MATHEMATICAL TRIPOS: PART II **Lent Term 2016**

OPTIMIZATION AND CONTROL **Richard Weber**

Example Sheet 1

1. Suppose that the matrix M_k is of dimension $n_k \times n_{k+1}$, $k \in \{1, \ldots, h\}$. We wish to compute the product $M_1 M_2 \cdots M_h$. Notice that the order of multiplication makes a difference. For example, if $(n_1, n_2, n_3, n_4) = (1, 10, 1, 10)$, 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. Indeed, multiplying a $m \times n$ matrix by a $n \times k$ matrix requires mnk scalar multiplications. Let $F(n_1, n_2, \ldots, n_{h+1}; h)$ be the minimal total number of scalar multiplications required to compute $M_1 M_2 \cdots M_h$. Explain why the dynamic programming equation is

$$F(n_1, n_2, \ldots, n_{k+1}; k) = \min_{1<i<k+1} \{n_{i-1} n_i n_{i+1} + F(n_1, \ldots, n_{i-1}, n_{i+1}, \ldots, n_{k+1}; k-1)\},$$

$k = 1, \ldots, h$. Hence describe an algorithm which finds the multiplication order requiring least scalar multiplications. Solve the problem for

(a) $h = 3$, $(n_1, n_2, n_3, n_4) = (2, 10, 5, 1)$;

(b) $h = 4$, $(n_1, n_2, n_3, n_4, n_5) = (2, 10, 1, 5, 1)$.

Show that as h increases the amount of effort required to find the optimal order increases faster than any polynomial function of h.

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 like, you may bet that the next card you turn over will be red. If 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.

Arguably, it should be possible to win the £1000 with a probability greater than 1/2 because you can wait until you have seen more black cards than red and then bet that the next card is red. Explain why this argument is wrong.

3. A gambler has the opportunity to bet on a sequence on N coin tosses. The probability of heads on the nth toss is known to be p_n, $n = 1, \ldots, N$. For the nth toss he may stake any non-negative amount not exceeding his current capital (which is his initial capital plus his winnings so far) and call 'heads' or 'tails'. If he calls correctly then he retains his stake and wins an amount equal to it, but if he calls incorrectly he loses his stake. Let $X_0 \geq 0$ denote his initial capital and X_N his capital after the final toss. Determine how the gambler should call and how much he should stake for each toss in order to maximize $E[\log X_N]$. How would your answer differ if the aim is to maximize $E[X_N]$?

4. A man stands in a queue waiting for service, with n people ahead of him. There is a constant probability p that the person at the head of the queue will complete service in the next unit of time (say, 1 minute) independently of what happens in all other units of time. He incurs a cost c for every unit of time spent waiting for his own service to begin. He may leave the queue at any time, but must pay a reneging cost r. The problem is to determine the policy that minimizes his expected cost.

Let $F(n)$ denote the expected return obtained by employing an optimal waiting policy when there are n people ahead. To what theorem can we appeal to justify the optimality equation

$$F(n) = \min[r, c + pF(n-1) + (1-p)F(n)], \quad n > 0, \tag{1}$$

with $F(0) = 0$? Show that (1) can be re-written as

$$F(n) = \min\left[r, F(n-1) + c/p\right], \quad n > 0. \tag{2}$$

Hence prove inductively that $F(n) \geq F(n-1)$. Why is this fact intuitive?

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

Give an alternative derivation of the optimal 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 ($i = 1, \ldots, n$) 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 ($= 1 - p_i - q_i$) he will find the passage to be a dead end and be forced to return to the room. Assume $0 < r_i < 1$. The fates associated with different passages are independent. He attempts each passsage at most once. Establish the order in which Theseus should attempt the passages if he wishes to maximize his probability of eventual escape.

6. 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 breakdown and need repair the following day, with probabilities p and p' respectively. 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 iff $(7p - 8p') \leq 1/\beta$.

