6.231 Dynamic Programming and Optimal Control
Midterm Exam, Fall 2015
Prof. Dimitri Bertsekas

Problem 1 (50 points)

This is a stopping problem, where before stopping is applied, the state of the system evolves according to

$$x_{k+1} = ax_k + bu_k, \tag{1}$$

where the nonzero scalars a and b are known. At each period k where the system has not yet stopped, we have the option of using a control u_k, incurring a cost $qx_k^2 + ru_k^2$, and moving to a new state x_{k+1} according to Eq. (1), or else stopping (i.e., moving to a special cost-free and absorbing termination state) and incurring a stopping cost tx_k^2 (we may have $t \neq q$). At the final period N, if not already stopped, we must stop and incur the stopping cost tx_N^2. We assume that the scalars q, r, and t are positive.

(a) Consider first the restricted optimization over policies that never stop, except at time N, so we obtain a standard linear quadratic problem, whose optimal cost function has the form

$$J_0(x_0) = Kx_0^2,$$

where K is a positive scalar that depends on N. Write the Riccati equation that yields K as well as the steady-state equation that has $\bar{K}$, the limit of K as $N \to \infty$, as its solution. Does the steady-state equation have any other solutions?

(b) Consider the unrestricted optimization where stopping is also allowed. Write the DP algorithm for this problem.

(c) Show that for the problem of part (b) there is a threshold value $\bar{t}$ such that if $t < \bar{t}$ immediate stopping is optimal at every state, and if $t > \bar{t}$ continuing at every state x_k and period k is optimal. How are the scalars $\bar{K}$ and $\bar{t}$ of parts (a) and (c) related?

(d) State an extension of the result of part (c) for the case of a multidimensional system.

Problem 2 (50 points)

Consider a situation involving a blackmailer and his victim. At each stage the blackmailer has a choice of:

(1) Retiring with his accumulated blackmail earnings.

(2) Demanding a payment of \$ 1, in which case the victim will comply with the demand (this happens with probability p, where $0 < p < 1$, independently of the past history), or will refuse to pay and denounce the blackmailer to the police (this happens with probability $1 - p$).

Once denounced to the police, the blackmailer loses all of his accumulated earnings and cannot blackmail again. Also, the blackmailer will retire once he reaches accumulated earnings of \$ n, where n is a given integer that may be assumed very large for the purposes of this problem. The blackmailer wants to maximize the expected amount of money he ends up with.

(a) Formulate the problem as a stochastic shortest path problem with states $i = 0, 1, \ldots, n$, plus a termination state,

(b) Write Bellman's equation and justify that its unique solution is the optimal value function $J^*(i)$.

(c) Use value iteration to show that $J^*(i)$ is monotonically strictly increasing with i, and that $J^*(i) = i$ for all i larger than a suitable scalar.

(d) Start policy iteration with the policy where the blackmailer retires at every i. Derive the sequence of generated policies and the optimal policy. How many iterations are needed for convergence?