TTIC 31070: Convex Optimization
Homework 2

Professor Zhiyuan Li

Spring 2026

Problem 1 (Strict convexity, inverse gradients, and smooth conjugates)**.** *Let $(E, \|\cdot\|)$ be a finite-dimensional normed space, and let $\|\cdot\|_*$ be the dual norm on E^*. Let $f : E \to \mathbb{R}$ be convex and differentiable, and define its convex conjugate by $f^*(g) := \sup_{x \in E} \{\langle g, x\rangle - f(x)\}$.*

In parts (a)–(c)*, assume f is strictly convex, i.e. $f((1-\theta)x + \theta y) < (1-\theta)f(x) + \theta f(y)$ for all distinct $x, y \in E$ and all $\theta \in (0,1)$.*

(a) *Show that $\nabla f : E \to E^*$ is injective.*

(b) *Show that for every $x \in E$ and every $g \in E^*$,*

$$g = \nabla f(x) \qquad \Longleftrightarrow \qquad f(x) + f^*(g) = \langle g, x\rangle \qquad \Longleftrightarrow \qquad x \in \partial f^*(g).$$

(c) *Let $\nabla f(E) := \{\nabla f(x) : x \in E\} \subseteq E^*$. Deduce that for every $g \in \nabla f(E)$, the set $\partial f^*(g)$ is a single point. Conclude that f^* is differentiable on $\nabla f(E)$ and*

$$\nabla f^* = (\nabla f)^{-1} \qquad \text{on } \nabla f(E).$$

Hint: you may use the fact that a convex function is differentiable at a point whenever its subdifferential there consists of a single point.

In parts (d)–(f)*, assume f is μ-strongly convex with respect to $\|\cdot\|$, where $\mu > 0$, i.e. $f((1-\theta)x + \theta y) \le (1-\theta)f(x) + \theta f(y) - \frac{\mu}{2}\theta(1-\theta)\|x-y\|^2$ for all $x, y \in E$ and all $\theta \in [0,1]$.*

(d) *Show that $\langle \nabla f(x) - \nabla f(y), x - y\rangle \ge \mu \|x-y\|^2$ for every $x, y \in E$.*

(e) *Show that for every $g \in E^*$, the maximization problem defining $f^*(g)$ has a unique maximizer, denoted by x_g. Using* (b)–(c)*, deduce that $\nabla f(E) = E^*$, i.e. $\mathrm{dom}(f^*) = E^*$.*

(f) *Deduce that $\|\nabla f^*(g) - \nabla f^*(h)\| \le \frac{1}{\mu}\|g-h\|_*$ for all $g, h \in E^*$. In other words, f^* is $1/\mu$-smooth with respect to the dual norm.*

Problem 2 (Maximum entropy under a moment constraint)**.** *Fix numbers $a_1, \ldots, a_m \in \mathbb{R}$ and a target moment $\alpha \in \mathbb{R}$. Consider the optimization problem*

$$\min_{p \in \mathbb{R}^m} \sum_{i=1}^m p_i \log p_i$$

with the convention $0 \log 0 := 0$, subject to

$$p_i \ge 0 \quad \forall i, \qquad \sum_{i=1}^m p_i = 1, \qquad \sum_{i=1}^m a_i p_i = \alpha.$$

Assume there exists a strictly feasible point, i.e. some $p \in \mathbb{R}^m$ with $p_i > 0$ for all i, $\sum_i p_i = 1$, and $\sum_i a_i p_i = \alpha$.

(a) *Explain why this is a convex optimization problem, and why strong duality holds.*

(b) *Derive the KKT conditions.*

(c) *Show that every optimal solution must satisfy*

$$p_i^\star = \exp(-1 - \nu - \lambda a_i) \qquad \forall i = 1, \ldots, m$$

for some scalars $\nu, \lambda \in \mathbb{R}$.

(d) *Deduce that every optimal solution has the form*

$$p_i^\star = \frac{e^{-\lambda a_i}}{\sum_{j=1}^m e^{-\lambda a_j}} \qquad \forall i = 1, \ldots, m,$$

where λ is chosen so that

$$\sum_{i=1}^m a_i p_i^\star = \alpha.$$

Problem 3 (Implementing a separation oracle for a near-optimal set)**.** *Consider the constrained convex optimization problem*

$$\min_{x \in \mathbb{R}^n} f(x) \qquad \text{subject to} \qquad c_i(x) \le 0, \quad i = 1, \ldots, m,$$

where $f, c_1, \ldots, c_m : \mathbb{R}^n \to \mathbb{R}$ are convex. Assume that the optimal value $f^\star$ is known, and that for each of the functions $f, c_1, \ldots, c_m$, a subgradient oracle is available: on input $x \in \mathbb{R}^n$, it returns the function value at x together with one subgradient at x. For $\delta \ge 0$, define the δ-approximate optimal set

$$K_\delta := \{x \in \mathbb{R}^n : \ c_i(x) \le 0 \ \forall i, \quad f(x) \le f^\star + \delta\}.$$

Consider the following ellipsoid method, where the ellipsoid update itself may be treated as a given subroutine.

