ISyE 6664 Stochastic Optimization

Fall 2012

Assignment 3

Issued: September 20, 2012

Due: October 4, 2012

Problem 1
Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a function $G : \mathbb{R} \times \Omega \mapsto \mathbb{R}$ such that $G(x, \cdot) : \Omega \mapsto \mathbb{R}$ is a random variable for each $x \in \mathbb{R}$. Suppose that $g : \mathbb{R} \mapsto \mathbb{R}$ is given by $g(x) := \mathbb{E}[G(x, \omega)] \in \mathbb{R}$ for all $x \in \mathbb{R}$. Suppose that we want to estimate the derivative $dg(x_0)/dx$ at a point x_0, for example, as part of a stochastic optimization algorithm. Suppose that

Condition A1: $dG(x_0, \omega)/dx$ exists with probability one, and

Condition A2: there is $\delta > 0$ and a random variable $K : \Omega \mapsto \mathbb{R}$ such that $\mathbb{E}[K(\omega)] < \infty$ and with probability one,

$$|G(x, \omega) - G(x_0, \omega)| \quad \leq \quad K(\omega)|x - x_0|$$

for all x such that $|x - x_0| < \delta$.

In class we used the dominated convergence theorem to show that Condition A1 and Condition A2 imply that

$$\frac{d\mathbb{E}[G(x_0, \omega)]}{dx} \quad = \quad \mathbb{E}\left[\frac{dG(x_0, \omega)}{dx}\right]$$

1. Sufficient conditions that are given in some textbooks are the following:

   Condition B1: For all x, $dG(x, \omega)/dx$ exists with probability one, and

   Condition B2: there is $\delta > 0$ and a random variable $K : \Omega \mapsto \mathbb{R}$ such that $\mathbb{E}[K(\omega)] < \infty$ and with probability one,

   $$\left|\frac{dG(x, \omega)}{dx}\right| \quad \leq \quad K(\omega)$$

   for all x such that $|x - x_0| < \delta$.

   (a) Either prove that Condition A2 implies Condition B2, or give a counterexample.

   (b) Either prove that Condition B2 implies Condition A2, or give a counterexample.

2. Another set of sufficient conditions are the following:

   Condition C1: $dG(x_0, \omega)/dx$ exists with probability one, and

Condition C2: for each $\varepsilon > 0$, there exists $C(\varepsilon)$ such that

$$\limsup_{x \to x_0} \int_{\left\{\left|\frac{G(x,\omega)-G(x_0,\omega)}{x-x_0}\right| > C(\varepsilon)\right\}} \left|\frac{G(x,\omega)-G(x_0,\omega)}{x-x_0}\right| \mathbb{P}(d\omega) \quad < \quad \varepsilon$$

that is, the family $\{(G(x,\omega)-G(x_0,\omega))/(x-x_0)\}$ of random variables for all x sufficiently close to x_0 is uniformly integrable.

(a) Show that if Conditions C1 and C2 hold, then

$$\frac{d\mathbb{E}[G(x_0,\omega)]}{dx} \quad = \quad \mathbb{E}\left[\frac{dG(x_0,\omega)}{dx}\right]$$

(b) Either prove that Condition A2 implies Condition C2, or give a counterexample.

(c) For bonus points: Either prove that Condition C2 implies Condition A2, or give a counterexample.

3. Next we construct an example in which Conditions A1, B1, and C1 hold, but Conditions A2, B2, and C2 do not hold, and it does not hold that

$$\frac{d\mathbb{E}[G(x_0,\omega)]}{dx} \quad = \quad \mathbb{E}\left[\frac{dG(x_0,\omega)}{dx}\right]$$

Suppose $\{N_\lambda(t,\omega) : t \geq 0, \omega \in \Omega\}$ is a homogeneous Poisson process with rate λ. For any fixed $t > 0$, let $g_t : [0,\infty) \mapsto [0,\infty)$ be defined by

$$g_t(\lambda) \quad := \quad \mathbb{E}[N_\lambda(t,\omega)]$$

(In terms of the notation of class, $G(\lambda,\omega) = N_\lambda(t,\omega)$ and $g(\lambda) = \mathbb{E}[G(\lambda,\omega)]$.) Suppose we want to estimate $g_t'(\lambda) = d\mathbb{E}[N_\lambda(t,\omega)]/d\lambda$. To support that, we need at least the following two things: (1) We need a single probability space for $\{N_\lambda(t,\omega) : t \geq 0, \omega \in \Omega\}$ over all $\lambda \geq 0$. (A single probability space for $G(\lambda,\omega)$ over all λ.) (2) We need to check whether $\mathbb{E}[dN_\lambda(t,\omega)/d\lambda] = d\mathbb{E}[N_\lambda(t,\omega)]/d\lambda$.

