MATHEMATICAL TRIPOS: PART II **Michaelmas Term 2013**

OPTIMIZATION AND CONTROL **Frank Kelly**

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. 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 (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. The problem is to determine the policy that maximizes his expected return, i.e. r (if he stays in the system until he completes service) minus the cost of waiting.

Let F_n denote the expected return obtained by employing an optimal waiting policy when there are n people ahead. Show that the optimality equation is

$$F_n = \max[-c + pF_{n-1} + (1-p)F_n, 0], \quad n > 0, \tag{1}$$

with $F_0 = r$. Show that (2) can be re-written as

$$F_n = \max\left[F_{n-1} - c/p, 0\right], \quad n > 0. \tag{2}$$

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

Show there exists an integer n^* such that the form of the optimal policy is to wait only if $n \le n^*$. 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. 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. 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, 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 $\beta \le 1/(7p - 8p')$.

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. What should be his stopping rule?

8. Consider a burglar who 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?

The next question is about solving the optimality equations using linear programming.

9. 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

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

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 (5). 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,

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

and hence that $F \geq G$.

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

The next question is about proving a structural property of an optimal policy. Many research papers in the field are about results like this.

10. 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* $x \geq a_s$, 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 $x \leq n(s)$ and to employ the assistant if and only if $y \geq m(s)$.*

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 $s \leq t$, and that F_t and G_{t-1} are concave functions of x. First show that for $x > 0$,

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

Now, by considering the values of terms on the right have side of this expression, separately in the three cases $x+1 \leq n(t)$, $x-1 > n(t)$ and $x-1 \leq n(t) < x+1$, 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.

The final question is optional but you might it interesting. It concerns a famous unsolved problem in dynamic programming. We do a bit of it that can be solved.

11. A trapper has been disabled with a broken leg in the Canadian wilderness and his aim is to survive t nights until his partner returns to camp. On certain nights, which occur independently with probability p, he hears wolves howl as he gets ready for bed. If on such a night he leaves u candles burning around his camp, then he will survive the night with probability $1-\theta^u$, $0<\theta<1$. On the other hand, with probability θ^u, the wolves will attack and he will be savaged to death while he sleeps. Unfortunately, he has only x candles left. Let $F_s(x)$ be his survival probability under an optimal policy. Explain why, with $q=1-p$,

$$F_s(x) = qF_{s-1}(x) + p \max_{u\in\{1,\ldots,x\}} (1-\theta^u)F_{s-1}(x-u), \quad 1\le s\le t,$$

where $F_0(x)=1$. Suppose that if he hears the wolves howl and there are s nights and x candles left then he burns $u_s(x)$ candles. Explain why it is intuitively reasonable to conjecture that an optimal policy lies in the class of policies having properties that

Conjecture A. *$u_s(x)$ is non-increasing in s,*

Conjecture B. *$u_s(x)$ is non-decreasing in x,*

Conjecture C. *$x-u_s(x)$ is non-decreasing in x.*

Conjecture C is easy to prove. You might like to try it. Klinger and Brown (1968) proved that Conjecture A is true if Conjecture B is true. Seeing as no one could prove Conjecture B, this is not much help. However, this unsatisfactory state of affairs was relieved by Samuel (1970) who showed that Conjecture A is true, whether or not Conjecture B is true.

Conjecture B has been verified numerically for a wide range of values for p, θ and t. Many top notch researchers have tried to prove Conjecture B, but no one has succeeded. If you can prove (or disprove) Conjecture B you will attain fame.

Let us examine Klinger and Brown's proof that Conjecture B $\Longrightarrow$ Conjecture A. Suppose π is the *uniquely* optimal policy, but that for some given x and s, $s>1$, we have $u_s(x)=a$, $u_{s-1}(x)=b$, $u_{s-1}(x-a)=c$ where $b<a<x$, $a+c<x$. In other words, π is a counterexample to Conjecture A. However, suppose Conjecture B holds for π, so that $b\ge c$.

Consider an alternative policy $\bar\pi$ with $\bar u_s(x)=b$, $\bar u_{s-1}(x)=a$, $\bar u_{s-1}(x-b)=a-b+c$ and $\bar u=u$ otherwise. Let the survival probabilities under the two policies be $P_s(x)$ and $\bar P_s(x)$ respectively. By writing $P_s(x)$ and $\bar P_s(x)$ in terms of P_{s-2}, prove that $\bar\pi$ is at least as good as π.

Hence deduce that if the optimal policy is unique and satisfies Conjecture B then it must also satisfy Conjecture A.

