### Paper 4, Section II
### 25K Optimization and Control

Consider the scalar system evolving as

$$x_t = x_{t-1} + u_{t-1} + \epsilon_t, \quad t = 1, 2, \ldots,$$

where $\{\epsilon_t\}_{t=1}^{\infty}$ is a white noise sequence with $E\epsilon_t = 0$ and $E\epsilon_t^2 = v$. It is desired to choose controls $\{u_t\}_{t=0}^{h-1}$ to minimize $E\left[\sum_{t=0}^{h-1}\left(\frac{1}{2}x_t^2 + u_t^2\right) + x_h^2\right]$. Show that for $h = 6$ the minimal cost is $x_0^2 + 6v$.

Find a constant $\lambda$ and a function $\phi$ which solve

$$\phi(x) + \lambda = \min_u \left[\tfrac{1}{2}x^2 + u^2 + E\phi(x + u + \epsilon_1)\right].$$

Let $P$ be the class of those policies for which every $u_t$ obeys the constraint $(x_t + u_t)^2 \leqslant (0.9)x_t^2$. Show that $E_\pi\phi(x_t) \leqslant x_0^2 + 10v$, for all $\pi \in P$. Find, and prove optimal, a policy which over all $\pi \in P$ minimizes

$$\lim_{h\to\infty} \frac{1}{h} E_\pi \left[\sum_{t=0}^{h-1}\left(\tfrac{1}{2}x_t^2 + u_t^2\right)\right].$$

### Paper 3, Section II
### 25K Optimization and Control

A burglar having wealth $x$ may retire, or go burgling another night, in either of towns 1 or 2. If he burgles in town $i$ then with probability $p_i = 1 - q_i$ he will, independently of previous nights, be caught, imprisoned and lose all his wealth. If he is not caught then his wealth increases by 0 or $2a_i$, each with probability 1/2 and independently of what happens on other nights. Values of $p_i$ and $a_i$ are the same every night. He wishes to maximize his expected wealth at the point he retires, is imprisoned, or $s$ nights have elapsed.

Using the dynamic programming equation

$$F_s(x) = \max\Big\{ x,\, q_1 E F_{s-1}(x + R_1),\, q_2 E F_{s-1}(x + R_2) \Big\}$$

with $R_j$, $F_0(x)$ appropriately defined, prove that there exists an optimal policy under which he burgles another night if and only if his wealth is less than $x^* = \max_i\{a_i q_i / p_i\}$.

Suppose $q_1 > q_2$ and $q_1 a_1 > q_2 a_2$. Prove that he should never burgle in town 2.

[*Hint: Suppose $x < x^*$, there are $s$ nights to go, and it has been shown that he ought not burgle in town 2 if less than $s$ nights remain. For the case $a_2 > a_1$, separately consider subcases $x + 2a_2 \geqslant x^*$ and $x + 2a_2 < x^*$. An interchange argument may help.*]

**Paper 2, Section II**

**26K Optimization and Control**

As a function of policy $\pi$ and initial state $x$, let

$$F(\pi, x) = E_\pi \left[ \sum_{t=0}^{\infty} \beta^t r(x_t, u_t) \;\middle|\; x_0 = x \right],$$

where $\beta \geqslant 1$ and $r(x,u) \geqslant 0$ for all $x, u$. Suppose that for a specific policy $\pi$, and all $x$,

$$F(\pi, x) = \sup_u \Big\{ r(x,u) + \beta E[F(\pi, x_1) \mid x_0 = x, u_0 = u] \Big\}.$$

Prove that $F(\pi, x) \geqslant F(\pi', x)$ for all $\pi'$ and $x$.

A gambler plays games in which he may bet 1 or 2 pounds, but no more than his present wealth. Suppose he has $x_t$ pounds after $t$ games. If he bets $i$ pounds then $x_{t+1} = x_t + i$, or $x_{t+1} = x_t - i$, with probabilities $p_i$ and $1 - p_i$ respectively. Gambling terminates at the first $\tau$ such that $x_\tau = 0$ or $x_\tau = 100$. His final reward is $(9/8)^{\tau/2} x_\tau$. Let $\pi$ be the policy of always betting 1 pound. Given $p_1 = 1/3$, show that $F(\pi, x) \propto x 2^{x/2}$.

Is $\pi$ optimal when $p_2 = 1/4$?

### Paper 4, Section II
### 28J Optimization and Control

A girl begins swimming from a point $(0,0)$ on the bank of a straight river. She swims at a constant speed $v$ relative to the water. The speed of the downstream current at a distance $y$ from the shore is $c(y)$. Hence her trajectory is described by

$$\dot{x} = v\cos\theta + c(y)\,, \quad \dot{y} = v\sin\theta\,,$$

where $\theta$ is the angle at which she swims relative to the direction of the current.

She desires to reach a downstream point $(1,0)$ on the same bank as she starts, as quickly as possible. Construct the Hamiltonian for this problem, and describe how Pontryagin's maximum principle can be used to give necessary conditions that must hold on an optimal trajectory. Given that $c(y)$ is positive, increasing and differentiable in $y$, show that on an optimal trajectory

$$\frac{d}{dt}\tan\big(\theta(t)\big) = -c'\big(y(t)\big)\,.$$

### Paper 3, Section II
### 28J Optimization and Control

A particle follows a discrete-time trajectory on $\mathbb{R}$ given by

$$x_{t+1} = Ax_t + \xi_t u_t + \epsilon_t$$

for $t = 1,2,\ldots,T$, where $T \geqslant 2$ is a fixed integer, $A$ is a real constant, $x_t$ is the position of the particle and $u_t$ is the control action at time $t$, and $(\xi_t, \epsilon_t)_{t=1}^{T}$ is a sequence of independent random vectors with $\mathbb{E}\,\xi_t = \mathbb{E}\,\epsilon_t = 0$, $\mathrm{var}(\xi_t) = V_\xi > 0$, $\mathrm{var}(\epsilon_t) = V_\epsilon > 0$ and $\mathrm{cov}(\xi_t, \epsilon_t) = 0$.

Find the closed-loop control, i.e. the control action $u_t$ defined as a function of $(x_1, \ldots, x_t; u_1, \ldots, u_{t-1})$, that minimizes

$$\sum_{t=1}^{T} x_t^2 + c\sum_{t=1}^{T-1} u_t\,,$$

where $c > 0$ is given. [Note that this function is quadratic in $x$, but linear in $u$.]

Does the closed-loop control depend on $V_\epsilon$ or on $V_\xi$? Deduce the form of the optimal open-loop control.

## Paper 2, Section II

### 29J Optimization and Control

Describe the elements of a discrete-time stochastic dynamic programming equation for the problem of maximizing the expected sum of non-negative rewards over an infinite horizon. Give an example to show that there may not exist an optimal policy. Prove that if a policy has a value function that satisfies the dynamic programming equation then the policy is optimal.

A squirrel collects nuts for the coming winter. There are plenty of nuts lying around, but each time the squirrel leaves its lair it risks being caught by a predator. Assume that the outcomes of the squirrel's journeys are independent, that it is caught with probability $p$, and that it returns safely with a random weight of nuts, exponentially distributed with parameter $\lambda$. By solving the dynamic programming equation for the value function $F(x)$, find a policy maximizing the expected weight of nuts collected for the winter. Here the state variable $x$ takes values in $\mathbb{R}_+$ (the weight of nuts so far collected) or $-1$ (a no-return state when the squirrel is caught).

