TTIC 31070: Convex Optimization
Homework 2 Solutions

Professor Zhiyuan Li

Spring 2026

Problem 1

(a) Suppose $\nabla f(x) = \nabla f(y) = g$. If $x \neq y$, then strict convexity implies the tangent inequality is strict:

$$f(y) > f(x) + \langle g, y - x \rangle, \qquad f(x) > f(y) + \langle g, x - y \rangle.$$

Adding these two inequalities gives $0 > 0$, a contradiction. Hence $x = y$, so $\nabla f : E \to E^*$ is injective.

(b) For convex conjugates, Fenchel–Young says

$$f(x) + f^*(g) \geq \langle g, x \rangle,$$

with equality iff $g \in \partial f(x)$ iff $x \in \partial f^*(g)$. Since f is differentiable,

$$g \in \partial f(x) \qquad \Longleftrightarrow \qquad g = \nabla f(x).$$

Thus

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

(c) Let

$$\nabla f(E) := \{\nabla f(x) : x \in E\} \subseteq E^*.$$

Now fix $g \in \nabla f(E)$, and write $g = \nabla f(x)$. If $y \in \partial f^*(g)$, then $g = \nabla f(y)$, so part (a) gives $y = x$. Hence $\partial f^*(g)$ consists of the single point x. Therefore f^* is differentiable at g, with gradient equal to the unique subgradient:

$$\nabla f^*(g) = x.$$

Equivalently,

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

(d) Strong convexity gives

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2} \|y - x\|^2,$$

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

Adding them yields

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu \|x - y\|^2 \qquad \forall x, y \in E.$$

(e) Fix $g \in E^*$. Consider

$$\phi_g(x) := f(x) - \langle g, x \rangle .$$

Since f is μ-strongly convex, so is ϕ_g. Also ϕ_g is continuous and, for any fixed x_0,

$$\phi_g(x) \geq \phi_g(x_0) + \langle \nabla f(x_0) - g, x - x_0 \rangle + \frac{\mu}{2} \|x - x_0\|^2 ,$$

so $\phi_g(x) \to +\infty$ as $\|x\| \to \infty$. Therefore ϕ_g has a unique minimizer x_g. Equivalently, the supremum defining $f^*(g)$ has a unique maximizer x_g.

$$x_g \in \partial f^*(g)$$

by the equality case of Fenchel–Young. Part (b) then gives

$$g = \nabla f(x_g),$$

so $g \in \nabla f(E)$. Since $g \in E^*$ was arbitrary,

$$\nabla f(E) = E^*.$$

Equivalently, $f^*(g) < \infty$ for every $g \in E^*$, so

$$\mathrm{dom}(f^*) = E^*.$$

By part (c),

$$\nabla f^*(g) = x_g = (\nabla f)^{-1}(g) \qquad \forall g \in E^*.$$

(f) Let

$$x_g := \nabla f^*(g), \qquad x_h := \nabla f^*(h).$$

Then $g = \nabla f(x_g)$ and $h = \nabla f(x_h)$. Applying part (d) with $x = x_g$ and $y = x_h$, we get

$$\langle g - h, \nabla f^*(g) - \nabla f^*(h) \rangle \geq \mu \|\nabla f^*(g) - \nabla f^*(h)\|^2 .$$

By Hölder,

$$\mu \|\nabla f^*(g) - \nabla f^*(h)\|^2 \leq \|g - h\|_* \|\nabla f^*(g) - \nabla f^*(h)\| .$$

If the last factor is nonzero, divide through by it. We obtain

$$\|\nabla f^*(g) - \nabla f^*(h)\| \leq \frac{1}{\mu} \|g - h\|_* .$$

So f^* is $1/\mu$-smooth with respect to $\|\cdot\|_*$.

Problem 2

(a) The feasible set is the intersection of the simplex with an affine hyperplane, so it is convex. The objective

$$\Phi(p) := \sum_{i=1}^{m} p_i \log p_i$$

is convex on the nonnegative orthant (with the convention $0 \log 0 = 0$). Because a strictly feasible point exists, Slater's condition holds, so strong duality holds.

