CS591 E2: Optimization Methods and their Applications – Fall 2016

Homework 1

Due date: Thursday, October 6, 2016

Homework Policy and Guidelines

You are encouraged to collaborate on the solution of the homeworks and to consult any materials, but you must write up your own answers and you must acknowledge all of your collaborators and sources.

The problems marked with a $(*)$ are more challenging. You may not be able to completely solve some of the more challenging problems, that is completely normal!

Some of the problems ask you to fill in a proof that we did not cover in class; the readings will often have these proofs, you are free to consult them but you must write up a complete proof in your own words. In general it may be good to keep in mind that some of the proofs in the textbooks may leave some steps to the reader, and it is very important to make sure that you know how to fill in those missing steps. Also, thinking about those proofs on your own will help you understand the material better.

Notation. For $x \in \mathbb{R}^n$ and $p \in \mathbb{R}$, we use $\|x\|_p$ to denote the ℓ_p-norm of x, i.e., $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$.

Problem 1 (Some famous inequalities) *The Cauchy-Schwartz inequality states that, for all vectors a and b in $\mathbb{R}^n$,*

$$|\langle a, b\rangle| \leq \|a\|_2 \|b\|_2.$$

(*a*) *Prove the Cauchy-Schwartz inequality.*
Hint. *Consider the following function $g : \mathbb{R} \to \mathbb{R}$, where $g(t) = \|a + tb\|_2^2$ for all $t \in \mathbb{R}$. This function is nonnegative for all t. What is the minimum value $\min_{t\in\mathbb{R}} g(t)$ of g?*

(*b*) *Use Cauchy-Schwartz to show that, for any $x \in \mathbb{R}^n$,*

$$\|x\|_1 \leq \sqrt{n}\|x\|_2.$$

Problem 2 (Equivalence of convexity definitions) *In class we have seen that a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if, for any $x, y \in \mathbb{R}^n$,*

$$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle .$$

In class we proved this equivalence for the one-dimensional case $n = 1$. Use this proof to show the equivalence for the general case.

Problem 3 (First-order optimality condition.) *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function and $\mathcal{X} \subseteq \mathbb{R}^n$ be a closed convex set over which f is differentiable. Then*

$$x^* \in \operatorname{argmin}_{x\in\mathcal{X}} f(x)$$

if and only if we have

$$\langle \nabla f(x^*), x^* - y\rangle \leq 0 \quad \forall y \in \mathcal{X}.$$

Problem 4 ($(*)$ Projection on the ℓ_1 and ℓ_2 balls) *Recall that the ℓ_p-ball of radius r centered at a point x_0 is the set $B_p(x_0, r) = \{x \in \mathbb{R}^n \mid \|x - x_0\|_p \leq r\}$. Recall that the ℓ_2 projection of a point z onto a set $\mathcal{X}$ is the point $\Pi_{\mathcal{X}}(z) = \operatorname{argmin}_{x\in\mathcal{X}} \|z - x\|_2^2$. Show how to compute the projection of a point onto the ℓ_1 and ℓ_2 balls $B_1(x_0, r)$ and $B_2(x_0, r)$.*

Problem 5 *Prove the following property of the projection $\Pi_{\mathcal{X}}$ onto $\mathcal{X}$ that we saw in class. Let x and y be two points in $\mathbb{R}^n$. Show that*

$$\|\Pi_{\mathcal{X}}(y) - x\| \leq \|y - x\|.$$

Problem 6 ((∗) Maximum of a convex function over a polyhedron) *A polyhedron $\mathcal{P} \subseteq \mathbb{R}^n$ is the set of all points that are convex combinations of finitely many points $v_1, v_2, \ldots, v_k \in \mathbb{R}^n$; we write $\mathcal{P} = \mathbf{conv}(\{v_1, \ldots, v_k\})$ and we refer to the points v_i as the vertices of $\mathcal{P}$. Show that the maximum of a convex function f over the polyhedron $\mathcal{P} = \mathbf{conv}(\{v_1, \ldots, v_k\})$ is achieved at one of its vertices, i.e.,*

$$\sup_{x \in \mathcal{P}} f(x) = \max_{i=1,\ldots,k} f(v_i).$$

(A stronger statement is: the maximum of a convex function over a closed bounded convex set is achieved at an extreme point, i.e., a point in the set that is not a convex combination of any other points in the set. You do not need to prove the stronger statement.)

Problem 7 (Bregman divergence) *A convex function f is strictly convex if $f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta)f(y)$ for all $x \neq y$ and all $0 < \theta < 1$ (the convex inequality holds with strict inequality). Let $f : \mathbb{R}^n \to \mathbb{R}$ be strictly convex and differentiable. The* Bregman divergence *associated with f is the function*

$$D_f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, D_f(x, y) = f(x) - f(y) - \langle \nabla f(y), x - y \rangle .$$

(a) *Show that $D_f(x, y) \geq 0$ for all $x, y \in \mathbf{dom}(f)$, where $\mathbf{dom}(f)$ is the domain of f.*

(b) *Show that, if $f = \| \cdot \|_2^2$, $D_f(x, y) = \|x - y\|_2^2$.*

(c) *Show that, if $f(x) = \sum_{i=1}^n x_i \log(x_i)$ with $\mathbf{dom}(f) = \mathbb{R}^n_+$ (with $0 \log 0$ taken to be 0), then*

$$D_f(x, y) = \sum_{i=1}^n (x_i \log(x_i/y_i) - x_i + y_i).$$

f is called the negative entropy *and D_f is the* Kullback-Leibler *divergence.*

(d) (Bregman projection). *The previous parts suggest that Bregman divergences can be viewed as generalized "distances," i.e., functions that measure how similar two vectors are. Explain whether the following problem is a convex optimization problem, where $\mathcal{X}$ is a convex set and y is a given point: $\min_{x \in \mathcal{X}} D_f(x, y)$.*
