060229 A policy π is to be chosen to maximize

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

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

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

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

where $\{\epsilon_t\}$ is a sequence of independent and positive random variables, such that $E\epsilon_t = \theta > 0$. Given x_0 and $0 < \beta \leq 1/(1 + \theta)$, show that $F(\pi, x)$ is maximized by taking $u_t = x_t$ for all t.

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

050328 A discrete-time controlled Markov process evolves according to

$$X_{t+1} = \lambda X_t + u_t + \epsilon_t\,, \quad t = 0, 1, 2, \dots\,,$$

where the ϵ_t are independent zero-mean random variables with common variance σ^2, and λ is a known constant. Consider the problem of minimizing

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

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 form $u_j = k_{T-t} x_j$, for certain constants k_i. Show also that the minimized value ofor $F_{t,T}(x)$ 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

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

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

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

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

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

050429 An investor has a (possibly negative) bank balance $x(t)$ at time t. For given positive $x(0)$, T, μ, A and r, he wishes to choose his spending rate $u(t) \geq 0$ 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 μ, show that there is a unique positive μ_* such that $x_*(T;\mu_*) = 0$. If the original problem is modified by setting $\mu = 0$, but requiring that $x(T) \geq 0$ show that the optimal control for this modified problem is $u(t) = u_*(t;\mu_*)$.

050229 Explain what is meant by a time-homogeonous 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 $r(i) \geq 0$, 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 $V(i) \geq r(i)$ for all i.

Describe an optimal policy.

Suppose $r(0), \ldots, r(N)$ are distinct numbers. Is the optimal policy necessarily unique?

050328 Consider the problem

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

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

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

$u(t)$ 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?

050429 A continuous-time control problem is defined in terms of state variable $x(t) \in \mathbb{R}^n$ and control $u(t) \in \mathbb{R}^m$, $0 \leq t \leq T$. We desire to minimize $\int_0^T c(x,t)\,dt + K(x(T))$, when 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 $\lambda_1, \ldots, \lambda_m$.

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 the 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

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

04215 A gambler is presented with a sequence of random numbers $N_1, N_2, \ldots, N_n$, one at a time. The N_k's are distributed so that

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

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 all the numbers, £2 if he chooses the largest of all the numbers, and £1 otherwise.

Let $r_0 = \lceil 1 + 1/p \rceil$. Show that the form of the optimal strategy is to choose the first number such that either (i) $N_k > 0$ and $k \geq n - r_0$, or (ii) $k = n - 1$.

04315 The strength of the economy evolves according to the equation

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

where $\dot{x}_0 = x_0 = 0$ 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

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

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

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

Hint: The general solution of the linear system

$$\begin{pmatrix} \dot{y}_1 \\ \dot{y}_2 \end{pmatrix} = \begin{pmatrix} 0 & \gamma^2 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

is given by

$$y = a \begin{pmatrix} \cos(\gamma t) \\ -\gamma^{-1} \sin(\gamma t) \end{pmatrix} + b \begin{pmatrix} \sin(\gamma t) \\ \gamma^{-1} \cos(\gamma t) \end{pmatrix}$$

where a and b are scalar constants.

04415 Consider the deterministic dynamical system

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

where A and B are constant matrices, $x_t \in R^n$ and u_t is the control variable, $u_t \in R^m$.

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 $\{u_s : 0 \le s \le t\}$ is allowed to vary, then V_t is a vector space.

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

1. $V_t^{\perp} := \left\{v \in R^n : v^{\top} y_t = 0 \ \forall y_t \in V_t\right\} = \{0\}$.
2. The matrix $H(t) := \int_0^t e^{-sA} BB^{\top} e^{-sA^{\top}} ds$ is (strictly) positive definite.
3. The matrix $M_n := [B \quad AB \quad A^2B \quad \cdots \quad A^{n-1}B]$ has rank n.

Consider the scalar system

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

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

03215 The owner of a put option may exercise it on any one of the days $1, \ldots, h$, 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 $\epsilon_1, \epsilon_2, \ldots$ 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

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

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

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

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

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 $x_t \leq a$.'

03314 State Pontryagin's Maximum Principle (PMP).

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

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

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 $0 \le u(t) \le 1.4$, and $u(t) = 0$ if $x(t) = 0$. Assume the PMP with an appropriate Hamiltonian of $H(x, u, t, \lambda)$. 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 $\eta(t)$, $0 \le t$, 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 $\eta(0) \ge 1$ is impossible and describe qualitatively what must happen under the optimal policy.

