Optimization, MATH 164
E. K. Ryu
Winter 2025

Midterm Exam
Friday, February 14, 2025, 10:00–10:50 am
50 minutes, 4 questions, 100 points, 6 pages

This exam is open-book in the sense that you may use any non-electronic resource.
While we don't expect you will need more space than provided,
you may continue on the back of the pages.

Name: ________________________________

# Do not turn to the next page until the start of the exam.

1. (25 points) *Strict convexity inequality.* Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be convex.

   (a) Let $x, y \in \mathbb{R}^n$. We claim
   $$h(\theta) = \frac{f(x+\theta(y-x)) - f(x)}{\theta}$$
   is a non-decreasing function of $\theta \in (0,1)$. Provide a geometric illustration justifying this claim.

   (b) Prove the claim of part (a).

   (c) Further, assume that $f$ is strictly convex and that $f$ is differentiable at $x \in \mathbb{R}^n$. Show that
   $$f(y) > f(x) + \langle \nabla f(x), y - x \rangle, \qquad \forall\, x, y \in \mathbb{R}^n,\, x \neq y.$$

   *Clarification.* Recall that $f\colon \mathbb{R} \to \mathbb{R}$ is strictly convex if
   $$f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta) f(y), \qquad \forall\, x, y \in \mathbb{R}^n,\, x \neq y,\, \theta \in (0,1).$$

   *Hint.* In (c), it is not enough to simply take the limit $\theta \to 0$ on
   $$f(y) - f(x) - \frac{f(x+\theta(y-x)) - f(x)}{\theta} > 0$$
   because the limit of strictly positive numbers is not necessarily strictly positive.

2. (25 points) *Method of Lagrange multipliers for inequality constraints.* Consider the inequality-constrained optimization problem

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & f(x) \\ \text{subject to} & g(x) \leq 0, \end{array}$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$ and $g\colon \mathbb{R}^n \to \mathbb{R}$ are differentiable.

(a) Show that if $x$ is a solution and $g(x) < 0$, then $\nabla f(x) = 0$.

(b) Show that if $x$ is a solution and $g(x) = 0$, then $x$ is in fact a solution to the optimization problem

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & f(x) \\ \text{subject to} & g(x) = 0. \end{array}$$

(c) Show that if $x$ is a solution and $g(x) = 0$, then there is a $\lambda \geq 0$ such that

$$\nabla f(x) + \lambda \nabla g(x) = 0.$$

(d) Show that if $x$ is a solution, then there exists a $\lambda$ such that $x$ and $\lambda$ satisfy

$$\nabla f(x) + \lambda \nabla g(x) = 0, \qquad g(x) \leq 0, \qquad \lambda \geq 0, \qquad \lambda g(x) = 0.$$

3. (25 points) *Majorization-minimization interpretation of projected GD.* Let $L > 0$ and $\alpha \in (0, 1/L]$. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be $L$-smooth and convex. Let $C \subset \mathbb{R}^n$ be nonempty closed convex. Recall that projected gradient descent has the form

$$x_{k+1} = \Pi_C\big(x_k - \alpha \nabla f(x_k)\big)$$

for $k = 0, 1, \dots$ For any $x_k \in \mathbb{R}^n$, define $\tilde{f}$ as

$$\tilde{f}(x; x_k, \alpha) = f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2\alpha}\|x - x_k\|^2.$$

(a) Show that

$$f(x) \le \tilde{f}(x; x_k, \alpha), \qquad \forall\, x, x_k \in \mathbb{R}^n.$$

(b) Show that the iterates of projected gradient descent $\{x_k\}_k$ satisfy

$$x_{k+1} = \operatorname*{argmin}_{x \in C} \tilde{f}(x; x_k, \alpha), \qquad \text{for } k = 0, 1, \dots.$$

(c) Use the characterization of (b) to conclude that

$$f(x_{k+1}) \le f(x_k), \qquad \text{for } k = 0, 1, \dots.$$

*Remark.* Majorization-minimization refers to the process of constructing an upper bound (majorize) and then minimizing the upper bound. The majorization-minimization view of projected GD makes it clear that it is a descent method.

4. (25 points) *Convergence rate of gradient norm for GD.* Let $L > 0$. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be $L$-smooth and convex. Assume $f$ has a minimizer $x_\star$ and write $f_\star = f(x_\star)$. Consider the gradient descent algorithm

$$x_{k+1} = x_k - \frac{1}{L}\nabla f(x_k), \qquad \text{for } k = 0, 1, \ldots.$$

(a) Show that

$$f(x_{k+1}) + \frac{1}{2L}\|\nabla f(x_{k+1})\|^2 + \frac{1}{2L}\|\nabla f(x_k)\|^2 \le f(x_k), \qquad \text{for } k = 0, 1, \ldots.$$

(b) Show that

$$-\frac{1}{L}\|\nabla f(x_k)\|^2 \le f(x_{k+1}) - f(x_k), \qquad \text{for } k = 0, 1, \ldots.$$

(c) Show that

$$\|\nabla f(x_{k+1})\|^2 \le \|\nabla f(x_k)\|^2, \qquad \text{for } k = 0, 1, \ldots.$$

(d) Show that

$$\mathcal{E}_k = (2k+1)L\big(f(x_k) - f_\star\big) + k(k+2)\|\nabla f(x_k)\|^2 + L^2\|x_k - x_\star\|^2$$

for $k = 0, 1, \ldots$ is a dissipative sequence.

(e) Show that

$$\|\nabla f(x_k)\|^2 \le \frac{1}{k(k+2)}\Big(L\big(f(x_0) - f_\star\big) + L^2\|x_0 - x_\star\|^2\Big), \qquad \text{for } k = 1, 2, \ldots.$$

*Clarification.* In class, we have shown

$$f(x_{k+1}) + \frac{1}{2L}\|\nabla f(x_k)\|^2 \le f(x_k)$$

and

$$f(x_{k+1}) - f(x_k) \le -\frac{1}{2L}\|\nabla f(x_k)\|^2,$$

which are related but different from the inequalities of Parts (a) and (b).