Convex Analysis and Optimization 16:711:558

Fall 2011 Rutgers University Prof. Eckstein

Homework 7

Due by 10:00am, Tuesday, December 20

Note: since this homework will be considered one of the "take-home exams", you may not collaborate with other students in any way. You may refer to the course textbook, earlier homework assignments and solutions, and your notes from class; you may not refer to other materials. You may treat any result proved in class or on an earlier homework as given.

I will be abroad December 7-19, but available to answer questions by e-mail (although there may be a substantial delay due to differing time zones). You may hand in this assignment to Clare Smietana in RUTCOR room 123-A between 9:30am and 3:00pm on any weekday before the due date, or give it to either me or Clare by 10:00am on December 20.

Note that the questions integrate material from the entire course.

1. *Directional Derivatives and Subgradients.* Recall that the directional derivative of the function f at x in the direction d is defined to be

$$f'(x;d) = \lim_{\alpha \searrow 0} \frac{f(x+\alpha d) - f(x)}{\alpha},$$

and that for convex $f : \mathbb{R}^n \to (\infty, +\infty]$ we established that this limit always exists (although it might be $+\infty$) and is equal to $\inf_{\alpha>0}\{\big(f(x+\alpha d) - f(x)\big)/\alpha$, since the expression $\big(f(x+\alpha d) - f(x)\big)/\alpha$ is a nondecreasing function of α (see Homework 2, problem 2).

   (a) Early in the semester, we proved that if f is convex and differentiable at x, then $\nabla f(x)$ is subgradient of f at x, that is, $f(x') \geq f(x) + \langle \nabla f(x), x' - x\rangle$ for all $x \in \mathbb{R}^n$. However, at several junctures we used without proof that at points x where f is differentiable, $\nabla f(x)$ is the *only* subgradient of f at x, that is, $\partial f(x) = \{\nabla f(x)\}$. Prove that this is the case. You may use without proof the standard calculus result that if f is differentiable at x, then $f'(x;d) = \langle \nabla f(x), d\rangle$ for all $d \in \mathbb{R}^n$.

   (b) Now assume that $f : \mathbb{R}^n \to (\infty, +\infty]$ is proper and convex, and consider a point $x \in \operatorname{dom} f$. Do not assume that f is differentiable at x. Show that $g \in \partial f(x)$ if and only if $f'(x;d) \geq \langle g, d\rangle$ for all $d \in \mathbb{R}^n$.

   (c) For some fixed x, let $h(d) = f'(x;d)$, that is, $h = f'(x;\,\cdot\,)$. Show that h is a convex and positively homogeneous function (see the definition of positive homogeity in Homework 5, problem 4).

   (d) Recall that in Homework 5, problem 4, we proved that the convex conjugate of a positively homogeneous convex function h is the indicator function of the closed convex set $C = \{u \mid \langle u, x\rangle \leq h(x)\ \forall\, x \in \mathbb{R}^n\}$ (the expression for C was actually implicit in the proof provided in the posted solution, but many of you proved it explicitly in your own solutions; here, you may take the expression as given). Combine this observation with part (b) to conclude that $\widehat{h} = \delta_{\partial f(x)}$, that is, the convex conjugate of the directional derivative function $f'(x;\,\cdot\,)$ is the indicator function of the set of subgradients of f at x.

(e) Show that if $h = f'(x;\,\cdot\,)$ is closed, then $f(x; d) = \sup\left\{\langle g, d\rangle \;\middle|\; g \in \partial f(x)\right\}$. *Hint*: use the properties of convex conjugate functions and the results of Homework 5, problem 4.

(f) Show that if $x \in \operatorname{ri}\operatorname{dom} f$, then $\operatorname{dom} h$ is a linear subpace of $\mathbb{R}^n$, and that it then follows that h must be closed and the "sup" formula from part (e) holds.

2. *Lagrangian Relaxation.* One common use of Lagrangian relaxation/subgradient algorithms is in combinatorial optimization. Suppose we have the problem

$$\begin{array}{ll} \min & c^\top x \\ \text{S.T.} & Ax = b \\ & x \in X, \end{array}$$

where X is a large but finite (and hence nonconvex) set of points over which one can "easily" optimize a linear function. However, with the addition of the constraints $Ax = b$, the optimization task becomes difficult. The Lagrangian relaxation approach attempts to reduce the harder task to a sequence of easier ones via the recursions

$$x^{k+1} \in \operatorname{Arg\,min}_{x \in X} \left\{c^\top x + \langle \lambda^k, Ax - b\rangle\right\} \tag{1}$$

$$\lambda^{k+1} = \lambda^k + \alpha_k(Ax^{k+1} - b), \tag{2}$$

where $\alpha_k = \tau_k \gamma_k$ is a stepsize of the form we discussed in our analysis of the subgradient method. Note that the step (1) reduces to just minimizing the linear objective given by $c + A^\top \lambda$ over X, which is assumed to be "easy".

