TTIC 31070: Convex Optimization
Homework 1 Solutions

Professor Zhiyuan Li

Spring 2026

Problem 1

(a) For $f(x_1, x_2) = \max\{x_1, x_2\}$,

$$\partial f(x_1, x_2) = \begin{cases} \{(1,0)\}, & x_1 > x_2, \\ \{(0,1)\}, & x_1 < x_2, \\ \{(\lambda, 1-\lambda) : \lambda \in [0,1]\}, & x_1 = x_2. \end{cases}$$

Indeed, away from the diagonal the function agrees locally with one coordinate projection, so the subgradient is the corresponding gradient. On the diagonal both affine functions x_1 and x_2 are active, and every convex combination of their gradients gives a valid supporting affine function.

(b) For $f(x) = \|x\|_1$,

$$\partial \|\cdot\|_1 (x) = \{g \in \mathbb{R}^n : g_i = \operatorname{sign}(x_i) \text{ if } x_i \neq 0,\ g_i \in [-1,1] \text{ if } x_i = 0\}.$$

This follows from the one-dimensional formula

$$\partial |t| = \begin{cases} \{1\}, & t > 0, \\ [-1,1], & t = 0, \\ \{-1\}, & t < 0, \end{cases}$$

and the separable decomposition $\|x\|_1 = \sum_i |x_i|$.

(c) Let

$$I(x) := \{i : |x_i| = \|x\|_\infty\}.$$

Then

$$\partial \|\cdot\|_\infty (x) = \begin{cases} \operatorname{conv}\{\operatorname{sign}(x_i) e_i : i \in I(x)\}, & x \neq 0, \\ \{g \in \mathbb{R}^n : \|g\|_1 \leq 1\}, & x = 0. \end{cases}$$

For $x \neq 0$, the function $\|x\|_\infty$ is the maximum of the affine functions $x \mapsto x_i$ and $x \mapsto -x_i$, so the subdifferential is the convex hull of the gradients of the active affine functions. At $x = 0$, this reduces to the dual unit ball of $\|\cdot\|_\infty$, namely the ℓ_1-unit ball.

(d) At $x = 0$, the subgradient condition reads

$$\forall y \in E, \qquad \|y\| \geq \langle g, y \rangle .$$

If $g \in \partial \|\cdot\| (0)$, then taking the supremum over $\|y\| \leq 1$ gives $\|g\|_* \leq 1$. Conversely, if $\|g\|_* \leq 1$, then $\langle g, y \rangle \leq \|g\|_* \|y\| \leq \|y\|$, so $g \in \partial \|\cdot\| (0)$. Hence

$$\partial \|\cdot\| (0) = \{g \in E^* : \|g\|_* \leq 1\} .$$

Now let $x \neq 0$. If $g \in \partial \|\cdot\| (x)$, then for every $y \in E$,

$$\|y\| \geq \|x\| + \langle g, y - x \rangle .$$

Taking $y = 0$ gives $\langle g, x \rangle \geq \|x\|$, while taking $y = 2x$ gives $\langle g, x \rangle \leq \|x\|$. Therefore

$$\langle g, x \rangle = \|x\| .$$

Next, for any $v \in E$ and any $t > 0$,

$$\|x + tv\| \geq \|x\| + t \langle g, v \rangle ,$$

so

$$\langle g, v \rangle \leq \frac{\|x + tv\| - \|x\|}{t} \leq \|v\| .$$

Thus $\|g\|_* \leq 1$. Since

$$\|x\| = \langle g, x \rangle \leq \|g\|_* \|x\| ,$$

we conclude that $\|g\|_* = 1$.