7. 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 0. The hunter wishes to maximize his net expected profit. The dynamic programming equation for this problem, posed to an infinite horizon, is intuitively

$$F(r) = \max\{0, -c + \lambda(r) + \lambda(r)F(r-1) + (1 - \lambda(r))F(r)\}.$$

Recall that it is a theorem that the dynamic programming equation holds in general cases of negative, positive or discounted programming. How can you reformulate this problem so that it becomes one of these cases? [Hint. Assume that initially $r = r_0$.] What should be his stopping rule?

8. A burglar loots some house every night. His profit from successive crimes forms a sequence of independent random variables, each having the exponential distribution with mean $1/\lambda$. Each night there is a probability q, $0 < q < 1$, 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? [Hint. Start by finding the optimal policy in a problem where the burgler must retire by the end of the sth day.]

9. This question shows that optimality equations can be solved with linear programming. Consider the following infinite-horizon discounted-cost optimality equation for a Markov decision process with, $0 < \beta < 1$, a finite state space, $x \in \{1, \ldots, N\}$, and $u \in \{1, \ldots, M\}$:

$$F(x) = \min_u \left[c(x, u) + \beta \sum_{x_1=1}^{N} F(x_1) P(x_1 \mid x_0 = x, u_0 = u) \right]. \tag{3}$$

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 (3). 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?

10. This question is about proving a structural property of an optimal policy. In lecture 2 we considered a 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* $x \ge a_s$, where a_s increases with the time-to-go, s. The following problem is of similar.

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 $x > 0$,

$$\begin{aligned} &G_t(x+1) - 2G_t(x) + G_t(x-1) \\ &\quad = (1-\theta)\Big\{F_t(x+1) - 2F_t(x) + F_t(x-1)\Big\} + \theta\Big\{\max[F_t(x+1), F_t(x+2)] \\ &\qquad -2\max[F_t(x), F_t(x+1)] + \max[F_t(x-1), F_t(x)]\Big\}. \end{aligned} \tag{4}$$

Now, by 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 $t + 1$.

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

MATHEMATICAL TRIPOS: PART II **Lent Term 2016**

OPTIMIZATION AND CONTROL **Richard Weber**

Example Sheet 2

1. Nine boxes are placed in front of you and money is put into each box: perhaps £2 in one box, £15 in another box, and so on. You are able to see each of the values and their sum is £100. A tenth box containing a Devil's penny is added to the mix. The boxes are then closed and mixed so that you can't tell which box contains which amount of money, or the Devil's penny. You may then open up, one by one, as many boxes as you'd like and keep the money inside, but if you open the box with the Devil's penny you lose everything. You wish to maximize the expected amount of money you can take home. Find your optimal strategy and prove that it is optimal.

2. A financial advisor can impress his clients if immediately following a week in which the FTSE 100 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 than 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?

3. 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, i.e. with $e^{-\alpha t}$ replacing β^t) to show that job 1 should be processed first (rather than job 2) if

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

Now replace r_i by a cost c_i. Suppose we modify the problem to one in which we initially we pay a cost $\sum_i c_i$, but then $c_i e^{-\alpha t_i}$ is refunded when job i completes at time t. Thus the net cost is $\sum_i [c_i - c_i e^{-\alpha t_i}] = \alpha \sum_i c_i t_i + o(\alpha)$, which is effectively minimized by maximizing $\sum_i c_i e^{-\alpha t_i}$.

Use this idea on a problem in which a waiting cost is incurred at rate c_i per unit of time until job i completes. Show that the total waiting cost is minimized by processing job 1 first (rather than 2) if

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

4. Recall Question 6 from Examples Sheet 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.*

Solve for the variables $f^0(0)$, $f^0(1)$ and λ,

$$f^0(0) + \lambda = 8c + f^0(1)$$
$$f^0(1) + \lambda = pf^0(0) + qf^0(1).$$

Explain the meaning of λ and $f^0(0) - f^0(1)$.

