Solution sheet

MATH 450 - 558: Smooth and Nonsmooth Optimization

Instructor: Emre Mengi

Fall Semester 2015
Final Exam
Tuesday January 5th, 2016
Duration: 180 minutes

NAME ____________________

STUDENT ID ____________________

SIGNATURE ____________________

#1	20	
#2	20	
#3	20	
#4	20	
#5	20	
Σ	100	

- Please put your name, student ID and signature in the space provided above.
- You may use the lecture notes and/or your own notes, but you should not be using any textbooks.

Question 1 (20 points) Express the dual problem of

$$\text{minimize}_{x \in \mathbb{R}^n}\ c^T x \quad \text{subject to} \quad f(x) \geq 0$$

with $c \neq 0$, in terms of the Fenchel conjugate f^*.

Lagrangian function

$$L(x;\lambda) = c^T x - \lambda f(x)$$

$$g(\lambda) = \inf_{x \in \mathbb{R}^n} c^T x - \lambda f(x)$$

Would like to maximize $g(\lambda)$ over $\lambda \geq 0$.

Without loss of generality $\lambda \neq 0$, since $g(0) = -\infty$. Otherwise,

$$g(\lambda) = -\lambda \sup_{x \in \mathbb{R}^n} \left(-\frac{c}{\lambda}\right)^T x - (-f(x))$$

$$= -\lambda \left[-f\left(\frac{-c}{\lambda}\right)\right]^*$$

Thus, the dual problem is given by

$$\sup_{\lambda \geq 0} g(\lambda)$$

$$=$$

$$\sup_{\lambda > 0} -\lambda \left[-f\left(\frac{-c}{\lambda}\right)\right]^*.$$

Question 2 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a Lipschitz continuous function with Lipschitz constant γ, that is

$$|f(x) - f(y)| \;\leq\; \gamma\|x - y\|_2$$

for all $x, y \in \mathbb{R}^n$.

(a) (10 points) Prove the following regarding the generalized directional derivative $f^{(0)}$ of f:

$$|f^\circ(x; p) - f^\circ(y; q)| \leq \gamma\,(\|x - y\|_2 + \|p - q\|_2) \quad \forall x, y \in \mathbb{R}^n, \quad \forall p, q \in \mathbb{R}^n.$$

(b) (10 points) Prove the following regarding the generalized gradient $\partial f(x)$ of f (at every x):

$$\|\Psi\|_2 \le \gamma \quad \forall \Psi \in \partial f(x).$$

Recall that

(++) $\psi \in \partial f(x) \iff f^\circ(x;p) \geq \psi^T p \quad \forall p$

Furthermore,

(+) $f^\circ(x;p) \leq \gamma \|p\|_2 \quad \forall p$

Now for each ψ s.t. $\|\psi\|_2 > \gamma$ we have

$$\psi^T\psi > \gamma\|\psi\|_2 \underset{\text{due to (+)}}{\geq} f^\circ(x;\psi)$$

Thus by (++)

$\psi \notin \partial f(x)$.

Question 3 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex Lipschitz continuous (but not necessarily differentiable) function. The penalized bundle method to find the global minimizer of f keeps track of two sequences $\{x^{(k)}\}$ and $\{y^{(k)}\}$ in $\mathbb{R}^n$. It requires the solution of the nonsmooth optimization problem

$$\text{minimize}_{x\in\mathbb{R}^n} \;\; P_k(x) := \phi_k(x) + \frac{\mu}{2}\|x - x^{(k)}\|_2^2 \tag{1}$$

where $\mu > 0$ is a penalty parameter, and

$$\phi_k(x) := \max\left\{ f(y^{(j)}) + s_j^T(x - y^{(j)}) \;\mid\; j = 0, \ldots, k \right\}$$

for some $s_j \in \partial f(y^{(j)})$ for $j = 0, \ldots, k$, repeatedly. The point $y^{(k+1)}$ is defined as the global minimizer of $P_k(x)$. Furthermore, $x^{(k+1)} := y^{(k+1)}$ if $f(y^{(k+1)})$ satisfies a sufficient decrease condition compared to $f(x^{(k)})$, otherwise $x^{(k+1)} := x^{(k)}$.