03414 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 $F(W_T) = \hat{x}_T^2$ 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?

02215 State Pontryagin's maximum principle (PMP) for the problem of minimizing

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

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

Consider the 2-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 1-dimensional problem.

Use PMP to solve this 1-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 $u(t) = k(t)x(t)$ for some $k(t)$.

Suppose that the initial state is perturbed by a small amount to $x(0) + (\epsilon_1, \epsilon_2)$. Give an expression (in terms of ϵ_1, ϵ_2, $x(0)$, q and T) for the increase in minimal cost.

02314 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 $a \leq 1$. 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 ξ_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, $t \geq 0$:

(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 $x^2 \leq 2c/3$.

[*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.*]

02414 A discrete-time decision process is defined on a finite set of spaces I as follows. Upon entry to state i_t at time t the decision-maker observes a variable ξ_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 λ 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 π 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 π 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 $c \geq 2$.

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

01215 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 $1-\beta$, $0<\beta<1$, that he will be rumbled 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 rumbled 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 rumbled. 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.

01314 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 γT^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 $u(t) = 1$ 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 γ is it optimal to transmit at a constant rate 1 between times 1/2 and 3/2?

01414 Consider the scalar system with plant equation $x_{t+1} = x_t + u_t$, $t = 0, 1, \dots$ and cost

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

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

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

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

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

Consider the stationary policy π_0 that has $u_t = -x_t$ for all t. What is the value of C_∞ 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 π_n, for which $u_t = k_n x_t$ for all t, determine the value of C_∞ under this policy as $V^{\pi_n} x_0^2$ by solving for V^{π_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 π_{n+1} as the policy for which $u_t = k_{n+1} x_t$ for all t.

Explain why π_{n+1} is guaranteed to be a better policy than π_n.

Let π_0 be the stationary policy with $u_t = -x_t$. Determine π_1 and verify that it minimizes C_∞ to within 0.2% of its optimum.

00215 Let H denote the Hamiltonian of Pontryagin's maximum principle. On an optimal trajectory H is maximized to $-\lambda_0(t)$. In what circumstances is $\lambda_0(t)$ constant?

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

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

where θ is the angle at which he swims relative to the direction of the current.

He desires to go as far as possible downstream in a given time T, ending upon the same bank as he starts. Given that $c(y)$ is increasing and differentiable in y, use Pontryagin's maximum principle to show that he should choose $\theta(t)$ to be a decreasing function of the time, t.

Show that if on an optimal trajectory he is at a distance y from the bank at two distinct times t_1 and t_2 then it must be that $\theta(t_1) = -\theta(t_2)$.

00314 In a television game show a contestant is successively asked questions $Q_1, \ldots, Q_9$. After correctly answering Q_i and hearing Q_{i+1} she has the option of either going home with 2^i pounds or attempting to answer Q_{i+1}. If she answers Q_{i+1} incorrectly then she goes home with nothing. If she answers Q_9 correctly then the game ends and she takes home 2^9 pounds.

Upon hearing Q_i she is able to classify it as either easy or hard; these types occur with probabilities 0.95 and 0.05 respectively, independently for each question. If Q_i is easy her probability of answering it correctly is $9/(9+i)$, but if it is hard the probability is only $6/(6+i)$. At most once in the game she may choose to 'phone a friend'; the effect is to increase her chance of correctly answering Q_i to $10/(10+i)$, if it is easy, and to $7/(7+i)$, if it is hard.

Let W_i (and V_i) denote her expected winnings if she plays optimally from the point that she has correctly answered $i-1$ questions and has (or has not yet) phoned a friend. Write down dynamic programming equations from which you could compute V_1.

Show that $W_9 = 2^8$ and $V_9 = (21/20)2^8$.

Suppose she has answered 7 questions correctly and has not yet phoned a friend. What should she do if Q_8 is an easy question?

The producers of the show are considering a new game in which everything is the same except that the potential number of questions is unlimited. The game ends only when the contestant answers incorrectly or chooses to retire. Quoting any theorem necessary to justify your answer, show that for a contestant who plays optimally the new game is the same as the old.