A classic case of this form is the Held-Karp algorithm for the traveling salesman problem, where X is taken to be the set of all "1-tree" (trees in a graph, plus one additional edge), and the constraints $Ax = b$ express that every node should have degree 2. The linear ordering problem (LOP) is another combinatorial problem that has been successfully attacked with methods of this form.

Suppose we apply use the standard duality-generating function

$$F(x, u) = \begin{cases} c^\top x, & \text{if } x \in X,\ Ax - b + u = 0 \\ +\infty, & \text{otherwise,} \end{cases}$$

even though X is not convex. Show that the optimal value of the dual function $F^*(0, \lambda)$ obtained in this case is the optimal value z^* of the linear programming problem

$$\begin{array}{ll} \min & c^\top x \\ \text{S.T.} & Ax = b \\ & x \in \operatorname{conv} X, \end{array} \tag{3}$$

Hint: show that $\operatorname{cl}\operatorname{conv} F$ is equal to the standard duality-generating function of problem (3) above, and that strong duality must hold for that problem.

3. Consider the convex optimization problem:

$$\begin{array}{ll} \min & f(x) \\ \text{ST} & Ax - b \in K \\ & x \in X, \end{array}$$

where $f : \mathbb{R}^n \to (-\infty, +\infty]$ is closed proper convex, A in an $n \times m$ matrix, $b \in \mathbb{R}^m$, $K \subseteq \mathbb{R}^m$ is a nonempty closed convex cone, and $X \subseteq \mathbb{R}^m$ is a nonempty closed convex set. Define the duality generating function for this problem to be

$$F(x, u) = \begin{cases} f(x), & \text{if } x \in X \text{ and } Ax - b + u \in K \\ +\infty, & \text{otherwise.} \end{cases}$$

In this problem, we will demonstrate that with F defined in this manner, the augmented Lagrangian method takes the form

$$x^{k+1} \in \text{Arg}\ \min_{x \in X} \left\{ f(x) + \left[\text{dist}_K\big(\lambda^k + c_k(Ax - b)\big)\right]^2 \right\} \tag{4}$$

$$\lambda^{k+1} = \text{proj}_{K^*}\big(\lambda^k + c_k(Ax^{k+1} - b)\big). \tag{5}$$

For this problem, we use the following notation: if C is any closed convex set, $\text{proj}_C(\cdot)$ denotes projection onto C, that is, $\text{proj}_C(v) = \arg\min_{w \in C}\{\|w - v\|\}$, and furthermore $\text{dist}_C(v) = \inf_{w \in C}\{\|w - v\|\} = \|\text{proj}_C(v) - v\|$, the distance from v to C. As usual, K^* denotes the polar of K.

(a) Derive an expression (in terms of f, A, b, X, and K^*) for the Lagrangian $L(x, \lambda)$ for the form of F given above.

(b) Let $C \subseteq \mathbb{R}^m$ be any convex cone, and $v \in \mathbb{R}^m$. Show that $w = \text{proj}_C(v)$ if and only if $w \in C$, $v - w \in C^*$, and $\langle v - w, w \rangle = 0$ (equivalently, $\langle v, w \rangle = \|w\|^2$). *Hint:* refer to Homework 5, problem 2(a).

(c) For $C \subseteq \mathbb{R}^m$ still defined to be any closed convex cone, show that for any vector $v \in \mathbb{R}^m$ and scalar $\alpha \geq 0$, we have $\text{proj}_C(\alpha v) = \alpha\, \text{proj}_C(v)$.

(d) Show that for any closed convex cone $C \subseteq \mathbb{R}^m$ and vector $v \in \mathbb{R}^m$, we have $\text{proj}_{C^*}(v) = v - \text{proj}_C(v)$.

(e) Show that for any scalar $c > 0$ and $\ell, t \in \mathbb{R}^m$, the solution to the problem $\min\left\{ \langle \ell, u \rangle + \frac{c}{2}\|u\|^2 \;\middle|\; u + t \in K \right\}$ is $u = \text{proj}_K(t - \frac{1}{c}\ell) - t$.

(f) Show that if the sequence $\{(x^k, \lambda^k)\} \subset \mathbb{R}^n \times \mathbb{R}^m$ obey the recursions (4)-(5), then the sequence $\{\lambda^k\}$ satisfies the proximal minimization recursions

$$\lambda^{k+1} = \arg\min_{\lambda \in \mathbb{R}^m} \left\{ -q(\lambda) + \tfrac{1}{2c}\|\lambda - \lambda^k\|^2 \right\},$$

where q denotes the dual function $q(\lambda) = \sup_{x \in \mathbb{R}^n}\{L(x, \lambda)\}$ for $L(\cdot\,, \cdot)$ as derived in part (a). Note: you may find parts (c)-(e) of this question helpful.