By considering a step of the poilcy improvement algorithm show that the policy of running the machine unattended minimizes the average-cost iff $(7p - 8p') \leq 1$.

Let $F_\beta(1)$ be the minimal expected β-discounted cost when starting on a day when the machine is working. Find $\lim_{\beta \to 1}(1-\beta)F_\beta(1)$ and compare it to the minimal average-cost.

5. 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.

You should check that you understand that the average-cost optimality equations are

$$\begin{aligned} \lambda + \phi(x) &= q\phi(x-1) + p\min[A + \phi(G), \phi(x-1)], \quad 1 < x \leq G, \\ \lambda + \phi(1) &= q[D + \phi(G)] + p[A + \phi(G)]. \end{aligned}$$

Consider the policy π, defined as '*On seeing a filling station, stop and fill the tank*'. Suppose we take $\phi(G) = 0$. If π is optimal then

$$\lambda + \phi(x) = q\phi(x-1) + pA, \quad 1 < x \leq G,$$

to which the general solution is of the form $\phi(x) = a + bq^x$.

Determine values for a and b. Thus prove that if the following condition holds:

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

then the π minimizes the expected long-run average cost. Show that when π is employed the average cost is $Ap + q^G Dp/(1-q^G)$.

6. Suppose that at time t a machine is in state x (where x is a non-negative integer.) The machine costs cx to run until time $t+1$. With probability $a = 1 - b$ the machine is serviced and so goes to state 0 at time $t+1$. If it is not serviced then the machine will be in states x or $x+1$ at time $t+1$ 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. Thus $F(x)$ obeys the equation

$$F(x) = cx + a\beta F(0) + b\beta[pF(x) + qF(x+1)].$$

Show that there is a solution $F(x) = \phi + \theta x$ and determine the coefficients ϕ, θ.

A maintenance engineer must divide his time between n such machines, the i the machine having parameters c_i, p_i and state variable x_i. Suppose he uses a non-deterministic stationary Markov policy in which he allocates his time randomly, in that he services machine i with a probability a_i at a given time, independently of machines states or of the previous history, $\sum_i a_i = 1$. 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 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 $c_i(x_i + q_i)/(1 - \beta b_i)$.

7. Customers arrive at a queue as a Poisson process of rate λ. They are served at rate $u = u(x)$, where x denotes the current size of the queue. Suppose that cost is incurred as rate $ax + bu$ where a, $b > 0$. The service rate u is the control variable. The dynamic programming equation can be viewed as the limit as $\delta \to 0$ of the discrete time average cost optimality equation

$$f(x) + \delta\gamma \quad = \quad \inf_u \left\{(ax + bu)\delta + \lambda\delta f(x+1) + 1_{x>0}u\delta f(x-1) + \Big(1 - \delta[\lambda + 1_{x>0}u]\Big) f(x)\right\}$$

where γ denotes the average rate at which cost is incurred under the optimal policy and where $f(x)$ denotes the extra cost associated with starting from state x. (Here $1_{x>0} = 0$ if $x = 0$, and $1_{x>0} = 1$ if $x = 1, 2, 3, \dots$) In the $\delta \to 0$ limit this gives

$$\gamma = \inf_u \{ax + bu + \lambda[f(x+1) - f(x)] + u1_{x>0}[f(x-1) - f(x)]\}\,.$$

Check that you understand the above derivation.

Show that under the constraint that $u(x)$ is a fixed constant, say $u(x) = \bar{u}$, independent of x, and greater that λ then, for some C, there is a solution of the form

$$\gamma = \frac{a\lambda}{\bar{u} - \lambda} + b\bar{u}, \quad f(x) = C + \frac{ax(x+1)}{2(\bar{u} - \lambda)}.$$

i.e., such that $f(x)$ does not grow exponentially in x (which is needed to ensure that $(1/t)Ef(x_t) \to 0$ as $t \to \infty$ and hence, similarly as in the proof for a discrete time model, that γ can be shown to be the time-average cost.)

