

# Problem Set 1 for EE227C (Spring 2018): Convex Optimization and Approximation

Instructor: Moritz Hardt

Email: `hardt+ee227c@berkeley.edu`

Graduate Instructor: Max Simchowitz

Email: `msimchow+ee227c@berkeley.edu`

January 18, 2018

## Problem 1: Subgradients and Gradients

**A)** Let  $\mathcal{X}$  be a closed set. Prove that that given any convex function  $f : \mathcal{X} \rightarrow \mathbb{R}$  and any  $x \in \mathcal{X}$ , there exists a nonempty set  $\partial f(x)$  such that, for all  $g \in \partial f(x)$ ,  $f(y) \geq f(x) + \langle g, y - x \rangle$ . Do so following the steps:

**i)** Define the *Epigraph* of  $f$ ,  $\text{Epi}(f) := \{(x, t) \in \mathcal{X} \times \mathbb{R} : f(x) \leq t\}$ . Prove the  $\text{Epi}(f)$  is convex. Let  $\text{Int}(\mathcal{C})$  and  $\text{Bd}(\mathcal{C})$  denote the interior and boundary of a set  $\mathcal{C}$ , respectively. Suppose that  $(x, y) \in \text{Bd}(\text{Epi}(f))$ . What can you say about the values of  $f(x)$  and  $y$ ?

**ii)** Using the separating hyperplane theorem from the notes (the full version, which applies to arbitrary convex sets not just compact ones), prove the supporting hyperplane theorem. (see Background)

**Theorem 1** (Supporting Hyperplane). Let  $\mathcal{C}$  be a convex set, and let  $x \in \text{Bd}(\mathcal{C})$ . Then, there exists a hyperplane  $\mathcal{H}$  passing through  $x$  such that  $\mathcal{C} \cap \mathcal{H} \subset \text{Bd}(\mathcal{C})$ .

*Hint: Consider  $\text{Int}(\mathcal{C}) := \{x \in \mathcal{C} | x \notin \text{Bd}(\mathcal{C})\}$ . This set is convex (you don't need to prove this). Find an appropriate other convex set to which you can apply the separating hyperplane theorem.*

**iii)** Using part *i)* and *ii)*, prove the existence of a subgradient.

iv) Let  $\{g_\alpha\}$  be a (possibly infinite, uncountable) family of convex functions, and suppose that  $g_\alpha(x) < \infty$  for all  $x \in \mathcal{X}$ . Show that  $g(x) := \sup_\alpha g_\alpha(x)$  is convex on  $\mathcal{X}$  (you may assume it's finite).

v) Using what we've proven about subgradients, prove that a function  $g : \mathcal{X} \rightarrow \mathbb{R}$  is convex if and only if it can be written as the supremum of affine functions (e.g. supremum of  $g_\alpha(x) = \langle a_\alpha, x \rangle + b_\alpha$ )

## B) Some facts about subgradients

i) Show that the subgradient is not unique, but that the set of all subgradients is convex.

ii) Show that if  $f$  is differentiable at  $x$ ,  $\partial f(x) = \{\nabla f(x)\}$ .

iii) Show that if  $g_1 \in \partial f_1(x)$  and  $g_2 \in \partial f_2(x)$ , then  $g_1 + g_2 \in \partial(f_1 + f_2)(x)$ .

iv) Danskin's Theorem States

**Theorem 2** (Danskin). Let  $\Omega$  be a convex domain, and let  $f(x) = \sup_\alpha g_\alpha(x)$ , where  $g_\alpha$  are convex functions and finite on  $\Omega$ . Suppose  $f$  is also finite on  $\Omega$ . Then for all  $x \in \Omega$ ,  $\partial f = \text{Conv}\{\partial g_\alpha(x) \mid \sup_x g_\alpha(x) = f(x)\}$

Prove one direction of the theorem,  $\partial f \subseteq \text{Conv}\{\partial g_\alpha(x) \mid \sup_x g_\alpha(x) = f(x)\}$

v) (Hard) Prove the other direction of Danskin's theorem. Here you may want to work in epigraph space.

