ISyE 6664 Stochastic Optimization

Fall 2012

Assignment 5

Issued: November 20, 2012

Due: December 6, 2012

Problem 1

Consider the following inventory control problem. An inventory manager uses a periodic inventory control system to control the inventory of a single product type. Consider a finite horizon $t = 0, 1, \ldots, T$. At the beginning of each period t, the controller observes the inventory level s_t. Then the controller chooses the amount $a_t \geq 0$ of inventory to add to the system. The demand for the product in time period t is a random variable D_t. The sequence $\{D_t\}_{t=0}^T$ is not necessarily i.i.d., that is, the inventory controller models the distribution of D_t as a function of the realization $d_0, \ldots, d_{t-1}$ of $D_0, \ldots, D_{t-1}$. The chosen amount of inventory a_t is added to the system before the demand D_t has to be satisfied.

The cost per unit product added to the inventory is a known constant $c > 0$. The inventory holding cost per period is a known constant $h > 0$ per unit of inventory remaining at the end of the period, and there is a known penalty per period $p > h$ per unit shortfall in inventory. The controller wants to determine a policy for choosing a_t to minimize

$$\mathbb{E}\left[\sum_{t=0}^{T} (ca_t + h\max\{0, s_t + a_t - D_t\} + p\max\{0, D_t - s_t - a_t\})\right]$$

1. Consider a convex function $G : \mathbb{R} \times \mathbb{R}_+ \mapsto \mathbb{R}$. Show that $g(s) := \min_{a \geq 0} G(s, a)$ is convex.

2. Suppose that unsatisfied demand is not backlogged. Write s_{t+1} as a function $f(s_t, a_t, D_t)$. Let $V_0(s) := \min \mathbb{E}\left[\sum_{t=0}^{T} (ca_t + h\max\{0, s_t + a_t - D_t\} + p\max\{0, D_t - s_t - a_t\}) \mid s_0 = s\right]$. Either prove or disprove (with a counterexample) the following statement: V_0 is convex. If it is convenient for you, you may assume that all the conditional distributions of D_t, $t = 0, \ldots, T$, have finite support.

3. Suppose that unsatisfied demand is backlogged. Write s_{t+1} as a function $f(s_t, a_t, D_t)$. Let $V_0(x) := \min \mathbb{E}\left[\sum_{t=0}^{T} (ca_t + h\max\{0, s_t + a_t - D_t\} + p\max\{0, D_t - s_t - a_t\}) \mid s_0 = s\right]$. Either prove or disprove (with a counterexample) the following statement: V_0 is convex. If it is convenient for you, you may assume that all the conditional distributions of D_t, $t = 0, \ldots, T$, have finite support.

4. Give sufficient conditions for optimality of a deterministic time-dependent base-stock policy, that is, for optimality of a policy with a deterministic base-stock level b_t for each time period t, such that $a_t = \max\{0, b_t - s_t\}$. Prove your claim.

Problem 2
Consider an infinite horizon, discrete time, discounted dynamic program:

$$\sup_{\pi \in \Pi^{\text{SD}}} E^{\pi}\left[\sum_{t=0}^{\infty} \alpha^t c(s_t, u_t) \middle| s_0\right]$$

with a finite state space S, $\alpha \in (0,1)$, and $u_t \in U(s_t)$ for all t. Suppose that the value iteration algorithm (that is, Jacobi value iteration) is used. Let V_n denote the value function approximation after n iterations, and let V^* denote the optimal value function. Suppose that the algorithm stops when

$$\|V_{n+1} - V_n\|_{\infty} \quad \leq \quad \frac{(1-\alpha)\varepsilon}{\alpha}$$