### Paper 4, Section II

### 28K Optimization and Control

Given $r, \rho, \mu, T$, all positive, it is desired to choose $u(t) > 0$ to maximize

$$\mu x(T) + \int_0^T e^{-\rho t} \log u(t)\, dt$$

subject to $\dot{x}(t) = rx(t) - u(t)$, $x(0) = 10$.

Explain what Pontryagin's maximum principle guarantees about a solution to this problem.

Show that no matter whether $x(T)$ is constrained or unconstrained there is a constant $\alpha$ such that the optimal control is of the form $u(t) = \alpha e^{-(\rho - r)t}$. Find an expression for $\alpha$ under the constraint $x(T) = 5$.

Show that if $x(T)$ is unconstrained then $\alpha = (1/\mu)e^{-rT}$.

### Paper 3, Section II

### 28K Optimization and Control

A particle follows a discrete-time trajectory in $\mathbb{R}^2$ given by

$$\begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_t \\ y_t \end{pmatrix} + \begin{pmatrix} t \\ 1 \end{pmatrix} u_t + \begin{pmatrix} \epsilon_t \\ 0 \end{pmatrix},$$

where $\{\epsilon_t\}$ is a white noise sequence with $E\epsilon_t = 0$ and $E\epsilon_t^2 = v$. Given $(x_0, y_0)$, we wish to choose $\{u_t\}_{t=0}^9$ to minimize $C = E\left[x_{10}^2 + \sum_{t=0}^9 u_t^2\right]$.

Show that for some $\{a_t\}$ this problem can be reduced to one of controlling a scalar state $\xi_t = x_t + a_t y_t$.

Find, in terms of $x_0, y_0$, the optimal $u_0$. What is the change in minimum $C$ achievable when the system starts in $(x_0, y_0)$ as compared to when it starts in $(0, 0)$?

Consider now a trajectory starting at $(x_{-1}, y_{-1}) = (11, -1)$. What value of $u_{-1}$ is optimal if we wish to minimize $5u_{-1}^2 + C$?

## Paper 2, Section II

### 29K Optimization and Control

Suppose $\{x_t\}_{t\geqslant 0}$ is a Markov chain. Consider the dynamic programming equation

$$F_s(x) = \max\Big\{r(x), \beta E\big[F_{s-1}(x_1) \mid x_0 = x\big]\Big\}, \quad s = 1, 2, \ldots,$$

with $r(x) > 0$, $\beta \in (0,1)$, and $F_0(x) = 0$. Prove that:

(i) $F_s(x)$ is nondecreasing in $s$;

(ii) $F_s(x) \leqslant F(x)$, where $F(x)$ is the value function of an infinite-horizon problem that you should describe;

(iii) $F_\infty(x) = \lim_{s\to\infty} F_s(x) = F(x)$.

A coin lands heads with probability $p$. A statistician wishes to choose between: $H_0 : p = 1/3$ and $H_1 : p = 2/3$, one of which is true. Prior probabilities of $H_1$ and $H_0$ in the ratio $x : 1$ change after one toss of the coin to ratio $2x : 1$ (if the toss was a head) or to ratio $x : 2$ (if the toss was a tail). What problem is being addressed by the following dynamic programming equation?

$$F(x) = \max\Big\{\tfrac{1}{1+x}, \tfrac{x}{1+x}, \beta\left[\left(\tfrac{1}{1+x}\tfrac{2}{3} + \tfrac{x}{1+x}\tfrac{1}{3}\right)F(x/2) + \left(\tfrac{1}{1+x}\tfrac{1}{3} + \tfrac{x}{1+x}\tfrac{2}{3}\right)F(2x)\right]\Big\}.$$

Prove that $G(x) = (1+x)F(x)$ is a convex function of $x$.

By sketching a graph of $G$, describe the form of the optimal policy.

**Paper 3, Section II**

**28J Optimization and Control**

A state variable $x = (x_1, x_2) \in \mathbb{R}^2$ is subject to dynamics

$$\begin{aligned}\dot{x}_1(t) &= x_2(t)\\ \dot{x}_2(t) &= u(t),\end{aligned}$$

where $u = u(t)$ is a scalar control variable constrained to the interval $[-1, 1]$. Given an initial value $x(0) = (x_1, x_2)$, let $F(x_1, x_2)$ denote the minimal time required to bring the state to $(0, 0)$. Prove that

$$\max_{u \in [-1,1]} \left\{ -x_2 \frac{\partial F}{\partial x_1} - u \frac{\partial F}{\partial x_2} - 1 \right\} = 0\,.$$

Explain how this equation figures in Pontryagin's maximum principle.

Use Pontryagin's maximum principle to show that, on an optimal trajectory, $u(t)$ only takes the values 1 and $-1$, and that it makes at most one switch between them.

Show that $u(t) = 1$, $0 \leqslant t \leqslant 2$ is optimal when $x(0) = (2, -2)$.

Find the optimal control when $x(0) = (7, -2)$.

**Paper 4, Section II**

**28J Optimization and Control**

A factory has a tank of capacity $3\,\mathrm{m}^3$ in which it stores chemical waste. Each week the factory produces, independently of other weeks, an amount of waste that is equally likely to be 0, 1, or 2 $\mathrm{m}^3$. If the amount of waste exceeds the remaining space in the tank then the excess must be specially handled at a cost of £$C$ per $\mathrm{m}^3$. The tank may be emptied or not at the end of each week. Emptying costs £$D$, plus a variable cost of £$\alpha$ for each $\mathrm{m}^3$ of its content. It is always emptied when it ends the week full.

It is desired to minimize the average cost per week. Write down equations from which one can determine when it is optimal to empty the tank.

Find the average cost per week of a policy $\pi$, which empties the tank if and only if its content at the end of the week is 2 or $3\,\mathrm{m}^3$.

Describe the policy improvement algorithm. Explain why, starting from $\pi$, this algorithm will find an optimal policy in at most three iterations.

Prove that $\pi$ is optimal if and only if $C \geqslant \alpha + (4/3)D$.

## Paper 2, Section II

### 29J Optimization and Control

Describe the elements of a generic stochastic dynamic programming equation for the problem of maximizing the expected sum of discounted rewards accrued at times $0, 1, \ldots$. What is meant by the *positive case*? What is specially true in this case that is not true in general?

An investor owns a single asset which he may sell once, on any of the days $t = 0, 1, \ldots$. On day $t$ he will be offered a price $X_t$. This value is unknown until day $t$, is independent of all other offers, and *a priori* it is uniformly distributed on $[0, 1]$. Offers remain open, so that on day $t$ he may sell the asset for the best of the offers made on days $0, \ldots, t$. If he sells for $x$ on day $t$ then the reward is $x\beta^t$. Show from first principles that if $0 < \beta < 1$ then there exists $\bar{x}$ such that the expected reward is maximized by selling the first day the offer is at least $\bar{x}$.

For $\beta = 4/5$, find both $\bar{x}$ and the expected reward under the optimal policy.

Explain what is special about the case $\beta = 1$.

## Paper 2, Section II

### 29K Optimization and Control

Consider an optimal stopping problem in which the optimality equation takes the form

$$F_t(x) = \max\{r(x), E[F_{t+1}(x_{t+1})]\}, \quad t = 1, \dots, N-1,$$

$F_N(x) = r(x)$, and where $r(x) > 0$ for all $x$. Let $S$ denote the stopping set of the *one-step-look-ahead rule.* Show that if $S$ is closed (in a sense you should explain) then the one-step-look-ahead rule is optimal.