(a) (10 points) Express (1) as a convex optimization problem with a quadratic objective function subject to linear constraints.

The problem (1) can be posed as follows:

$$\underset{\varepsilon\in\mathbb{R},\, x\in\mathbb{R}^n}{\text{minimize}} \quad \varepsilon + \frac{\mu}{2}(x - x^{(k)})^T(x - x^{(k)})$$

$$f(y^{(j)}) + s_j^T(x - y^{(j)}) \leq \varepsilon \quad j = 0, \ldots, k$$

equivalently

$$\underset{\varepsilon\in\mathbb{R},\, x\in\mathbb{R}^n}{\text{minimize}} \quad \varepsilon + \frac{\mu}{2}(x - x^{(k)})^T(x - x^{(k)})$$

$$\varepsilon - f(y^{(j)}) - s_j^T(x - y^{(j)}) \geq 0 \quad j = 0, \ldots, k$$

(b) (10 points) Write down the centrality conditions (that is the KKT conditions but with the complementarity condition replaced by a centering equation in the primal-dual space) for the convex optimization problem in part **(a)**.

(1) $$\begin{bmatrix} 1 \\ \mu (x - x^{(k)}) \end{bmatrix} = \sum_{j=0}^{k} \lambda_j \begin{bmatrix} 1 \\ -s_j \end{bmatrix}$$

(2) centrality equation $$\lambda_j \left(\varepsilon - f(y^{(j)}) - s_j^T (x - y^{(j)}) \right) = \sigma \cdot \tau \qquad j = 0, \dots, k$$

$$\lambda_j \geq 0 \qquad j = 0, \dots, k$$

$$\varepsilon - f(y^{(j)}) - s_j^T (x - y^{(j)}) \geq 0 \qquad j = 0, \dots, k$$

(Bonus) (10 points) (~~5 points~~) Write down one iteration of Newton's method for the solution of the centrality conditions in part **(b)**.

Apply Newton's method to solve (1) & (2) over $\varepsilon \in \mathbb{R}$, $\lambda \in \mathbb{R}^{k+1}$, $x \in \mathbb{R}^n$,

$$\begin{array}{rc} & \begin{array}{ccc} \lambda & x & \varepsilon \end{array} \\ 1 - \sum_{j=0}^{k} \lambda_j = 0 \leftarrow & \\ \mu(x - x^{(k)}) + \sum_{j=0}^{k} \lambda_j s_j = 0 \leftarrow & \begin{bmatrix} -e^T & 0 & 0 \\ S & \mu I & 0 \\ \varepsilon I - F_j - S_{xy} & -\Lambda S^T & \Lambda e \end{bmatrix} \\ \leftarrow & \end{array} \begin{bmatrix} \Delta \lambda^{(\ell)} \\ \Delta x^{(\ell)} \\ \Delta \varepsilon^{(\ell)} \end{bmatrix} = \begin{bmatrix} -1 + \sum_{j=0}^{k} [\lambda_j]^{(\ell)} \\ -\mu([x]^{(\ell)} - x^{(k)}) - \sum_{j=0}^{k} [\lambda_j]^{(\ell)} s_j \\ -\Lambda(\varepsilon I - F_j - S_{xy}) e + \sigma \end{bmatrix}$$

$$\lambda_j \left(\varepsilon - f(y^{(j)}) - s_j^T (x - y^{(j)}) \right) - \sigma \cdot \tau = 0 \qquad j = 0, \dots, k$$

$e \in \mathbb{R}^{k+1}$ - vector of ones, $S = [s_0 \dots s_k]$ $F_j = \mathrm{diag}(f(y^{(0)}), \dots, f(y^{(k)}))$

$S_{xy} = \mathrm{diag}(s_0^T([x]^{(\ell)} - y^{(j)}), \dots, s_k^T([x]^{(\ell)} - y^{(j)}))$ $\Lambda = \mathrm{diag}([\lambda_0]^{(\ell)}, \dots, [\lambda_k]^{(\ell)})$

