Problem Set 2

1. Consider an MDP with a "goal state" $\bar{x}$, and suppose that for every other state y and policy u, there is $k \in \{1, 2, \dots\}$ such that $P_u^k(y, \bar{x}) > 0$. We will analyze the problem known as *stochastic shortest path*, defined as follows. For every state x, let $T(x)$ denote the first time stage t such that $x_t = \bar{x}$, given that $x_0 = x$. Then the objective is to choose a policy u that minimizes

$$\mathrm{E}\left[\sum_{t=0}^{T(x)} g_u(x_t) | x_0 = x\right].$$

   (a) Define a cost-to-go function for this problem and write the corresponding Bellman's equation.

   (b) Show that $P_u(T(x) > t) \le \rho^{\lfloor t/|\mathcal{S}|\rfloor}$, for some $0 < \rho < 1$ and all x.

   (c) Show that Bellman's equation has a unique solution corresponding to the optimal cost-to-go function and leads to an optimal policy.

   (d) Consider finding the policy that minimizes the average cost in this MDP, and assume that we have chosen the differential cost function h^* such that $h^*(\bar{x}) = 0$. Show that h^* may be interpreted as the cost-to-go function in a stochastic shortest path problem.

   (e) Define operators T_u and T for the stochastic shortest path problem. Show that they satisfy the monotonicity property. Is the offset property satisfied? Why?

   (f) (bonus) Define the *weighted maximum norm* to be given by

$$\|J\|_{\infty,\nu} = \max_x \nu(x)|J(x)|,$$

   for some positive vector ν. Show that T_u and T are contractions with respect to some weighted maximum norm contraction.

2. Consider the problem of controlling the service rate in a single queue, based on the current queue length x. In any time stage, at most one of two types of events may happen: a new job arrives at the queue with probability λ, or a job departs the queue with probability $\mu_1 + a\mu_2$, where $a \in \{0, 1\}$ is the current action. At each time stage, a cost $g_a(x) = (1 + a)x$ is incurred. The objective is to choose a policy so as to minimize the average cost.

   (a) Model this problem as an MDP. Write Bellman's equation and define the operator T.

   (b) Show that the differential cost function h^* is convex.

   (c) Show that the optimal policy is of the form $u^*(x) = 0$ if and only if $x \le \bar{x}$ for some $\bar{x}$.

   (d) Take the differential cost function h^* such that $h^*(0) = 0$. Show that there is $\gamma \in \Re$ such that $\gamma x^2 \le h^* \le \gamma^{-1} x^2$.

3. Consider an MDP with two states 0 and 1. Upon entering state 0 the system stays there permanently at no cost. In state 1 there is a choice of staying there at no cost or moving to state 0 at cost 1. Show that every policy is average cost optimal, but the only stationary policy that is Blackwell optimal is the one that keeps the system in the state it currently is.

4. (bonus) Let P be a transition probability matrix. Suppose that there is a state $\bar{x}$ such that, for every other state x, there is $k \in \{1, 2, \dots\}$ such that $P^k(x, \bar{x}) > 0$.

   (a) Let $T(x)$ be the smallest time stage t such that $x_t = \bar{x}$, conditioned on $x_0 = x$. Show that $T(x) < \infty$ with probability 1.

   (b) Let
   $$\lambda(x) = \mathrm{E}\left[\sum_{t=0}^{T(\bar{x})-1} \delta(x_t = x) | x_0 = \bar{x} \right]$$
   where $\delta(x_t = x) = 1$ if $x_t = x$ and 0 otherwise. Show that $\lambda P = \lambda$. Also show that
   $$\lim_{N \to \infty} \frac{1}{N} \sum_{t=0}^{N-1} P^k(x, y) \to \frac{\lambda(y)}{\mathrm{E}[T(\bar{x})]}$$
   and that any solution of $\pi P = \pi$ is of the form $\pi = m\lambda$ for some scalar m.

   (c) Suppose that $P(x, x) > 0$. Show that
   $$\lim_{k \to \infty} P^k(x, y) = \pi(y)$$
   for some $\pi : \mathcal{S} \mapsto [0, 1]$.

   (d) Provide examples of transition probability matrices illustrating the two following situations: (i) $\pi P = \pi$ has at least two solutions that are not multiple of each other, and (ii) $\pi P = \pi$ has all solutions of the form $\pi = m\lambda$ for some scalar m but $P^k(x, y)$ does not converge.

2