$N$ biased coins are to be tossed successively. The probability that the $i$th coin toss will show a head is known to be $p_i$ ($0 < p_i < 1$). At most once, after observing a head, and before tossing the next coin, you may guess that you have just seen the last head (i.e. that all subsequent tosses will show tails). If your guess turns out to be correct then you win £1.

Suppose that you have not yet guessed 'last head', and the $i$th toss is a head. Show that it cannot be optimal to guess that this is the last head if

$$\frac{p_{i+1}}{q_{i+1}} + \cdots + \frac{p_N}{q_N} > 1,$$

where $q_j = 1 - p_j$.

Suppose that $p_i = 1/i$. Show that it is optimal to guess that the last head is the first head (if any) to occur after having tossed at least $i^*$ coins, where $i^* \approx N/e$ when $N$ is large.

## Paper 3, Section II

### 28K Optimization and Control

An observable scalar state variable evolves as $x_{t+1} = x_t + u_t$, $t = 0, 1, \ldots$. Let controls $u_0, u_1, \ldots$ be determined by a policy $\pi$ and define

$$C_s(\pi, x_0) = \sum_{t=0}^{s-1} (x_t^2 + 2x_t u_t + 7u_t^2) \quad \text{and} \quad C_s(x_0) = \inf_{\pi} C_s(\pi, x_0)\,.$$

Show that it is possible to express $C_s(x_0)$ in terms of $\Pi_s$, which satisfies the recurrence

$$\Pi_s = \frac{6(1+\Pi_{s-1})}{7+\Pi_{s-1}}\,, \qquad s = 1, 2, \ldots,$$

with $\Pi_0 = 0$.

Deduce that $C_\infty(x_0) \geqslant 2x_0^2$. [$C_\infty(x_0)$ is defined as $\lim_{s\to\infty} C_s(x_0)$.]

By considering the policy $\pi^*$ which takes $u_t = -(1/3)(2/3)^t x_0$, $t = 0, 1, \ldots$, show that $C_\infty(x_0) = 2x_0^2$.

Give an alternative description of $\pi^*$ in closed-loop form.

## Paper 4, Section II

### 28K Optimization and Control

Describe the type of optimal control problem that is amenable to analysis using Pontryagin's Maximum Principle.

A firm has the right to extract oil from a well over the interval $[0, T]$. The oil can be sold at price £$p$ per unit. To extract oil at rate $u$ when the remaining quantity of oil in the well is $x$ incurs cost at rate £$u^2/x$. Thus the problem is one of maximizing

$$\int_0^T \left[ pu(t) - \frac{u(t)^2}{x(t)} \right] dt\,,$$

subject to $dx(t)/dt = -u(t)$, $u(t) \geqslant 0$, $x(t) \geqslant 0$. Formulate the Hamiltonian for this problem.

Explain why $\lambda(t)$, the adjoint variable, has a boundary condition $\lambda(T) = 0$.

Use Pontryagin's Maximum Principle to show that under optimal control

$$\lambda(t) = p - \frac{1}{1/p + (T-t)/4}$$

and

$$\frac{dx(t)}{dt} = -\frac{2px(t)}{4 + p(T-t)}\,.$$

Find the oil remaining in the well at time $T$, as a function of $x(0)$, $p$, and $T$,

**Paper 2, Section II**
**29J Optimization and Control**

(a) Suppose that

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left( \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} V_{XX} & V_{XY} \\ V_{YX} & V_{YY} \end{pmatrix} \right).$$

Prove that conditional on $Y = y$, the distribution of $X$ is again multivariate normal, with mean $\mu_X + V_{XY} V_{YY}^{-1}(y - \mu_Y)$ and covariance $V_{XX} - V_{XY} V_{YY}^{-1} V_{YX}$.

(b) The $\mathbb{R}^d$-valued process $X$ evolves in discrete time according to the dynamics

$$X_{t+1} = AX_t + \varepsilon_{t+1},$$

where $A$ is a constant $d \times d$ matrix, and $\varepsilon_t$ are independent, with common $N(0, \Sigma_\varepsilon)$ distribution. The process $X$ is not observed directly; instead, all that is seen is the process $Y$ defined as

$$Y_t = CX_t + \eta_t,$$

where $\eta_t$ are independent of each other and of the $\varepsilon_t$, with common $N(0, \Sigma_\eta)$ distribution.

If the observer has the prior distribution $X_0 \sim N(\hat{X}_0, V_0)$ for $X_0$, prove that at all later times the distribution of $X_t$ conditional on $\mathcal{Y}_t \equiv (Y_1, \ldots, Y_t)$ is again normally distributed, with mean $\hat{X}_t$ and covariance $V_t$ which evolve as

$$\begin{aligned} \hat{X}_{t+1} &= A\hat{X}_t + M_t C^T (\Sigma_\eta + C M_t C^T)^{-1} (Y_{t+1} - CA\hat{X}_t), \\ V_{t+1} &= M_t - M_t C^T (\Sigma_\eta + C M_t C^T)^{-1} C M_t, \end{aligned}$$

where

$$M_t = AV_t A^T + \Sigma_\varepsilon.$$

(c) In the special case where both $X$ and $Y$ are one-dimensional, and $A = C = 1$, $\Sigma_\varepsilon = 0$, find the form of the updating recursion. Show in particular that

$$\frac{1}{V_{t+1}} = \frac{1}{V_t} + \frac{1}{\Sigma_\eta}$$

and that

$$\frac{\hat{X}_{t+1}}{V_{t+1}} = \frac{\hat{X}_t}{V_t} + \frac{Y_{t+1}}{\Sigma_\eta}.$$

Hence deduce that, with probability one,

$$\lim_{t \to \infty} \hat{X}_t = \lim_{t \to \infty} t^{-1} \sum_{j=1}^{t} Y_j.$$

## Paper 3, Section II

### 28J Optimization and Control

Consider an infinite-horizon controlled Markov process having per-period costs $c(x,u) \geqslant 0$, where $x \in \mathcal{X}$ is the state of the system, and $u \in \mathcal{U}$ is the control. Costs are discounted at rate $\beta \in (0,1]$, so that the objective to be minimized is

$$\mathbb{E}\Bigg[\sum_{t \geqslant 0} \beta^t c(X_t, u_t) \mid X_0 = x \Bigg].$$

What is meant by a *policy* $\pi$ for this problem?

Let $\mathcal{L}$ denote the dynamic programming operator

$$\mathcal{L}f(x) \equiv \inf_{u \in \mathcal{U}} \big\{ \; c(x,u) + \beta \mathbb{E}\big[ f(X_1) \,\big|\, X_0 = x, u_0 = u \big] \; \big\}.$$

Further, let $F$ denote the value of the optimal control problem:

$$F(x) = \inf_{\pi} \, \mathbb{E}^{\pi}\Bigg[\sum_{t \geqslant 0} \beta^t c(X_t, u_t) \mid X_0 = x \Bigg],$$

where the infimum is taken over all policies $\pi$, and $\mathbb{E}^{\pi}$ denotes expectation under policy $\pi$. Show that the functions $F_t$ defined by

$$F_{t+1} = \mathcal{L}F_t \quad (t \geqslant 0), \qquad F_0 \equiv 0$$

increase to a limit $F_\infty \in [0, \infty]$. Prove that $F_\infty \leqslant F$. Prove that $F = \mathcal{L}F$.

