CS 726: Homework #2

Posted: 09/24/2020, due: 10/05/2020 by 5pm CT on Canvas

Please typeset or write your solutions neatly! If we cannot read it, we cannot grade it.

Note: You can use the results we have proved in class – no need to prove them again.

Q 1. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a twice continuously differentiable function. Prove that f is m-strongly convex if and only if

$$\nabla^2 f(\mathbf{x}) \succeq m\mathbf{I}, \quad \forall \mathbf{x} \in \mathbb{R}^d,$$

that is, $(\nabla^2 f(\mathbf{x}) - m\mathbf{I})$ is positive semidefinite where $\mathbf{I}$ is the identity matrix. [20pts]

Q 2. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a continuously differentiable function that satisfies the following:

$$(\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d) : \quad f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{m}{2}\|\mathbf{y} - \mathbf{x}\|_2^2,$$

where $m > 0$ is some constant.

Prove that f cannot be Lipschitz continuous on the entire $\mathbb{R}^d$. Would it be possible for f to be Lipschitz continuous if we take the domain to be the unit Euclidean ball (i.e., the set: $\{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_2 \leq 1\}$)? Explain. [20pts]

Q 3. Let $f : \mathbb{R}^d \to \mathbb{R}$ be an L-smooth function, where $L \in (0, \infty)$ is a parameter you are given. Consider the following randomized coordinate descent update rule:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \nabla_{i_k} f(\mathbf{x}_k)\mathbf{e}_{i_k},$$

where i_k is chosen uniformly at random from the set $\{1, 2, ..., d\}$ (and independently from any prior random choices), $\mathbf{e}_{i_k}$ is the vector that has 0 in all coordinates except for i_k, where it equals 1 (it is the ${i_k}^{\text{th}}$ standard basis vector), and α_k is the step size you are asked to determine. Prove that there exists the choice of the step size $\alpha_k > 0$ and a constant $\beta > 0$ such that:

$$\mathbb{E}_{i_k \sim \texttt{Unif}(\{1,...,d\})}[f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k)] \leq -\frac{\beta}{2}\|\nabla f(\mathbf{x}_k)\|_2^2.$$

How would you choose α_k?

Hint: You should use Lemma 2.2 to solve this question. [20pts]

Q 4 (Bregman Divergence)**.** Bregman divergence of a continuously differentiable function $\psi : \mathbb{R}^d \to \mathbb{R}$ is a function of two points defined by

$$D_\psi(\mathbf{x}, \mathbf{y}) = \psi(\mathbf{x}) - \psi(\mathbf{y}) - \langle \nabla \psi(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle .$$

Equivalently, you can view Bregman divergence as the error in the first-order approximation of a function:

$$\psi(\mathbf{x}) = \psi(\mathbf{y}) + \langle \nabla \psi(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle + D_\psi(\mathbf{x}, \mathbf{y}).$$

(i) What is the Bregman divergence of a simple quadratic function $\psi(\mathbf{x}) = \frac{1}{2}\|\mathbf{x} - \mathbf{x}_0\|_2^2$, where $\mathbf{x}_0 \in \mathbb{R}^d$ is a given point? [5pts]

(ii) Given $\mathbf{x}_0 \in \mathbb{R}^d$ and some continuously differentiable $\psi : \mathbb{R}^d - \mathbb{R}$, what is the Bregman divergence of function $\phi(\mathbf{x}) = \psi(\mathbf{x}) + \langle \mathbf{x}_0, \mathbf{x} \rangle$? [5pts]

(iii) Use Part (ii) and the definition of Bregamn divergence to prove the following 3-point identity:

$$(\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^d) : \quad D_\psi(\mathbf{x}, \mathbf{y}) = D_\psi(\mathbf{z}, \mathbf{y}) + \langle \nabla \psi(\mathbf{z}) - \nabla \psi(\mathbf{y}), \mathbf{x} - \mathbf{z} \rangle + D_\psi(\mathbf{x}, \mathbf{z}).$$ [5pts]

(iv) Suppose I give you the following function: $m_k(\mathbf{x}) = \sum_{i=0}^k a_i \psi_i(\mathbf{x})$, where $\psi_i(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}-\mathbf{x}_i\|_2^2 + \langle \mathbf{b}_i, \mathbf{x} - \mathbf{x}_i \rangle$, where $\{a_i\}_{i\geq 0}$ is a sequence of positive reals and $\{\mathbf{b}_i\}_{i=0}^k$, $\{\mathbf{x}_i\}_{i=0}^k$ are fixed vectors from $\mathbb{R}^d$. Define $\mathbf{v}_k = \operatorname{argmin}_{\mathbf{x}\in\mathbb{R}^d} m_k(\mathbf{x})$ and $A_k = \sum_{i=0}^k a_i$. Using what you have proved so far, prove the following inequality:

$$(\forall \mathbf{x} \in \mathbb{R}^d): \quad m_{k+1}(\mathbf{x}) = m_k(\mathbf{v}_k) + a_{k+1}\psi_{k+1}(\mathbf{x}) + \frac{A_k}{2}\|\mathbf{x}-\mathbf{v}_k\|_2^2. \qquad \text{[5pts]}$$

Q 5 (Gradient descent with ℓ_p norms)**.** Let $p > 1$ be a parameter and let $q = \frac{p}{p-1}$ (so that $\frac{1}{p} + \frac{1}{q} = 1$). Prove that the following function:

$$h_{\mathbf{z}}(\mathbf{x}) = \langle \mathbf{z}, \mathbf{x} \rangle + \frac{1}{2}\|\mathbf{x}\|_p^2$$

is minimized for $\mathbf{x} = -\nabla(\frac{1}{2}\|\mathbf{z}\|_q^2)$ and that $\min_{\mathbf{x}\in\mathbb{R}^d} h_{\mathbf{z}}(\mathbf{x}) = -\frac{1}{2}\|\mathbf{z}\|_q^2$.

Now let $f : \mathbb{R}^d \to \mathbb{R}$ be a function that is L-smooth w.r.t. $\|\cdot\|_p$, for some L, i.e.,

$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d: \quad \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_q \leq L\|\mathbf{x} - \mathbf{y}\|_p.$$

Consider the following update rule:

$$\mathbf{x}_{k+1} = \operatorname*{argmin}_{\mathbf{u}\in\mathbb{R}^d} \left\{ f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \mathbf{u} - \mathbf{x}_k \rangle + \frac{L}{2}\|\mathbf{u} - \mathbf{x}_k\|_p^2 \right\}.$$

Use the first part of the question to argue that:

$$f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) - \frac{1}{2L}\|\nabla f(\mathbf{x}_k)\|_q^2.$$

Assuming that f is bounded below, derive the bound for $\min_{0\leq i\leq k}\|\nabla f(\mathbf{x}_i)\|_q$ similar to the one that was derived in class for $p = 2$. What is the best bound you could have gotten for $\min_{0\leq i\leq k}\|\nabla f(\mathbf{x}_i)\|_q$ if instead of the approach used in this question, you used standard gradient descent (w.r.t. $\|\cdot\|_2$) that we analyzed in class? [20pts]