6.231 Dynamic Programming

Midterm Exam, Fall 2002

Prof. Dimitri Bertsekas

Problem 1 (20 points)

We have a set of N objects, denoted $1, 2, \ldots, N$, which we want to group in clusters that consist of consecutive objects. For each cluster $i, i+1, \ldots, j$, there is an associated cost a_{ij}. We want to find a grouping of the objects in clusters such that the total cost is minimum. Formulate the problem as a shortest path problem, and write a DP algorithm for its solution. (Note: An example of this problem arises in typesetting programs, such as TEX/LATEX, that break down a paragraph into lines in a way that optimizes the paragraph's appearance.)

Problem 2 (40 points)

The latest casino sensation is a slot machine with N arms, labeled $1, \ldots, N$. A single play with arm i costs C_i dollars, and has two possible outcomes: a "win," which occurs with probability p_i and pays a reward R_i, and a "loss," which occurs with probability $1 - p_i$. The rule is that each arm may be played at most once, and play must stop at the first loss or after playing all arms once, whichever comes first. The objective is to find the arm-playing order that maximizes the total expected reward minus the total expected cost.

(a) Write a DP algorithm for solving the problem.

(b) Show that it is optimal to play the arms in order of nonincreasing $(p_i R_i - C_i)/(1 - p_i)$. *Note*: This may be shown with or without using the DP algorithm of part (a).

(c) Assume that at any time, there is the option to stop playing, in addition to selecting a new arm to play. Write a DP algorithm for solving this variant of the problem, and find an optimal policy.

(d) Suppose that in the context of part (c), you may play an arm as many times as you want, but each time the reward to be obtained diminishes by a factor β with $0 < \beta < 1$. Assuming that $C_i > 0$, find an optimal policy.

Problem 3 (40 points)

Consider an inventory control problem where stock evolves according to

$$x_{k+1} = x_k + u_k - w_k,$$

and the cost of stage k is

$$cu_k + h\max(0, w_k - x_k - u_k) + p\max(0, x_k + u_k - w_k),$$

where c, h, and p are positive scalars with $p > c$. There is no terminal cost. The stock x_k is perfectly observed at each stage. The demands w_k are independent, identically distributed, nonnegative random variables. However, the (common) distribution of the w_k is unknown. Instead it is known that this distribution is one out of two known distributions F_1 and F_2, and that the a priori probability that F_1 is the correct distribution is a given scalar q, with $0 < q < 1$. You may assume for convenience that w_k can take a finite number of values under each of F_1 and F_2.

(a) Formulate this as an imperfect state information problem, and identify the state, control, system disturbance, observation, and observation disturbance.

(b) Write a DP algorithm in terms of a suitable sufficient statistic.

(c) Characterize as best as you can the optimal policy.