6.231 Dynamic Programming and Stochastic Control
Fall 2008

6.231 Dynamic Programming and Optimal Control

Midterm Exam, Fall 2004

Prof. Dimitri Bertsekas

Problem 1: (30 points)

Air transportation is available between all pairs of n cities, but because of a perverse fare structure, it may be more economical to go from one city to another through intermediate stops. A cost-minded traveller wants to find the minimum cost fare to go from an origin city s to a destination city t. The airfare between cities i and j is denoted by a_{ij}, and for the mth intermediate stop, there is a stopover cost c_m (a_{ij} and c_m are assumed positive). Thus, for example, to go from s to t directly it costs a_{st}, while to go from s to t with intermediate stops at cities i_1 and i_2, it costs $a_{si_1} + c_1 + a_{i_1 i_2} + c_2 + a_{i_2 t}$.

(a) Formulate the problem as a shortest path problem, and identify the nodes, arcs, and arc costs.

(b) Formulate the problem as a stopping problem, and identify the state space, control space, system, cost per stage, and terminal cost.

(c) Write a corresponding DP algorithm that finds an optimal solution in $n - 2$ stages.

(d) Assume that c_m is the same for all m. Devise a rule for detecting that an optimal solution has been found before iteration $n - 2$ of the DP algorithm.

Solution: (a) We introduce a node for each pair (i, m), where i is a city other than s and t, and m is the number of stopovers thus far, where $m = 1, 2, \ldots, n - 2$. Thus, when at node (i, m), the implication is that we are at city i after m stopovers. The two other nodes are the origin and destination cities s and t. The arcs of the graph are:

$$s \text{ to } t \text{ with cost } a_{st},$$

$$s \text{ to } (i, 1) \text{ with cost } c_1 + a_{si},\ i \neq s, t,$$

$$(i, m) \text{ to } t \text{ with cost } a_{it},\ i \neq s, t,$$

$$(i, m) \text{ to } (j, m + 1) \text{ with cost } c_{m+1} + a_{ij},\ i \neq s, t,\ j \neq i, s, t.$$

Evidently, the shortest path from s to t gives the least cost path with stopovers.

(b) We introduce a stopping state corresponding to the destination city t, and an initial state corresponding to the origin city s. There are $n-1$ stages (stage 0 corersponds to being at the initial state s). At the kth stage, $k = 1, \ldots, n-2$, the states (other than t) are the cities $i \neq s, t$, and state $i_k = i$ corresponds to being at city i after k stopovers. The stopping action at state s or i_k has cost a_{st} or $a_{i_k t}$, respectively. The continuation action at s chooses as next state $i_1 = i \neq s, t$ with cost $c_1 + a_{si}$, and at $i_k = i$, $i \neq s, t$, chooses as next state $i_{k+1} = j \neq i, s, t$ with cost $c_{k+1} + a_{ij}$. Stopping is mandatory at stage $n-2$. The problem is deterministic, and evidently the minimal cost starting at s gives the least cost from s to t with stopovers.

(c) The DP algorithm for the stopping problem of part (b) is

$$J_{n-2}(i_{n-2}) = a_{i_{n-2}t}, \qquad i_{n-2} \neq s, t,$$

$$J_k(i_k) = \min\left\{ a_{i_k t},\ \min_{j \neq i_k, s, t} \{ c_{k+1} + a_{i_k j} + J_{k+1}(j) \} \right\}, \qquad k = 1, 2, \ldots, n-3, \quad i_k \neq s, t,$$

$$J_0(s) = \min\left\{ a_{st},\ \min_{j \neq s, t} \{ c_1 + a_{sj} + J_1(j) \} \right\},$$

and requires $n-2$ stages.

(d) If c_m is the same for all m, the DP algorithm of part (c) is stationary. Thus, if for some k, we have $J_k(i) = J_{k+1}(i)$ for all $i \neq s, t$, we will have $J_{k'}(i) = J_{k+1}(i)$ for all $i \neq s, t$ and $k' \leq k$, so the computation of $J_{k'}(i)$ for $k' < k$ is unnecessary, and the DP algorithm can be terminated. The meaning of $J_k(i) = J_{k+1}(i)$ for all $i \neq s, t$ is that the minimum cost path from s to t requires no more than $n - k - 2$ stopovers.