00414 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 $\Pi = 0$, 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 $t \to \infty$.

99215

A linear system on a straight line has the plant equation of the form

$$x_{n+1} = Ax_n + \xi_n u_n + \delta_n \,.$$

Here, $x_1, x_2, \ldots$ are subsequent states of the system, $u_1, u_2, \ldots$ are control variables, and (ξ_1, δ_1), $(\xi_1, \delta_1), \ldots$ are independent random vectors with $\mathbb{E}\,\xi_j = \mathbb{E}\,\delta_j = 0$, $\mathrm{var}\xi_j = \mathrm{var}\delta_j = \nu$ and and $\mathrm{cov}(\xi, \delta) = 0$. Find the closed loop controls over the time horizon N minimizing the expected value of the following cost function

$$c \sum_{j=1}^{N} x_j^2 + d \sum_{j=1}^{N-1} u_j \,,$$

that is linear in u and quadratic in x. State the certainty-equivalence principle and check whether your solution satisfies it.

99314

A rodent in a northern forest collects hazelnuts to eat during the coming winter. Every time the rodent forays from its lair it collects a random weight of nuts, but risks being caught by a predator. The weights of nuts collected in successive forays form a sequence of independent random variables, each having an exponential distribution with mean l/λ. On each foray there is a probability p, $0 < p < 1$, of the rodent being caught, and hence of eating no nuts. Find a policy maximizing the expected weight of nuts eaten by the rodent during the coming winter, and justify your answer.

99414

A butterfly enjoys fluttering among blossoming trees on a sunny June morning. A wind begins to blow as the butterfly flies between trees A and B at a distance L apart. Assume that the wind blows at a constant non-zero speed u_0 in the direction from tree A to tree B, and the equation of motion is $\dot{x} = u$, or $dx = udt$, $0 < x < L$. Here, x is the butterfly's distance from tree A, and velocity u is a control variable. When the butterfly reaches tree A or tree B, it rests there until the wind dies. Find a trajectory from x to one of the trees, minimising the cost

$$\int_0^T (u - u_0)^2 dt + c(x(T)) \, ,$$

where T is the time it reaches tree A or B, and $c(A)$ and $c(B)$ are given terminal costs (related to an 'attraction', viz. the scent or the brightness of blossoms of a particular tree).

[*Hint: The value function $F(x)$ is piecewise linear.*]

If the butterfly panics, the result is erratic movement, and the equation of motion becomes stochastic: $dx = udt + vdB(t)$, where $v > 0$ is a constant and $B(t)$ is a standard Brownian motion. This means that the butterfly follows the trajectory of a controlled diffusion process, with the infinitesimal generator

$$\Lambda(u)\phi(x) = u\phi'(x) + \frac{v}{2}\phi''(x).$$

Solve the corresponding dynamic programming equation for the value function $F(x)$:

$$0 = \inf_u \left[(u - u_o)^2 + \Lambda(u)F \right], \quad 0 < x < L \, ,$$

with $F(0) = c(A)$, $F(L) = c(B)$.

[*Hint: Use a standard substitution $F(z) = \alpha \log \phi(x)$ and adjust the boundary conditions.*]

98212

A discrete-time Markov decision process has discount factor $\beta < 1$, and one-step rewards, $r(x, u)$, with $0 \leq r(x, u) \leq 1$, for all x, u. Starting in state x, let $F_s(x)$ and $F(x)$ denote respectively the maximal expected discounted rewards that can be obtained over s steps and over the infinite horizon. Prove that $F_s(x) \to F(x)$ as $s \to \infty$.

A hunter has limited time and bullets. Each day that he goes hunting he encounters exactly one target and then assesses accurately the probability he can hit it. On the basis of this assessment he decides to shoot once or not at all. The probabilities with which targets can be hit are independent samples from the uniform distribution on $[0, 1]$. Each night there is a probability $1 - \beta$ that all his remaining bullets will be stolen. Given that on a particular day he has i bullets left and must retire from hunting within s days, or when he has no bullets left, show that he maximizes the expected number of targets he hits by shooting if and only if he can hit the target he encounters that day with probability exceeding some $z(s, i)$.

Prove that $z(s, i)$ is a nonincreasing function of i.

