南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Course Name: Algorithms for Convex Optimization **Department:** Mathematics

Exam Duration: 120 minutes **Instructor:** Jin Zhang

Question No.	1	2	3	4	5
Score	25	15	15	15	30

This exam paper contains 5 exercises and the score is 100 in total. (Please hand in your answer sheet and your scrap paper to the proctor when the exam ends.)

Please note that all statements are based on the vectorial l_2−norm without special instructions.

1. Answer True or False for each of the following statements and justify your answer (5*5=25).

   (a) Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex, and $\varphi : \mathbb{R} \to \mathbb{R}$ be the absolute value function. Then the composite $h = \varphi \circ f$ is also convex.

   (b) Let $\{\Omega_k\}_{k=0}^{\infty} \subset \mathbb{R}^n$ be convex with $\Omega_k \subset \Omega_{k+1}$ for $k = 0, 1, 2, \cdots$. Then the union $\Omega = \bigcup_{k=0}^{\infty} \Omega_n$ is also convex.

   (c) Consider the linear programming (LP):

   $$\min_{x \in \mathbb{R}^n} \quad c^T x, \qquad \text{s.t. } Ax \le b,$$

   where $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. The dual of the LP problem is

   $$\max_{y \in \mathbb{R}^m} b^T y, \qquad \text{s.t. } A^T y + c = 0, \ y \ge 0.$$

   (d) Recall the conjugate of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined by

   $$f^*(y) = \sup_{x \in \mathbb{R}^n} \{\langle y, x\rangle - f(x)\}.$$

   Let $A \in \mathbb{R}^{n \times n}$ be an invertible matrix, $a \in \mathbb{R}^n$. Then the conjugate of the function $g(x) = f(A(x - a))$ is given by

   $$g^*(y) = f^*\left(\left(A^T\right)^{-1} y\right) + \langle a, y\rangle.$$

   (e) Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and $\bar{x} \in \mathbb{R}^n$. Then there exists $L > 0$ such that the directional derivative of f at $\bar{x}$ satisfies

   $$|f'(\bar{x}; d)| \le L \|d\| \text{ for any } d \in \mathbb{R}^n.$$

   Hint: the local Lipschitz continuity of convex functions.

2. **(optimality conditions)** (5+10=15). Consider the problem

$$(P) \qquad \min f(x), \text{ s.t. } g_1(x) \le 0, \cdots, g_m(x) \le 0,$$

where $f, g_1, \cdots, g_m : \mathbb{R}^n \to \mathbb{R}$ are real-valued convex functions. Define $g = \max\{g_1, \cdots, g_m\}$ and it's easy to see the feasible region $\mathcal{F} = \{x : g(x) \le 0\}$. Suppose that x^* is a local optimal solution of (P).

(a) Write down the Fritz-John condition and KKT condition.

(b) We say the Cottle constraint qualification (CQ) holds at some point $x \in \mathcal{F}$ if either $g(x) < 0$ or $0 \notin \partial g(x)$. Show that the problem (P) satisfies the Slater condition if and only if it satisfies the Cottle CQ at every $x \in \mathcal{F}$.

3. **(subdifferential and directional derivative)** (5*3=15). Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and $\varepsilon > 0$. We say the ε-subdifferential of f at x is the set

$$\partial_\varepsilon f(x) = \left\{ g \in \mathbb{R}^n : f(y) \ge f(x) + g^T(y - x) - \varepsilon \text{ for all } y \in \mathbb{R}^n \right\}.$$

We say the ε-directional derivative of f at x in the direction $d \in \mathbb{R}^n$ is defined by

$$f'_\varepsilon(x; d) = \inf_{\alpha > 0} \frac{f(x + \alpha d) - f(x) + \varepsilon}{\alpha}.$$

We claim both the nonempty set $\partial_\varepsilon f(x)$ and the real-valued function $d \mapsto f'_\varepsilon(x; d)$ are well-defined for any $x \in \mathbb{R}^n$.