A continuous-time version of this problem was originally posed by researchers at the RAND Corporation in 1965, motivated by a cold-war problem about a nuclear bomber that is armed with anti-aircraft missiles and wishes to reach its target while fending off potential attacks by enemy fighter jets. For this reason the problem is commonly known as 'the bomber problem'. You can read more about it in these slides for a talk that Professor Richard Weber gave to the Adams Society:
`www.statslab.cam.ac.uk/~rrw1/talks/adams.pdf`

MATHEMATICAL TRIPOS: PART II **Michaelmas Term 2013**

OPTIMIZATION AND CONTROL **Frank Kelly**

Example Sheet 2

1. 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?

2. 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

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

3. 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.

4. 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. Show that $F(x)$ has the linear form $\phi + \theta x$ 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 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 $c_i(x_i + q_i)/(1-\beta b_i)$.

5. 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 in the infinite horizon limit is then

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

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 $x = 1, 2, 3, \dots$.) Give a brief justification of this equation.

Show that under the constraint that u is a fixed constant, independent of x, and greater that λ then, for some C, there is a solution of the form

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

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.) What is the optimal constant service rate?

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 the optimal constant service policy?

6. 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?

7. 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.

8. Consider the scalar-state control problem with plant equation $x_{t+1} = x_t + u_t + \epsilon_t$ and cost function $\sum_{t=0}^{h-1} u_t^2 + Dx_h^2$. Here current state is observable, the horizon point h is prescribed, and the disturbances ϵ_t are i.i.d. with zero mean and variance v. Show that the open-loop form of the optimal control in the deterministic case $v = 0$ is $u_t = -Dx_0/(1 + hD)$ and that the closed-loop form of the optimal control is $u_t = -Dx_t/[1 + (h - t)D]$, whatever v.

Show that if the open-loop control is used in the stochastic case then a total expected cost $Dx_0^2/(1+hD) + hDv$ is incurred, while use of the closed-loop control leads to a smaller expected cost of

$$F(x_0, 0) = \frac{Dx_0^2}{1 + hD} + Dv \sum_{s=0}^{h-1} \frac{1}{1 + sD}.$$

9. Consider the real-valued system defined by

$$x_{n+1} = ax_n + \xi_n u_n \quad (n = 0, 1, \dots),$$

where u_t is the control at time n and $\{\xi_n; n = 0, 1, \dots\}$ is a sequence of independent random variables with mean b and variance σ^2. Suppose that the cost incurred at time n is $x_n^2 + u_n^2$, and that there are no terminal costs. Find the recursions satisfied by the finite-horizon optimal cost function. Is the optimal control certainty-equivalent control?

$$\left[\text{Hint: The answer is } F_s(x) = \Pi_s x^2, \text{ where } \; \Pi_s = 1 + \frac{a^2 \Pi_{s-1}(1 + \sigma^2 \Pi_{s-1})}{1 + (b^2 + \sigma^2)\Pi_{s-1}}.\right]$$

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.]

11. 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 that 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 ($k = 2, \ldots, n$) 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 $(\dot{x}_1, x_1, \dot{x}_2, x_2, \ldots, \dot{x}_n, x_n)^\top$. Example F.]

MATHEMATICAL TRIPOS: PART II **Michaelmas Term 2013**

OPTIMIZATION AND CONTROL **Frank Kelly**

Example Sheet 3

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

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

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 $x_1, x_2, \ldots$ cannot be observed. Instead, we observe $y_1, y_2, \ldots \in \mathbb{R}^r$, 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 $t \to \infty$.

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, calculate the optimal control under steady-state (stationary) conditions and find the expected cost per unit time incurred when this control is used.

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. Show that the law of $\hat{x}_t$ conditional on $(y_1, \ldots, y_t)$ has steady-state variance 12.

Find the optimal control and a recursion for the optimal state estimate under stationary conditions.

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, $\int_0^\infty e^{-\alpha t} up(u)\, dt$. 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 (proposed in 1931) 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.

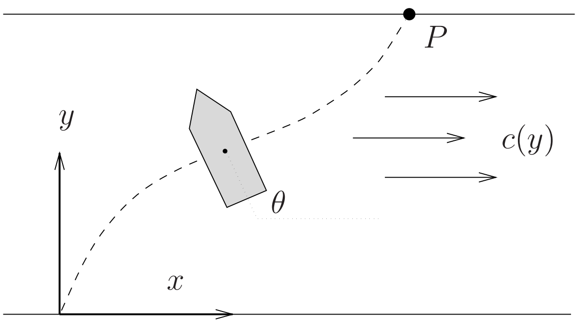

Figure 1: Zermelo's navigation problem

(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

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

(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.