Let π be the policy $u(x) = \bar{u}$, with optimal value of $\bar{u}$. What is $\bar{u}$?

Suppose now that we allow u to vary with x, subject to the constraint $m \le u \le M$, where $M > \lambda$. What is the policy which results if we carry out one stage of policy improvement to π.

Questions 8–10 are about optimal control in LQ models, which is the subject of Lectures 10. However, you should be able to do all the following 3 questions using only what you know about dynamic programming from Lectures 1 and 2.

8. Successive attempts are made to regulate the speed of a clock, but these introduce also a random change whose size tends to increase with the size of the intended change. Explicitly, let x_n be the error in the speed of the clock after n corrections. On the basis of the observed value of x_n one attempts to correct the speed by an amount u_n. The actual error in speed then becomes

$$x_{n+1} = x_n - u_n + \epsilon_{n+1}$$

where, conditional on events up to the choice of u_n, the variable ϵ_{n+1} is normally distributed with zero mean and variance αu_n^2. If, after all attempts at regulation, one leaves the clock with an error x, then there is a cost x^2.

Suppose exactly s attempts are to be made to regulate the clock with initial error x. Determine the optimal policy and the minimal expected cost.

9. Consider a scalar deterministic linear system, $x_t = Ax_{t-1} + Bu_{t-1}$, with cost function $\sum_{t=0}^{h-1} Qu_t^2 + x_h^2$. Show from first principles (i.e., not simply by substituting values into the Riccati equation), that in terms of the time to go s, Π_s^{-1} obeys a linear recurrence and that

$$\Pi_s = \left[\frac{B^2}{Q(A^2-1)} + \left(1 - \frac{B^2}{Q(A^2-1)}\right) A^{-2s}\right]^{-1}.$$

Under what conditions does Π_s tend to a limit as $s \to \infty$? What are the limiting forms of Π_s and Γ_s?

10. Consider the linear system, $(x_t, v_t) \in \mathbb{R}^2$,

$$\begin{aligned} x_{t+1} &= x_t + v_t \\ v_{t+1} &= v_t + u_t + \epsilon_t, \end{aligned}$$

whose state represents the position and velocity of a body, $\{u_t\}$ is a sequence of control variables and $\{\epsilon_t\}$ is a sequence of independent zero-mean disturbances, with variance N. We wish to minimize the expected value of $\sum_{t=0}^{T-1} u_t^2 + P_0 x_T^2$. Show that the optimal choice of u_t from state (x_t, v_t) is

$$u_t = -(s-1)P_s(x_t + sv_t),$$

where $s = T - t$ and

$$P_s^{-1} = P_0^{-1} + \tfrac{1}{6}s(s-1)(2s-1).$$

[Hint: reduce this problem to LQ regulation of the scalar variable $z_t = x_t + (T-t)v_t$. Re-write the plant equation and cost in terms of this quantity and in terms of time to go.]

MATHEMATICAL TRIPOS: PART II **Lent Term 2016**

OPTIMIZATION AND CONTROL **Richard Weber**

Example Sheet 3

1. A one-dimensional model of the problem faced by a juggler trying to balance a light stick with a weight on top is given by the equation

$$\ddot{x}_1 = \alpha(x_1 - u)$$

where x_1 is the horizontal displacement of the top of the stick from some fixed point and u is the horizontal displacement of the bottom. (The stick is assumed to be nearly upright and stationary and $\alpha > 0$ is inversely proportional to the length.) Show the juggler can control x_1 by manipulating u.

If he tries to balance n such weighted sticks on top of one another, the equations governing stick k ($k = 2, \dots, n$) are (provided the weights on the sticks get smaller fast enough as n increases)

$$\ddot{x}_k = \alpha(x_k - x_{k-1})$$

Show that the n-stick system is controllable. [You may find it helpful to take the state vector as $(\dot{x}_1, x_1, \dot{x}_2, x_2, \dots, \dot{x}_n, x_n)^\top$.]

