Name: ______________ Student ID: ______________ Scores: ______________

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

Notice: You can invoke results in the lecture and homework 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).

Solution: By L_2-smoothness of f and strong convexity of ϕ (strong monotonicity of $\nabla\phi$), we have

$$\begin{aligned} f(x_{k+1}) - f(x_k) &\leq \langle \nabla f(x_k), x_{k+1} - x_k\rangle + \frac{L_2}{2}\|x_{k+1} - x_k\|^2 \\ &= \frac{1}{\alpha}\langle \nabla\phi(x_k) - \nabla\phi(x_{k+1}), x_{k+1} - x_k\rangle + \frac{L_2}{2}\|x_{k+1} - x_k\|^2 \\ &\leq \left(-\frac{\mu_1}{\alpha} + \frac{L_2}{2}\right)\|x_{k+1} - x_k\|^2. \end{aligned}$$

(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.

Solution: Since f is μ_2-strongly convex

$$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \frac{\mu_2}{2}\|y - x\|^2.$$

Minimize over y on both sides, the LHS is $f(x^*)$ while the RHS is minimized at y^* where

$$\nabla f(x) + \mu_2(y^* - x) = 0 \implies y^* = x - \frac{1}{\mu_2}\nabla f(x).$$

Thus, we obtain

$$f(x^*) \geq f(x) - \frac{1}{\mu_2}\|\nabla f(x)\|^2 + \frac{1}{2\mu_2}\|\nabla f(x)\|^2 = f(x) - \frac{1}{2\mu_2}\|\nabla f(x)\|^2.$$

(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).

Solution: Using part (a) and (b), we have

$$\begin{aligned}
f(x_{k+1}) - f(x_k) &\leq \left(-\frac{\mu_1}{\alpha} + \frac{L_2}{2}\right)\|x_{k+1} - x_k\|^2 \\
&\leq \left(-\frac{\mu_1}{\alpha} + \frac{L_2}{2}\right)\frac{1}{L_1^2}\|\phi(x_{k+1}) - \phi(x_k)\|^2 \\
&= \left(-\frac{\mu_1}{\alpha} + \frac{L_2}{2}\right)\frac{\alpha^2}{L_1^2}\|\nabla f(x_k)\|^2 \\
&\leq \left(-\frac{\mu_1}{\alpha} + \frac{L_2}{2}\right)\frac{\alpha^2}{L_1^2}2\mu_2(f(x_k) - f(x^*)).
\end{aligned}$$

Thus, the desired inequality follows.

(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).

Solution: The minimizer of the quadratic function g, i.e.,

$$g(\alpha) := 1 - 2\alpha\frac{\mu_2\mu_1}{L_1^2} + \alpha^2\frac{\mu_2 L_2}{L_1^2},$$

is given by $\alpha^* = \frac{\mu_1}{L_2}$. Also $g(0) = g\left(\frac{2\mu_1}{L_2}\right) = 1$. By property of quadratic function,

$$0 \leq 1 - \frac{\mu_1^2\mu_2}{L_1^2 L_2} = g(\alpha^*) \leq g(\alpha) < 1$$

for all $\alpha \in (0, \frac{2\mu_1}{L_2})$, which gives linear convergence of $f(x_k) - f(x^*)$.

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.

Solution: Recall that $\Pi_C(\bullet)$ is 1-Lipschitz. For any $x^* \in X^*$ and all $k \geq 1$, by convexity,

$$\begin{aligned}
\|x_{k+1} - x^*\|^2 &= \left\| \Pi_C \left(x_k - \frac{f(x_k) - f^*}{\|s_k\|^2} s_k \right) - \Pi_C(x^*) \right\|^2 \\
&\leq \left\| x_k - \frac{f(x_k) - f^*}{\|s_k\|^2} s_k - x^* \right\|^2 \\
&= \|x_k - x^*\|^2 + \frac{(f(x_k) - f^*)^2}{\|s_k\|^2} - 2\frac{f(x_k) - f^*}{\|s_k\|^2} \langle s_k, x_k - x^* \rangle \\
&\leq \|x_k - x^*\|^2 + \frac{(f(x_k) - f^*)^2}{\|s_k\|^2} - 2\frac{f(x_k) - f^*}{\|s_k\|^2} (f(x_k) - f(x^*)) \\
&= \|x_k - x^*\|^2 - \frac{(f(x_k) - f^*)^2}{\|s_k\|^2}.
\end{aligned}$$

This implies that $(\|x_k - x^*\|)_{k\in\mathbb{N}}$ is a decreasing sequence and $\|x_k - x^*\| \leq \|x_0 - x^*\|$ for all $k \geq 0$, which means $\|x_k\| \leq \|x^*\| + \|x_0 - x^*\|$. Let $c = \|x^*\| + \|x_0 - x^*\|$ and the result follows.

(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.

Solution: From (a), there exists $L > 0$ such that $\|s_k\| \leq L$ for all $s_k \in \partial f(x_k)$ and $k \geq 0$. Also,

$$\frac{1}{L^2} \sum_{i=0}^{k} (f(x_i) - f^*)^2 \leq \sum_{i=0}^{k} \frac{(f(x_i) - f^*)^2}{\|s_i\|^2} \leq \sum_{i=0}^{k} (\|x_i - x^*\|^2 - \|x_{i+1} - x^*\|^2) = \|x_0 - x^*\|^2.$$

Let $k \to \infty$, and $f(x_k) - f^* \to 0$ as $k \to \infty$.