Suppose that there is no limit on how many days he hunts. Show that with i bullets left an optimal policy is to shoot if and only if the probability he can hit the target exceeds $z(i)$, where $z(i)$ is also some nonincreasing function of i.

98312

The position, x_1, and speed, x_2, of a particle moving on the line obey

$$\dot{x}_1 = x_2, \qquad \dot{x}_2 = u,$$

where u is an applied control force. Initially the particle is at rest and $x_1(0) = X$, where $X > 0$. It is desired to bring the particle to rest at the origin while minimizing the cost

$$J = \int_0^T \tfrac{1}{2}\left(k^2 + u^2\right) dt.$$

Here k is a constant, T is the time that the particle reaches the origin and T is unconstrained.

Use the maximum principle to show that under optimal control the particle reaches the origin at time $T = (6X/k)^{1/2}$ and that u varies linearly with time from $-k$ to k.

98416

Consider the discrete-time controlled system $x_t = x_{t-1} + u_{t-1} + \epsilon_t$, where the ϵ_t are independent $N(0,4)$ variables. At time t one observes not x_t, but $y_t = x_{t-1} + \eta_t$, where the η_t are independent $N(0,3)$. One also has $\hat{x}_0$, an estimate of x_0, such that $\hat{x}_0 \sim N(x_0, 6)$. Consider state estimates obtained from $\hat{x}_t = \hat{x}_{t-1} + u_{t-1} + H_t(y_t - \hat{x}_{t-1})$. Show that the variances of these estimates are minimized by $H_1 = \cdots = H_h = 2/3$; find these variances and show that they do not depend on $u_0, \ldots, u_{h-1}$.

Solve the problem of minimizing

$$E\left[\sum_{t=0}^{h-1}(x_t^2 + 6u_t^2) + 3x_h^2 \,\middle|\, \hat{x}_0\right].$$

Show that $u_t = -\hat{x}_t/3$ is optimal for all t and that the minimized value is $3\hat{x}_0^2 + 18(h+1)$.

In giving your answer, mention at relevant points the terms 'Kalman filter', 'separation principle', 'Riccati equation', 'stationary policy' and 'certainty equivalence control.'

97211

A scalar linear system obeys the plant equation

$$x_{t+1} = x_t + u_t + \epsilon_t, \quad t = 0, \ldots, h-1\,,$$

where $\epsilon_0, \ldots, \epsilon_{h-1}$ are independent random variables, all having mean 0 and variance 1, $x_0 = -1$, and at the time the control u_t must be chosen the state x_t is known. It is desired to minimize the expected value of the cost function

$$\sum_{t=0}^{h-1} u_t^2 + \Pi x_h^2\,.$$

Find the optimal closed-loop control and the minimal expected cost as functions of Π.

Suppose that at each time t it is possible to apply a control to the noise that reduces the variance of ϵ_t to α, $0 \le \alpha < 1$. The control may be applied at some times, but not at others, and the decision whether or not to apply it at time t may be deferred until x_t is known and u_t is about to be chosen. For each t for which the variance of the noise is so reduced a cost c is added to the total cost above. Find the policy which minimizes the expected value of the sum of all costs. Show that noise reduction control should be used at least once provided $c < (1-\alpha)\Pi$.

97311

(a) Suppose an optimal control problem has a terminal cost K which depends only on the final state $x(T)$ and this state is unconstrained. How can one determine boundary conditions for the adjoint equations of Pontryagin's maximum principle?

(b) For t in the interval $[0, 1]$ the price of gold has been $p(t)$ pounds per ounce. A gold trader wishes to evaluate his trading in this interval. At the start of the interval he had £$x_1(0)$ and $x_2(0)$ ounces of gold. Gold which he held in storage was charged at £γ per ounce per unit time. Purchases, but not sales, of gold were taxed at £δ per ounce, $\delta > 0$. Let $u(t)$ be the rate at which he sold gold at time t (with $u(t) < 0$ corresponding to purchase of gold). Then supposing his stocks of money $x_1(t)$ and gold $x_2(t)$ were always positive, these obeyed

$$\dot{x}_1(t) = p(t)u(t) - \gamma x_2(t) + \delta \min\{0, u(t)\} \quad \text{and} \quad \dot{x}_2(t) = -u(t)\,.$$

The rate at which he could trade was constrained by $-1 \le u(t) \le 1$.