Algorithm 1 Ellipsoid method for finding a point in K_δ

Require: An initial ellipsoid $E_0 \subseteq \mathbb{R}^n$ containing K_δ.

for $t = 0, 1, 2, \ldots$ **do**
    Let x_t be the center of E_t.
    Query a subroutine $\mathrm{Separate}_{K_\delta}(x_t)$.
    if $\mathrm{Separate}_{K_\delta}(x_t)$ certifies that $x_t \in K_\delta$ **then**
        Output x_t and stop.
    else
        Receive a halfspace H_t such that $K_\delta \subseteq H_t$ but $x_t \notin H_t$.
        Let E_{t+1} be the central-cut ellipsoid update obtained from E_t and H_t.
    end if
end for

(a) *Using only the subgradient oracles for $f, c_1, \ldots, c_m$, implement the subroutine* $\mathrm{Separate}_{K_\delta}(x)$*: given $x \in \mathbb{R}^n$, it should either certify that $x \in K_\delta$, or return an explicit halfspace H such that $K_\delta \subseteq H$ but $x \notin H$.*

(b) *Prove that your subroutine is correct.*

Problem 4 (Reducing the smooth convex case to the strongly convex case)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and M-smooth with respect to the Euclidean norm. Let $x^{(0)} \in \mathbb{R}^n$, and assume that f has a minimizer $x^\star$ satisfying*

$$\left\| x^\star - x^{(0)} \right\|_2 \le R.$$

For $\lambda > 0$, define the regularized objective

$$f_\lambda(x) := f(x) + \frac{\lambda}{2} \left\| x - x^{(0)} \right\|_2^2 .$$

(a) *Show that f_λ is $(M + \lambda)$-smooth and λ-strongly convex.*

(b) *Let $x_\lambda^\star$ minimize f_λ. Show that*

$$f(x_\lambda^\star) - f(x^\star) \le \frac{\lambda R^2}{2}.$$

(c) *Show that if $x \in \mathbb{R}^n$ satisfies*

$$f_\lambda(x) \le f_\lambda(x_\lambda^\star) + \frac{\varepsilon}{2},$$

and if one chooses

$$\lambda = \frac{\varepsilon}{R^2},$$

then x is ε-optimal for the original problem $\min_x f(x)$.

(d) *Consider running gradient descent on f_λ with step size*

$$\eta = \frac{1}{M + \lambda}.$$

Use the linear-rate result for smooth strongly convex functions from Lecture 7 to derive an iteration bound guaranteeing an ε-optimal point for $\min_x f(x)$.

Problem 5 (Smoothness, gradient Lipschitzness, and the factor 2 gap)**.** *Let $(E, \|\cdot\|)$ be a finite-dimensional normed space, and let $\|\cdot\|_*$ be the dual norm on E^*. Let $f : E \to \mathbb{R}$ be twice continuously differentiable, i.e. $f \in C^2(E)$. Define*

$$\mathrm{S}_f := \inf \left\{ S \ge 0 : |f(y) - f(x) - \langle \nabla f(x), y - x \rangle| \le \frac{S}{2} \|y - x\|^2 \ \textit{for all } x, y \in E \right\},$$

$$\mathrm{L}_f := \inf \left\{ L \ge 0 : \|\nabla f(y) - \nabla f(x)\|_* \le L \|y - x\| \ \textit{for all } x, y \in E \right\}.$$

Thus S_f is the smallest smoothness constant in the remainder form, and L_f is the smallest gradient-Lipschitz constant. In Euclidean spaces they coincide, while in general normed spaces they are not identical but remain very close: the later parts of this problem show that they differ by at most a factor of 2.

For each $x \in E$, define

$$Q_{f,x} := \sup_{\|h\| \le 1} \left| \left\langle \nabla^2 f(x) h, h \right\rangle \right|, \qquad S_{f,x} := \sup_{\|u\| \le 1,\ \|v\| \le 1} \left| \left\langle \nabla^2 f(x) u, v \right\rangle \right|.$$

(a) *Show that*

$$\mathrm{S}_f = \sup_{x\in E} Q_{f,x}, \qquad \mathrm{L}_f = \sup_{x\in E} S_{f,x}.$$

(b) *Assume* $E = \mathbb{R}^n$ *with the Euclidean norm. Prove that*

$$\mathrm{S}_f = \mathrm{L}_f.$$

Hint: for each symmetric matrix A*,*

$$\sup_{\|u\|_2\le 1,\ \|v\|_2\le 1} |\langle Au, v\rangle| = \sup_{\|h\|_2\le 1} |\langle Ah, h\rangle| .$$

(c) *Show that the equality* $\mathrm{S}_f = \mathrm{L}_f$ *with the same constant fails for general norms. Consider*

$$E = \mathbb{R}^2, \qquad \|\cdot\| = \|\cdot\|_1 , \qquad f(x_1, x_2) = x_1 x_2.$$

Compute S_f *and* L_f*, and show that*

$$\mathrm{L}_f > \mathrm{S}_f.$$

(d) *Let* $B : E \times E \to \mathbb{R}$ *be a symmetric bilinear form. Define*

$$Q(B) := \sup_{\|h\|\le 1} |B(h,h)|, \qquad M(B) := \sup_{\|u\|\le 1,\ \|v\|\le 1} |B(u,v)|.$$

Prove that

$$Q(B) \le M(B) \le 2Q(B).$$

Hint:

$$4B(u,v) = B(u+v, u+v) - B(u-v, u-v).$$

(e) *Deduce that*

$$\mathrm{S}_f \le \mathrm{L}_f \le 2\mathrm{S}_f.$$

Show also that the factor 2 *is sharp.*