(b) Introduce multipliers $\nu, \lambda \in \mathbb{R}$ for the equality constraints and $\mu_i \geq 0$ for the inequalities $p_i \geq 0$, written as $-p_i \leq 0$. The Lagrangian is

$$\mathcal{L}(p, \nu, \lambda, \mu) = \sum_{i=1}^{m} p_i \log p_i + \nu\Big(\sum_{i=1}^{m} p_i - 1\Big) + \lambda\Big(\sum_{i=1}^{m} a_i p_i - \alpha\Big) - \sum_{i=1}^{m} \mu_i p_i.$$

The KKT conditions are:

$$\begin{aligned}
&p_i \geq 0, \qquad \sum_i p_i = 1, \qquad \sum_i a_i p_i = \alpha,\\
&\mu_i \geq 0,\\
&1 + \log p_i + \nu + \lambda a_i - \mu_i = 0 \qquad (i = 1, \ldots, m),\\
&\mu_i p_i = 0 \qquad (i = 1, \ldots, m).
\end{aligned}$$

(c) We first show that every optimal solution is strictly positive. Suppose $p^\star$ is optimal and $p_i^\star = 0$ for some i. Let q be a strictly feasible point. Then

$$p(t) := (1-t)p^\star + tq$$

is feasible for all sufficiently small $t > 0$, and the i-th coordinate of $p(t)$ equals $tq_i > 0$. Since $t \log t$ has derivative $-\infty$ at 0^+, the one-sided derivative of $\Phi(p(t))$ at $t = 0^+$ is $-\infty$, contradicting optimality of $p^\star$. Thus $p_i^\star > 0$ for all i.

Hence complementary slackness gives $\mu_i = 0$ for all i, and the stationarity condition becomes

$$1 + \log p_i^\star + \nu + \lambda a_i = 0.$$

Therefore

$$\log p_i^\star = -1 - \nu - \lambda a_i, \qquad p_i^\star = \exp(-1 - \nu - \lambda a_i).$$

(d) Summing the formula from part (c) over i and using $\sum_i p_i^\star = 1$, we get

$$1 = e^{-1-\nu} \sum_{j=1}^{m} e^{-\lambda a_j}, \qquad e^{-1-\nu} = \frac{1}{\sum_{j=1}^{m} e^{-\lambda a_j}}.$$

Thus

$$p_i^\star = \frac{e^{-\lambda a_i}}{\sum_{j=1}^{m} e^{-\lambda a_j}}.$$

The scalar λ is then chosen so that the moment constraint holds:

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

Problem 3

(a) Given $x \in \mathbb{R}^n$, query the value/subgradient oracle for each c_i, obtaining $c_i(x)$ and $g_i \in \partial c_i(x)$, and query the oracle for f, obtaining $f(x)$ and $g_f \in \partial f(x)$.

Then define $\mathrm{Separate}_{K_\delta}(x)$ as follows:

- If $c_i(x) > 0$ for some i, return the halfspace

$$H = \{y \in \mathbb{R}^n : \ c_i(x) + \langle g_i, y - x \rangle \leq 0\} .$$

- Else, if $f(x) \leq f^\star + \delta$, certify that $x \in K_\delta$.
- Else return the halfspace

$$H = \{y \in \mathbb{R}^n : \ f(x) + \langle g_f, y - x \rangle \leq f^\star + \delta\} .$$

(b) If the procedure certifies $x \in K_\delta$, then by construction $c_i(x) \leq 0$ for all i and $f(x) \leq f^\star + \delta$, so indeed $x \in K_\delta$.

Now suppose the procedure returns the constraint halfspace associated to some i with $c_i(x) > 0$. For any $y \in K_\delta$, convexity gives

$$c_i(y) \geq c_i(x) + \langle g_i, y - x \rangle .$$

Since $c_i(y) \leq 0$, it follows that

$$c_i(x) + \langle g_i, y - x \rangle \leq 0,$$

so $y \in H$. Thus $K_\delta \subseteq H$. But

$$c_i(x) + \langle g_i, x - x \rangle = c_i(x) > 0,$$

so $x \notin H$.

Finally, suppose all constraints are satisfied but $f(x) > f^\star + \delta$. For any $y \in K_\delta$, convexity gives

$$f(y) \geq f(x) + \langle g_f, y - x \rangle .$$

Since $f(y) \leq f^\star + \delta$,

$$f(x) + \langle g_f, y - x \rangle \leq f^\star + \delta,$$

so again $K_\delta \subseteq H$. But

$$f(x) + \langle g_f, x - x \rangle = f(x) > f^\star + \delta,$$

hence $x \notin H$. Therefore the oracle is correct in every case.

