Dynamic Programming and Optimal Control **Autumn 2014**

Richard Weber

Problems Sheet 2

1. At the beginning of each day a machine can be in either a working or broken state. If it is broken then the whole day is spent repairing it, and this costs $8c$ in labour and lost production. If the machine is working, then it may be run unattended or attended, at costs of 0 or c respectively. In either case there is a chance that the machine will breakdown and need repair the following day, with probabilities p and p' respectively, where $p' < (7/8)p$. Costs are discounted by factor β, $0 < \beta < 1$, and it is desired to minimize the total-expected discounted-cost over the infinite horizon. Let $F(0)$ and $F(1)$ denote the minimal value of such cost, starting from a morning on which the machine is broken or working respectively. Show that it is optimal to run the machine unattended only if $\beta \le 1/(7p - 8p')$.

2. A hunter earns £1 for each member of an animal population captured, but hunting costs him £c per unit time. The number r, of animals remaining uncaptured is known, and will not change by natural causes on the relevant time scale. The probability of a single capture, in the next time unit, is $\lambda(r)$, where λ is a known increasing function. The probability of more than one capture per unit time is negligible. What stopping rule will maximize the hunter's net expected profit?

3. A financial advisor can impress his clients if immediately following a week in which the FTSE index moves by more than 5% in some direction he correctly predicts that this is the last week during the calendar year that it moves more that 5% in that direction.

Suppose that in each week the market change is up > 5%, down > 5%, or neither of these, with probabilities p, p, $1 - 2p$, respectively, ($p < 1/2$). He makes at most one prediction this year. With what strategy does he maximize the probability of impressing his clients?

4. *This question shows you how to optimality equations using linear programming.*

Consider the following infinite-horizon discounted-cost optimality equation for a Markov decision process with, $0 < \beta < 1$, a finite state space, $x \in \{1, \dots, N\}$, and M actions available in each state, $u \in \{1, \dots, M\}$:

$$F(x) = \min_u \left[c(x, u) + \beta \sum_{x_1=1}^{N} F(x_1) P(x_1 \mid x_0 = x, u_0 = u) \right]. \tag{1}$$

Consider also the linear programming problem

$$\textbf{LP:} \qquad \underset{G(1),\dots,G(N)}{\text{maximize}} \sum_{i=1}^{N} G(i)$$

with

$$G(x) \le c(x, u) + \beta \textstyle\sum_{x_1=1}^{N} G(x_1) P(x_1 \mid x_0 = x, u_0 = u), \quad \text{for all } x,\, u.$$

This **LP** has N variables and $N \times M$ constraints. Suppose F is a solution to (1). Show that F is a feasible solution to **LP**. Suppose G is also a feasible solution to **LP**. Show that for each x there exists a u such that,

$$F(x) - G(x) \ge \beta E[F(x_1) - G(x_1) \mid x_0 = x, u_0 = u],$$

and hence that $F \ge G$.

Argue finally, that F is the unique optimal solution to **LP**. What is the use of this result?

5. A ball is hidden in one of n boxes. A search of the ith box costs $C_i > 0$ and finds the ball with probability α_i if the ball is in the box. Suppose that we are given prior probabilities P_i^0, $i = 1, 2, \ldots, n$ that the ball is in the ith box. Show that the policy which searches a box with maximal value of $\alpha_i P_i / C_i$ minimizes the expected searching cost, where P_i is the posterior probability (given everything that has occurred up to that time) that the ball is in box i. Hint: use an interchange argument.

6. Jobs $1, 2, 3, 4$ are to be processed in some order by a single machine. Once a job has been started its processing cannot be interrupted. Job i has a known processing time s_i. If it completes at time t_i then a discounted reward of $r_i e^{-\alpha t_i}$ is obtained, $\alpha > 0$. There are precedence constraints amongst jobs such that job i cannot be started until job $i-2$ is complete, $i = 3, 4$. We wish to maximize the total discounted reward obtained from the 4 jobs. E.g. a possible schedule is 1, 2, 4, 3, with reward

$$r_1 e^{-\alpha s_1} + r_2 e^{-\alpha(s_1+s_2)} + r_4 e^{-\alpha(s_1+s_2+s_4)} + r_3 e^{-\alpha(s_1+s_2+s_4+s_3)}$$

