CS 726: Homework #2

Posted: Feb 4, 2025. Due: Feb 17, 2025 on Canvas

Please typeset your solutions.

You should provide sufficient justification for the steps of your solution. The level of detail should be such that your fellow students can understand your solution without asking you for further explanation.

Q1.1	Q1.2	Q2.1	Q2.2	Q2.3	Q3.1	Q3.2	Q3.3	Q3.4	Q4.1	Q4.2	Q4.3	Total
10	10	10	10	10	5	5	5	5	10	10	10	100 pts

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

Q 1 Smoothness

Suppose that $f : \mathbb{R}^d \to \mathbb{R}$ is a convex and L-smooth with respect to the Euclidean norm $\|\cdot\|_2$, and f has a minimizer x^* with function value $f^* = f(x^*)$.

Q 1.1

Show that for any $x \in \mathbb{R}^d$ we have

$$f(x) - f^* \geq \frac{1}{2L} \|\nabla f(x)\|_2^2 .$$

Hint: Recall that an L-smooth f satisfies $f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2} \|y - x\|_2^2 , \forall x, y$.

Q 1.2

Prove the following: for any $x, y \in \mathbb{R}^n$, we have

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \frac{1}{L} \|\nabla f(x) - \nabla f(y)\|_2^2 .$$

Hint: Apply the Part 1 to the two functions h_x and h_y defined as

$$h_x(z) := f(z) - \langle \nabla f(x), z \rangle , \qquad h_y(z) := f(z) - \langle \nabla f(y), z \rangle .$$

To apply Part 1, you need to first argue that h_x and h_y (as functions of z) are convex and L-smooth.

Q 2 Randomized coordinate descent

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a continuously differentiable function, let $\{L_1, \ldots, L_d\}$ be positive constants, and suppose that for all $i \in \{1, \ldots, d\}$, all $\delta \in \mathbb{R}$, and all $\mathbf{x} \in \mathbb{R}^d$, you have

$$|\nabla_i f(\mathbf{x} + \delta \mathbf{e}_i) - \nabla_i f(\mathbf{x})| \leq L_i |\delta|, \tag{1}$$

where $\mathbf{e}_i$ is the i^{th} standard basis vector (i.e., the vector with all zeros except for the i^{th} entry, which equals one) and ∇_i denotes the i^{th} entry of the gradient.

Q 2.1 Quadratic upper bound

Prove that for all $i \in \{1, \ldots, d\}$, all $\delta \in \mathbb{R}$, and all $\mathbf{x} \in \mathbb{R}^d$,

$$f(\mathbf{x} + \delta \mathbf{e}_i) - f(\mathbf{x}) \leq \delta \nabla_i f(\mathbf{x}) + \frac{L_i}{2} |\delta|^2. \tag{2}$$

Q 2.2 Expected descent

Now consider the following randomized coordinate descent update rule:

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

where i_k is chosen uniformly at random from the set $\{1, 2, ..., d\}$ (and independently from any prior random choices) and $\alpha_{i_k} > 0$ is the step size you are asked to determine. Prove that there exists a choice of the step sizes α_i, $i \in \{1, \ldots, d\}$, and a constant $\beta > 0$ such that:

$$\mathbb{E}_{i_k \sim \mathrm{Unif}(\{1,\ldots,d\})} \big[f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k) \big] \leq -\frac{\beta}{2} \|\nabla f(\mathbf{x}_k)\|_2^2. \tag{3}$$

How would you choose α_{i_k}'s? What is the largest β you can get this way?

(The expectation in (3) should be understood as a conditional expectation: we are treating $\mathbf{x}_k$ as fixed, and the expectation is with respect to the randomness in sampling i_k uniformly at random from $\{1, \ldots, d\}$.)

Q 2.3 Convergence rate

Prove that if the function f is bounded below by some number $f^* > -\infty$, then

$$\min_{0 \leq k \leq K} \mathbb{E} \left[\|\nabla f(\mathbf{x}_k)\|_2^2 \right] \leq \frac{2(f(\mathbf{x}_0) - f^*)}{\beta(K+1)}, \tag{4}$$

where the expectation is taken w.r.t. all the random choices the algorithm takes (i.e., over all $i_1, i_2, \ldots, i_K$).