2. Consider the controlled system $x_{t+1} = x_t + u_t + 3\epsilon_{t+1}$, where the ϵ_t are independent $N(0,1)$ variables. The instantaneous cost at time t is $x_t^2 + 2u_t^2$. Assuming that x_t is observable at time t, show that the optimal control under steady-state (stationary) conditions is $u_t = -(1/2)x_t$, and that when this control is used the average-cost incurred per unit time is 18.

Suppose now that at time t one observes, not x_t, but $y_t = x_{t-1} + 2\eta_t$, where the η_t are again independent $N(0,1)$ variables independent of the ϵ_t. The law of $\hat{x}_{t+1}$ conditional on $(y_1, \dots, y_{t+1})$ is Gaussian, with a mean that is a the linear function of $\hat{x}_t$, u_t and y_{t+1} having minimum variance. Find under steady-state conditions this linear function, and show that $\hat{x}_t$ has steady-state variance 12.

Assuming stationary conditions, express the optimal control, u_t, as a function of $\hat{x}_t$.

3. Consider the continuous-time system with scalar state variable, plant equation $\dot{x} = u$ and cost function $Q\int_0^h u^2 dt + x(h)^2$. By writing the DP equation in infinitesimal form and taking the appropriate limit, show that the value function satisfies

$$0 = \frac{\partial F}{\partial t} + \inf_u \left[Qu^2 + \frac{\partial F}{\partial x} u \right], \quad s > 0.$$

Show that F and the optimal control with time s to go are

$$F = \frac{Qx^2}{Q+s}, \qquad u = -\frac{x}{Q+s}.$$

4. Miss Prout holds the entire remaining stock of Cambridge elderberry wine for the vintage year 1959. If she releases it at rate u (in continuous time) she realises a unit price $p(u)$. She holds an amount x at time 0 and wishes to release this in such a way as to maximize her total discounted return, $\int_0^\infty e^{-\alpha t} u p(u)\, dt$. Consider the particular case $p(u) = u^{-\gamma}$, where the constant γ is positive and less than one. Show that the value function is proportional to a power of x and determine the optimal release rule in closed-loop form (i.e., as a function of the present stock level.)

[Hint: The answers are $F(x) = (\gamma/\alpha)^\gamma x^{1-\gamma}$, $u = \alpha x/\gamma$. Try to derive these answers from the DP equation; not simply substitute them into the DP equation and check that they work.]

5. Let the vector x denote the Cartesian co-ordinates of a particle moving in $\mathbb{R}^d$. When at position x the particle moves with speed $v(x)$ and in a direction that can be chosen. The equation of motion is thus $\dot{x} = v(x)u$, where u is a unit vector to be chosen afresh at each position x. Let $F(x)$ denote the minimal time taken for the particle to reach a set $\mathcal{D}$ from a point x outside it. Show that after minimizing over u the dynamic programming equation for F implies that $|\nabla F(x)| = v(x)^{-1}$; i.e.,

$$\sum_{j=1}^{d} \left(\frac{\partial F}{\partial x_j} \right)^2 = v(x)^{-2}.$$

This is the *eikonal equation* of geometric optics; a short-wavelength form of the wave equation. How is the optimal direction at a given point determined from F?

[Hint: Show that the DP equation is $\inf_{u:|u|=1}[1 + v(x)u^\top \nabla F] = 0$. Then use a Cauchy-Schwartz inequality to show that the infimum is achieved by $u = -\nabla F/|\nabla F|$.]

6. Consider the optimal control problem:

$$\text{minimize} \quad \int_0^T \tfrac{1}{2}u(t)^2\, dt \quad \text{subject to} \quad \dot{x}_1 = -x_1 + x_2\,, \quad \dot{x}_2 = -2x_2 + u\,,$$

where u is unrestricted, $x_1(0)$ and $x_2(0)$ are known, T is given and $x_1(T)$ and $x_2(T)$ are to be made to vanish. Rewrite the problem in terms of new variables, $z_1 = (x_1 + x_2)e^t$ and $z_2 = x_2 e^{2t}$ and then show that the optimal control takes the form $u = \lambda_1 e^t + \lambda_2 e^{2t}$, for some constants λ_1 and λ_2. Find equations for $x_1(0)$, $x_2(0)$ in terms of λ_1, λ_2, and T, which you could in principle solve for λ_1, λ_2 in terms of $x_1(0)$, $x_2(0)$ and T.