Conversely, if $\|g\|_* = 1$ and $\langle g, x \rangle = \|x\|$, then for every $y \in E$,

$$\langle g, y \rangle \leq \|g\|_* \|y\| = \|y\| ,$$

and therefore

$$\|y\| \geq \langle g, y \rangle = \langle g, x \rangle + \langle g, y - x \rangle = \|x\| + \langle g, y - x \rangle .$$

So $g \in \partial \|\cdot\| (x)$. Hence

$$\partial \|\cdot\| (x) = \{g \in E^* : \|g\|_* = 1,\ \langle g, x \rangle = \|x\|\} \qquad (x \neq 0).$$

Problem 2

(a) We argue by induction on m. The case $m = 2$ is exactly the definition of convexity. Assume the statement holds for $m - 1$ points. Let

$$\mu := \lambda_1 + \cdots + \lambda_{m-1}.$$

If $\mu = 0$, then $\lambda_m = 1$ and there is nothing to prove. Otherwise set

$$\widetilde{x} := \sum_{i=1}^{m-1} \frac{\lambda_i}{\mu} x_i.$$

Then

$$\sum_{i=1}^{m} \lambda_i x_i = \mu \widetilde{x} + (1-\mu) x_m.$$

By convexity,

$$f\Big(\sum_{i=1}^{m} \lambda_i x_i\Big) \le \mu f(\widetilde{x}) + (1-\mu) f(x_m).$$

By the induction hypothesis,

$$f(\widetilde{x}) \le \sum_{i=1}^{m-1} \frac{\lambda_i}{\mu} f(x_i).$$

Substituting gives

$$f\Big(\sum_{i=1}^{m} \lambda_i x_i\Big) \le \sum_{i=1}^{m} \lambda_i f(x_i).$$

(b) If X takes the value x_i with probability λ_i, then

$$\mathbb{E}[X] = \sum_i \lambda_i x_i, \qquad \mathbb{E}[f(X)] = \sum_i \lambda_i f(x_i).$$

So part (a) gives

$$f(\mathbb{E}[X]) \le \mathbb{E}[f(X)].$$

Problem 3

(a) Choose such a representation with m minimal and with all $\lambda_i > 0$ (discard any zero coefficients). We claim that $s_1, \dots, s_m$ are affinely independent.

Suppose not. Then there exist scalars $\alpha_1, \dots, \alpha_m$, not all zero, such that

$$\sum_{i=1}^{m} \alpha_i s_i = 0 \qquad \text{and} \qquad \sum_{i=1}^{m} \alpha_i = 0.$$

Let

$$I_+ := \{i : \alpha_i > 0\}.$$

This set is nonempty; otherwise all $\alpha_i \le 0$ and their sum could not be zero unless all $\alpha_i = 0$. Define

$$t := \min_{i \in I_+} \frac{\lambda_i}{\alpha_i} > 0.$$

Set

$$\lambda_i' := \lambda_i - t\alpha_i.$$

Then $\lambda_i' \ge 0$ for all i, and at least one λ_i' is zero. Also

$$\sum_{i=1}^{m} \lambda_i' = \sum_{i=1}^{m} \lambda_i - t \sum_{i=1}^{m} \alpha_i = 1,$$

and

$$\sum_{i=1}^{m} \lambda_i' s_i = \sum_{i=1}^{m} \lambda_i s_i - t \sum_{i=1}^{m} \alpha_i s_i = x.$$

So x has a convex representation using fewer than m points, contrary to minimality.

(b) Affinely independent points in $\mathbb{R}^d$ can have cardinality at most $d+1$. Hence $m \le d+1$.

(c) Combining the definition of $\operatorname{conv}(S)$ with parts (a)–(b), every $x \in \operatorname{conv}(S)$ can be written as a convex combination of at most $d+1$ points of S. This is Carathéodory's theorem.