Newton update

$$([\lambda]^{(\ell+1)}, x^{(\ell+1)}, \varepsilon^{(\ell+1)}) = ([\lambda]^{(\ell)}, [x]^{(\ell)}, [\varepsilon]^{(\ell)}) + \alpha (\Delta\lambda^{(\ell)}, \Delta x^{(\ell)}, \Delta\varepsilon^{(\ell)})$$

Question 4 Let A be a matrix-valued function defined by

$$A(x) := A_0 + x_1 A_1 + \cdots + x_d A_d$$

where $A_0, A_1, \ldots, A_d$ are given $n \times n$ symmetric positive semidefinite matrices, and let $\lambda_{\max} : \mathbb{R}^d \to \mathbb{R}$, $\lambda_{\max}(x) := \lambda_{\max}(A(x))$. Furthermore, assume $A_0 \neq 0$.

(a) (10 points) Express the unconstrained eigenvalue optimization problem

$$\text{minimize}_{x \in \mathbb{R}^d} \lambda_{\max}(x) \tag{2}$$

as a constrained optimization problem with a linear objective subject to a positive semidefiniteness constraint.

(2) can be rewritten as follows:

$$\underset{x \in \mathbb{R}^d,\ \varepsilon \in \mathbb{R}}{\text{minimize}} \quad \varepsilon$$

$$\underbrace{\lambda_{max}(x) \leq \varepsilon}$$

i.e., $\varepsilon - \lambda_{max}(x) \geq 0$

$\iff \varepsilon - \lambda_j(x) \geq 0$ for each eigenvalue $\lambda_j(x)$ of $A(x)$

$\iff \varepsilon I - A(x) \in S^n_+$

that is

(D) $$\underset{x \in \mathbb{R}^d,\ \varepsilon \in \mathbb{R}}{\text{minimize}} \quad \varepsilon$$

$$\varepsilon I - A_0 - x_1 A_1 - \cdots - x_d A_d \in S^n_+$$

(b) (10 points) Derive a semidefinite program that yields an upper bound for (2). State also conditions that guarantee that this semidefinite program is equivalent to (2).
(Hint: Try to view the constrained optimization problem in part (a) as the dual of a semidefinite program.)

Recall that the standard SDP is of the form

$$\begin{aligned} &\text{minimize } \langle C, X\rangle \\ &\langle A_j, X\rangle = b_j \quad j=1,\dots,m \\ &X \in S^n_+ \end{aligned}$$

its dual is given by

$$\begin{aligned} &\underset{\mu\in\mathbb{R}^m}{\text{maximize}} \ b^T\mu \\ &C - \sum_{j=1}^m \mu_j A_j \in S^n_+ \end{aligned}$$

Express (D) as a maximization problem, i.e., consider

$$\text{(D2)} \quad \begin{aligned} &\underset{x\in\mathbb{R}^d,\ \varepsilon\in\mathbb{R}}{\text{maximize}} \ -\varepsilon \\ &\varepsilon I - A_0 - x_1 A_1 - \dots - x_d A_d \in S^n_+ \end{aligned}$$

This is the dual of an SDP with $\mu = [\varepsilon \ x_1 \dots x_d]^T$, $b = [-1, 0 \dots 0]$, $C = -A_0$. Corresponding SDP is given by

$$\begin{aligned} &\text{minimize } \langle -A_0, X\rangle \\ &\langle I, X\rangle = -1 \\ &\langle -A_j, X\rangle = 0 \quad j=1,\dots,d \\ &X \in S^n_+ \end{aligned}$$

equivalently

$$\text{(SDP)} \quad \begin{aligned} &-\text{maximize } \langle A_0, X\rangle \\ &\langle I, X\rangle = -1 \\ &\langle A_j, X\rangle = 0 \quad j=1,\dots,d \\ &X \in S^n_+ \end{aligned}$$

Weak duality

(SDP) $\geqslant$ (D2)

always holds.

Strong duality (SDP) = (D2) holds if $\exists X \in S^n_{++}$ s.t.