Compare a linear feedback controller of the form $u(t) = -k_1 x_1(t) - k_2 x_2(t)$, where k_1 and k_2 are constants. Show that with this controller x_1 and x_2 cannot be made to vanish in finite time. Discuss the choice of optimal control with a quadratic performance criterion as opposed to linear feedback control, indicating which is likely to be more appropriate in given circumstances.

7. A princess is jogging with speed r in the counterclockwise direction around a circular running track of radius r, and so has a position whose horizontal and vertical components at time t are $(r\cos t, r\sin t)$, $t \geq 0$. A monster who is initially located at the centre of the circle can move with velocity u_1 in the horizontal direction and u_2 in the vertical direction, where both velocities have a maximum magnitude of 1. The monster wishes to catch the princess in minimal time.

Analyse the monster's problem using Pontryagin's maximum principle. By considering feasible values for the adjoint variables, show that whatever the value of r the monster should always set at least one of $|u_1|$ or $|u_2|$ equal to 1. Show that if $r = \pi/\sqrt{8}$ then the monster catches the princess in minimal time by adopting the uniquely optimal policy $u_1 = 1$, $u_2 = 1$. Is the optimal policy always unique?

[Hint: Let x_1 and x_2 be the differences in the horizontal and vertical directions between the positions of the monster and princess.]

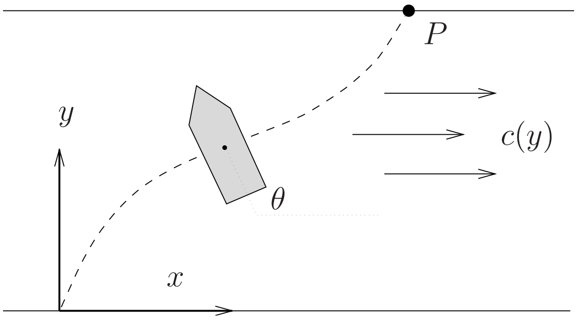
8. An aircraft flies in straight and level flight at height h, so that lift L balances weight mg. The mass rate of fuel consumption is proportional to the drag, and may be taken as $q = av^2 + b(Lv)^{-2}$, where a and b are constants and v is the speed. Thus

$$\dot{m} = -q = -av^2 - \frac{b}{(mgv)^2}.$$

Find a rule for determining v in terms of m if (i) the distance flown is to be maximized, (ii) the time spent flying at height h (until fuel is exhausted) is to be maximized.

$$\left[\text{Hint: Answers are (i) } v = \left(\frac{3b}{a(mg)^2}\right)^{1/4}, \text{ and (ii) } v = \left(\frac{b}{a(mg)^2}\right)^{1/4}\right].$$

9. In Zermelo's navigation problem a straight river has current $c(y)$, where y is the distance from the bank from which a boat is leaving. A boat is to cross the river at constant speed v relative to the water, so that its position (x, y) satisfies $\dot{x} = v\cos\theta + c(y)$, $\dot{y} = v\sin\theta$, where θ is the heading angle indicated in the diagram.



(i) Suppose $c(y) > v$ for all y and the boatman wishes to be carried downstream as little as possible in crossing. Show that he should follow the heading

$$\theta = \cos^{-1}(-v/c(y)).$$

(ii) Suppose the boatman wishes to reach a given point P on the opposite bank in minimal time. Show that he should follow the heading

$$\theta = \cos^{-1}\left(\frac{\lambda_1 v}{1 - \lambda_1 c(y)}\right),$$

where λ_1 is a parameter chosen to make his path pass through the target point.

