Optimization, MATH 164
E. K. Ryu
Winter 2025

Homework 3
Due on Wednesday, February 12, 2025.

Problem 1: *Analysis of non-convex projected gradient.* Let $L > 0$ and $\alpha \in (0, 2/L)$. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be L-smooth and $C \subseteq \mathbb{R}^n$ be nonempty closed convex. Consider the projected GD method

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

for $k = 0, 1, \dots$ with $x_0 \in C$.

(a) Show that
$$\|G_\alpha(x_k)\|^2 \le \langle G_\alpha(x_k), \nabla f(x_k)\rangle$$
for $k = 0, 1, \dots$.

(b) Show that
$$f(x_{k+1}) \le f(x_k) - \alpha\big(1 - \tfrac{L\alpha}{2}\big)\|G_\alpha(x_k)\|^2.$$

(c) Assuming $\inf_{x \in C} f(x) > -\infty$, show that $G_\alpha(x_k) \to 0$.

Hint. For (a), use the projection theorem. For (b), use the L-smoothness lemma.

Problem 2: *A gradient provides a cutting plane for solutions.* Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be convex.

(a) Show that if $\nabla f(x) \neq 0$, then
$$\operatorname{argmin} f \subseteq \{y \in \mathbb{R}^n \,|\, \langle \nabla f(x), y - x\rangle < 0\}.$$

(b) The above inclusion implies that a non-zero gradient at $x \in \mathbb{R}^n$ defines a half-space (which boundary goes through x) in which minimizers don't lie. Draw a 2D depiction of this.

Problem 3: *Simple projections.* Provide formulae for the projections onto the following sets.

(a) $\{x \in \mathbb{R}^n \,|\, x_i \ge 0,\ i = 1, \dots, n\}$.

(b) $\{x \in \mathbb{R}^n \,|\, \|x\| \le D\}$.

(c) $\{x \in \mathbb{R}^n \,|\, l_i \le x_i \le u_i,\ i = 1, \dots, n\}$, where $l, u \in \mathbb{R}^n$.

Problem 4: *Projection onto a subspace.* In this problem, we will compute the projection onto $C = \{x \in \mathbb{R}^n \mid Ax = b\}$, where $A \in \mathbb{R}^{m \times n}$ has $\operatorname{rank}(A) = m < n$ and $b \in \mathbb{R}^m$. Use the notation

$$A = \begin{bmatrix} — a_1^\intercal — \\ — a_2^\intercal — \\ \vdots \\ — a_m^\intercal — \end{bmatrix}, \qquad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

The optimization problem of interest is

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & \tfrac{1}{2}\|x - x_0\|^2, \\ \text{subject to} \quad & g_1(x) = 0 \\ & \vdots \\ & g_m(x) = 0, \end{aligned}$$

where $g_i(x) = a_i^\intercal x - b_i$ for $i = 1, \ldots, m$.

(a) Using the method of Lagrange multipliers, show that an optimal x satisfies

$$x - x_0 = A^T \lambda, \qquad Ax = b$$

for some $\lambda \in \mathbb{R}^m$.

(b) Show that the corresponding λ is given by

$$\lambda = (AA^\intercal)^{-1}(b - Ax_0).$$

(c) Show that

$$\Pi_C(x_0) = x_0 + A^\intercal (AA^\intercal)^{-1}(b - Ax_0).$$

Problem 5: *Proximal interpretation of projected GD.* Let $f \colon \mathbb{R}^n \to \mathbb{R}$ is differentiable, $C \subset \mathbb{R}^n$ be non-empty closed convex, and $\alpha > 0$. Recall that the projected gradient descent has the update

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

Show that x_{k+1} can be equivalently defined as

$$x_{k+1} = \underset{y \in C}{\operatorname{argmin}} \left\{ f(x_k) + \langle \nabla f(x_k), y - x_k \rangle + \frac{1}{2\alpha}\|y - x_k\|^2 \right\}$$

Remark. The interpretation is that the k-th projected GD step is minimizing over the constraint set C the first-order Taylor expansion of f about x_k with an added "proximal term" that penalizes moving away from x_k too much. So it is minimizing a simpler approximation of f while staying in proximity to x_k, which is the region where the approximation is accurate.