Use the Gittins index theorem (appropriately generalized to continuous time) to show that job 1 should be processed first (rather than job 2) if

$$\max\left\{\frac{r_1 e^{-\alpha s_1}}{1 - e^{-\alpha s_1}}, \frac{r_1 e^{-\alpha s_1} + r_3 e^{-\alpha(s_1+s_3)}}{1 - e^{-\alpha(s_1+s_3)}}\right\} \geq \max\left\{\frac{r_2 e^{-\alpha s_2}}{1 - e^{-\alpha s_2}}, \frac{r_2 e^{-\alpha s_2} + r_4 e^{-\alpha(s_2+s_4)}}{1 - e^{-\alpha(s_2+s_4)}}\right\}.$$

Let us modify the problem so that initially we pay a fee of $\sum_i r_i$, but that $r_i e^{-\alpha t_i}$ is refunded when job i completes. Thus the net cost is $\sum_i [r_i - r_i e^{-\alpha t_i}] = \alpha \sum_i r_i t_i + o(\alpha)$.

Use this idea to address a problem in which there are no rewards, but a waiting cost c_i is incurred per unit of time until job i completes. Show that the total waiting cost is minimized by processing job 1 first (rather than job 2) if

$$\max\left\{\frac{c_1}{s_1}, \frac{c_1 + c_3}{s_1 + s_3}\right\} \geq \max\left\{\frac{c_2}{s_2}, \frac{c_2 + c_4}{s_2 + s_4}\right\}.$$

7. *This question is about proving a structural property of an optimal policy. Many research papers in the field have been written about results like this.*

Recall the problem about exercising a call option. We proved the the value function $F_s(\cdot)$ has the property that $F_s(x) - x$ is non-decreasing in x. We used this to prove that the optimal policy is of threshold type, i.e. *exercise the option if* $x \geq a_s$, where a_s increases with the time-to-go, s. The following problem is of similar type.

Each morning at 9 am a barrister has a meeting with his instructing solicitor. With probability θ, independently of other mornings, he will be offered a new case, which he may either decline or accept it. If he accepts it he will be paid R when it is complete. However, for each day that the case is unfinished he will incur a charge of c and so it is expensive to have too many cases outstanding. Following the meeting he spends the rest of the day working on a single case, which he finishes by the end of the day with probability p, $p < 1/2$. If he wishes he can hire a temporary assistant for the day, at cost a, and by working on a case together they can finish it with probability $2p$.

The barrister wishes to maximize his expected total-profit over s days. Let $G_s(x)$ and $F_s(y)$ be the maximal such profit he can obtain, given that his number of outstanding cases are x and $y \in \{x, x+1\}$ respectively, just before and just after the meeting on the first day. It is a reasonable to conjecture that the optimal policy is a 'threshold policy', i.e., Conjecture C. *There exist integers $n(s)$ and $m(s)$ such that it is optimal to accept a new case if and only if $x \le n(s)$ and to employ the assistant if and only if $y \ge m(s)$.*

By writing G_s in terms of F_s, and writing F_s in terms of G_{s-1}, show that the optimal decisions do indeed take this form provided both $F_s(x)$ and $G_{s-1}(x)$ are concave functions of x.

Now suppose that conjecture C is true for all $s \le t$, and that F_t and G_{t-1} are concave functions of x. First show that for $x > 0$,

$$\begin{aligned} & G_t(x+1) - 2G_t(x) + G_t(x-1) \\ &= (1-\theta)\Big\{F_t(x+1) - 2F_t(x) + F_t(x-1)\Big\} + \theta\Big\{\max[F_t(x+1), F_t(x+2)] \\ &\quad -2\max[F_t(x), F_t(x+1)] + \max[F_t(x-1), F_t(x)]\Big\}. \end{aligned} \tag{2}$$

By carefully considering the values of terms on the right have side of this expression, separately in the three cases $x+1 \le n(t)$, $x-1 > n(t)$ and $x-1 \le n(t) < x+1$, show that G_t is also concave and hence that it is also true that the optimal hiring policy is of threshold form when the horizon is $t+1$.

I am not asking you to do it, but in a similar manner, one could show that F_{t+1} is concave, and so inductively push through a proof of Conjecture C for all finite-horizon problems.