$\langle I, X\rangle = -1$ and $\langle A_j, X\rangle = 0$ $j=1,\dots,d$

Question 5 (20 points) Let $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ with $m > n$ be given. Write down a necessary and sufficient condition for a point to be a global minimizer for the following problem using generalized gradients:

$$\text{minimize}_{x\in\mathbb{R}^n} \ \|b - Ax\|_\infty.$$

For simplicity, assume the following:

(i) $\text{rank}(A) = n$;

(ii) every set consisting of n rows of A is linearly independent;

(iii) letting $r(x) := b - Ax$, at every $x \in \mathbb{R}^n$ no more than n components of $r(x)$ are identical in absolute value.

Note that the function $f(x) := \|b - Ax\|_\infty$ is differentiable everywhere excluding a set Ω of (Lebesgue) measure zero; you can use this fact in your answer.

Recall (since $\|b-Ax\|_\infty$ is convex, i.e., maximum of linear functions)

(OPT) x_* is a global minimizer of $f(x) := \|b-Ax\|_\infty \iff 0 \in \partial f(x_*)$

We will make use of the following result.

(R) $\partial f(x) = \text{Co}\left\{ \lim_{k\to\infty} \nabla f(x^{(k)}) \mid \exists \{x^{(k)}\} \text{ such that } x^{(k)} \notin \Omega \ \forall k, \ \lim_{k\to\infty} x^{(k)} = x \text{ and } \{\nabla f(x^{(k)})\} \text{ is convergent} \right\}$

Let

$$A(x_*) := \left\{ i \mid |b_i - \bar{a}_i \cdot x| = \|b - Ax\|_\infty \right\}$$

where $\bar{a}_i$ is the ith row of A. Furthermore, for each $j \in A(x_*)$

$$s_j := \begin{cases} 1 & \bar{a}_j \cdot x_* > b_j \\ -1 & \bar{a}_j \cdot x_* < b_j \end{cases}$$

Note that $|A(x_*)| \leq n$ due to (iii). Also, without loss of generality, we can assume $r(x_*) = 0$ (since this implies $|A(x_*)| = m > n$).

Observe that for any sequence $\{x^{(k)}\}$ such that $\lim_{k\to\infty} x^{(k)} = x_*$, $x^{(k)} \notin \Omega$ $\forall k$ and $\{\nabla f(x^{(k)})\}$ is convergent, we must have

$$\lim_{k\to\infty} \nabla f(x^{(k)}) = s_j \bar{a}_j^T \quad \exists j \in A(x_*).$$

Consequently, from (R) (see page 9)

(INC) $\partial f(x_*) \subseteq Co\{ s_j \bar{a}_j^T \mid j \in A(x_*)\}$.

Furthermore, for each $j \in A(x_*)$, $\exists p \in \mathbb{R}^n$ s.t.

(i) $s_j \bar{a}_j \cdot p > 0$ and (ii) $\bar{a}_k \cdot p = 0$ $\forall k \in A(x_*)$, $k \neq j$

(due to $|A(x_*)| \leq n$ and $\{\bar{a}_k^T \mid k \in A(x_*)\}$ is linearly independent by assumption) Consider the sequence $\{x_* + \frac{1}{k} p\}$

Observe $\lim_{k\to\infty} x_* + \frac{1}{k} p = x_*$, $x_* + \frac{1}{k} p \notin \Omega$ $\forall k$ (to be precise for sufficiently large k, but truncate finitely many initial terms if they are in Ω) and $\nabla f(x_* + \frac{1}{k} p) = \bar{a}_j^T s_j$ $\forall k$ (sufficiently large). Thus $\lim_{k\to\infty} \nabla f(x_* + \frac{1}{k} p) = s_j \bar{a}_j^T \in \partial f(x_*)$ $\forall j \in A(x_*)$.

Combining this with (INC) yields

$$\partial f(x_*) = Co\{ s_j \bar{a}_j^T \mid j \in A(x_*)\}.$$

It follows from (OPT) (page 9) that

x_* is a global minimizer of $f(x)$

$$\Longleftrightarrow$$

$\exists \lambda_j$ for each $j \in A(x_*)$ such that

(i) $\lambda_j \geq 0$ $\forall j \in A(x_*)$,

(ii) $\sum_{j \in A(x_*)} \lambda_j = 1$

(iii) $\sum_{j \in A(x_*)} \lambda_j \bar{a}_j^T = 0$.