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

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

Solution: From part (b), we know that

$$(k+1) \min_{i=0,\ldots,k} (f(x_i) - f^*)^2 \leq \sum_{i=0}^{k} (f(x_i) - f^*)^2 \leq L^2 \|x_0 - x^*\|^2.$$

Thus, we obtain

$$\min_{i=0,\ldots,k} f(x_i) - f^* \leq \frac{L\|x_0 - x^*\|}{\sqrt{k+1}}.$$

(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.

Solution: From part (a) and (b), we have

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \frac{(f(x_k) - f^*)^2}{\|s_k\|^2} \leq \|x_k - x^*\|^2 - \frac{\mu^2}{L^2}\|x_k - x^*\|^2.$$

Thus, by letting $q = \sqrt{1 - \mu^2/L^2}$, we have

$$\|x_{k+1} - x^*\| \leq q\|x_k - x^*\| \implies \|x_k - x^*\| \leq q^k\|x_0 - x^*\|.$$

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

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

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

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

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

Solution: Since $i \sim \text{Unif}\{1, \dots, n\}$,

$$\mathbb{E}_i[\nabla f_i(x)] = \frac{1}{n}\sum_{i=1}^n \nabla f_i(x) = \nabla f(x).$$

For the second moment, use the variance decomposition with $g_i(x) := \nabla f_i(x)$:

$$\begin{aligned}\mathbb{E}_i\|g_i(x)\|_2^2 &= \Big\|\mathbb{E}_i[g_i(x)]\Big\|_2^2 + \mathbb{E}_i\Big\|g_i(x) - \mathbb{E}_i[g_i(x)]\Big\|_2^2 \\ &= \|\nabla f(x)\|_2^2 + \mathbb{E}_i\|\nabla f_i(x) - \nabla f(x)\|_2^2 \leq \|\nabla f(x)\|_2^2 + \sigma^2.\end{aligned}$$

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

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

Solution: Expand the square:

$$\begin{aligned}\|x_{k+1} - x^*\|_2^2 &= \|x_k - \alpha_k\nabla f_{i_k}(x_k) - x^*\|_2^2 \\ &= \|x_k - x^*\|_2^2 - 2\alpha_k\langle x_k - x^*, \nabla f_{i_k}(x_k)\rangle + \alpha_k^2\|\nabla f_{i_k}(x_k)\|_2^2.\end{aligned}$$

Taking conditional expectation given x_k and using part (a),

$$\mathbb{E}[\nabla f_{i_k}(x_k) \mid x_k] = \nabla f(x_k), \qquad \mathbb{E}[\|\nabla f_{i_k}(x_k)\|_2^2 \mid x_k] \leq \|\nabla f(x_k)\|_2^2 + \sigma^2.$$

By convexity of f and optimality of x^*,

$$f(x_k) - f^* \leq \langle \nabla f(x_k), x_k - x^* \rangle.$$

Combining the above inequalities gives

$$\begin{aligned}\mathbb{E}[\|x_{k+1} - x^*\|_2^2 \mid x_k] &\le \|x_k - x^*\|_2^2 - 2\alpha_k\langle x_k - x^*, \nabla f(x_k)\rangle + \alpha_k^2(\|\nabla f(x_k)\|_2^2 + \sigma^2) \\ &\le \|x_k - x^*\|_2^2 - 2\alpha_k(f(x_k) - f^*) + \alpha_k^2(\|\nabla f(x_k)\|_2^2 + \sigma^2).\end{aligned}$$

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

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

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

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

Solution: From part (b) with $\alpha_k \equiv \alpha$ and taking total expectation,

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

By L-smoothness of f, for any x we have the standard inequality

$$f\Big(x - \frac{1}{L}\nabla f(x)\Big) \le f(x) - \frac{1}{2L}\|\nabla f(x)\|_2^2.$$

Since $f^* \le f\big(x - \frac{1}{L}\nabla f(x)\big)$, it follows that

$$\|\nabla f(x)\|_2^2 \le 2L\big(f(x) - f^*\big).$$

Apply this to $x = x_k$ and substitute into $(\star)$:

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

Rearranging gives

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

Using $\alpha \le \frac{1}{2L}$ yields $1 - \alpha L \ge \frac{1}{2}$, hence

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

Sum over $k = 0, 1, \ldots, T-1$ (telescoping the distance terms):

$$\alpha\sum_{k=0}^{T-1}\mathbb{E}[f(x_k) - f^*] \le \|x_0 - x^*\|_2^2 + \alpha^2\sigma^2 T.$$

Divide by αT:

$$\frac{1}{T}\sum_{k=0}^{T-1}\mathbb{E}[f(x_k) - f^*] \le \frac{\|x_0 - x^*\|_2^2}{\alpha T} + \alpha\sigma^2.$$

Finally, by convexity of f and Jensen, $f(\bar{x}_T) \le \frac{1}{T}\sum_{k=0}^{T-1} f(x_k)$, so

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

Replacing α by $\alpha/2$ in the above bound (equivalently tightening the constants by keeping $1 - \alpha L \ge 1/2$ throughout) yields the stated form:

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

Page 5

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

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

where $R := \|x_0 - x^*\|_2$.

Solution: Let $R := \|x_0 - x^*\|_2$. By part (c),

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

With $T = \frac{R^2}{\alpha\varepsilon}$, the first term equals $\frac{\varepsilon}{2}$:

$$\frac{R^2}{2\alpha T} = \frac{R^2}{2\alpha \cdot \frac{R^2}{\alpha\varepsilon}} = \frac{\varepsilon}{2}.$$

Also, since $\alpha \leq \varepsilon/\sigma^2$, the second term satisfies

$$\frac{\alpha\sigma^2}{2} \leq \frac{\varepsilon}{2}.$$

Therefore $\mathbb{E}[f(\bar{x}_T) - f^*] \leq \varepsilon/2 + \varepsilon/2 \leq \varepsilon$.

Page 6