(a) for any x, there exists $L > 0$ with

$$\partial f(z) \subseteq \partial_\varepsilon f(x), \text{ for all } z \in \mathbb{B}\left(x; \frac{\varepsilon}{2L}\right).$$

(b) for any $x, y \in \mathbb{R}^n$,

$$f(y) \ge f(x) + f'_\varepsilon(x; y - x) - \varepsilon.$$

(c) for any $x, d \in \mathbb{R}^n$,

$$f'_\varepsilon(x; d) = \max\left\{ g^T d : g \in \partial_\varepsilon f(x) \right\}.$$

Hint: you can apply the convexity and homogeneity of $d \mapsto f'_\varepsilon(x; d)$ directly.

4. **(generalized smoothness and generalized strong convexity correspondence)** (5+10=15). Let the univariate function $\rho : \mathbb{R}_+ \to \overline{\mathbb{R}}_+$ satisfies $\rho(0) = 0$. We say the function $f : \mathbb{R}^n \to \mathbb{R}$ is ρ-smooth if for any $x_1, x_2 \in \mathbb{R}^n$ and $0 \le \lambda \le 1$,

$$f(\lambda x_1 + (1 - \lambda) x_2) \ge \lambda f(x_1) + (1 - \lambda) f(x_2) - \lambda(1 - \lambda)\rho(\|x_1 - x_2\|).$$

We say the function $g : \mathbb{R}^n \to \mathbb{R}$ is ρ-convex if for any $x_1, x_2 \in \mathbb{R}^n$ and $0 \le \lambda \le 1$,

$$g(\lambda x_1 + (1 - \lambda) x_2) \le \lambda g(x_1) + (1 - \lambda) g(x_2) - \lambda(1 - \lambda)\rho(\|x_1 - x_2\|).$$

(a) Let $\rho(t) = \frac{\sigma}{2}t^2$ with $\sigma > 0$, if g is ρ-convex then the conjugate g^* is ρ^*-smooth.

(b) If f is ρ-convex then the conjugate f^* is ρ^*-smooth.

Hint: you may use the identity

$$\lambda\langle x_1, y_1\rangle + (1-\lambda)\langle x_2, y_2\rangle = \langle \lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2\rangle + \lambda(1-\lambda)\langle x_1 - x_2, y_1 - y_2\rangle.$$

5. **(proximal gradient method)** (5*6=30). Consider the problem

$$\min_{x\in\mathbb{R}^n} F(x) = f(x) + g(x), \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is L-smooth and $g : \mathbb{R}^n \to \mathbb{R}$ is convex. Let F_{opt} be the optimal value of (1). Let $\{x^k\}_{k\geq 0}$ be the sequence generated by the proximal gradient method for problem (1) with a constant stepsize $t_k \equiv \frac{1}{L}$.

(a) For any $x, y \in \mathbb{R}^n$, it holds that

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

(b) For any $x, y \in \mathbb{R}^n$, it holds that

$$\langle x - y, \text{prox}_g(x) - \text{prox}_g(y)\rangle \geq \|\text{prox}_g(x) - \text{prox}_g(y)\|^2.$$

(c) For any $x \in \mathbb{R}^n$, it holds that

$$F(x) - F(T_L(x)) \geq \frac{L}{2}\|x - T_L(x)\|^2.$$

where $T_L(x) = \text{prox}_{\frac{1}{L}g}\left(x - \frac{1}{L}\nabla f(x)\right)$.

(d) Write down the scheme of the proximal gradient method with a constant stepsize $t_k \equiv \frac{1}{L}$.

(e)

$$\min_{i=0,1,\cdots,k} \|x^i - x^{i+1}\| \leq \frac{\sqrt{F(x^0) - F_{\text{opt}}}}{\sqrt{\frac{L}{2}(k+1)}}.$$

(f) If there exists $0 < \mu < L$ such that for any $x \in \mathbb{R}^n$,

$$\mu\left(F(x) - F_{\text{opt}}\right) \leq \frac{L^2}{2}\|x - T_L(x)\|^2.$$

Then

$$F(x^k) - F_{\text{opt}} \leq \left(1 - \frac{\mu}{L}\right)^k \left(F(x^0) - F_{\text{opt}}\right).$$