6.231 Dynamic Programming and Optimal Control
Midterm Exam, Fall 2011
Prof. Dimitri Bertsekas

Problem 1: (50 points)

Alexei plays a game that starts with a deck with b "black" cards and r "red" cards. Alexie knows b and r. At each time period he draws a random card and decides between the following two options:

(1) Without looking at the card, "predict" that it is black, in which case he wins the game if the prediction is correct and loses if the prediction is incorrect.

(2) "Discard" the card, after looking at its color, and continue the game with one card less.

If the deck has only black cards he wins the game, while if the deck has only red cards he loses the game. Alexei wants to find a policy that maximizes his probability of a win.

(a) Formulate Alexei's problem into the format of the finite-horizon basic problem with perfect state information. Identify states, controls, disturbances, etc.

(b) Write the DP algorithm. Use induction to show that the optimal probability of a win starting with b black cards and r red cards is $\dfrac{b}{b+r}$.

(c) Characterize the optimal policies.

(d) Suppose that Alexei is given the additional option to randomize his decision at each time period. In particular, he may choose a probability $p \in [0,1]$, flip a coin that has probability of head equal to p, and decide upon option 1 or 2 above depending on the outcome of the flip. What would then be the optimal policies?

Solution: (a) The state is the current pair (b,r) plus a termination state, to which we move upon selecting option 1. If option 2 is selected, then we move to state $(b-1,r)$ with probability $b/(b+r)$ and to state $(b,r-1)$ with probability $r/(b+r)$. At states with either $b=0$ or $r=0$, only option 1 is available. The reward per stage is 1 if Alexei chooses option 1 and wins the game, and 0 otherwise.

(b,c,d) The DP algorithm is

$$J^*(b,r) = \min_{p\in\{0,1\}} \left[p\frac{b}{b+r} + (1-p)\left(\frac{b}{b+r}J^*(b-1,r) + \frac{r}{b+r}J^*(b,r-1)\right)\right],$$

where $p = 0$ and $p = 1$ correspond to options 1 and 2, respectively. We show by induction that the optimal reward-to-go starting with b black and r red cards is

$$J^*(b,r) = \frac{b}{b+r},$$

and that all strategies give the same probability of a win.

This formula is correct for all (b,r) with $b + r = 1$, based on the assumption that Alexei wins when the deck has only black cards and loses if the deck has only red cards. Assuming that the formula is correct for all (b,r) with $b + r = k - 1$, we will show that it is true for all (b,r) with $b + r = k$. For any $p \in [0,1]$, the right-hand side of the DP algorithm is

$$\begin{aligned}
&p\frac{b}{b+r} + (1-p)\left(\frac{b}{b+r}J^*(b-1,r) + \frac{r}{b+r}\,J^*(b,r-1)\right) \\
&\qquad = p\,\frac{b}{b+r} + (1-p)\,\left(\frac{b}{b+r}\,\frac{b-1}{b+r-1} + \frac{r}{b+r}\,\frac{b}{b+r-1}\right) \\
&\qquad = p\,\frac{b}{b+r} + (1-p)\,\frac{b}{b+r} \\
&\qquad = \frac{b}{b+r},
\end{aligned}$$

where the second equality makes use of the induction hypothesis. Thus

$$J^*(b,r) = \frac{b}{b+r},$$

and the induction is complete. For the case $p \in [0,1]$, this calculation also yields the same formula for J^*, regardless of the value of p chosen.

Problem 2: (50 points)

An ambitious engineer, currently employed in a recently impoverished country at a salary c, seeks employment at another country. He receives a job offer at each time period, which she may accept or reject. The offered salary takes one of n possible values $w^1, \ldots, w^n$ with given probabilities $\xi^1, \ldots, \xi^n$, independent of preceding offers. If she accepts the offer, she keeps the job for the rest of her life. If she rejects, she continues at her current salary c for one more period and is eligible to accept offers in future periods. Assume that income is discounted by a factor $\alpha < 1$.

(a) Formulate the engineer's problem as a discounted infinite horizon problem. Identify states, controls, disturbances, etc.

(b) Formulate the Bellman equation and characterize the optimal policy.

(c) Consider the variant of the problem where there is a given probability p^i that the worker will be fired from her job at any one period if the salary is w^i. Once fired, she returns to her own country and resumes her former employment and is eligible to accept offers in future periods. Formulate the Bellman equation and characterize the optimal policy.

(d) Do part (c) for the case when income is not discounted and the worker maximizes the average income per period.

(e) Assume that for the job with salary w^i, the expected number of periods for keeping that job is t^i, and afterwards she returns to her former employment as in part (c). Formulate the Bellman equation for the problem of maximizing her average return using t^i instead of p^i. Characterize the optimal policy.

