gets to make its single decision. If the agent believes with high probability that the tiger is on the left, then the best action is to open the right door; if it believes that the tiger is on the right, the best action is to open the left door. Calculate the corresponding belief states. What if the agent is highly uncertain about the tiger's location? Is it always better choose a door, or could it the best thing to do listen although there is no chance to use the gained information within the game? Why?*

*b. What are the optimal belief based policies for $T = 2$? and $T = 3$?*

**. (Voluntary:) If you want, you can proceed with $T = 4$, $T = 5$, etc. Do you see regularities?*

**Exercise 3 (2 points)**

*We have always looked at these problems in terms of finite horizon problems? What would change if they are regarded as infinite horizon problems? (Note that the policy depends on your belief and the belief depends on the observations.)*