Problem 2: (35 points)

Consider an inventory control problem where the stock x_k is perfectly observed at each stage and evolves according to

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

The demands w_k are independent, identically distributed, nonnegative random variables with known distribution. The control u_k is nonnegative. There is no terminal cost. The cost of stage k is

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

where c, h, and p are positive scalars with $p > c$, and t_k, $k = 0, 1, \ldots, N-1$, are independent identically distributed nonnegative random variables that take values in some bounded interval. The common distribution of the t_k is unknown, except for the fact that it is one out of two known distributions, F_1 and F_2. The a priori probability that F_1 is the correct distribution is a given scalar q, with $0 < q < 1$. The exact value of t_k is known once the controller reaches stage k, but not before.

(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.

Solution: (a) The state is (x_k, t_k, d_k), where d_k takes the value 1 or 2 depending on whether the common distribution of the t_k is F_1 or F_2. The variable d_k stays constant (i.e., satisfies $d_{k+1} = d_k$ for all k), but is not observed perfectly. Instead, the sample values $t_0, t_1, \ldots$ are observed and provide information regarding the value of d_k. In particular, given the a priori probability q and the demand values $t_0, \ldots, t_{k-1}$, we can calculate the conditional probability that t_k will be generated according to F_1.

(b) A suitable sufficient statistic is (x_k, t_k, q_k), where

$$q_k = P(d_k = 1 \mid t_0, \ldots, t_{k-1}).$$

The conditional probability q_k evolves according to

$$q_{k+1} = \frac{q_k F_1(t_k)}{q_k F_1(t_k) + (1 - q_k) F_2(t_k)}, \qquad q_0 = q.$$

The initial step of the DP algorithm in terms of this sufficient statistic is

$$J_{N-1}(x_{N-1}, t_{N-1}, q_{N-1}) = \min_{u_{N-1} \geq 0} \Big[c u_{N-1} + E_{w_{N-1}} \big\{ p \max(0, w_{N-1} - x_{N-1} - u_{N-1} - t_{N-1}) + h \max(0, x_{N-1} + u_{N-1} - w_{N-1}) \big\} \Big].$$

The typical step of the DP algorithm for $k = 0, 1, \ldots, N-1$ is

$$J_k(x_k, t_k, q_k) = \min_{u_k \geq 0} \Big[c u_k + E_{w_k, t_{k+1}} \big\{ p \max(0, w_k - x_k - u_k - t_k) + h \max(0, x_k + u_k - w_k) + J_{k+1}\big(x_k + u_k - w_k,\ t_{k+1}, \phi(q_k, t_k)\big) \big\} \Big]$$

where

$$\phi(q_k, t_k) = \frac{q_k F_1(t_k)}{q_k F_1(t_k) + (1 - q_k) F_2(t_k)},$$

and t_{k+1} has distribution F_1 with probability $\phi(q_k, t_k)$ and distribution F_2 with probability $1 - \phi(q_k, t_k)$.

(c) Notice that the cost-per-stage, for fixed finite-valued u_k, w_k, and t_k, is convex and coercive in x_k. Therefore, it can be shown inductively, as in the text, that $J_k(x_k, t_k, q_k)$ for $k = 0, 1, \ldots, N-1$ is convex and coercive as a function of x_k for fixed t_k and q_k. It follows that for each value of t_k and q_k, there is a threshold $S_k(t_k, q_k)$ such that it is optimal to order an amount $S_k(t_k, q_k) - x_k$, if $S_k(t_k, q_k) > x_k$, and to order nothing otherwise. In particular, $S_k(t_k, q_k)$ minimizes over y the function

$$cy + E_{w_k, t_{k+1}}\{p \max(0, w_k - y - t_k) + h \max(0, y - w_k) + J_{k+1}\big(y - w_k, t_{k+1}, \phi(q_k, t_k)\big)\}$$

Problem 3: (35 points)

You decide not to use your car for N days, which raises the issue of where to park it. At the beginning of each day you may either park it in a garage, which costs G per day, or on the street for free. However, in the latter case, you run the risk of getting a parking ticket, which costs T, with probability p_j, where j is the number of consecutive days that the car has been parked on the street (e.g., on the first day you park on the street, you have probability p_1 of getting a ticket, on the second successive day you park on the street, you have probability p_2, etc). Assume that p_j is monotonically increasing in j, and that you may receive at most one ticket per day when parked on the street.

(a) Formulate this as a DP problem, identify the state space, control space, system, cost per stage, and terminal cost, and write the corresponding DP algorithm.

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

(c) Consider the variant of the problem whereby once you decide to park in the garage, you must stay parked in the garage for the remaining days at a cost of G per day. Formulate this as a DP problem, and characterize as best as you can the optimal policy.