10. In the neoclassical economic growth model, x is the existing capital per worker and u is consumption of capital per worker. The plant equation is

$$\dot{x} = f(x) - \gamma x - u, \tag{5}$$

where $f(x)$ is the production per worker, and $-\gamma x$ represents depreciation of capital and change in the size of the workforce. We wish to choose u to maximize

$$\int_{t=0}^{T} e^{-\alpha t} g(u)\, dt,$$

where $g(u)$ measures utility, is strictly increasing and concave, and T is prescribed. It is convenient to take a Hamiltonian

$$H = e^{-\alpha t}\left[g(u) + \lambda(f(x) - \gamma x - u)\right],$$

thereby including a discount factor in the definition of λ and expressing F in terms of present value.

Show that the optimal control satisfies $g'(u) = \lambda$ (assuming the maximum is at a stationary point) and

$$\dot{\lambda} = (\alpha + \gamma - f')\lambda. \tag{6}$$

Hence show that the optimal consumption obeys

$$\dot{u} = \frac{1}{\sigma(u)}[f'(x) - \alpha - \gamma], \qquad \text{where} \qquad \sigma(u) = -\frac{g''(u)}{g'(u)} > 0. \tag{7}$$

(σ is called the 'elasticity of marginal utility.')

Characterise an equilibrium solution, i.e., an $x(0) = \bar{x}$ such that the optimal trajectory is $x(t) = \bar{x}$, $t \geq 0$, and show that this $\bar{x}$ is independent of g.

11. This is a starred question, which might be interesting for discussion. It points up the duality between the questions of controllability and observability.

Consider the system $x_{t+1} = Ax_t + Bu_t$, $x_t \in \mathbb{R}^n$, $u_t \in \mathbb{R}^m$, and let

$$F_t(x_0) = \min_{u_0,\dots,u_{t-1}} \sum_{s=0}^{t-1} x_s^\top R x_s \ + x_t^\top \Pi_0 x_t, ,$$

where R is positive definite. Assuming that the optimal control is of the form $u_s = K_s x_s$, and $F_t(x) = x^\top \Pi_t x$, show that

$$\Pi_t = f(R, A, B, \Pi_{t-1}) \equiv \min_K \left\{R + (A + BK)^\top \Pi_{t-1}(A + BK)\right\} .$$

Explain what is meant by saying the system is controllable.

State necessary and sufficient condition for controllability in terms of A and B.

Show that if the system is controllable and $\Pi_0 = 0$, then $F_t(x)$ is monotone increasing in t and tends to the finite limit $x^\top \Pi x$, where $\Pi = f(R, A, B, \Pi)$.

Suppose now that $x_{t+1} = Ax_t + Bu_t + \epsilon_t$, where $\{\epsilon_t\}$ is noise, $E\epsilon_t = 0$, $E\epsilon_t\epsilon_t^\top = N$, and ϵ_s and ϵ_t are independent for $s \neq t$. Moreover, x_0 is known, but $x_1, x_2, \dots$ cannot be observed. Instead, we observe $y_1, y_2, \dots \in \mathbb{R}^r$, where $y_t = Cx_{t-1}$. Consider the estimate of x_t given by

$$\hat{x}_t = A\hat{x}_{t-1} + Bu_{t-1} - H_t(y_t - C\hat{x}_{t-1})$$

where $\hat{x}_0 = x_0$ and H_t is chosen to minimize, V_t, the covariance matrix of $\hat{x}_t$. Show that $\hat{x}_t$ is unbiased and that, with $V_0 = 0$,

$$V_t = f(N, A^\top, C^\top, V_{t-1}) = \min_H \left\{N + (A + HC)V_{t-1}(A + HC)^\top\right\} .$$

Hence, quoting a condition in terms of A and C for the noiseless system to be observable, show that observability is a sufficient condition for V_t to tend to a finite limit as $t \to \infty$.