Suppose that $\Phi = \mathcal{L}\Phi \geqslant 0$. Prove that $\Phi \geqslant F$.

[You may assume that there is a function $u_* : \mathcal{X} \to \mathcal{U}$ such that

$$\mathcal{L}\Phi(x) = c(x, u_*(x)) + \beta \mathbb{E}\big[ \Phi(X_1) \,\big|\, X_0 = x, u_0 = u_*(x) \big],$$

though the result remains true without this simplifying assumption.]

## Paper 4, Section II

## 28J Optimization and Control

Dr Seuss' wealth $x_t$ at time $t$ evolves as

$$\frac{dx}{dt} = rx_t + \ell_t - c_t,$$

where $r > 0$ is the rate of interest earned, $\ell_t$ is his intensity of working ($0 \leqslant \ell \leqslant 1$), and $c_t$ is his rate of consumption. His initial wealth $x_0 > 0$ is given, and his objective is to maximize

$$\int_0^T U(c_t, \ell_t)\, dt,$$

where $U(c, \ell) = c^\alpha (1 - \ell)^\beta$, and $T$ is the (fixed) time his contract expires. The constants $\alpha$ and $\beta$ satisfy the inequalities $0 < \alpha < 1$, $0 < \beta < 1$, and $\alpha + \beta > 1$. At all times, $c_t$ must be non-negative, and his final wealth $x_T$ must be non-negative. Establish the following properties of the optimal solution $(x^*, c^*, \ell^*)$:

(i) $\beta c_t^* = \alpha(1 - \ell_t^*)$;

(ii) $c_t^* \propto e^{-\gamma r t}$, where $\gamma \equiv (\beta - 1 + \alpha)^{-1}$;

(iii) $x_t^* = Ae^{rt} + Be^{-\gamma r t} - r^{-1}$ for some constants $A$ and $B$.

Hence deduce that the optimal wealth is

$$x_t^* = \frac{(1 - e^{-\gamma r T}(1 + rx_0))e^{rt} + ((1 + rx_0)e^{rT} - 1)e^{-\gamma r t}}{r(e^{rT} - e^{-\gamma r T})} - \frac{1}{r}.$$

## Paper 2, Section II

### 29I Optimization and Control

In the context of stochastic dynamic programming, explain what is meant by an *average-reward optimal policy.*

A player has a fair coin and a six-sided die. At each epoch he may choose either to toss the coin or to roll the die. If he tosses the coin and it shows heads then he adds 1 to his total score, and if it shows tails then he adds 0. If he rolls the die then he adds the number showing. He wins a reward of £1 whenever his total score is divisible by 3.

Suppose the player always tosses the coin. Find his average reward per toss.

Still using the above policy, and given that he starts with a total score of $x$, let $F_s(x)$ be the expected total reward over the next $s$ epochs. Find the value of

$$\lim_{s\to\infty}\big[F_s(x)-F_s(0)\big].$$

Use the policy improvement algorithm to find a policy that produces a greater average reward than the policy of only tossing the coin.

Find the average-reward optimal policy.

## Paper 3, Section II

### 28I Optimization and Control

Two scalar systems have dynamics

$$x_{t+1}=x_t+u_t+\epsilon_t, \qquad y_{t+1}=y_t+w_t+\eta_t,$$

where $\{\epsilon_t\}$ and $\{\eta_t\}$ are independent sequences of independent and identically distributed random variables of mean 0 and variance 1. Let

$$F(x)=\inf_{\pi}\mathbb{E}\left[\sum_{t=0}^{\infty}\Big(x_t^2+u_t^2\Big)(2/3)^t \,\middle|\, x_0=x\right],$$

where $\pi$ is a policy in which $u_t$ depends on only $x_0,\ldots,x_t$.

Show that $G(x)=Px^2+d$ is a solution to the optimality equation satisfied by $F(x)$, for some $P$ and $d$ which you should find.

Find the optimal controls.

State a theorem that justifies $F(x)=G(x)$.

For each of the two cases (a) $\lambda=0$ and (b) $\lambda=1$, find controls $\{u_t,w_t\}$ which minimize

$$\mathbb{E}\left[\sum_{t=0}^{\infty}\Big(x_t^2+2\lambda x_ty_t+y_t^2+u_t^2+w_t^2\Big)(2/3+\lambda/12)^t \,\middle|\, x_0=x\,,\ y_0=y\right].$$

## Paper 4, Section II

### 28I Optimization and Control

Explain how *transversality conditions* can be helpful when employing Pontryagin's Maximum Principle to solve an optimal control problem.

A particle in $\mathbb{R}^2$ starts at (0, 0.5) and follows the dynamics

$$\dot{x} = u\sqrt{|y|}\,, \qquad \dot{y} = v\sqrt{|y|}\,, \qquad t \in [0,T]\,,$$

where controls $u(t)$ and $v(t)$ are to be chosen subject to $u^2(t) + v^2(t) = 1$.

Using Pontryagin's maximum principle do the following:

(a) Find controls that minimize $-y(1)$;

(b) Suppose we wish to choose $T$ and the controls $u, v$ to minimize $-y(T) + T$ under a constraint $(x(T), y(T)) = (1,1)$. By expressing both $dy/dx$ and $d^2y/dx^2$ in terms of the adjoint variables, show that on an optimal trajectory,

$$1 + \left(\frac{dy}{dx}\right)^2 + 2y\,\frac{d^2y}{dx^2} = 0\,.$$

2/II/29I **Optimization and Control**

Consider a stochastic controllable dynamical system $P$ with action-space $A$ and countable state-space $S$. Thus $P = (p_{xy}(a) : x, y \in S, a \in A)$ and $p_{xy}(a)$ denotes the transition probability from $x$ to $y$ when taking action $a$. Suppose that a cost $c(x, a)$ is incurred each time that action $a$ is taken in state $x$, and that this cost is uniformly bounded. Write down the dynamic optimality equation for the problem of minimizing the expected long-run average cost.

State in terms of this equation a general result, which can be used to identify an optimal control and the minimal long-run average cost.

A particle moves randomly on the integers, taking steps of size 1. Suppose we can choose at each step a control parameter $u \in [\alpha, 1-\alpha]$, where $\alpha \in (0, 1/2)$ is fixed, which has the effect that the particle moves in the positive direction with probability $u$ and in the negative direction with probability $1-u$. It is desired to maximize the long-run proportion of time $\pi$ spent by the particle at 0. Show that there is a solution to the optimality equation for this example in which the relative cost function takes the form $\theta(x) = \mu\,|x|\,$, for some constant $\mu\,$.

Determine an optimal control and show that the maximal long-run proportion of time spent at 0 is given by

$$\pi \,=\, \frac{1-2\alpha}{2\,(1-\alpha)}\,.$$

You may assume that it is valid to use an unbounded function $\theta$ in the optimality equation in this example.

3/II/28I **Optimization and Control**

Let $Q$ be a positive-definite symmetric $m \times m$ matrix. Show that a non-negative quadratic form on $\mathbb{R}^d \times \mathbb{R}^m$ of the form

$$c(x,a) = x^T Rx + x^T S^T a + a^T Sx + a^T Qa, \quad x \in \mathbb{R}^d, \quad a \in \mathbb{R}^m,$$

is minimized over $a$, for each $x$, with value $x^T(R - S^T Q^{-1} S)x$, by taking $a = Kx$, where $K = -Q^{-1}S$.