Use the maximum principle to show that with hindsight, and with $x_1(0)$ and $x_2(0)$ sufficiently large, the trader would have minimized his trading loss of

$$K = [x_1(0) + p(0)x_2(0)] - [x_1(1) + p(1)x_2(1)]$$

by taking $u(t) = -1$, 0, or 1 as $\Delta(t) := p(t) - p(1) + \gamma(1 - t)$ was in the interval $(-\infty, -\delta]$, $(-\delta, 0]$, or $(0, \infty)$ respectively.

Comment on the reason for the requirement that $x_1(0)$ and $x_2(0)$ be sufficiently large.

97415

Each day a manufacturer receives an order with probability p and no order with probability $q = 1 - p$. If i orders are outstanding, he can choose either to process them all, at cost k, or hold them over until the next day at a cost of ci, $c > 0$, except that if $i = N$ he must process them all. He desires to minimize the expected infinite-horizon discounted-cost, with discount rate $\alpha < 1$. Explain the relevance to this problem of V, where for $i = 0, \dots, N-1$,

$$V(i) = \min\bigl\{k + \alpha q V(0) + \alpha p V(1)\,,\; ci + \alpha q V(i) + \alpha p V(i+1)\bigr\},$$

and $V(N) = k + \alpha q V(0) + \alpha p V(1)$.

For each $j = 1, \dots, N$, let π_j denote the 'threshold policy' which processes the waiting orders if and only if their number is at least j. Let $V_j(i)$ denote the expected infinite-horizon discounted-cost under π_j when starting with i orders. Write down a set of simultaneous linear equations in $V_j(0), \dots, V_j(N)$ and use these to show that if i' is the greatest integer not exceeding j such that $V_j(i'-1) < V_j(i')$, then $V_j(0) < \cdots < V_j(i')$ and $V_j(i') \geq \cdots \geq V_j(N)$.

Let $g(i) = ci + \alpha q V_j(i) + \alpha p V_j(i+1)$. Show that $g(0) < \cdots < g(i')$ and that if $i' < j$ then $g(i) \geq V_j(N)$ for all $i \geq i'$. Hence show that if the policy improvement algorithm is started with π_j then the policy obtained after one iteration of the algorithm is also a threshold policy.

Let $p = 0.5$, $\alpha = 0.8$, $c = 1$. Show that π_1 is optimal if and only if $k \leq 5/3$. Suppose additionally that $N = 4$, $k = 3.6$ and the algorithm begins with policy π_1. Use the data in the table below to find the number of iterations taken by the algorithm before it terminates.

j	$V_j(0)$	$V_j(1)$	$V_j(2)$	$V_j(3)$	$V_j(4)$
1	7.2	10.8	10.8	10.8	10.8
2	4.88	7.32	8.48	8.48	8.48
3	5.2	7.8	9.2	8.8	8.8
4	5.96	8.94	10.92	11.38	9.56

96210

A fantasy kingdom has populations of x vampires and y humans. At the start of each of the four yearly seasons the king takes a census and then intervenes to admit or expel some humans. Suppose that the population dynamics of the kingdom are governed by the plant equations

$$x_{t+1} = x_t + (y_t - Y), \qquad y_{t+1} = y_t - (x_t - X) + u_t,$$

where x_t and y_t are integers representing the respective populations of vampires and humans in the t-th season, $t = 0, 1, 2, 3, \ldots$; X and Y represent equilibrium values in the population without regal intervention; and u_t denotes the effect of the king's intervention in the t-th season. Show, by defining z_t by an appropriate translation of $(x_t, y_t)^\top$, that the plant equations can be written in the form

$$z_{t+1} = Az_t + Bu_t, \quad \text{where } A = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}, \text{ and } B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

One spring, denoted $t = 0$, the king finds $x_0 = X + c$ and $y_0 = Y$, where c is a positive integer. Show that the equilibrium population (X, Y) will not be regained without the king's intervention. What condition would have to be satisfied by A in order that $z_t \to 0$ as $t \to \infty$ without any interaction by the king?

Can the king regain equilibrium by only expelling humans?