Let $\{N(t,\omega) : t \geq 0, \omega \in \Omega\}$ denote a homogeneous Poisson process with rate 1 on a probability space $(\Omega, \mathcal{F}, P)$. From this Poisson process we can construct a family $\{N_\lambda(t,\omega) : \lambda \geq 0, t \geq 0, \omega \in \Omega\}$ of Poisson processes on the same probability space $(\Omega, \mathcal{F}, P)$, by defining

$$N_\lambda(t,\omega) \quad := \quad N(\lambda t, \omega)$$

(a) Show that $\{N_\lambda(t,\omega) : t \geq 0, \omega \in \Omega\}$ as constructed above is a homogeneous Poisson process with rate λ.

(b) Calculate $g_t'(\lambda)$.

(c) Show that, with probability 1, $dN_\lambda(t,\omega)/d\lambda = 0$.

(d) Show that Conditions A2, B2, and C2 do not hold for $\{G(\lambda,\omega) = N_\lambda(t,\omega) : \lambda \geq 0, \omega \in \Omega\}$ for a fixed $t \geq 0$.

Problem 2
Consider the use of a diminishing stepsize algorithm to minimize the function $f(x) := x^4$. Specifically, consider the sequence $\{x^n\}$ generated by

$$x^{n+1} \quad := \quad x^n - \frac{1}{n+1} f'(x^n)$$

1. Suppose that $|x^0| \geq 1$. Show that $|x^n| > n^2$ for all n.
2. Suppose that $|x^0| \leq 1/2$. Show that $x^n \to 0$ as $n \to \infty$.

Problem 3
Consider the following stochastic approximation sequence $\{x^n\} \subset \mathbb{R}^d$:

$$x^{n+1} \quad := \quad x^n + \alpha^n s^{n+1}$$

where $\{\alpha^n\} \subset \mathbb{R}_+$ is a deterministic sequence such that

$$\sum_{n=0}^{\infty} \alpha^n \quad = \quad \infty \qquad \text{and} \qquad \sum_{n=0}^{\infty} (\alpha^n)^2 \quad < \quad \infty$$

Let $\mathcal{F}^n := \sigma(x^0, s^1, \ldots, s^n) = \sigma(x^0, x^1, \ldots, x^n)$. Suppose that there is a potential function (Lyapunov function) $L : \mathbb{R}^d \mapsto \mathbb{R}_+$ such that the following conditions hold: L is Lipschitz continuously differentiable, that is, there is a constant K such that

$$\|\nabla L(x_2) - \nabla L(x_1)\|_2 \quad \leq \quad K\|x_2 - x_1\|_2$$

for all $x_1, x_2 \in \mathbb{R}^d$. There is a constant C, and for each $x \in \mathbb{R}^d$ there is a probability distribution $F(x; \cdot)$ on $\mathbb{R}^d$, such that

$$\mathbb{E}[\|s^{n+1}\|_2^2 | \mathcal{F}^n] \quad = \quad \int_{\mathbb{R}^d} \|s\|_2^2 F(x^n; ds) \quad \leq \quad C[1 + L(x^n)]$$

Let

$$y(x^n) \quad := \quad \int_{\mathbb{R}^d} s F(x^n; ds) \quad = \quad \mathbb{E}[s^{n+1} | \mathcal{F}^n]$$

denote the conditional expected direction of movement from x^n. Suppose there is a $x^* \in \mathbb{R}^d$ such that for any $\varepsilon > 0$ it holds that

$$\sup_{\{x \,:\, \|x - x^*\|_2 \geq \varepsilon\}} \nabla L(x)^T y(x) \quad < \quad 0$$

1. Show that x^* is the unique minimizer of L, and for any sequence $\{x^{(k)}\} \subset \mathbb{R}^d$, if $L(x^{(k)}) \to L(x^*)$ as $k \to \infty$, then $x^{(k)} \to x^*$ as $k \to \infty$.
2. Show that, w.p.1, $x^n \to x^*$ as $n \to \infty$.