MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Fall 2000 | 6.231
Midterm exam | Tuesday 10/24/00, 7:30-9:30 pm

Problem 1. (25 points)
Consider the standard finite–horizon, discrete–time, linear quadratic problem, with perfect state information, and finite time horizon N. However, we impose the following additional constraint:

$$u_0 = u_{N-1}.$$

(a) Put the problem into a form to which dynamic programming can be applied.

(b) Prove that the optimal cost–to–go function is a quadratic function of the initial state. *Hint:* The proof need not repeat any lengthy calculations like the ones in the text.

Problem 2. (25 points)
We are given a directed graph with nodes $1, \ldots, n$, and with a set $\mathcal{A}$ of directed arcs (i, j). Node 1 is an origin node, and node n is a destination node. Each arc (i, j) has a length c_{ij}. For each node i, the lengths of the arcs (i, j) originating from i are given, except for one special arc (i, j_i) whose length is 1 with probability p, and 0 with probability $1 - p$. (The lenghts of different arcs are assumed independent.)

A vehicle starts at node 1 and wishes to travel through this graph until it reaches its destination n. The vehicle is not allowed to visit the same node twice. The vehicle can learn the status of an arc (i, j_i) only by visiting node i. The objective is to minimize the **expected** sum of the lengths of the arcs traversed by the vehicle.

(a) Express the problem as one with perfect state information (and finite state space), by defining an appropriate state. Be precise in specifying the state, control, and the evolution equation.

(b) Write down the dynamic programming equation.

Problem 3. (50 points)
A shipping company starts (at time 0) with an empty container of integer size K. During each period k ($k = 0, 1, \ldots, N-1$) a potential customer shows up and offers to pay a (positive integer) price p_k to have an item of (positive integer) size s_k included in the container. Assume that the offers (p_k, s_k) at different periods k are i.i.d., with known probability distributions. Each time, the company can choose to either accept (if there is available room in the container) or reject a customer. Its objective is to maximize expected revenue.

(a) Provide a complete dynamic programming formulation of the problem (state, evolution equation, etc.) as well as a dynamic programming algorithm.

(b) Show that there is an optimal policy with the following property: for any fixed state and time, if an offer (p_k, s_k) is accepted, then any offer (p'_k, s_k) with a higher price ($p'_k > p_k$) and the same size s_k is also accepted.

(c) Show that there is an optimal policy with the following property: for any fixed state and time, if an offer (p_k, s_k) is accepted, then any offer (p_k, s'_k) with the same price p_k and a smaller size $s'_k < s_k$ is also accepted.

(d) Suppose now that the size of the items are not accurately known: when the company accepts an item of declared size s_k, the actual size turns out to be $s_k + w_k$, where the w_k are unobserved independent normal random variable with mean zero and variance σ^2. (In principle, this may result in items with negative sizes. But assuming that σ^2 is fairly small, this is very unlikely and let us not be concerned with this possibility.) At time N all items are measured and if the total size turns out to be more than K, the company needs a second container and suffers a cost of C. Provide a dynamic programming formulation (state, evolution equation, costs etc.) of the problem of maximizing expected revenue minus expected costs.

(e) *Extra credit question, in case you have time to spare. Not required.*
Let us go back to the perfect information problem in parts (a)-(c), and assume that offered items always have size 1 ($s_k = 1$). We claim the following. If it is optimal to accept an offer (p_k, s_k) when the total size of past accepted offers is a, then it is also optimal to accept the same offer when the total size of past accepted offers is less than a.
(i) Which property of the value functions would suffice to prove the claim?
(ii) Prove the property in (i).
(iii) Show, by means of an example, that the claimed property of optimal policies fails to hold without the assumption $s_k = 1$.

6.231, Midterm Exam Solutions, Fall 2000

1.

(a) Introduce an additional state vector y_k. State equations:

$$x_{k+1} = A_k x_k + B_k u_k + w_k, \qquad k < N-1,$$

$$x_N = A_{N-1} x_{N-1} + B_{N-1} y_{N-1} + w_{N-1}.$$

$$y_0 = 0, \quad y_1 = u_0, \qquad y_{k+1} = y_k, \quad k > 0.$$

Leave the standard quadratic cost per stage unchanged ($x_k' Q_k x_k + u_k' R_k u_k$), except that at stage $N-1$, the cost per stage should be $x_{N-1}' Q_{N-1} x_{N-1} + y_{N-1}' R_{N-1} y_{N-1}$.

(b) This is still a linear system with costs per stage that are quadratic in the state variables and the controls. Hence, the cost to go is quadratic in (x_0, y_0). Since $y_0 = 0$, the cost-to-go at time zero is quadratic in x_0.

2.

(a) State of the system is a triple (i, S, c): Here i is the present node, S is a subset of $\{1, \ldots, n\}$, which indicates which nodes were visited in the past, and $c \in \{0, 1\}$ indicates the cost of the special arc (i, j_i). The control u is some j that selects the next node to be visited (unless $i = n$ in which case we are the terminal state). The evolution equation is

$$(i_{t+1}, S_{t+1}, c_{t+1}) = (u_t, S_t \cup \{i_t\}, w_t),$$

where w_t is a random variable that is 1 or 0 with probability p or $1-p$.

Alternatively, we can take the state to be just (i, S); this is the state before we learn the status of the special arc.

