Optimization and Control: Examples Sheet 1

Dynamic programming

1. Given a sequence of matrix multiplications

$$M_1 M_2 \cdots M_k M_{k+1} \cdots M_n,$$

where each M_k is a matrix of dimension $d_{k-1} \times d_k$, the order in which the multiplications are done makes a difference. For example, if $d_0 = 1$, $d_1 = 10$, $d_2 = 1$ and $d_3 = 10$, the calculation $(M_1 M_2) M_3$ requires 20 scalar multiplications, but the calculation $M_1(M_2 M_3)$ requires 200 scalar multiplications. Generally, multiplying a $d \times d'$ matrix by a $d' \times d''$ matrix requires $dd'd''$ scalar multiplications. Let $F(d_0, d_1, \ldots, d_n)$ be the minimal number of scalar multiplications required. Explain why F satisfies the recursion

$$F(d_0, d_1, \ldots, d_n) = \min_{1 \le i \le n-1} \{d_{i-1} d_i d_{i+1} + F(d_0, \ldots, d_{i-1}, d_{i+1}, \ldots, d_n)\}.$$

Describe an algorithm for finding the optimal order of multiplication. Solve the problem for $n = 3$, $d_0 = 2$, $d_1 = 10$, $d_2 = 5$ and $d_3 = 1$.

Solve it also for $n = 4$, $d_0 = 2$, $d_1 = 10$, $d_2 = 1$, $d_3 = 5$ and $d_4 = 1$. Obtain an estimate of the number of prior evaluations of F needed to find $F(d_0, d_1, \ldots, d_n)$.

2. A deck of cards is thoroughly shuffled and placed face down on the table. You turn over cards one by one, counting the numbers of reds and blacks you have seen so far. Exactly once, whenever you choose, you may bet that the next card you turn over will be red. If you are correct you win £1000.

Let $F(r, b)$ be the probability of winning if you play optimally, beginning from a point at which you have not yet bet and you know that exactly r red and b black cards remain in the face-down pack. Find $F(26, 26)$ and your optimal strategy. How does this answer compare with your intuition?

3. A gambler has the opportunity to bet on a sequence on n coin tosses. The probability of heads on the kth toss is known to be p_k, $k = 1, \ldots, n$. On each toss he may stake any non-negative amount not exceeding his current capital (which is his initial capital plus his winnings to date) and call 'heads' or 'tails'. If his call is correct he retains his stake and wins an equal amount, while if the call is incorrect he loses his stake. Suppose the gambler has initial capital x and write X_n for his capital after the final toss. Determine how he should call and how much he should stake for each toss in order to maximize $E(\log X_n)$.

4. A man is standing in a queue waiting for service, with m people ahead of him. He knows the utility of waiting out the queue, r, and the constant probability p that the person at the head of the queue will complete service in the next unit of time (independently of what happens in all other units of time). On the other hand he incurs a cost c for every unit of time spent waiting for his own service to begin. The problem is to determine the waiting policy that maximizes his expected return.

Let F_m denote the expected return obtained by employing an optimal waiting policy when there are m people ahead. Show that the dynamic programming equation is

$$F_m = \max\{-c + pF_{m-1} + (1-p)F_m, 0\}, \quad m \ge 1, \tag{1}$$

with $F_0 = r$. Show that this can be re-written as

$$F_m = \max\{F_{m-1} - c/p, 0\}, \quad m \geq 1. \tag{2}$$

Hence prove inductively that $F_n \leq F_{n-1}$. Why is this fact intuitive?

Show there exists an integer m^* such that the form of the optimal policy is to wait only if $m \leq m^*$. Find expressions for F_m and m^* in terms of r, c and p.

Give an alternative derivation of the optimal waiting policy, without recourse to dynamic programming.

5. The Greek adventurer Theseus is trapped in a room from which lead n passages. Theseus knows that if he enters passage i one of three fates will befall him: he will escape with probability p_i, he will be killed with probability q_i, and with probability r_i he will find the passage to be a dead end and be forced to return to the room. The fates associated with different passages are independent. Establish the order in which Theseus should attempt the passages if he wishes to maximize his probability of eventual escape.

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

The barrister wishes to maximize her expected total profit over the next n days. Let $G_n(x)$ and $F_n(y)$ be the maximal such profit she can obtain, given that her 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 reasonable to conjecture that the optimal policy is a 'threshold policy', i.e.,

C: *There exist integers $a(n)$ and $b(n)$ such that it is optimal to accept a new case if and only if $x \leq a(n)$ and to employ the assistant if and only if $y \geq b(n)$.*

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

Now suppose that conjecture **C** is true for all $n \leq m$, and that F_m and G_{m-1} are concave functions of x. First show that for $x > 0$,

$$\begin{aligned}
&G_m(x+1) - 2G_m(x) + G_m(x-1) \\
&\quad = (1-\theta)(F_m(x+1) - 2F_m(x) + F_m(x-1)) \\
&\qquad + \theta(\max\{F_m(x+1), F_m(x+2)\} \\
&\qquad\quad - 2\max\{F_m(x), F_m(x+1)\} + \max\{F_m(x-1), F_m(x)\}).
\end{aligned}$$

Now, by considering the values of terms on the right hand side of this expression, separately in the three cases $x+1 \leq a(m)$, $x-1 > a(m)$ and $x-1 \leq a(m) < x+1$, show that G_m is also concave and hence that it is also true that the optimal hiring policy is of threshold form when the horizon is $m+1$.