Consider for $k \leqslant n$ the controllable stochastic linear system in $\mathbb{R}^d$

$$X_{j+1} = AX_j + BU_j + \varepsilon_{j+1}, \quad j = k, k+1, \ldots, n-1,$$

starting from $X_k = x$ at time $k$, where the control variables $U_j$ take values in $\mathbb{R}^m$, and where $\varepsilon_{k+1}, \ldots, \varepsilon_n$ are independent, zero-mean random variables, with $\operatorname{var}(\varepsilon_j) = N_j$. Here, $A$, $B$ and $N_j$ are, respectively, $d \times d$, $d \times m$ and $d \times d$ matrices. Assume that a cost $c(X_j, U_j)$ is incurred at each time $j = k, \ldots, n-1$ and that a final cost $C(X_n) = X_n^T \Pi_0 X_n$ is incurred at time $n$. Here, $\Pi_0$ is a given non-negative-definite symmetric matrix. It is desired to minimize, over the set of all controls $u$, the total expected cost $V^u(k,x)$. Write down the optimality equation for the infimal cost function $V(k,x)$.

Hence, show that $V(k,x)$ has the form

$$V(k,x) \;=\; x^T \Pi_{n-k} x + \gamma_k$$

for some non-negative-definite symmetric matrix $\Pi_{n-k}$ and some real constant $\gamma_k$. Show how to compute the matrix $\Pi_{n-k}$ and constant $\gamma_k$ and how to determine an optimal control.

4/II/29I **Optimization and Control**

State Pontryagin's maximum principle for the controllable dynamical system with state-space $\mathbb{R}^+$, given by

$$\dot{x}_t = b(t, x_t, u_t), \quad t \geqslant 0,$$

where the running costs are given by $c(t, x_t, u_t)$, up to an unconstrained terminal time $\tau$ when the state first reaches 0, and there is a terminal cost $C(\tau)$.

A company pays a variable price $p(t)$ per unit time for electrical power, agreed in advance, which depends on the time of day. The company takes on a job at time $t = 0$, which requires a total amount $E$ of electrical energy, but can be processed at a variable level of power consumption $u(t) \in [0,1]$. If the job is completed by time $\tau$, then the company will receive a reward $R(\tau)$. Thus, it is desired to minimize

$$\int_0^\tau u(t)p(t)dt - R(\tau),$$

subject to

$$\int_0^\tau u(t)dt = E, \quad u(t) \in [0,1],$$

with $\tau > 0$ unconstrained. Take as state variable the energy $x_t$ still needed at time $t$ to complete the job. Use Pontryagin's maximum principle to show that the optimal control is to process the job on full power or not at all, according as the price $p(t)$ lies below or above a certain threshold value $p^*$.

Show further that, if $\tau^*$ is the completion time for the optimal control, then

$$p^* + \dot{R}(\tau^*) \ = \ p(\tau^*)\,.$$

Consider a case in which $p$ is periodic, with period one day, where day 1 corresponds to the time interval $[0,2]$, and $p(t) = (t-1)^2$ during day 1. Suppose also that $R(\tau) = 1/(1+\tau)$ and $E = 1/2$. Determine the total energy cost and the reward associated with the threshold $p^* = 1/4$.

Hence, show that any threshold low enough to carry processing over into day 2 is suboptimal.

Show carefully that the optimal price threshold is given by $p^* = 1/4$.

2/II/29I **Optimization and Control**

State Pontryagin's maximum principle in the case where both the terminal time and the terminal state are given.

Show that $\pi$ is the minimum value taken by the integral

$$\tfrac{1}{2}\int_0^1 (u_t^2 + v_t^2)\, dt$$

subject to the constraints $x_0 = y_0 = z_0 = x_1 = y_1 = 0$ and $z_1 = 1$, where

$$\dot{x}_t = u_t, \quad \dot{y}_t = v_t, \quad \dot{z}_t = u_t y_t - v_t x_t, \quad 0 \leqslant t \leqslant 1.$$

[*You may find it useful to note the fact that the problem is rotationally symmetric about the $z$-axis, so that the angle made by the initial velocity $(\dot{x}_0, \dot{y}_0)$ with the positive $x$-axis may be chosen arbitrarily.*]

3/II/28I **Optimization and Control**

Let $P$ be a discrete-time controllable dynamical system (or Markov decision process) with countable state-space $S$ and action-space $A$. Consider the $n$-horizon dynamic optimization problem with instantaneous costs $c(k, x, a)$, on choosing action $a$ in state $x$ at time $k \leqslant n-1$, with terminal cost $C(x)$, in state $x$ at time $n$. Explain what is meant by a Markov control and how the choice of a control gives rise to a time-inhomogeneous Markov chain.

Suppose we can find a bounded function $V$ and a Markov control $u^*$ such that

$$V(k,x) \leqslant (c + PV)(k,x,a), \quad 0 \leqslant k \leqslant n-1, \quad x \in S, \quad a \in A,$$

with equality when $a = u^*(k,x)$, and such that $V(n,x) = C(x)$ for all $x$. Here $PV(k,x,a)$ denotes the expected value of $V(k+1, X_{k+1})$, given that we choose action $a$ in state $x$ at time $k$. Show that $u^*$ is an optimal Markov control.

A well-shuffled pack of cards is placed face-down on the table. The cards are turned over one by one until none are left. Exactly once you may place a bet of £1000 on the event that the next *two* cards will be red. How should you choose the moment to bet? Justify your answer.

4/II/29I **Optimization and Control**

Consider the scalar controllable linear system, whose state $X_n$ evolves by

$$X_{n+1} = X_n + U_n + \varepsilon_{n+1},$$

with observations $Y_n$ given by

$$Y_{n+1} = X_n + \eta_{n+1}.$$

Here, $U_n$ is the control variable, which is to be determined on the basis of the observations up to time $n$, and $\varepsilon_n, \eta_n$ are independent $N(0,1)$ random variables. You wish to minimize the long-run average expected cost, where the instantaneous cost at time $n$ is $X_n^2 + U_n^2$. You may assume that the optimal control in equilibrium has the form $U_n = -K\hat{X}_n$, where $\hat{X}_n$ is given by a recursion of the form

$$\hat{X}_{n+1} = \hat{X}_n + U_n + H(Y_{n+1} - \hat{X}_n),$$

and where $H$ is chosen so that $\Delta_n = X_n - \hat{X}_n$ is independent of the observations up to time $n$. Show that $K = H = (\sqrt{5}-1)/2 = 2/(\sqrt{5}+1)$, and determine the minimal long-run average expected cost. You are not expected to simplify the arithmetic form of your answer but should show clearly how you have obtained it.

**2/II/29I** **Optimization and Control**

A policy $\pi$ is to be chosen to maximize

$$F(\pi, x) = \mathbb{E}_\pi \left[ \sum_{t \geqslant 0} \beta^t r(x_t, u_t) \,\middle|\, x_0 = x \right],$$

where $0 < \beta \leqslant 1$. Assuming that $r \geqslant 0$, prove that $\pi$ is optimal if $F(\pi, x)$ satisfies the optimality equation.

An investor receives at time $t$ an income of $x_t$ of which he spends $u_t$, subject to $0 \leqslant u_t \leqslant x_t$. The reward is $r(x_t, u_t) = u_t$, and his income evolves as

$$x_{t+1} = x_t + (x_t - u_t)\varepsilon_t,$$

where $(\varepsilon_t)_{t \geqslant 0}$ is a sequence of independent random variables with common mean $\theta > 0$. If $0 < \beta \leqslant 1/(1+\theta)$, show that the optimal policy is to take $u_t = x_t$ for all $t$.