Suppose, the census is taken during the day, so the king can only count humans, and thus u_t is allowed only to depend on y_{t-1} and Y. Quoting freely from the theory of LQ regulation and Kalman filtering, show that there are constants a, b, c, a', b', c' (which you should not determine) such that controls of the form $u_0 = 0$, $u_1 = 0$, $u_2 = a + by_0 + cy_1$, $u_3 = a' + b'y_0 + c'y_1$ will regain equilibrium at $t = 4$, whatever the value of x_0 and the observed values y_0, y_1.

[Hint: *You may find helpful the powers of A, tabulated below.*]

t	2	3	4	5	6
A^t	$\begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix}$	$\begin{pmatrix} -2 & 2 \\ -2 & -2 \end{pmatrix}$	$\begin{pmatrix} -4 & 0 \\ 0 & -4 \end{pmatrix}$	$\begin{pmatrix} -4 & -4 \\ 4 & -4 \end{pmatrix}$	$\begin{pmatrix} 0 & -8 \\ 8 & 0 \end{pmatrix}$

96311

An oil company wishes to develop *exactly one* of n potential sites. If a test drilling is made at site i then an estimate, x_i, is obtained for the profit that will arise if that site is selected for development. Prior to testing at site i, and independent of information obtained about other sites, the distribution of x_i has density function $f_i(x)$, $x \geq 0$. Test drillings are to be made at the sites sequentially, in whatever order is best, but these may be stopped at any point and then the site with the greatest x_i selected for development. Each test drilling costs c and the aim is to maximize the expected profit from development minus the cost of all the test drillings.

Suppose that U is the set of all sites that have not yet been tested. Let $x(U) = \max_{j \notin U} x_j$ be the value of the best site tested. Show that following a test drilling, it is optimal to make no further test drillings and develop the best site so far, if and only if

$$c \geq \max_{j \in U} \int_{x(U)}^{\infty} [y - x(U)] f_j(y) dy.$$

Show that if the densities f_i are equal then the site ultimately selected will also be the location of the final test drilling. Show, on the other hand that is the f_i are not equal it can be optimal to select one of the previously tested sites.

96414

(i) Define the notions of open-loop and closed-loop control. Discuss their relative advantages and describe circumstances in which they are the same or different.

The ranger at the Trumpington Safari Park has a problem with his prized population of lesser striped gnus. The population of animals has deviated by x (which can be positive or negative) from the sustainable equilibrium level. If x is negative, the population is in danger of becoming extinct, whereas if x is positive, the gnus are considered a pest, intimidating passing cyclists. Control of the gnu population is possible, either by culling to reduce their numbers, or importing from other parks. At time t, the excess of the animal population above its equilibrium is denoted by $x(t)$, and evolves according to the plant equation

$$\dot{x} = -x(t) + u(t), \quad t \geq 0,$$

with $x(0) = x$. The ranger wishes to return the population to equilibrium in finite time, whilst achieving the minimal cost function

$$V(x) = \inf_u \int_0^{T(u)} (1 + |u(t)|)\, dt,$$

where the infimum is taken over all controls u which return the population to equilibrium in finite time, and such that $-1 \leq u(s) \leq 1 \ \forall s \geq 0$. $T(u)$ denoted the time taken to return the population to equilibrium using the control u.

(ii) Find the dynamic programming equation, show that it is solved by

$$V(x) = \begin{cases} \log|x| + 2\log 2, & |x| \geq 1, \\ 2\log(1 + |x|), & |x| \leq 1. \end{cases},$$

and find the optimal control in closed-loop form.

(iii) Suppose $x(0) > 0$. Find functions $\lambda(t)$ and $H(x(t), u(t), \lambda(t))$ such that for the optimal control $u^*(t)$, corresponding optimal trajectory $x^*(t)$, and for all $0 \leq t \leq T(u^*)$,

$$H(x^*(t), u^*(t), \lambda(t)) = 0,$$

with

$$H(x^*(t), u(t), \lambda(t)) \leq 0,$$

for all other permissible controls u.

(iv) Hence find the optimal control in closed-loop form.

(v) Verify that the closed-loop and open-loop controls are the same.

95210

Consider an optimality equation

$$F(x) = \max_u \{ \, r(x,u) + \beta E[F(x_1) \mid x_0 = x, u_0 = u)] \, \}, \quad x \in X,$$

where $\beta > 0$. Suppose a policy has a value function which satisfies this equation. State conditions under which this fact implies that the policy is optimal. Give an example to show that these conditions cannot be arbitrarily relaxed.