Problem 4

(a) By definition,

$$\pi_i(x) := \frac{e^{x_i}}{\sum_{j=1}^m e^{x_j}}.$$

Since the numerator and denominator are positive, $\pi_i(x) \ge 0$. Also,

$$\sum_{i=1}^m \pi_i(x) = \frac{\sum_{i=1}^m e^{x_i}}{\sum_{j=1}^m e^{x_j}} = 1.$$

(b) Differentiate coordinatewise:

$$\frac{\partial F}{\partial x_i}(x) = \frac{e^{x_i}}{\sum_{j=1}^m e^{x_j}} = \pi_i(x).$$

Hence

$$\nabla F(x) = (\pi_1(x), \ldots, \pi_m(x)).$$

(c) Since $\pi_i(x) \ge 0$ and $\sum_i \pi_i(x) = 1$, the formula in part (b) shows that each gradient component is smooth. Differentiating once more gives

$$\frac{\partial \pi_i}{\partial x_j}(x) = \pi_i(x)(\delta_{ij} - \pi_j(x)),$$

so

$$\nabla^2 F(x) = \operatorname{Diag}(\pi(x)) - \pi(x)\pi(x)^\top,$$

where $\pi(x) = (\pi_1(x), \ldots, \pi_m(x))$. For any $v \in \mathbb{R}^m$,

$$v^\top \nabla^2 F(x)\, v = \sum_{i=1}^m \pi_i(x) v_i^2 - \Big(\sum_{i=1}^m \pi_i(x) v_i\Big)^2 \ge 0$$

by the variance identity. Hence $\nabla^2 F(x) \succeq 0$, so F is convex.

(d) Let

$$M := \max_{1 \le i \le m} x_i.$$

Then

$$\sum_{i=1}^m e^{x_i} \ge e^M,$$

so

$$F(x) \ge \log(e^M) = M.$$

Also each term satisfies $e^{x_i} \le e^M$, hence

$$\sum_{i=1}^m e^{x_i} \le m e^M.$$

Taking logs gives

$$F(x) \le M + \log m.$$

Problem 5

(a) Consider the $d+2$ vectors

$$(x_1, 1), \ldots, (x_{d+2}, 1) \in \mathbb{R}^{d+1}.$$

Since there are $d+2$ of them in a $(d+1)$-dimensional space, they are linearly dependent. Hence there exist scalars $\alpha_1, \ldots, \alpha_{d+2}$, not all zero, such that

$$\sum_{i=1}^{d+2} \alpha_i (x_i, 1) = 0.$$

Reading off the first d coordinates and the last coordinate gives

$$\sum_{i=1}^{d+2} \alpha_i x_i = 0 \qquad \text{and} \qquad \sum_{i=1}^{d+2} \alpha_i = 0.$$

(b) If all $\alpha_i \geq 0$, then $\sum_i \alpha_i = 0$ forces $\alpha_i = 0$ for all i, contrary to the choice of the coefficients. Thus some $\alpha_i > 0$, so $I_+ \neq \varnothing$. The same argument shows that not all $\alpha_i \leq 0$, hence $I_- \neq \varnothing$.

(c) For $i \in I_+$, we have $\lambda_i > 0$, and

$$\sum_{i \in I_+} \lambda_i = \frac{\sum_{i \in I_+} \alpha_i}{\sum_{j \in I_+} \alpha_j} = 1.$$

Similarly, for $i \in I_-$,

$$\mu_i > 0, \qquad \sum_{i \in I_-} \mu_i = 1.$$

So both families are convex coefficients.

Now let

$$T := \sum_{j \in I_+} \alpha_j = \sum_{j \in I_-} (-\alpha_j),$$

where the equality follows from $\sum_i \alpha_i = 0$. Since $\sum_i \alpha_i x_i = 0$,

$$\sum_{i \in I_+} \alpha_i x_i = \sum_{i \in I_-} (-\alpha_i) x_i.$$

Dividing by $T > 0$ gives

$$\sum_{i \in I_+} \lambda_i x_i = \sum_{i \in I_-} \mu_i x_i.$$

(d) Set

$$A := \{x_i : i \in I_+\}, \qquad B := \{x_i : i \in I_-\}.$$

Then $A \cap B = \varnothing$, and part (c) shows that the point

$$\sum_{i \in I_+} \lambda_i x_i = \sum_{i \in I_-} \mu_i x_i$$

lies in both $\operatorname{conv}(A)$ and $\operatorname{conv}(B)$. Therefore

$$\operatorname{conv}(A) \cap \operatorname{conv}(B) \neq \varnothing.$$

This is the desired Radon partition.

