Dynamic Programming and Optimal Control **Autumn 2014**

Richard Weber

Problems Sheet 3

1. Consider again the problem from sheet 2.

At the beginning of each day a machine can be in either a working or broken state. If it is broken then the whole day is spent repairing it, and this costs $8c$ in labour and lost production. If the machine is working, then it may be run unattended or attended, at costs of 0 or c respectively. In either case there is a chance that the machine will breakdown and need repair the following day, with probabilities p and p' respectively, where $p' < (7/8)p$.

Find the deterministic stationary Markov policy that minimizes the time average-cost.

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

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

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

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

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

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

4. Customers arrive at a queue as a Poisson process of rate λ. They are served at rate $u = u(x)$, where x denotes the current size of the queue. Suppose that cost is incurred as rate $ax + bu$ where

a, $b > 0$. The service rate u is the control variable. The dynamic programming equation in the infinite horizon limit is then

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

where γ denotes the average rate at which cost is incurred under the optimal policy and where $f(x)$ denotes the extra cost associated with starting from state x. (Here $1_{x>0} = 0$ if $x = 0$, and $1_{x>0} = 1$ if $x = 1, 2, 3, \dots$.) Give a brief justification of this equation.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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