Suppose $\beta = 1$ and the problem is a stopping problem, with continuation and stopping actions of $u = 0$ and $u = 1$ respectively; $r(x,u) = ur(x)$, $r(x) \geq 0$. Under $u = 0$ the state evolves according to a Markov process; under $u = 1$ it evolves to a special absorbing state from which no further rewards can be obtained.

A function f is said to be *excessive* if $f(x) \geq E[f(x_1) \mid x_0 = x]$ for all $x \in X$. Show that if r is excessive then $u = 1$ is optimal in every state.

Suppose $X = \{0, 1, 2, 3, 4\}$. Under the continuation option the state performs one further step of a symmetric random walk on the integers, except that it becomes trapped in states 0 and 4. Find the solution for find the solution when $r(0) = 0$, $r(1) = 3$, $r(2) = 2$, $r(3) = 5$, $r(4) = 4$.

95311

State Pontryagin's maximum principle (PMP) as it applies to the problem of identifying a control u which minimizes

$$C = \int_0^T c(x, u)\, dt,$$

where x and u are functions of t, with $x(t) \in \mathbb{R}^2$, $u(t) \geq 0$, $\dot{x} = a(x, u)$; $x(0)$ is specified and T is the first time that x reaches a set $\mathcal{S}$.

The height of the water in a reservoir is to be raised by h through pumping in fresh water. The added water must also compensate for a linearly increasing rate of water loss. Let $x_1(t)$, $x_2(t)$ and $u(t)$ denote respectively the height of the water above its initial level, the rate of water loss, and the pumping rate at time t. The pumping cost is proportional to the square of the pumping rate, so the problem is

$$\text{minimize} \int_0^T \frac{1}{2} u(t)^2\, dt$$

under the constraints

$$x_1(0) = 0, \quad x_1(T) = h, \quad x_2(0) = 0, \quad \dot{x}_1 = u(t) - x_2(t), \quad \dot{x}_2 = 1.$$

Show that the adjoint variables in PMP can be written $\lambda_1(t) = A$ and $\lambda_2(t) = At + B$ for constants A and B to be determined.

By identifying $T = 1$ with a constraint on the terminal value of x_2 show that the optimal control under this constraint is $u(t) = h + 1/2$, $0 \leq t \leq 1$. Deduce that if the initial value of x_2 is increased by a small amount ϵ then the optimal cost increases by approximately $(1/2)(h + 1/2)^2 \epsilon$.

Also find the optimal u when T is unconstrained.

95414

Consider scalar system and cost function,

$$x_{t+1} = x_t + b_t u_t, \quad C = \sum_{t=0}^{h-1} u_t^2 + x_h^2 \Pi_h.$$

Show that C is minimized by a control

$$u_t = -(b_t^2 + \cdots + b_{h-1}^2 + \Pi_h^{-1})^{-1} b_t x_t.$$

Frogs, labelled $1, \ldots, k$, play the following game on a line. At time t frog i has position $x_{i,t}$, $t = 0, 1, \ldots$. Frog k leaps frog $k-1$, traveling in an arc over frog $k-1$ and to the point equally distant on the opposite side. Then frog $k-1$ leaps frog $k-2$ in the same manner, and so on. Thus $x_{i,t+1} = 2x_{i-1,t} - x_{i,t}$, $i = 2, 3, \ldots, k$. Finally, frog 1, the queen, who can jump however she likes, jumps by u_t, so that $x_{1,t+1} = x_{1,t} + u_t$.

Define and discuss the concept of *controllability*, illustrating the key ideas in the context of the above system when $k = 4$. You should show that the queen can jump in such a way as to cause all four frogs to meet at any given point and explain how she could work out the minimum time and sequence of jumps required to achieve this. Explain how the queen can do this if the only things she can observe are the initial and subsequent positions of frog 4 and the magnitude and direction of her own jumps.

Suppose the queen desires to minimize a cost comprising jumping effort and her final squared distance from frog k.

$$\sum_{t=0}^{h} u_t^2 + (x_{1,h} - x_{k,h})^2 \Pi_h.$$

Show that the problem may be reduced to control of a scalar variable z_t by a transformation of the form $z_t = \alpha^\top A^{h-t} x_t$, for appropriately defined A and α. Interpret z_t. Explain what further calculations would be needed to compute the optimal control completely.