## Problem 2: Extensions for Gradient Descent

### A) Generalizations of SGD

i)) Prove the following statement:

**Proposition 1.** Let  $\Omega$  be a convex domain of radius  $R$ , and let  $f$  be a convex function on  $\Omega$ . Let  $x_0 \in \Omega$ , and let  $x_t = \Pi_\Omega(x_{t-1} - \eta g_t)$ , where  $\text{Exp}[g_t \mid g_1, \dots, g_{t-1}] \in \partial f(x_t)$ , and  $\sup_t \text{Exp}[\|g_t\|^2] \leq L$ . Prove that

$$f\left(\frac{1}{T} \sum_{t=1}^T x_t\right) \leq \inf_{x \in \Omega} f(x) + \dots \quad (1)$$

You fill in the ....

ii) Prove the following statement:

**Proposition 2.** Let  $\Omega$  be a convex domain of radius  $R$ , Let  $f_1, f_2, \dots, f_T$  be  $L$ -Lipschitz, convex functions on  $\Omega$ . Given any  $x_0 \in \Omega$ , let  $x_t = \Pi_\Omega(x_{t-1} - \eta g_t)$ , where  $g_t \in \partial f_t(x_t)$ , and  $\eta = \frac{LR}{\sqrt{T}}$ . Prove that

$$\frac{1}{T} \sum_{t=1}^T f\left(\frac{1}{T} \sum_{t=1}^T x_t\right) \leq \frac{1}{T} \sum_{t=1}^T f_t(x_t) \leq \inf_{x \in \Omega} \frac{1}{T} \sum_{t=1}^T f_t(x_t) + \dots \quad (2)$$

You fill in the ....

**B)** In this problem we show that in the stochastic setting, smoothness of the function  $f$  does not help. Let  $\Omega = [-1, 1]$ , let  $\sigma$  be a random variable with  $\mathbb{Pr}[\sigma = 1] = \mathbb{Pr}[\sigma = -1] = 1/2$ , fix an  $\epsilon \in (0, 1/4)$ . Let  $z_1, z_2, \dots, z_T$  be  $T$  i.i.d random variables, such that  $z_i | \sigma$  are mutually independent, and

$$\mathbb{Pr}[z_i = 1 | \sigma] = 1/2 + \sigma\epsilon \text{ and } \mathbb{Pr}[z_i = -1 | \sigma] = 1/2 - \sigma\epsilon \quad (3)$$

You will need the following information

**Lemma 1.** Let  $\sigma$  and  $z_1, z_2, \dots, z_T$  be as above. Then there exists a universal constant  $C$  such that, if  $T \leq C\epsilon^2$ , any algorithm which returns an estimate  $\hat{\sigma}$  of  $\sigma$  from observing  $z_1, z_2, \dots, z_T$  satisfies  $\mathbb{Pr}[\hat{\sigma} \neq \sigma] \geq \frac{1}{4}$ .

i) Construct a function on  $f_\sigma$  such that  $\mathbb{E}[z_i | \sigma] = \nabla f_\sigma(x)$  for all  $x \in \Omega$ . What is the optimum  $x_\sigma^*$  of  $f_\sigma$ ? What is the “smoothness” of  $f_\sigma$ .

ii) Show that there universal constants  $c$  such that the following hold: Fix a  $T \in \mathbb{N}$ , and let  $\mathcal{A}$  be an algorithm which is allowed to make  $T$  queries  $x_t \in [-1, 1]$  and  $g_t$ , where  $\mathcal{A}$  decides  $x_t$ , and receives responses  $g_t$  such that  $\mathbb{E}[g_t] = \nabla f(x_t)$  and  $|g_t| \leq 1$  as.s. Then, there is a 0-smooth, 1-Lipschitz function  $f$  and a mechanism generating responses  $g_t$  such that the iterate  $x_T$  satisfies

$$\mathbb{E}[f(x_1)] - \inf_{x \in [-1, 1]} f(x) \geq c/\sqrt{T} \quad (4)$$

iii) Answer to Problem 1(B)(iii) here.

C) Answer to Problem 1(C) here.

D) Answer to Problem 1(D) here.

E) Answer to Problem 1(E) here.

## **Bacgkround**