What can you say about the problem if $\beta > 1/(1+\theta)$?

3/II/28I **Optimization and Control**

A discrete-time controlled Markov process evolves according to

$$X_{t+1} = \lambda X_t + u_t + \varepsilon_t, \quad t = 0, 1, \ldots,$$

where the $\varepsilon$ are independent zero-mean random variables with common variance $\sigma^2$, and $\lambda$ is a known constant.

Consider the problem of minimizing

$$F_{t,T}(x) = \mathbb{E}\left[\sum_{j=t}^{T-1} \beta^{j-t} C(X_j, u_j) + \beta^{T-t} R(X_T)\right],$$

where $C(x,u) = \frac{1}{2}(u^2 + ax^2)$, $\beta \in (0,1)$ and $R(x) = \frac{1}{2}a_0 x^2 + b_0$. Show that the optimal control at time $j$ takes the form $u_j = k_{T-j} X_j$ for certain constants $k_i$. Show also that the minimized value for $F_{t,T}(x)$ is of the form

$$\tfrac{1}{2}a_{T-t}x^2 + b_{T-t}$$

for certain constants $a_j, b_j$. Explain how these constants are to be calculated. Prove that the equation

$$f(z) \equiv a + \frac{\lambda^2 \beta z}{1 + \beta z} = z$$

has a unique positive solution $z = a_*$, and that the sequence $(a_j)_{j \geqslant 0}$ converges monotonically to $a_*$.

Prove that the sequence $(b_j)_{j \geqslant 0}$ converges, to the limit

$$b_* \equiv \frac{\beta \sigma^2 a_*}{2(1-\beta)} .$$

Finally, prove that $k_j \to k_* \equiv -\beta a_* \lambda / (1 + \beta a_*)$.

4/II/29I **Optimization and Control**

An investor has a (possibly negative) bank balance $x(t)$ at time $t$. For given positive $x(0), T, \mu, A$ and $r$, he wishes to choose his spending rate $u(t) \geqslant 0$ so as to maximize

$$\Phi(u;\mu) \equiv \int_0^T e^{-\beta t} \log u(t)\, dt + \mu e^{-\beta T} x(T),$$

where $dx(t)/dt = A + rx(t) - u(t)$. Find the investor's optimal choice of control $u(t) = u_*(t;\mu)$.

Let $x_*(t;\mu)$ denote the optimally-controlled bank balance. By considering next how $x_*(T;\mu)$ depends on $\mu$, show that there is a unique positive $\mu_*$ such that $x_*(T;\mu_*) = 0$. If the original problem is modified by setting $\mu = 0$, but requiring that $x(T) \geqslant 0$, show that the optimal control for this modified problem is $u(t) = u_*(t;\mu_*)$.

### 2/II/29I **Optimization and Control**

Explain what is meant by a time-homogeneous discrete time Markov decision problem.

What is the positive programming case?

A discrete time Markov decision problem has state space $\{0, 1, \ldots, N\}$. In state $i$, $i \neq 0, N$, two actions are possible. We may either stop and obtain a terminal reward $r(i) \geqslant 0$, or may continue, in which case the subsequent state is equally likely to be $i-1$ or $i+1$. In states 0 and $N$ stopping is automatic (with terminal rewards $r(0)$ and $r(N)$ respectively). Starting in state $i$, denote by $V_n(i)$ and $V(i)$ the maximal expected terminal reward that can be obtained over the first $n$ steps and over the infinite horizon, respectively. Prove that $\lim_{n\to\infty} V_n = V$.

Prove that $V$ is the smallest concave function such that $V(i) \geqslant r(i)$ for all $i$.

Describe an optimal policy.

Suppose $r(0), \ldots, r(N)$ are distinct numbers. Show that the optimal policy is unique, or give a counter-example.

### 3/II/28I **Optimization and Control**

Consider the problem

$$\text{minimize } E\left[x(T)^2 + \int_0^T u(t)^2\, dt\right]$$

where for $0 \leqslant t \leqslant T$,

$$\dot{x}(t) = y(t) \quad \text{and} \quad \dot{y}(t) = u(t) + \epsilon(t)\,,$$

$u(t)$ is the control variable, and $\epsilon(t)$ is Gaussian white noise. Show that the problem can be rewritten as one of controlling the scalar variable $z(t)$, where

$$z(t) = x(t) + (T-t)y(t)\,.$$

By guessing the form of the optimal value function and ensuring it satisfies an appropriate optimality equation, show that the optimal control is

$$u(t) = -\frac{(T-t)z(t)}{1 + \frac{1}{3}(T-t)^3}\,.$$

Is this certainty equivalence control?

4/II/29I **Optimization and Control**

A continuous-time control problem is defined in terms of state variable $x(t) \in \mathbb{R}^n$ and control $u(t) \in \mathbb{R}^m$, $0 \leqslant t \leqslant T$. We desire to minimize $\int_0^T c(x,t)\,dt + K(x(T))$, where $T$ is fixed and $x(T)$ is unconstrained. Given $x(0)$ and $\dot{x} = a(x,u)$, describe further boundary conditions that can be used in conjunction with Pontryagin's maximum principle to find $x$, $u$ and the adjoint variables $\lambda_1, \ldots, \lambda_n$.

Company 1 wishes to steal customers from Company 2 and maximize the profit it obtains over an interval $[0,T]$. Denoting by $x_i(t)$ the number of customers of Company $i$, and by $u(t)$ the advertising effort of Company 1, this leads to a problem

$$\text{minimize } \int_0^T \big[x_2(t) + 3u(t)\big]\,dt\,,$$

where $\dot{x}_1 = ux_2$, $\dot{x}_2 = -ux_2$, and $u(t)$ is constrained to the interval $[0,1]$. Assuming $x_2(0) > 3/T$, use Pontryagin's maximum principle to show that the optimal advertising policy is bang-bang, and that there is just one change in advertising effort, at a time $t^*$, where

$$3\,e^{t^*} = x_2(0)(T - t^*)\,.$$

B2/15 **Optimization and Control**

A gambler is presented with a sequence of $n \geqslant 6$ random numbers, $N_1, N_2, \ldots, N_n$, one at a time. The distribution of $N_k$ is

$$P(N_k = k) = 1 - P(N_k = -k) = p\,,$$

where $1/(n-2) < p \leq 1/3$. The gambler must choose exactly one of the numbers, just after it has been presented and before any further numbers are presented, but must wait until all the numbers are presented before his payback can be decided. It costs £1 to play the game. The gambler receives payback as follows: nothing if he chooses the smallest of all the numbers, £2 if he chooses the largest of all the numbers, and £1 otherwise.

Show that there is an optimal strategy of the form "Choose the first number $k$ such that either (i) $N_k > 0$ and $k \geq n - r_0$, or (ii) $k = n-1$", where you should determine the constant $r_0$ as explicitly as you can.

