6.231 Dynamic Programming

Final Exam, Fall 2001

Prof. Dimitri Bertsekas

Problem 1 (15 points)

An energetic salesman works every day of the week. He can work in only one of two towns A and B on each day. For each day he works in town A (or B) his expected reward is r_A (or r_B, respectively). The cost for changing towns is c. Assume that $c > r_A > r_B$ and that there is a discount factor $\alpha < 1$.

(a) Show that for α sufficiently small, the optimal policy is to stay in the town he starts in, and that for α sufficiently close to 1, the optimal policy is to move to town A (if not starting there) and stay in A for all subsequent times.

(b) Solve the problem for $c = 3$, $r_A = 2$, $r_B = 1$, and $\alpha = 0.9$ using policy iteration.

Problem 2 (50 points)

Suppose that you want to travel from a start point S to a destination point D in minimum expected time. There are two options:

(1) Use a direct route that requires a time units.

(2) Take a potential shortcut that requires b time units to go to an intermediate point I. From I you can either go to the destination D in c time units or return to the start (this will take an additional b time units). You will find out the value of c once you reach the intermediate point I. What you know a priori is that c has one of the m values $c_1, \ldots, c_m$ with corresponding probabilities $p_1, \ldots, p_m$. For each of the following questions, consider two cases: (i) The value of c is constant over time, and (ii) The value of c changes each time you return to the start independently of the value at the previous time periods.

We assume that $a, b, c_1, \ldots, c_m$ are all positive.

(a) Formulate the problem as a stochastic shortest path problem and write Bellman's equation.

(b) Are all policies proper? If not, why does Bellman's equation hold?

(c) Characterize the optimal stationary policies as best as you can in terms of the given problem data.

(d) Solve the problem for the case $a = 2$, $b = 1$, $c_1 = 0$, $c_2 = 5$, $p_1 = 0.5$, $p_2 = 0.5$.

(e) For case (ii) only, formulate as a stochastic shortest path problem the variation where once you reach the intermediate point I, you can wait there. Each d time units the value of c changes to one of the values $c_1, \ldots, c_m$ with probabilities $p_1, \ldots, p_m$, independently of its earlier values. Each time the value of c changes, you have the option of waiting for an extra d units, returning to the start, or going to the destination. Characterize the optimal stationary policies as best as you can.

(f) Assume that at the start point S, it is optimal to go to the intermediate point I. For the variation in part (e), show that by selecting an appropriate μ_0, the policy iteration algorithm will terminate after at most m iterations.

Problem 3 (35 points)

The latest casino sensation is a slot machine with three arms, labeled 1, 2, and 3. A single play with arm i, where $i = 1, 2, 3$, costs c_i dollars, and has two possible outcomes: a "win," which occurs with probability p_i, or a "loss," which occurs with probability $1 - p_i$. The slot machine pays you m dollars each time you complete a sequence of three successive "wins," with each win obtained using a different arm.

(a) Consider the problem of finding the arm-playing order that minimizes the expected cost if we are restricted to stop at the first time the machine pays you. Formulate this problem as a stochastic shortest path problem where arm-playing orders are identified with stationary policies and write Bellman's equation for each stationary policy.

(b) Show that the expected cost of the arm-playing order ijk (a permutation of 123) is

$$\frac{c_i + p_i c_j + p_i p_j c_k - p_i p_j p_k m}{p_i p_j p_k}.$$

Show that it is optimal to play the arms in order of increasing $c_i/(1 - p_i)$.

(c) Consider the problem of finding the arm-playing order that minimizes the average expected cost per play, assuming you play infinitely many times. Formulate this problem as an average cost per stage problem, where arm-playing orders are identified with stationary policies, and write Bellman's equation for each stationary policy.

(d) Show that the expected cost per play of the arm-playing order ijk is

$$\frac{c_i + p_i c_j + p_i p_j c_k - p_i p_j p_k m}{1 + p_i + p_i p_j}.$$

Is it possible that the optimal playing order is different than the one of part (b)? If this is so, how do you explain it?