(b)

$$J(i, S, c) = \min_j \Big\{ c_{ij} + pJ(j, S \cup \{i\}, 1) + (1-p) J(j, S \cup \{i\}, 0) \Big\},$$

$$J(n, S, c) = 0.$$

With the alternative choice of the state, the DP equations are

$$\begin{aligned} J(i, S) = & p \min \Big\{ 1 + J(j_i, S \cup \{i\}), \min_{j \neq j_i} \Big\{ c_{ij} + J(j, S \cup \{i\}) \Big\} \Big\} \\ & + (1-p) \min \Big\{ J(j_i, S \cup \{i\}), \min_{j \neq j_i} \Big\{ c_{ij} J(j, S \cup \{i\}) \Big\} \Big\}, \end{aligned}$$

$$J(n, S) = 0.$$

3.

(a) The state is (x_k, p_k, s_k), where x_k is the sum of the sizes of previously accepted offers. Let $u_k = 1$ if (p_k, s_k) is accepted, zero otherwise. The evolution equation is $x_{k+1} = x_k + u_k s_k$, $(p_{k+1}, s_{k+1}) = w_{k+1}$, where w_{k+1} is a random offer.

$$J_k(x, p, s) = \min\Big\{E[J_{k+1}(x, p_{k+1}, s_{k+1})], p + E[J_{k+1}(x + s, p_{k+1}, s_{k+1})]\Big\}, \qquad k < N.$$

The expectations are with respect to the distribution of $(p_{k+1}, s_{k+1}$. The second possibility is available only if $x + s \le K$. (An easy way of handling this is to define $J(x, p, s) = -\infty$ if $x > K$.) Also, $J_N(x) = 0$ for any $x \le K$.

Alternatively, we can use a value function $J_k(x)$ which is the optimal expected revenue **before** we see the offer (p_k, s_k). The DP equation becomes $J_N(x) = 0$ and

$$J_k(x) = E\left[\min\{J_{k+1}(x), p_k + J_{k+1}(x + s_k)\}\right].$$

This is a little easier to work with.

(b) An offer is accepted if $p_k \ge J_{k+1}(x_k) - J_{k+1}(x_k + s_k)$. If this is the case and $p'_k > p_k$, then $p'_k > J_{k+1}(x_k) - J_{k+1}(x_k + s_k)$, and the offer (p'_k, s_k) should be acceptted.

(c) We claim that the value function is monotonic (nonincreasing): if $x \le y$, then $J_k(x) \ge J_k(y)$. We can see this intuitively: if the state is reduced from y to x, we can do (starting from x) everything that we could do before (starting from y) and achieve the same revenue. Mathematically, this is proved by induction. Monotonicity is true for J_N. Assume that $J_{k+1}(x)$ is nonincreasing in x. Then, $J_{k+1}(x)$ and $p_k + J_{k+1}(x + s_k)$ are also nonincreasing functions of x. The minimum of two nonincreasing functions is nonincreasing. Taking the expectation amounts to forming a weighted average of nonincreasing functions, which shows that $J_k(x)$ is also nonincreasing.

Using this monotonicity, and assuming that $s'_k < s_k$, we see that $J_{k+1}(x_k + s'_k) \ge J_{k+1}(x_k + s_k)$. If (p_k, s_k) is accepted, then $p_k + J_{k+1}(x + s_k) \ge J_{k+1}(x)$, which implies that $p_k + J_{k+1}(x + s'_k) \ge J_{k+1}(x)$, and (p_k, s'_k) can also be accepted.

(d) Here, we have imperfect information, and the total size of past accepted offers is unknown. However, this total size is $\sum_{i=0}^{k-1} u_k(s_k + w_k)$, which is a normal random variable with mean $\sum_{i=0}^{k-1} u_k s_k$ and variance $\sum_{i=0}^{k-1} u_k \sigma^2$. Therefore, a sufficient statistic is the mean and the variance. We can therefore use a two-dimensional state (m_k, v_k) which evolves as follows:

$$m_0 = 0, \qquad m_{k+1} = m_k + u_k s_k,$$

$$v_0 = 0, \qquad v_{k+1} = v_k + u_k \sigma^2.$$

The rewards per stage are again $p_k u_k$, and there is also a terminal cost $g(m_N, v_N)$ equal to the probability that a normal random variable with mean m_N and variance v_N exceeds K.

(e)

(i) What we need is the following property: for every p, and a, x, with $x \leq a$, $p + J_{k+1}(a+1) \geq J_{k+1}(a)$, then $p + J_{k+1}(x+1) \geq J_{k+1}(x)$. This will be satisfied as long as we require

$$J_{k+1}(a) - J_{k+1}(a-1) \geq J_{k+1}(a+1) - J_{k+1}(a).$$

The value function is of interest only at integer points. The above property requires that the "slope" (change from one integer point to the next) of the function J_{k+1} be nonincreasing. This is a discrete counterpart of concavity.

(ii) The proof is by induction. Clearly J_N has the desired property. Assuming that J_{k+1} has the property, it is not hard to check that J_k also has it. (Rather than doing this algebraically, draw a picture to see it; the assumption $s = 1$, or more generally that there is only one possible size, is crucial.)

(iii) Suppose that there is a lot of time left and many items will arrive, some with $(p, s) = (1, 1)$ and some with $(p, s) = (10, 2)$. If the state is $K - 1$, an item of the form $(p, s) = (1, 1)$ should be accepted, since there is no better alternative. But if the state is $K - 2$, the item $(p, s) = (1, 1)$ should be rejected to leave open the possibility of accepting a much more profitable item of the form $(p, s) = (10, 2)$.