B3/14 **Optimization and Control**

The strength of the economy evolves according to the equation

$$\ddot{x}_t = -\alpha^2 x_t + u_t\,,$$

where $x_0 = \dot{x}_0 = 0$ and $u_t$ is the effort that the government puts into reform at time $t$, $t \geq 0$. The government wishes to maximize its chance of re-election at a given future time $T$, where this chance is some monotone increasing function of

$$x_T - \frac{1}{2}\int_0^T u_t^2 dt\,.$$

Use Pontryagin's maximum principle to determine the government's optimal reform policy, and show that the optimal trajectory of $x_t$ is

$$x_t = \frac{t}{2}\alpha^{-2}\cos(\alpha(T-t)) - \frac{1}{2}\alpha^{-3}\cos(\alpha T)\sin(\alpha t)\,.$$

B4/14 **Optimization and Control**

Consider the deterministic dynamical system

$$\dot{x}_t = Ax_t + Bu_t$$

where $A$ and $B$ are constant matrices, $x_t \in \mathbb{R}^n$, and $u_t$ is the control variable, $u_t \in \mathbb{R}^m$. What does it mean to say that the system is *controllable*?

Let $y_t = e^{-tA}x_t - x_0$. Show that if $V_t$ is the set of possible values for $y_t$ as the control $\{u_s : 0 \leq x \leq t\}$ is allowed to vary, then $V_t$ is a vector space.

Show that each of the following three conditions is equivalent to controllability of the system.

(i) The set $\{v \in \mathbb{R}^n : v^\top y_t = 0 \text{ for all } y_t \in V_t\} = \{0\}$.

(ii) The matrix $H(t) = \int_0^t e^{-sA}BB^\top e^{-sA^\top}\,ds$ is (strictly) positive definite.

(iii) The matrix $M_n = [B \;\; AB \;\; A^2B \;\; \cdots \;\; A^{n-1}B]$ has rank $n$.

Consider the scalar system

$$\sum_{j=0}^{n} a_j \left(\frac{d}{dt}\right)^{n-j} \xi_t = u_t\,,$$

where $a_0 = 1$. Show that this system is controllable.

B2/15 **Optimization and Control**

The owner of a put option may exercise it on any one of the days $1, \dots, h$, or not at all. If he exercises it on day $t$, when the share price is $x_t$, his profit will be $p - x_t$. Suppose the share price obeys $x_{t+1} = x_t + \epsilon_t$, where $\epsilon_1, \epsilon_2, \dots$ are i.i.d. random variables for which $E|\epsilon_t| < \infty$. Let $F_s(x)$ be the maximal expected profit the owner can obtain when there are $s$ further days to go and the share price is $x$. Show that

(a) $F_s(x)$ is non-decreasing in $s$,

(b) $F_s(x) + x$ is non-decreasing in $x$, and

(c) $F_s(x)$ is continuous in $x$.

Deduce that there exists a non-decreasing sequence, $a_1, \dots, a_h$, such that expected profit is maximized by exercising the option the first day that $x_t \leqslant a_t$.

Now suppose that the option never expires, so effectively $h = \infty$. Show by examples that there may or may not exist an optimal policy of the form 'exercise the option the first day that $x_t \leqslant a$.'

B3/14 **Optimization and Control**

State Pontryagin's Maximum Principle (PMP).

In a given lake the tonnage of fish, $x$, obeys

$$dx/dt = 0.001(50 - x)x - u\,, \quad 0 < x \leqslant 50\,,$$

where $u$ is the rate at which fish are extracted. It is desired to maximize

$$\int_0^\infty u(t)e^{-0.03t}\,dt\,,$$

choosing $u(t)$ under the constraints $0 \leqslant u(t) \leqslant 1.4$, and $u(t) = 0$ if $x(t) = 0$. Assume the PMP with an appropriate Hamiltonian $H(x, u, t, \lambda)$. Now define $G(x, u, t, \eta) = e^{0.03t}H(x, u, t, \lambda)$ and $\eta(t) = e^{0.03t}\lambda(t)$. Show that there exists $\eta(t)$, $0 \leqslant t$ such that on the optimal trajectory $u$ maximizes

$$G(x, u, t, \eta) = \eta[0.001(50 - x)x - u] + u,$$

and

$$d\eta/dt = 0.002(x - 10)\eta\,.$$

Suppose that $x(0) = 20$ and that under an optimal policy it is not optimal to extract all the fish. Argue that $\eta(0) \geqslant 1$ is impossible and describe qualitatively what must happen under the optimal policy.

B4/14 **Optimization and Control**

The scalars $x_t$, $y_t$, $u_t$, are related by the equations

$$x_t = x_{t-1} + u_{t-1}\,, \quad y_t = x_{t-1} + \eta_{t-1}\,, \quad t = 1, \ldots, T\,,$$

where $\{\eta_t\}$ is a sequence of uncorrelated random variables with means of 0 and variances of 1. Given that $\hat{x}_0$ is an unbiased estimate of $x_0$ of variance 1, the control variable $u_t$ is to be chosen at time $t$ on the basis of the information $W_t$, where $W_0 = (\hat{x}_0)$ and $W_t = (\hat{x}_0, u_0, \ldots, u_{t-1}, y_1, \ldots, y_t)$, $t = 1, 2, \ldots, T-1$. Let $\hat{x}_1, \ldots, \hat{x}_T$ be the Kalman filter estimates of $x_1, \ldots, x_T$ computed from

$$\hat{x}_t = \hat{x}_{t-1} + u_{t-1} + h_t(y_t - \hat{x}_{t-1})$$

by appropriate choices of $h_1, \ldots, h_T$. Show that the variance of $\hat{x}_t$ is $V_t = 1/(1+t)$.

Define $F(W_T) = E\left[x_T^2 \mid W_T\right]$ and

$$F(W_t) = \inf_{u_t, \ldots, u_{T-1}} E\left[\sum_{\tau=t}^{T-1} u_\tau^2 + x_T^2 \,\middle|\, W_t\right]\,, \quad t = 0, \ldots, T-1\,.$$

Show that $F(W_t) = \hat{x}_t^2 P_t + d_t$, where $P_t = 1/(T-t+1)$, $d_T = 1/(1+T)$ and $d_{t-1} = V_{t-1}V_tP_t + d_t$.

How would the expression for $F(W_0)$ differ if $\hat{x}_0$ had a variance different from 1?

B2/15 **Optimization and Control**

State Pontryagin's maximum principle (PMP) for the problem of minimizing

$$\int_0^T c(x(t), u(t))\, dt + K(x(T))\,,$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, $dx/dt = a(x(t), u(t))$; here, $x(0)$ and $T$ are given, and $x(T)$ is unconstrained.

Consider the two-dimensional problem in which $dx_1/dt = x_2$, $dx_2/dt = u$, $c(x, u) = \frac{1}{2}u^2$ and $K(x(T)) = \frac{1}{2}qx_1(T)^2$, $q > 0$. Show that, by use of a variable $z(t) = x_1(t) + x_2(t)(T - t)$, one can rewrite this problem as an equivalent one-dimensional problem.

Use PMP to solve this one-dimensional problem, showing that the optimal control can be expressed as $u(t) = -qz(T)(T - t)$, where $z(T) = z(0)/(1 + \frac{1}{3}qT^3)$.

Express $u(t)$ in a feedback form of $u(t) = k(t)z(t)$ for some $k(t)$.

Suppose that the initial state $x(0)$ is perturbed by a small amount to $x(0) + (\epsilon_1, \epsilon_2)$. Give an expression (in terms of $\epsilon_1$, $\epsilon_2$, $x(0)$, $q$ and $T$) for the increase in minimal cost.