1. Show that

$$\|V_{n+1} - V^*\|_{\infty} \quad \leq \quad \varepsilon$$

2. Suppose that a stationary deterministic policy π is chosen that satisfies

$$c(s,\pi(s)) + \alpha \sum_{s' \in S} P[s'|s,\pi(s)] V_{n+1}(s') + \delta \quad \geq \quad \sup_{u \in U(s)} \left\{ c(s,u) + \alpha \sum_{s' \in S} P[s'|s,u] V_{n+1}(s') \right\}$$

for all $s \in S$. Let V^{π} denote the value function of policy π. Show that

$$V^{\pi} + \frac{2\alpha\varepsilon + \delta}{1-\alpha} \quad \geq \quad V^*$$

or

$$V^{\pi} + 2\varepsilon + \frac{\delta}{1-\alpha} \quad \geq \quad V^*$$

3. Suppose that V_0 is chosen so that

$$V_0(s) \quad \leq \quad T(V_0)(s) \quad := \quad \sup_{u \in U(s)} \left\{ c(s,u) + \alpha \sum_{s' \in S} P[s'|s,u] V_0(s') \right\}$$

for each state $s \in S$. Suppose that the Gauss-Seidel version of value iteration is used, and let V'_n denote the corresponding value function approximation after n iterations, with $V'_0 = V_0$. Show that

$$V_n(s) \quad \leq \quad V'_n(s) \quad \leq \quad V^*(s)$$

for each n and each $s \in S$, that is, Gauss-Seidel value iteration converges at least as fast as Jacobi value iteration.

4. Suppose that the modified policy iteration algorithm is used, with m_n Gauss-Seidel policy evaluation steps for policy π_n. Let $V''_{n,j}$ denote the corresponding value function approximation for policy π_n after j Gauss-Seidel policy evaluation iterations. Suppose that V_0 is chosen so that

$$V_0(s) \quad \leq \quad T(V_0)(s) \quad := \quad \sup_{u \in U(s)} \left\{ c(s,u) + \alpha \sum_{s' \in S} P[s'|s,u] V_0(s') \right\}$$

Let

$$V''_{1,0}(s) \;=\; \max_{u \in U(s)} \left\{ c(s,u) + \alpha \sum_{\{s' \in S \,:\, s' < s\}} P[s'|s,u] V''_{1,0}(s') + \alpha \sum_{\{s' \in S \,:\, s' \geq s\}} P[s'|s,u] V_0(s') \right\}$$

Let the initial policy $\pi_1 \in \Pi^{\mathrm{SD}}$ be chosen to satisfy

$$\pi_1(s) \;\in\; \arg\max_{u \in U(s)} \left\{ c(s,u) + \alpha \sum_{\{s' \in S \,:\, s' < s\}} P[s'|s,u] V''_{1,0}(s') + \alpha \sum_{\{s' \in S \,:\, s' \geq s\}} P[s'|s,u] V_0(s') \right\}$$

For all $n \geq 1$ and all $j = 0, 1, \ldots, m_n - 1$, let

$$V''_{n,j+1}(s) \;=\; c(s,\pi_n(s)) + \alpha \sum_{\{s' \in S \,:\, s' < s\}} P[s'|s,\pi_n(s)] V''_{n,j+1}(s') + \alpha \sum_{\{s' \in S \,:\, s' \geq s\}} P[s'|s,\pi_n(s)] V''_{n,j}(s')$$

For all $n \geq 1$, let

$$V''_{n+1,0}(s) \;=\; \max_{u \in U(s)} \left\{ c(s,u) + \alpha \sum_{\{s' \in S \,:\, s' < s\}} P[s'|s,u] V''_{n+1,0}(s') + \alpha \sum_{\{s' \in S \,:\, s' \geq s\}} P[s'|s,u] V''_{n,m_n}(s') \right\}$$

and let $\pi_{n+1} \in \Pi^{\mathrm{SD}}$ be chosen to satisfy

$$\pi_{n+1}(s) \;\in\; \arg\max_{u \in U(s)} \left\{ c(s,u) + \alpha \sum_{\{s' \in S \,:\, s' < s\}} P[s'|s,u] V''_{n+1,0}(s') + \alpha \sum_{\{s' \in S \,:\, s' \geq s\}} P[s'|s,u] V''_{n,m_n}(s') \right\}$$

Show that

$$V'_n(s) \;\leq\; V''_{n,m_n}(s) \;\leq\; V^*(s)$$

for each n and each $s \in S$, that is, modified policy iteration converges at least as fast as value iteration.