Q 3 Bregman divergence

Given a continuously differentiable function $\psi : \mathbb{R}^d \to \mathbb{R}$, the *Bregman divergence* of ψ is a function $D_\psi(\cdot, \cdot)$ of two points defined by

$$D_\psi(\mathbf{x}, \mathbf{y}) = \psi(\mathbf{x}) - \psi(\mathbf{y}) - \langle \nabla \psi(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle. \tag{5}$$

One may view the Bregman divergence as the error in the first-order Taylor's approximation of the function ψ:

$$\psi(\mathbf{x}) = \psi(\mathbf{y}) + \langle \nabla \psi(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle + D_\psi(\mathbf{x}, \mathbf{y}). \tag{6}$$

Q 3.1

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?

Q 3.2

Given a fixed vector $\mathbf{z} \in \mathbb{R}^d$ and a continuously differentiable $\psi : \mathbb{R}^d \to \mathbb{R}$, what is the Bregman divergence D_ϕ of the function ϕ given by $\phi(\mathbf{x}) = \psi(\mathbf{x}) + \langle \mathbf{z}, \mathbf{x} \rangle$?

Q 3.3

Prove the following 3-point identity:

$$(\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^d): \quad D_\psi(\mathbf{x}, \mathbf{y}) = D_\psi(\mathbf{z}, \mathbf{y}) + \langle \nabla \psi(\mathbf{z}) - \nabla \psi(\mathbf{y}), \mathbf{x} - \mathbf{z} \rangle + D_\psi(\mathbf{x}, \mathbf{z}). \tag{7}$$

Hint: Use Q3.2, or use the definition of Bregman divergence.

Q 3.4

Let $\mathbf{z} \in \mathbb{R}^d$ and $\bar{\mathbf{x}} \in \mathbb{R}^d$ be given, fixed vectors. Consider the function: $h(\mathbf{x}) = \langle \mathbf{z}, \mathbf{x} \rangle + D_\psi(\mathbf{x}, \bar{\mathbf{x}})$. Let $\mathcal{X}$ be a closed convex set. Define $\mathbf{y} = \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} h(\mathbf{x})$. Prove that

$$\forall \mathbf{x} \in \mathcal{X} : \quad h(\mathbf{x}) \geq \langle \mathbf{z}, \mathbf{y} \rangle + D_\psi(\mathbf{y}, \bar{\mathbf{x}}) + D_\psi(\mathbf{x}, \mathbf{y}). \tag{8}$$

Hint: You may use the following first-order necessary condition for constrained optimization: if $\mathbf{y} = \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} h(\mathbf{x})$, then $\langle \nabla h(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \geq 0, \forall \mathbf{x} \in \mathcal{X}$.

Q 4 Gradient descent with ℓ_p norms

Let $p \in (1, \infty)$ be a parameter and let $q = \frac{p}{p-1}$ (so that $\frac{1}{p} + \frac{1}{q} = 1$).

Q 4.1

For a given $\mathbf{z} \in \mathbb{R}^d$, consider a function $h_{\mathbf{z}} : \mathbb{R}^d \to \mathbb{R}$ defined by

$$h_{\mathbf{z}}(\mathbf{x}) = \langle \mathbf{z}, \mathbf{x} \rangle + \frac{1}{2} \|\mathbf{x}\|_p^2. \tag{9}$$

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

Hint: HW1 Q2 may be useful.

Q 4.2

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:

$$\mathbf{x}_{k+1} = \operatorname*{argmin}_{\mathbf{u} \in \mathbb{R}^d} \Big\{ f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \mathbf{u} - \mathbf{x}_k \rangle + \frac{L}{2}\|\mathbf{u} - \mathbf{x}_k\|_p^2 \Big\}. \tag{10}$$

Use the first part of the question to argue that:

$$f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) - \frac{1}{2L}\|\nabla f(\mathbf{x}_k)\|_q^2. \tag{11}$$

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 $p = 2$.

Q 4.3

Still assume that $f : \mathbb{R}^d \to \mathbb{R}$ is L-smooth w.r.t. $\|\cdot\|_p$, where $p > 1$. What is L_2, defined as the smoothness parameter of f w.r.t. the ℓ_2 norm? (Hint: use HW1 Q1.)

Instead of using the update rule in Q4.2, let us now apply the standard gradient descent algorithm that we analyzed in class. Using the convergence result derived in class, what is the best bound you can get for $\min_{0 \leq i \leq k} \|\nabla f(\mathbf{x}_i)\|_q$?