Problem 6

(a) With the notation

$$E^+(v) := \{(v,w) \in E\}, \qquad E^-(v) := \{(u,v) \in E\},$$

the max-flow LP is

$$\max\ \mathrm{val}(f) \quad \text{s.t.} \quad 0 \le f_{uv} \le c_{uv} \quad ((u,v) \in E), \qquad \sum_{e \in E^+(v)} f_e = \sum_{e \in E^-(v)} f_e \quad (v \in V \setminus \{s,t\}),$$

where

$$\mathrm{val}(f) := \sum_{e \in E^+(s)} f_e - \sum_{e \in E^-(s)} f_e.$$

(b) Introduce free dual variables y_v for the flow-conservation constraints at $v \in V \setminus \{s,t\}$, and nonnegative dual variables z_{uv} for the upper-bound constraints $f_{uv} \le c_{uv}$. After the usual LP duality calculation, one obtains the dual problem

$$\min \sum_{(u,v) \in E} c_{uv}\, z_{uv} \quad \text{s.t.} \quad y_s - y_t = 1, \qquad z_{uv} \ge y_u - y_v, \qquad z_{uv} \ge 0 \quad ((u,v) \in E).$$

Equivalently,

$$z_{uv} \ge \max(y_u - y_v, 0) \qquad ((u,v) \in E).$$

(c) Let (y,z) be feasible for the dual problem, and let f be any feasible flow. We begin with the identity

$$\sum_{e=(u,v) \in E} f_{uv}(y_u - y_v) = \sum_{v \in V} y_v \Big(\sum_{e \in E^+(v)} f_e - \sum_{e \in E^-(v)} f_e \Big).$$

For $v \notin \{s,t\}$, the term in parentheses vanishes by flow conservation. At $v = s$ it equals $\mathrm{val}(f)$, and at $v = t$ it equals $-\mathrm{val}(f)$. Hence

$$\sum_{e=(u,v) \in E} f_{uv}(y_u - y_v) = (y_s - y_t)\, \mathrm{val}(f) = \mathrm{val}(f).$$

Since $z_{uv} \ge y_u - y_v$ and $f_{uv} \ge 0$,

$$\mathrm{val}(f) \le \sum_{e=(u,v) \in E} f_{uv} z_{uv}.$$

Using $f_{uv} \le c_{uv}$ and $z_{uv} \ge 0$, we get

$$\mathrm{val}(f) \le \sum_{e=(u,v) \in E} c_{uv} z_{uv}.$$

(d) For the proposed y, we have $y_s - y_t = 1$. If $u \in S$ and $v \notin S$, then $y_u - y_v = 1$ and $z_{uv} = 1$. In every other case, $y_u - y_v \le 0$ and $z_{uv} = 0$. Thus $z_{uv} \ge y_u - y_v$ for every edge, and clearly $z_{uv} \ge 0$. Hence (y,z) is feasible.

Its objective value is

$$\sum_{e=(u,v) \in E} c_{uv} z_{uv} = \sum_{\substack{e=(u,v) \in E \\ u \in S,\ v \notin S}} c_{uv} = \mathrm{cap}(S).$$

(e) Replacing y by $y - y_t\mathbf{1}$ does not change any difference $y_u - y_v$, so feasibility is preserved. After this shift we indeed have $y_t = 0$, and then the constraint $y_s - y_t = 1$ gives $y_s = 1$.

Now let $\alpha \in [0,1)$. Since $y_s = 1 > \alpha$, we have $s \in S_\alpha$. Since $y_t = 0 \le \alpha$, we have $t \notin S_\alpha$. Therefore S_α is an s-t cut.