Problem 4

(a) We have

$$\nabla f_\lambda(x) = \nabla f(x) + \lambda(x - x^{(0)}).$$

Hence for all $x, y \in \mathbb{R}^n$,

$$\|\nabla f_\lambda(y) - \nabla f_\lambda(x)\|_2 \leq \|\nabla f(y) - \nabla f(x)\|_2 + \lambda \|y - x\|_2 \leq (M + \lambda) \|y - x\|_2 .$$

So f_λ is $(M + \lambda)$-smooth.

For strong convexity, convexity of f gives

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle .$$

Also

$$\frac{\lambda}{2} \left\|y - x^{(0)}\right\|_2^2 = \frac{\lambda}{2} \left\|x - x^{(0)}\right\|_2^2 + \lambda \left\langle x - x^{(0)}, y - x \right\rangle + \frac{\lambda}{2} \|y - x\|_2^2 .$$

Adding these two identities yields

$$f_\lambda(y) \geq f_\lambda(x) + \langle \nabla f_\lambda(x), y - x \rangle + \frac{\lambda}{2} \|y - x\|_2^2 .$$

Thus f_λ is λ-strongly convex.

(b) By optimality of $x_\lambda^\star$,

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

Since $f(x_\lambda^\star) \leq f_\lambda(x_\lambda^\star)$, we conclude

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

(c) We have

$$f(x) \leq f_\lambda(x) \leq f_\lambda(x_\lambda^\star) + \frac{\varepsilon}{2} \leq f(x^\star) + \frac{\lambda R^2}{2} + \frac{\varepsilon}{2}.$$

With $\lambda = \varepsilon / R^2$, this becomes

$$f(x) \leq f(x^\star) + \varepsilon.$$

Hence x is ε-optimal for $\min_x f(x)$.

(d) Let

$$L := M + \lambda, \qquad \mu := \lambda.$$

Gradient descent with step size $1/L$ on the L-smooth, μ-strongly convex function f_λ satisfies

$$f_\lambda(x_t) - f_\lambda(x_\lambda^\star) \leq \left(1 - \frac{\mu}{L}\right)^t \big(f_\lambda(x^{(0)}) - f_\lambda(x_\lambda^\star)\big).$$

It is enough to make the right-hand side at most $\varepsilon/2$. Since $f_\lambda(x^{(0)}) = f(x^{(0)})$, a sufficient condition is

$$\left(1 - \frac{\lambda}{M + \lambda}\right)^t \big(f(x^{(0)}) - f_\lambda(x_\lambda^\star)\big) \leq \frac{\varepsilon}{2}.$$

Using $1 - s \leq e^{-s}$, it suffices to take

$$t \geq \frac{M + \lambda}{\lambda} \log \frac{2\big(f(x^{(0)}) - f_\lambda(x_\lambda^\star)\big)}{\varepsilon}.$$

Since $f_\lambda(x_\lambda^\star) \geq f(x^\star)$, we may simplify this to

$$t \geq \frac{M + \lambda}{\lambda} \log \frac{2\big(f(x^{(0)}) - f(x^\star)\big)}{\varepsilon}.$$

With $\lambda = \varepsilon / R^2$, this becomes

$$t \geq \left(\frac{M R^2}{\varepsilon} + 1\right) \log \frac{2\big(f(x^{(0)}) - f(x^\star)\big)}{\varepsilon}.$$

By part (c), this guarantees that x_t is ε-optimal for the original problem.

Problem 5

(a) Let

$$Q := \sup_{x\in E} Q_{f,x}, \qquad M := \sup_{x\in E} S_{f,x}.$$

We first prove $\mathrm{S}_f = Q$. Let $d := y - x$. Since $f \in C^2(E)$,

$$f(y) - f(x) - \langle \nabla f(x), d\rangle = \int_0^1 (1-t)\left\langle \nabla^2 f(x+td)d, d\right\rangle\, dt.$$

Hence

$$|f(y) - f(x) - \langle \nabla f(x), d\rangle| \le \int_0^1 (1-t)Q\,\|d\|^2\, dt = \frac{Q}{2}\,\|d\|^2\,.$$

Therefore $\mathrm{S}_f \le Q$.

Conversely, if

$$|f(y) - f(x) - \langle \nabla f(x), y-x\rangle| \le \frac{S}{2}\,\|y-x\|^2 \qquad \forall x, y \in E,$$

then applying this with $y = x + th$, dividing by $t^2/2$, and letting $t \to 0$ gives

$$\left|\left\langle \nabla^2 f(x)h, h\right\rangle\right| \le S\,\|h\|^2 \qquad \forall x, h.$$

Thus $Q_{f,x} \le S$ for every x, so $Q \le S$. Taking the infimum over such S gives $Q \le \mathrm{S}_f$. Hence $\mathrm{S}_f = Q$.

