The University of Texas at Austin
Department of Electrical and Computer Engineering

EE381K: Large Scale Optimization — Fall 2016

Problem Set Two

Constantine Caramanis Due: Wednesday, September 21, 2016.

Reading Assignments

1. (?) Reading: Boyd & Vandenberghe: Chapters 9.1 - 9.5.

Written Problems

1. **Condition Number**. We saw in class that a fixed step size is able to guarantee linear convergence when the function is strongly convex and smooth. The choice of step size we gave in class, however, depended on the function f. Show that it is not possible to choose a fixed step size t, that gives convergence for any strongly convex function. That is, for any fixed step size t, show that there exists (by finding one!) a smooth (twice continuously-differentiable) strongly convex function with bounded Hessian, such that a fixed-stepsize gradient algorithm starting from some point x_0, does not converge to the optimal solution.

2. **Decreasing Stepsize.**[1] The previous problem shows that no constant step-size works for every strongly convex function. Consider now, a decreasing step size. Thus, at time k, you use step size $t_k \geq 0$. Show that if this sequence of step sizes satisfies:

$$\lim_k t_k = 0, \qquad \sum_{k=0}^{\infty} t_k = \infty,$$

then gradient descent converges to the global optimal solution. Hint: Recall that strong convexity implies lower and upper bounds on the Hessian. Each of these bounds in turn gives lower and upper bounds on the value of $f(y)$ with respect to $f(x)$. Use one of these to show that for k large enough,

$$f(x_{k+1}) \leq f(x_k) - \frac{1}{2} t_k \|\nabla f(x_k)\|_2^2.$$

Use the other inequality to get (lower) bound on $\|\nabla f(x_k)\|$ in terms of the optimality gap. Then put these together to conclude that gradient descent must converge.

3. **Convex functions**

   (a) If f_i are convex functions, show that $f(x) := \sup_i f_i(x)$ is also convex.

   (b) Show that the largest eigenvalue of a matrix is a convex function of the matrix (i.e. $\lambda_{\max}(M)$ is a convex function of M). Is the same true for the eigenvalue of largest magnitude ?

---

[1]This problem borrowed from Nati Srebro.

(c) Consider a weighted graph with edge weight vector w. Fix two nodes a and b. The *weighted shortest path* from a to b is the path whose sum of edge weights is the minimum, among all paths with one endpoint at a and another at b. Let $f(w)$ be the weight of this path. Show that f is a concave function of w.

4. **Convex functions: Jensen's Inequality**. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be any function. Its epigraph is defined as the set:

$$\text{epi}(f) = \{(x, y) \in \mathbf{R}^{n+1} \, : \, y \geq f(x)\}.$$

(a) Show that if f is convex, then epi(f) is also convex.

(b) Prove (the finite version of) Jensen's inequality. Jensen's inequality says that if p is a distribution on $\{x_1, \ldots, x_m\}$ with weights $p_1, \ldots, p_m$, and f is any concave function, then

$$\mathbb{E}[f(X)] \leq f(\mathbb{E}(X)).$$

5. **Projection**.

(a) Suppose that $\mathcal{X} \subseteq \mathbb{R}^d$ is a closed and bounded convex set. Let $\mathbf{y} \in \mathcal{X}$ be any point in $\mathcal{X}$. The projection of $\mathbf{y}$ on $\mathcal{X}$ is defined by

$$\Pi_{\mathcal{X}}(\mathbf{y}) = \arg\min_{\mathbf{x} \in \mathcal{X}} \|\mathbf{y} - \mathbf{x}\|_2^2.$$

Show that the solution to the optimization problem is unique.

(b) Show that for $\mathcal{X}$ as above, if $\mathbf{y} \notin \mathcal{X}$, there exists a hyperplane with $\mathcal{X}$ on one side, and $\mathbf{y}$ strictly on the other side. That is, show that there is a vector $\mathbf{s}$ and a scalar b, such $\langle \mathbf{s}, \mathbf{x} \rangle \leq b$ for all $\mathbf{x} \in \mathcal{X}$, and $\langle \mathbf{s}, \mathbf{y} \rangle > b$.

(c) We discussed (briefly) projected gradient descent, where the update is given by:

$$x^{(k+1)} = \text{Proj}_{\mathcal{X}}(x^{(k)} - t_k \nabla f(x^{(k)})).$$

Show that this is equivalent to the update:

$$x^{(k+1)} = \arg\min_{x \in \mathcal{X}} \left\{ \langle x, \nabla f(x^{(k)}) \rangle + \frac{1}{2t_k} \|x - x^{(k)}\|_2^2 \right\}.$$

6. **Computing Projections**. For the given convex set $\mathcal{X}$, compute the projection of a point z.

(a) $\mathcal{X} = \mathbb{R}^n_+$.

(b) Euclidean ball: $\{x \, : \, \|x\|_2 \leq 1\}$.

(c) Positive semidefinite cone: $S^n_+ = \{M \in S^n \, : \, x^\top M x \geq 0, \; \forall x \in \mathbb{R}^n\}$.

(d) $\mathcal{X}$ is a rectangle defined by vectors L and U that satisfy $U_i \geq L_i$. Thus, $\mathcal{X} = \{x \, : \, L_i \leq x_i \leq U_i, \; i = 1, \ldots, n\}$.

(e) 1-norm ball: $\{x \, : \, \sum_i |x_i| \leq 1\}$.

(f) Probability simplex: $\mathcal{X} = \{x \, : \, \sum_i x_i = 1, \; x_i \geq 0, \; i = 1, \ldots, n\}$.