Solution: (a) Let the state be the number of consecutive days that the car is parked on the street, so the initial state is 0. Because there are N days in total, the state space is $\{0, 1, \ldots, N\}$. At the end of each day and at state j, the controller chooses to either park on the street, which increases the state to $j+1$ and incurs a cost T with probability p_{j+1}, or in the garage, which returns the state to 0 and incurs a cost G. There is no terminal cost. We have the following DP algorithm:

$$J_N(j) = 0$$

$$J_k(j) = \min[\underbrace{G + J_{k+1}(0)}_{garage}, \underbrace{p_{j+1}T + J_{k+1}(j+1)}_{street}], \qquad k = 0, 1, \ldots, N-1$$

(b) We show by induction that $J_k(j)$ is monotonically nondecreasing in j for $k = 0, 1, \ldots, N-1$, which simultaneously shows that the optimal policy at each stage k is to park on the street if and only if the state j is less than some threshold j_k. At stage $N-1$, we have $J_{N-1}(j) = \min[G, p_{j+1}T]$. Because p_{j+1} is monotonically increasing in j, we have

$$J_{N-1}(j) = \begin{cases} G & \text{if } j \geq j_{N-1}, \\ p_{j+1}T & \text{if } j < j_{N-1} \end{cases}$$

where j_{N-1} is the smallest integer j such that $p_{j+1}T \geq G$. Notice that $J_{N-1}(j)$ is monotonically nondecreasing in j and corresponds to the optimal policy:

$$\mu^*_{N-1}(j) = \begin{cases} garage & \text{if } j \geq j_{N-1}, \\ street & \text{if } j < j_{N-1} \end{cases}$$

Assume for induction that $J_{k+1}(j)$ is monotonically nondecreasing in j. Then the right-hand term in the minimization of the DP algorithm, $p_{j+1}T + J_{k+1}(j+1)$, is monotonically nondecreasing in j. Since the left-hand term in the minimization, $G + J_{k+1}(0)$, is constant with respect to j, we know that $J_k(j)$ is monotonically nondecreasing in j, which corresponds to the following optimal policy:

$$\mu^*_k(j) = \begin{cases} garage & \text{if } j \geq j_k, \\ street & \text{if } j < j_k \end{cases}$$

where j_k is the smallest integer j such that

$$p_{j+1}T + J_{k+1}(j+1) \geq G + J_{k+1}(0).$$

Notice that if j satisfies $p_{j+1}T \geq G$, then j satisfies $p_{j+1}T + J_{k+1}(j+1) \geq G + J_{k+1}(0)$, meaning $j_k \leq j_{N-1}$ for all k.

The optimal policy has one of two forms: 1) Alternate between parking in the street for a number of days, and parking in the garage for one day, or 2) Alternate between parking in the street for a number of days, and parking in the garage for one day, up to some point, and then park in the garage permanently.

(c) We rewrite the DP algorithm from part (a), replacing the cost-to-go for parking in the garage with $(N-k)G$, where k is the current stage.

$$J_N(j) = 0$$
$$J_k(j) = \min[\underbrace{(N-k)G}_{garage}, \underbrace{p_{j+1}T + J_{k+1}(j+1)}_{street}]$$

In order to have the stopping cost functions be stationary and equivalent to the terminal cost, define $V_k(j) = J_k(j) - (N-k)G$. Rewriting the DP algorithm in terms of $V_k(j)$, we have:

$$V_N(j) = 0$$
$$V_k(j) = \min[\underbrace{0}_{garage}, \underbrace{p_{j+1}T + V_{k+1}(j+1) - G}_{street}]$$

The problem now follows the format for a basic stopping problem as defined in the text, where parking in the garage is considered stopping. We first find the one-step stopping set T_{N-1}. At stage $N-1$, we have:

$$V_{N-1}(j) = \min[\underbrace{0}_{garage}, \underbrace{p_{j+1}T - G}_{street}]$$

which corresponds to the following one-step stopping set:

$$T_{N-1} = \{j \mid p_{j+1}T \geq G\} = \{j \mid j \geq j_{N-1}\}$$

We now show T_{N-1} is absorbing. For any $j \in T_{N-1}$ and if we do not stop, the next state is $j+1$. Because $j+1 > j \geq j_{N-1}$, we have $j+1 \in T_{N-1}$, meaning T_{N-1} is absorbing. Therefore, $T_k = T_{N-1}$ for all k.

Thus the optimal policy is to park in the street for the minimum number of days needed to get into the one-step stopping set, and then park in the garage permanently thereafter. Intuitively this policy makes sense. The cost-per-day of parking in the garage is constant, while the cost-per-day of parking on the street is increasing. Irrespective of the stage index, we stop once the expected cost of street parking exceeds the cost of the garage.