Next we prove $\mathrm{L}_f = M$. Let $d := y - x$. For every $v \in E$,

$$\langle \nabla f(y) - \nabla f(x), v\rangle = \int_0^1 \left\langle \nabla^2 f(x+td)d, v\right\rangle\, dt,$$

so

$$|\langle \nabla f(y) - \nabla f(x), v\rangle| \le M\,\|d\|\,\|v\|\,.$$

Taking the supremum over $\|v\| \le 1$ yields

$$\|\nabla f(y) - \nabla f(x)\|_* \le M\,\|y-x\|\,,$$

so $\mathrm{L}_f \le M$.

Conversely, if

$$\|\nabla f(y) - \nabla f(x)\|_* \le L\,\|y-x\| \qquad \forall x, y \in E,$$

then for every $x, u \in E$,

$$\frac{\|\nabla f(x+tu) - \nabla f(x)\|_*}{|t|} \le L\,\|u\|\,.$$

Letting $t \to 0$ gives

$$\left\|\nabla^2 f(x)u\right\|_* \le L\,\|u\|\,.$$

Hence

$$\left|\left\langle \nabla^2 f(x)u, v\right\rangle\right| \le L\,\|u\|\,\|v\| \qquad \forall x, u, v,$$

so $S_{f,x} \le L$ for every x, hence $M \le L$. Taking the infimum over such L gives $M \le \mathrm{L}_f$. Therefore $\mathrm{L}_f = M$.

(b) In Euclidean space, $\nabla^2 f(x)$ is symmetric for every x. By the hint,

$$S_{f,x} = Q_{f,x} \qquad \forall x \in \mathbb{R}^n.$$

Taking the supremum over x and using part (a) gives

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

(c) Write $\Delta = y - x = (a, b)$. Then

$$f(y) - f(x) - \langle \nabla f(x), y - x \rangle = ab,$$

because $\nabla f(x_1, x_2) = (x_2, x_1)$. Therefore S_f is the smallest constant S such that

$$|ab| \le \frac{S}{2}(|a| + |b|)^2 \qquad \forall a, b \in \mathbb{R}.$$

Since

$$2\,|ab| \le \frac{(|a| + |b|)^2}{2},$$

with equality at $|a| = |b|$, we get

$$\mathrm{S}_f = \frac{1}{2}.$$

Also

$$\nabla f(y) - \nabla f(x) = (b, a), \qquad \|\nabla f(y) - \nabla f(x)\|_\infty = \max\{|a|, |b|\}.$$

Thus L_f is the smallest constant L such that

$$\max\{|a|, |b|\} \le L(|a| + |b|) \qquad \forall a, b \in \mathbb{R}.$$

Taking $(a, b) = (1, 0)$ shows $L \ge 1$, and $L = 1$ clearly works. Hence

$$\mathrm{L}_f = 1 > \frac{1}{2} = \mathrm{S}_f.$$

(d) The first inequality $Q(B) \le M(B)$ is immediate by setting $u = v = h$.

For the second inequality, let $\|u\| \le 1$ and $\|v\| \le 1$. Since B is symmetric,

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

Therefore

$$4\,|B(u, v)| \le |B(u + v, u + v)| + |B(u - v, u - v)|\,.$$

Now $\|u \pm v\| \le \|u\| + \|v\| \le 2$, so by the definition of $Q(B)$,

$$|B(u \pm v, u \pm v)| \le 4Q(B).$$

Hence $4\,|B(u, v)| \le 8Q(B)$, i.e.

$$|B(u, v)| \le 2Q(B).$$

Taking the supremum over $\|u\|, \|v\| \le 1$ gives

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

(e) For each $x \in E$, apply part (d) to the symmetric bilinear form

$$B_x(u, v) := \left\langle \nabla^2 f(x)u, v \right\rangle.$$

Then

$$Q_{f,x} \leq S_{f,x} \leq 2Q_{f,x} \qquad \forall x \in E.$$

Taking the supremum over x and using part (a) yields

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

Part (c) gives

$$\mathrm{L}_f = 1 = 2 \cdot \frac{1}{2} = 2\mathrm{S}_f,$$

so the factor 2 is sharp.