(In a similar manner, one can next show that F_{m+1} is concave, and so inductively push through a proof of **C** for all finite-horizon problems.)

7. At the beginning of each day a certain machine can be either working or broken. 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 break down and need repair the following day, with probabilities p and p' respectively, where $p' < (7/8)p$. Costs are discounted by a factor $\beta \in (0, 1)$, 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 this 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 $\beta \leq 1/(7p - 8p')$.

8. Consider the infinite-horizon discounted-cost dynamic programming equation for a stochastic controllable dynamical system P, with state-space $S = \{1, \ldots, N\}$, action-space $A = \{1, \ldots, M\}$, cost function c and discount rate $\beta \in (0, 1)$:

$$F(x) = \min_{a \in A}(c + \beta PF)(x, a), \quad x \in S. \tag{3}$$

Write F for the unique solution. Consider also the linear programming problem

$$\begin{aligned} &\text{maximize} \quad \sum_{x \in S} G(x) \\ &\text{subject to} \quad G(x) \leq (c + \beta PG)(x, a), \quad \text{for all } a \in A. \end{aligned}$$

Show that F is a feasible solution to the linear programming problem and that, for any such feasible solution G, for all $x \in S$, there is an $a \in A$ such that

$$F(x) - G(x) \geq \beta P(F - G)(x, a).$$

Deduce that $F \geq G$ and hence that F is the unique optimal solution to the linear programming problem. What is the use of this result?

9. A hunter receives a unit bounty for each member of an animal population captured, but hunting costs him an amount 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.

The hunter wishes to maximize his net expected profit. What should be his stopping rule?

10. Consider a burglar who loots some house every night. His profits from successive crimes form a sequence of independent random variables, each having the exponential distribution with mean $1/\lambda$. Each night there is a probability $q \in (0, 1)$, of his being caught and forced to return his whole profit. If he has the choice, when should the burglar retire so as to maximize his total expected profit?

11. A motorist has to travel an enormous distance along a newly open motorway. Regulations insist that filling stations can be built only at sites at distances $1, 2, \ldots$ from his starting point. The probability that there is a filling station at any particular point is p, independently of the situation at other sites. On a full tank of petrol, the motorist's car can travel a distance of exactly G units (where G is an integer greater than 1), so that it can just reach site G when starting full at site 0. The petrol gauge on the car is extremely accurate. Since he has to pay for the petrol anyway, the motorist ignores its cost. Whenever he stops to fill his tank, he incurs an 'annoyance' cost A. If he arrives with an empty tank at a site with no filling station, he incurs a 'disaster' cost D and has to have the tank filled by a motoring organization. Prove that if the following condition holds:

$$(1 - q^G)A < pq^{G-1}D,$$

then the policy: '*On seeing a filling station, stop and fill the tank*' minimizes the expected long-run average cost. Calculate this cost when the policy is employed.

12. Suppose that at time n a machine is in state x (where x is a non-negative integer.) The machine costs cx to run until time $n+1$. With probability $a = 1-b$ the machine is serviced and so goes to state 0 at time $n+1$. If it is not serviced then the machine will be in states x or $x+1$ at time $n+1$ with respective probabilities p and $q = 1-p$. Costs are discounted by a factor β per unit time. Let $F(x)$ be the expected discounted cost over an infinite future for a machine starting from state x. Show that $F(x)$ has the linear form $\phi + \theta x$ and determine the coefficients ϕ and θ.

A maintenance engineer must divide his time between M such machines, the ith machine having parameters c_i, p_i and state variable x_i. Suppose he allocates his time randomly, in that he services machine i with a probability a_i at a given time, independently of machine states or of the previous history, where $\sum_i a_i = 1$. The expected cost starting from state variables x_i under this policy will be $\sum_i F_i(x_i) = \sum_i (\phi_i + \theta_i x_i)$ if one neglects the coupling of machine-states introduced by the fact that the engineer can only be in one place at once (a coupling which vanishes in the limit of continuous time.)

Consider one application of the policy improvement algorithm. Show that under the improved policy the engineer should next service the machine whose label i maximizes $c_i(x_i + q_i)/(1 - \beta b_i)$.