B3/14 **Optimization and Control**

Consider a scalar system with $x_{t+1} = (x_t + u_t)\xi_t$, where $\xi_0, \xi_1, \ldots$ is a sequence of independent random variables, uniform on the interval $[-a, a]$, with $a \leqslant 1$. We wish to choose $u_0, \ldots, u_{h-1}$ to minimize the expected value of

$$\sum_{t=0}^{h-1}(c + x_t^2 + u_t^2) + 3x_h^2\,,$$

where $u_t$ is chosen knowing $x_t$ but not $\xi_t$. Prove that the minimal expected cost can be written $V_h(x_0) = hc + x_0^2\Pi_h$ and derive a recurrence for calculating $\Pi_1, \ldots, \Pi_h$.

How does your answer change if $u_t$ is constrained to lie in the set $\mathcal{U}(x_t) = \{u : |u + x_t| < |x_t|\}$?

Consider a stopping problem for which there are two options in state $x_t$, $t \geqslant 0$:

(1) stop: paying a terminal cost $3x_t^2$; no further costs are incurred;

(2) continue: choosing $u_t \in \mathcal{U}(x_t)$, paying $c + u_t^2 + x_t^2$, and moving to state $x_{t+1} = (x_t + u_t)\xi_t$.

Consider the problem of minimizing total expected cost subject to the constraint that no more than $h$ continuation steps are allowed. Suppose $a = 1$. Show that an optimal policy stops if and only if either $h$ continuation steps have already been taken or $x^2 \leqslant 2c/3$.

[*Hint: Use induction on h to show that a one-step-look-ahead rule is optimal. You should not need to find the optimal $u_t$ for the continuation steps.*]

B4/14 **Optimization and Control**

A discrete-time decision process is defined on a finite set of states $I$ as follows. Upon entry to state $i_t$ at time $t$ the decision-maker observes a variable $\xi_t$. He then chooses the next state freely within $I$, at a cost of $c(i_t, \xi_t, i_{t+1})$. Here $\{\xi_0, \xi_1, \ldots\}$ is a sequence of integer-valued, identically distributed random variables. Suppose there exist $\{\phi_i : i \in I\}$ and $\lambda$ such that for all $i \in I$

$$\phi_i + \lambda = \sum_{k \in \mathbb{Z}} P(\xi_t = k) \min_{i' \in I} \left[ c(i, k, i') + \phi_{i'} \right] .$$

Let $\pi$ denote a policy. Show that

$$\lambda = \inf_{\pi} \limsup_{t \to \infty} E_{\pi} \left[ \frac{1}{t} \sum_{s=0}^{t-1} c(i_s, \xi_s, i_{s+1}) \right] .$$

At the start of each month a boat manufacturer receives orders for 1, 2 or 3 boats. These numbers are equally likely and independent from month to month. He can produce $j$ boats in a month at a cost of $6+3j$ units. All orders are filled at the end of the month in which they are ordered. It is possible to make extra boats, ending the month with a stock of $i$ unsold boats, but $i$ cannot be more than 2, and a holding cost of $ci$ is incurred during any month that starts with $i$ unsold boats in stock. Write down an optimality equation that can be used to find the long-run expected average-cost.

Let $\pi$ be the policy of only ever producing sufficient boats to fill the present month's orders. Show that it is optimal if and only if $c \geqslant 2$.

Suppose $c < 2$. Starting from $\pi$, what policy is obtained after applying one step of the policy-improvement algorithm?

B2/15 **Optimization and Control**

A street trader wishes to dispose of $k$ counterfeit Swiss watches. If he offers one for sale at price $u$ he will sell it with probability $ae^{-u}$. Here $a$ is known and less than 1. Subsequent to each attempted sale (successful or not) there is a probability $1-\beta$ that he will be arrested and can make no more sales. His aim is to choose the prices at which he offers the watches so as to maximize the expected values of his sales up until the time he is arrested or has sold all $k$ watches.

Let $V(k)$ be the maximum expected amount he can obtain when he has $k$ watches remaining and has not yet been arrested. Explain why $V(k)$ is the solution to

$$V(k) = \max_{u>0} \left\{ ae^{-u}[u + \beta V(k-1)] + (1 - ae^{-u})\beta V(k) \right\} .$$

Denote the optimal price by $u_k$ and show that

$$u_k = 1 + \beta V(k) - \beta V(k-1)$$

and that

$$V(k) = ae^{-u_k}/(1-\beta) .$$

Show inductively that $V(k)$ is a nondecreasing and concave function of $k$.

B3/14 **Optimization and Control**

A file of $X$ Mb is to be transmitted over a communications link. At each time $t$ the sender can choose a transmission rate, $u(t)$, within the range $[0,1]$ Mb per second. The charge for transmitting at rate $u(t)$ at time $t$ is $u(t)p(t)$. The function $p$ is fully known at time 0. If it takes a total time $T$ to transmit the file then there is a delay cost of $\gamma T^2$, $\gamma > 0$. Thus $u$ and $T$ are to be chosen to minimize

$$\int_0^T u(t)p(t)dt + \gamma T^2 ,$$

where $u(t) \in [0,1]$, $dx(t)/dt = -u(t)$, $x(0) = X$ and $x(T) = 0$. Quoting and applying appropriate results of Pontryagin's maximum principle show that a property of the optimal policy is that there exists $p^*$ such that $u(t) = 1$ if $p(t) < p^*$ and $u(t) = 0$ if $p(t) > p^*$.

Show that the optimal $p^*$ and $T$ are related by $p^* = p(T) + 2\gamma T$.

Suppose $p(t) = t + 1/t$ and $X = 1$. For what value of $\gamma$ is it optimal to transmit at a constant rate 1 between times 1/2 and 3/2?

B4/14 **Optimization and Control**

Consider the scalar system with plant equation $x_{t+1} = x_t + u_t$, $t = 0, 1, \ldots$ and cost

$$C_s(x_0, u_0, u_1, \ldots) = \sum_{t=0}^{s} \left[ u_t^2 + \frac{4}{3} x_t^2 \right] .$$

Show from first principles that $\min_{u_0, u_1, \ldots} C_s = V_s x_0^2$, where $V_0 = 4/3$ and for $s = 0, 1, \ldots$,

$$V_{s+1} = 4/3 + V_s/(1 + V_s) .$$

Show that $V_s \to 2$ as $s \to \infty$.

Prove that $C_\infty$ is minimized by the stationary control, $u_t = -2x_t/3$ for all $t$.

Consider the stationary policy $\pi_0$ that has $u_t = -x_t$ for all $t$. What is the value of $C_\infty$ under this policy?

Consider the following algorithm, in which steps 1 and 2 are repeated as many times as desired.

1. For a given stationary policy $\pi_n$, for which $u_t = k_n x_t$ for all $t$, determine the value of $C_\infty$ under this policy as $V^{\pi_n} x_0^2$ by solving for $V^{\pi_n}$ in

$$V^{\pi_n} = k_n^2 + 4/3 + (1 + k_n)^2 V^{\pi_n} .$$

2. Now find $k_{n+1}$ as the minimizer of

$$k_{n+1}^2 + 4/3 + (1 + k_{n+1})^2 V^{\pi_n}$$

and define $\pi_{n+1}$ as the policy for which $u_t = k_{n+1} x_t$ for all $t$.

Explain why $\pi_{n+1}$ is guaranteed to be a better policy than $\pi_n$.

Let $\pi_0$ be the stationary policy with $u_t = -x_t$. Determine $\pi_1$ and verify that it minimizes $C_\infty$ to within 0.2% of its optimum.