Mathematical Machine Learning Theory, M1407.002700
E. Ryu
Spring 2024

Homework 5
Due 5pm, Thursday, May 09, 2024

Problem 1: *Quadratic objectives in standard form II.* Let

$$F(\theta) = \frac{1}{2}\theta^\intercal H\theta + b^\intercal\theta + c,$$

where $H = H^\intercal \in \mathbb{R}^{d\times d}$ is positive <u>semi</u>definite, $b \in \mathbb{R}^d$, and $c \in \mathbb{R}$.

(a) Assume $b \in \mathcal{R}(H)$. Show that there exists some $\theta^\star \in \mathbb{R}^d$ and $c' \in \mathbb{R}$ such that

$$F(\theta) = \frac{1}{2}(\theta - \theta^\star)^\intercal H(\theta - \theta^\star) + c'.$$

(b) Assume $b \notin \mathcal{R}(H)$. Show that $\inf_{\theta\in\mathbb{R}^d} F(\theta) = -\infty$.

Problem 2: *Explicit parameterization of affine sets I.* Let $B \in \mathbb{R}^{m\times d}$ and $b \in \mathbb{R}^m$. Show that

$$A = \{x \in \mathbb{R}^d \mid Bx = b\}$$

is an affine set.

Problem 3: *Explicit parameterization of affine sets II.* Let $A \subset \mathbb{R}^d$ be a nonempty affine set such that $A \neq \mathbb{R}^d$. Show that there is a $B \in \mathbb{R}^{m\times d}$ with full row rank, $b \in \mathbb{R}^m$, and $x_0 \in \mathcal{R}(B^\intercal)$ such that $Bx_0 = b$ and

$$A = \{x \in \mathbb{R}^d \mid Bx = b\} = x_0 + \mathcal{N}(B).$$

Problem 4: *Affine hull and span.* Let $C \subseteq \mathbb{R}^d$ be nonempty and $x_0 \in C$. Recall

$$\begin{aligned} \operatorname{aff} C &= \{\theta_1 x_1 + \cdots + \theta_k x_k \mid x_1, \dots, x_k \in C,\ \theta_1 + \cdots + \theta_k = 1,\ k \geq 1\} \\ \operatorname{span} C &= \{\alpha_1 x_1 + \cdots + \alpha_k x_k \mid x_1, \dots, x_k \in C,\ \alpha_1, \dots, \alpha_k \in \mathbb{R},\ k \geq 1\}. \end{aligned}$$

(a) Show that $\operatorname{aff} C = x_0 + \operatorname{aff}(C - x_0)$.

(b) Show that $\operatorname{aff}(C - x_0) = \operatorname{span}(C - x_0)$.

Conclude that

$$\operatorname{aff} C = x_0 + \operatorname{aff}(C - x_0) = x_0 + \operatorname{span}(C - x_0).$$

Hint. Note that

$$\operatorname{aff}(C - x_0) = \{\theta_1(x_1 - x_0) + \cdots + \theta_k(x_k - x_0) \mid x_1, \dots, x_k \in C,\ \theta_1 + \cdots + \theta_k = 1,\ k \geq 1\}.$$

Problem 5: *Affine hull is the smallest affine set containing C.* Let $C \subseteq \mathbb{R}^d$ be nonempty.

(a) Show that $\operatorname{aff} C$ is an affine set.

(b) Show that if A is an affine set such that $C \subseteq A$, then $\operatorname{aff} C \subseteq A$.

Hint. Note that (a) is immediate from the previous problem. For (b), let $A = v_0 + V$, where $v_0 \in \mathbb{R}^d$ and V is a subspace. Assume for contradiction that there is an

$$x = \theta_1 x_1 + \cdots + \theta_k x_k \in \operatorname{aff} C$$

with $x_1, \ldots, x_k \in C \subseteq A$, $\theta_1 + \cdots + \theta_k = 1$, and $k \geq 1$, but $x \notin A$.

Problem 6: *A closed convex set is the intersection of all supporting half-planes containing it.* Let $C \subseteq \mathbb{R}^d$ be a nonempty closed convex set. We say $(p, v) \in \partial C \times \mathbb{R}^d$ defines a supporting hyperplane if $v \neq 0$ and $v^\intercal x \leq v^\intercal p$ for all $x \in C$. Show that

$$C = \bigcap_{\text{supporting hyperplane } (v,p)} \{x \in \mathbb{R}^d \mid v^\intercal x \leq v^\intercal p\}.$$

Problem 7: *Indicator function is CCP.* Let $C \subseteq \mathbb{R}^d$ be a nonempty closed convex set. Let $\delta_C \colon \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ be defined as

$$\delta_C(x) = \begin{cases} 0 & \text{if } x \in C \\ \infty & \text{if } x \notin C. \end{cases}$$

Show that δ_C is CCP.

Problem 8: *Non-closed function.* Let

$$C = \{x \in \mathbb{R}^2 \mid \|x\|_2 \leq 1\}, \qquad \partial C = \{x \in \mathbb{R}^2 \mid \|x\|_2 = 1\}.$$

Let

$$f(x) = \begin{cases} 0 & \text{if } x \in \operatorname{int} C \\ \infty & \text{if } x \notin C \\ \text{any value in } [0, \infty) & \text{if } x \in \partial C. \end{cases}$$

(a) Show that f is convex and proper.

(b) What choice of $f|_{\partial C}$ makes f closed?

Problem 9: *Gradient descent is a descent algorithm.* Let $F\colon \mathbb{R}^d \to \mathbb{R}$ be L-smooth convex. Show that if $\alpha \in [0, 2/L]$, then

$$F(\theta - \alpha \nabla F(\theta)) \le F(\theta).$$

Problem 10: *Subgradient descent is <u>not</u> a descent algorithm.* Consider the convex function $F\colon \mathbb{R}^2 \to \mathbb{R}$ defined as

$$F(\theta_1, \theta_2) = |\theta_1| + 2|\theta_2|.$$

Let $\theta = (1, 0)$.

(a) Show that $g = (1, 2) \in \partial F(\theta)$.

(b) Show that $F(\theta - \alpha g) > F(\theta)$ for all $\alpha \neq 0$.

Problem 11: *Subgradient descent with "any-time" guarantee.* Let $F\colon \mathbb{R}^d \to \mathbb{R}$ be a G-Lipschitz continuous convex function. Assume F has a minimizer $\theta^\star$. Let $\theta^0 \in \mathbb{R}^d$ be a starting point, and let $R > 0$ satisfy $\|\theta^0 - \theta^\star\|_2 \le R$. Consider subgradient descent with the non-constant stepsize

$$\alpha_k = \frac{R}{G\sqrt{k+1}}$$

for $k = 0, 1, \ldots$. Consider subgradient method

$$\begin{aligned} g^k &\in \partial F(\theta^k) \\ \theta^{k+1} &= \theta^k - \alpha_k g^k, \end{aligned}$$

for $k = 0, 1, \ldots$. Show that

$$\min_{0 \le s \le k} F(\theta^s) - f(\theta^\star) \le \frac{GR(2 + \log(k+1))}{2\sqrt{k+1}}, \qquad \text{for } k = 0, 1, \ldots$$

and

$$F(\bar{\theta}^k) - F(\theta^\star) \le \frac{GR(2 + \log(k+1))}{2\sqrt{k+1}}, \qquad \text{for } k = 0, 1, \ldots,$$

where

$$\bar{\theta}^k = \frac{1}{\sum_{s=0}^k \alpha_s} \sum_{s=0}^k \alpha_s \theta^s, \qquad \text{for } k = 0, 1, \ldots.$$