Solution: (a) Let the states be s^i, $i = 1, \ldots, n$, corresponding to the worker being unemployed and being offered a salary w^i, and $\bar{s}^i$, $i = 1, \ldots, n$, corresponding to the worker being employed at a salary level w^i. Suppose the current state is s_i, the control is to either accept and go to state $\bar{s}_i$, or to reject and wait for the next offer. The reward is w_i if she accepts, or c if she rejects. Suppose the state is $\bar{s}_i$, there is no more decision to make and the reward is always w_i.

(b) The Bellman equation is

$$J(s^i) = \max \left[c + \alpha \sum_{j=1}^{n} \xi^j J(s^j),\ w^i + \alpha J(\bar{s}^i) \right], \qquad i = 1, \ldots, n, \tag{1}$$

$$J(\bar{s}^i) = w^i + \alpha J(\bar{s}^i), \qquad i = 1, \ldots, n. \tag{2}$$

From Eq. (2), we have

$$J(\bar{s}^i) = \frac{w^i}{1-\alpha} \qquad i = 1, \ldots, n,$$

so that from Eq. (1) we obtain

$$J(s^i) = \max \left[c + \alpha \sum_{j=1}^{n} \xi^j J(s^j),\ \frac{w^i}{1-\alpha} \right],$$

Thus it is optimal to accept salary w^i if

$$w^i \geq (1-\alpha) \left(c + \alpha \sum_{j=1}^{n} \xi^j J(s^j) \right).$$

The right-hand side of the above relation gives the threshold for acceptance of an offer.

(c) In this case the Bellman equation becomes

$$J(s^i) = \max\left[c + \alpha \sum_{j=1}^{n} \xi^j J(s^j),\ w^i + \alpha\left((1-p^i)J(\bar{s}^i) + p^i \sum_{j=1}^{n} \xi^j J(s^j)\right)\right] \tag{3}$$

$$J(\bar{s}^i) = w^i + \alpha\left((1-p^i)J(\bar{s}^i) + p^i \sum_{j=1}^{n} \xi_j J(s^j)\right). \tag{4}$$

Let us assume without loss of generality that

$$w^1 < w^2 < \cdots < w^n.$$

Let us assume further that $p^i = p$ for all i. From Eq. (4), we have

$$J(\bar{s}^i) = \frac{w^i + p\sum_{j=1}^{n} \xi^j J(s^j)}{1-\alpha(1-p)},$$

so it follows that

$$J(\bar{s}^1) < J(\bar{s}^2) < \cdots < J(\bar{s}^n). \tag{5}$$

We thus obtain that the second term in the maximization of Eq. (3) is monotonically increasing in i, implying that there is a salary threshold above which the offer is accepted.

In the case where p^i is not independent of i, salary level is not the only criterion of choice. There must be consideration for job security (the value of p^i). However, if p^i and w^i are such that Eq. (5) holds, then there still is a salary threshold above which the offer is accepted.

(d) The Bellman equation has the form

$$\lambda + h(s^i) = \max\left[c + \sum_{j=1}^{n} \xi^j h(s^i),\ w^i + (1-p^i)h(\bar{s}^i) + p^i \sum_{j=1}^{n} \xi^j h(s^j)\right], \qquad i = 1, \ldots, n, \tag{3}$$

$$\lambda + h(\bar{s}^i) = w^i + (1-p^i)h(\bar{s}^i) + p^i \sum_{j=1}^{n} \xi^j h(s^j), \qquad i = 1, \ldots, n. \tag{4}$$

From these equations, we have

$$\lambda + h(s^i) = \max\left[c + \sum_{j=1}^{n} \xi^j h(s^j),\ \lambda + h(\bar{s}^i)\right], \qquad i = 1, \ldots, n,$$

so it is optimal to accept a salary offer w^i if $h(\bar{s}^i)$ is no less that the threshold

$$c - \lambda + \sum_{j=1}^{n} \xi^j h(s^j).$$

Here λ is the optimal average salary per period (over an infinite horizon). If $p^i = p$ for all i and $w^1 < w^2 < \cdots < w^n$, then from Eq. (4) it follows that $h(\bar{s}^i)$ is monotonically increasing in i, and the optimal policy is to accept a salary offer if it exceeds a certain threshold.

(e) Let us denote by i the state corresponding to the offer (w^i, t^i). We have the following Bellman's equation:

$$h(i) = \max \left\{ w^i t^i - \lambda t^i + \sum_{j=1}^{n} p^j h(j), c - \lambda + \sum_{j=1}^{n} p^j h(j) \right\}$$

The optimal policy is to accept offer (w^i, t^i) if

$$w^i - \frac{c}{t^i} \geq \lambda,$$

where λ is the optimal average income per unit time.

6.231 Dynamic Programming and Stochastic Control
Fall 2015