Assignment 2

00103335: Deep Learning and Reinforcement Learning

Fall 2026

Give complete derivations and state any additional assumptions you use. Unless otherwise stated, all norms are Euclidean.

Problem 1. Let $f : \mathbb{R}^d \to \mathbb{R}$ be differentiable. Prove that the following conditions are equivalent.

(i) For every $x, y \in \mathbb{R}^d$ and every $\lambda \in (0,1)$,

$$f\big(\lambda x + (1-\lambda)y\big) \leq \lambda f(x) + (1-\lambda) f(y).$$

(ii) For every $x, y \in \mathbb{R}^d$,

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x).$$

(iii) For every $x, y \in \mathbb{R}^d$,

$$\big(\nabla f(y) - \nabla f(x)\big)^\top (y - x) \geq 0.$$

Problem 2. Let $f : \mathbb{R}^d \to \mathbb{R}$ be differentiable and let $\mu > 0$. In this problem, condition (i) is the definition of μ-strong convexity. Prove that the following conditions are equivalent.

(i) The function

$$x \longmapsto f(x) - \frac{\mu}{2} \|x\|^2$$

is convex; that is, f is μ-strongly convex.

(ii) For every $x, y \in \mathbb{R}^d$ and every $\alpha \in (0,1)$,

$$f\big(\alpha x + (1-\alpha)y\big) \leq \alpha f(x) + (1-\alpha) f(y) - \frac{\mu}{2}\alpha(1-\alpha) \|y - x\|^2 .$$

(iii) For every $x, y \in \mathbb{R}^d$,

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) + \frac{\mu}{2} \|y - x\|^2 .$$

(iv) For every $x, y \in \mathbb{R}^d$,

$$\big(\nabla f(y) - \nabla f(x)\big)^\top (y - x) \geq \mu \|y - x\|^2 .$$

Problem 3. Let $f : \mathbb{R}^d \to \mathbb{R}$ be differentiable and μ-strongly convex. Prove the following statement.

(ii) For every $x, y \in \mathbb{R}^d$,

$$\|\nabla f(y) - \nabla f(x)\| \geq \mu \|y - x\| .$$

(iii) For every $x, y \in \mathbb{R}^d$,

$$f(y) \leq f(x) + \nabla f(x)^\top (y - x) + \frac{1}{2\mu} \|\nabla f(y) - \nabla f(x)\|^2 .$$

(iv) For every $x, y \in \mathbb{R}^d$,

$$\big(\nabla f(y) - \nabla f(x)\big)^\top (y - x) \leq \frac{1}{\mu} \|\nabla f(y) - \nabla f(x)\|^2 .$$

Problem 4. Let $f : \mathbb{R}^d \to \mathbb{R}$ be differentiable and μ-strongly convex. Let $x^\star$ minimize f, let $x_1 \in \mathbb{R}^d$, and run gradient descent

$$x_{t+1} = x_t - \eta_t \nabla f(x_t), \qquad \eta_t = \frac{2}{\mu(t+1)}, \qquad t = 1, 2, \ldots, T.$$

Assume that the gradients encountered by the algorithm are uniformly bounded:

$$\|\nabla f(x_t)\| \le L, \qquad t = 1, 2, \ldots, T.$$

No Lipschitz continuity of ∇f is assumed.

(a) Let

$$d_t := \|x_t - x^\star\|, \qquad \Delta_t := f(x_t) - f(x^\star).$$

Prove that

$$\nabla f(x_t)^\top (x_t - x^\star) \ge \Delta_t + \frac{\mu}{2} d_t^2.$$

(b) Prove the one-step inequality

$$d_{t+1}^2 \le (1 - \mu\eta_t) d_t^2 - 2\eta_t \Delta_t + \eta_t^2 L^2.$$

(c) Prove that

$$\sum_{t=1}^{T} t\Delta_t \le \frac{TL^2}{\mu}.$$

(d) Let the linearly weighted average be

$$\overline{x}_T := \frac{2}{T(T+1)} \sum_{t=1}^{T} t x_t.$$

Prove that

$$f(\overline{x}_T) - f(x^\star) \le \frac{2L^2}{\mu(T+1)} = O\left(\frac{1}{T}\right).$$