Name: ____________ Student ID: ____________ Scores: ____________

AIE6003 Midterm
Instructor: Xiaopeng Li **Time Limit**: 2 hours

Notice: You can invoke results in the lecture, homework and hint without proving it.

1. (100 points) Let $\phi : \mathbb{R}^n \to \mathbb{R}$ be a μ_1-strongly convex L_1-smooth function and $f : \mathbb{R}^n \to \mathbb{R}$ be a μ_2-strongly convex L_2-smooth function. Consider the following algorithm:

$$\nabla\phi(x_{k+1}) = \nabla\phi(x_k) - \alpha\nabla f(x_k), \quad \forall\, k \geq 0,$$

with given initial point $x_0 \in \mathbb{R}^n$ and step size $\alpha > 0$. Let $x^* = \arg\min_{x\in\mathbb{R}^n} f(x)$.

   (a) (20 points) Prove that for all $k \geq 0$,

$$f(x_{k+1}) - f(x_k) \leq \left(-\frac{\mu_1}{\alpha} + \frac{L_2}{2}\right)\|x_{k+1} - x_k\|^2.$$

   Hint: the gradient of a strongly convex function ϕ is strongly monotone (Lecture 2).

   (b) (30 points) Prove that for any $x \in \mathbb{R}^n$,

$$\|\nabla f(x)\|^2 \geq 2\mu_2(f(x) - f(x^*)).$$

   Hint: use strong convexity of f and try to minimize a quadratic function.

   (c) (30 points) Set $\alpha < 2\mu_1/L_2$. Prove that any $k \geq 0$,

$$f(x_{k+1}) - f(x^*) \leq \left(1 - 2\alpha\frac{\mu_2\mu_1}{L_1^2} + \alpha^2\frac{\mu_2 L_2}{L_1^2}\right)(f(x_k) - f(x^*)).$$

   Hint: try to use part (a) and (b).

   (d) (20 points) Conclude that the above algorithm has linear convergence rate in terms of function value for all $\alpha \in (0, 2\mu_1/L_2)$. What is the optimal α?
   Hint: try to use part (c).

2. (100 points) Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and $X^* = \arg\min_{x\in C} f(x)$ be nonempty, where $C \subset \mathbb{R}^n$ is nonempty convex closed. Denote $f^* := f(x^*)$ for all $x^* \in X^*$. Consider the following algorithm:

$$x_{k+1} = \Pi_C\left(x_k - \frac{f(x_k) - f^*}{\|s_k\|^2}s_k\right), \text{ where } s_k \in \partial f(x_k),\ s_k \neq 0, \quad \forall\, k \in \mathbb{N},$$

and $x_0 \in \mathbb{R}^n$ is the initial point.

   (a) (30 points) Show that there is a constant $c > 0$ such that $\|x_k\| \leq c$ for all $k \geq 0$.
   Hint: try to show sequence $(\|x_k - x^*\|^2)_{k\in\mathbb{N}}$ is decreasing.

   (b) (30 points) Show that $f(x_k) \to f^*$ as $k \to \infty$.
   Hint: try to use the fact that ∂f maps any bounded sets to bounded sets.

   (c) (20 points) Show that the convergence rate is given by

$$\min_{i=0,\ldots,k} f(x_i) - f^* = O\left(\frac{1}{\sqrt{k}}\right).$$

(d) (20 points) If each minimum of f is sharp, i.e., for each $x^* \in X^*$, there exists $\mu > 0$ such that $f(x) - f^* \geq \mu \|x - x^*\|$ for all x, then

$$\|x_k - x^*\| = O\left(q^k\right),$$

where $q \in (0, 1)$ is a constant.

3. (100 points) Let $f(x) = \frac{1}{n}\sum_{i=1}^{n} f_i(x)$, where each $f_i : \mathbb{R}^d \to \mathbb{R}$ is convex and L-smooth. Assume f is convex and attains its minimum $f^* = \min_x f(x)$. Consider the following SGD update

$$x_{k+1} = x_k - \alpha_k \nabla f_{i_k}(x_k), \quad i_k \sim \mathrm{Unif}\{1, \ldots, n\} \text{ i.i.d.}$$

Assume that $\mathbb{E}_i\left[\|\nabla f_i(x) - \nabla f(x)\|_2^2\right] \leq \sigma^2$ for all $x \in \mathbb{R}^d$.

(a) (20 points) Show that $\mathbb{E}_i[\nabla f_i(x)] = \nabla f(x)$ and

$$\mathbb{E}_i[\|\nabla f_i(x)\|_2^2] \leq \|\nabla f(x)\|_2^2 + \sigma^2.$$

(b) (30 points) Let $x^* \in \arg\min f$. Show that for any step size $\alpha_k > 0$,

$$\mathbb{E}[\|x_{k+1} - x^*\|_2^2 \mid x_k] \leq \|x_k - x^*\|_2^2 - 2\alpha_k(f(x_k) - f^*) + \alpha_k^2(\|\nabla f(x_k)\|_2^2 + \sigma^2).$$

(c) (30 points) Assume a constant step size $\alpha_k \equiv \alpha \leq \frac{1}{2L}$ and define $\bar{x}_T := \frac{1}{T}\sum_{k=0}^{T-1} x_k$. Show that

$$\mathbb{E}[f(\bar{x}_T) - f^*] \leq \frac{\|x_0 - x^*\|_2^2}{2\alpha T} + \frac{\alpha\sigma^2}{2}.$$

Hint: try to first show that $\|\nabla f(x)\|_2^2 \leq 2L(f(x) - f^*)$ and

$$2\alpha(1 - \alpha L)\,\mathbb{E}[f(x_k) - f^*] \leq \mathbb{E}\|x_k - x^*\|_2^2 - \mathbb{E}\|x_{k+1} - x^*\|_2^2 + \alpha^2\sigma^2.$$

(d) (20 points) Show that $\mathbb{E}[f(\bar{x}_T) - f^*] \leq \varepsilon$ given that

$$\alpha := \min\left\{\frac{1}{2L}, \frac{\varepsilon}{\sigma^2}\right\}, \quad T := \frac{R^2}{\alpha\varepsilon}.$$

Page 2