(f) For an edge $e = (u,v)$,

$$\mathbf{1}_{\{u\in S_\alpha,\ v\notin S_\alpha\}} = \mathbf{1}_{\{y_u > \alpha \ge y_v\}}.$$

Therefore

$$\int_0^1 \mathbf{1}_{\{u\in S_\alpha,\ v\notin S_\alpha\}}\, d\alpha = \big|[0,1) \cap [y_v, y_u)\big| \le (y_u - y_v)_+.$$

Since $z_{uv} \ge 0$ and $z_{uv} \ge y_u - y_v$, we have

$$(y_u - y_v)_+ \le z_{uv}.$$

This proves the first claim.

Now sum over all edges with weights $c_{uv} \ge 0$:

$$\int_0^1 \operatorname{cap}(S_\alpha)\, d\alpha = \sum_{e=(u,v)\in E} c_{uv} \int_0^1 \mathbf{1}_{\{u\in S_\alpha,\ v\notin S_\alpha\}}\, d\alpha \le \sum_{e=(u,v)\in E} c_{uv} z_{uv}.$$

The left-hand side is the average of $\operatorname{cap}(S_\alpha)$ over $\alpha \in [0,1)$. Therefore there exists some $\alpha \in [0,1)$ such that

$$\operatorname{cap}(S_\alpha) \le \sum_{e=(u,v)\in E} c_{uv} z_{uv}.$$

Thus every feasible (y,z) dominates the capacity of some cut. It follows that

$$\min\{\operatorname{cap}(S) : S \text{ is an } s\text{-}t \text{ cut}\} \le \inf\left\{ \sum_{e=(u,v)\in E} c_{uv} z_{uv} : (y,z) \text{ feasible} \right\}.$$

On the other hand, part (b) shows that every cut produces a feasible (y,z) with the same objective value, so

$$\inf\left\{ \sum_{e=(u,v)\in E} c_{uv} z_{uv} : (y,z) \text{ feasible} \right\} \le \min\{\operatorname{cap}(S) : S \text{ is an } s\text{-}t \text{ cut}\}.$$

Hence the optimum value of the second optimization problem is exactly the minimum cut capacity.

Let (y,z) be an optimal feasible pair for the second optimization problem. By part (e), there exists an s-t cut S such that

$$\operatorname{cap}(S) \le \sum_{e=(u,v)\in E} c_{uv} z_{uv}.$$

Since (y,z) is optimal and part (b) constructs a feasible pair $(\widetilde{y}, \widetilde{z})$ from any cut S with objective value exactly $\operatorname{cap}(S)$, we must in fact have equality throughout:

$$\operatorname{cap}(S) = \sum_{e=(u,v)\in E} c_{uv} z_{uv}.$$

Now part (b) gives an optimal feasible pair $(\widetilde{y}, \widetilde{z})$ with

$$\widetilde{y}_v \in \{0, 1\} \qquad (v \in V),$$

and

$$\widetilde{z}_{uv} \in \{0, 1\} \qquad ((u, v) \in E).$$

So the second optimization problem admits an optimal 0/1 solution.

(g) By part (c), every feasible dual point (y, z) gives an upper bound on the value of every flow, so

$$\max \{\mathrm{val}(f) : f \text{ is a flow}\} \leq \inf \left\{ \sum_{e=(u,v)\in E} c_{uv} z_{uv} : (y, z) \text{ feasible} \right\}.$$

Since the second optimization problem is the LP dual of the max-flow LP from part (b), LP duality implies equality between these two optimal values. Part (e) identified the dual optimum with the minimum cut capacity. Therefore

$$\max \{\mathrm{val}(f) : f \text{ is a flow}\} = \min \{\mathrm{cap}(S) : S \text{ is an } s\text{-}t \text{ cut}\}.$$

This is the max-flow min-cut theorem.