Homework #1 - Solution

Due Feb 15 (Wednesday) at the beginning of class
Part of solutions are provided by Xiaobo Dong.

Problem 1: Convex Sets

A real-valued function on $\mathbf{R}^n$ is called a *norm*, denoted as $\|\cdot\|$, if it satisfies the three properties:

- (*positivity*): $\forall x \in \mathbf{R}^n, \|x\| \geq 0$; $\|x\| = 0$ if an only if $x = 0$;
- (*homogeneity*): $\forall x \in \mathbf{R}^n, \alpha \in \mathbf{R}$: $\|\alpha x\| = |\alpha| \cdot \|x\|$;
- (*triangle inequality*): $\forall x, y \in \mathbf{R}^n : \|x + y\| \leq \|x\| + \|y\|$.

The standard norms on $\mathbf{R}^n$ are the ℓ_p-norms ($p \geq 1$):

$$\|x\|_p := \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

- [Euclidean norm]: $p = 2$, $\|x\|_2 := \sqrt{\sum_{i=1}^n x_i^2}$;
- [Manhattan norm]: $p = 1$, $\|x\|_1 := \sum_{i=1}^n |x_i|$;
- [Maximum norm]: $p = \infty$, $\|x\|_\infty := \max_{1 \leq i \leq n} |x_i|$.

Exercise 1.1 (Dual Norm) The *dual norm* of $\|\cdot\|$ on $\mathbf{R}^n$ is defined as

$$\|x\|_* = \max_{y \in \mathbf{R}^n : \|y\| \leq 1} x^T y$$

(a) Prove that $\|\cdot\|_*$ is a valid norm.

(b) Denote $\|x\|_{p,*}$ as the dual norm to the ℓ_p-norm ($p \geq 1$). Show that

$$\|x\|_{p,*} = \|x\|_q, \text{ where } \frac{1}{p} + \frac{1}{q} = 1.$$

Hence, $\|\cdot\|_{2,*} = \|\cdot\|_2$, $\|\cdot\|_{1,*} = \|\cdot\|_\infty$, $\|\cdot\|_{\infty,*} = \|\cdot\|_1$.

Solution

(a) The positivity property is true, because we know

$$||x||_* = \max_{y \in R^n: ||y|| \leq 1} x^T y \geq x^T 0 = 0$$

The homogeneity property is true, because we know

$$||\alpha x||_* = \max_{y \in R^n: ||y|| \leq 1} \alpha x^T y = |\alpha| \cdot \max_{y \in R^n: ||y|| \leq 1} x^T y = |\alpha| \cdot ||x||_*$$

The triangle inequality property is true, because we know

$$||x+z||_* = \max_{y \in R^n: ||y|| \leq 1} (x+z)^T y = \max_{y \in R^n: ||y|| \leq 1} (x^T y + z^T y)$$

Let y^* be the optimal solution of the above. Then, we have

$$||x+z||_* = x^T y^* + z^T y^* \leq \max_{y \in R^n: ||y|| \leq 1} x^T y + \max_{y \in R^n: ||y|| \leq 1} z^T y = ||x||_* + ||z||_*$$

(b) For $1 < p, q < \infty$, we have

$$||x||_{p,*} = \max_{y \in R^n: ||y||_p \leq 1} x^T y$$

Now consider the following with Hölder's inequality

$$x^T y \leq \sum_{i=1}^{n} |x_i y_i| \leq ||y||_p \cdot ||x||_q \leq ||x||_q$$

Now the remaining part is to find a y such that the equality can hold. Let $z_i = \text{sign}(x_i)|x_i|^{q-1}$ for all $i \in [1, n]$. We calculate

$$\sum_{i=1}^{n} x_i z_i = \sum_{i=1}^{n} x_i \text{sign}(x_i)|x_i|^{q-1} = \sum_{i=1}^{n} |x_i|^q = ||x||_q^q$$

Further, we calculate

$$||z||_p^p = \sum_{i=1}^{n} |z_i|^p = \sum_{i=1}^{n} \left| \text{sign}(x_i)|x_i|^{q-1} \right|^p = \sum_{i=1}^{n} \left|x_i\right|^{(q-1)p} = \sum_{i=1}^{n} \left|x_i\right|^q = ||x||_q^q$$

Now constuct $y = \frac{z}{||z||_p}$, and we have $||y||_p = 1$. Then we have

$$\sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} x_i \frac{z_i}{||z||_p} = \frac{1}{||z||_p} \sum_{i=1}^{n} x_i z_i = \frac{1}{||x||_q^{q/p}} ||x||_q^q = ||x||_q$$

Therefore,

$$||x||_{p,*} = ||x||_q$$

For $p = 1$ and $q = \infty$ or $p = \infty$ and $q = 1$, they are the trivial cases. The proof is similar but make 1 and ∞ be the pairs accordingly.

Exercise 1.2 (Unit Ball) The *unit ball* of any norm $\|\cdot\|$ is the set

$$B_{\|\cdot\|} = \{x \in \mathbf{R}^n : \|x\| \leq 1\}.$$

One can easily see that $B_{\|\cdot\|}$ is symmetric w.r.t. the origin ($x \in B_{\|\cdot\|}$ if and only $-x \in B_{\|\cdot\|}$) and compact (closed and bounded) with nonempty interior. Show that the set $B_{\|\cdot\|}$ is convex.

Solution For any $x, y \in B_{||\cdot||}$, and for $\lambda \in [0,1]$. Then we have

$$||\lambda x + (1-\lambda)y|| \leq ||\lambda x|| + ||(1-\lambda)y|| \leq \lambda||x|| + (1-\lambda)||y|| \leq \lambda + (1-\lambda) = 1$$

Therefore, $\lambda x + (1-\lambda)y \in B_{||\cdot||}$. Hence, $B_{||\cdot||}$ is a convex set.

Exercise 1.3 (Vice Versa) Let B be convex, symmetric w.r.t. the origin, and compact with nonempty interior. Show that the following function $\|\cdot\|_B$:

$$\|x\|_B = \inf\{t > 0 : \frac{x}{t} \in B\}$$

is a valid norm. Moreover, its unit ball is exactly the set B.

Solution The positivity property is true, because we know

$$||x||_B = \inf\{t > 0 : \frac{x}{t} \in B\} \geq 0$$

If$||x||_B = 0$, then $x \in tB, \forall t > 0$, so $x = 0$.

The homogeneity property is true, because we know

$$||\alpha x||_B = \inf\{t > 0 : \frac{\alpha x}{t} \in B\} = \inf\{\alpha t' > 0 : \frac{\alpha x}{\alpha t'} \in B\} = \alpha \inf\{t' > 0 : \frac{x}{t'} \in B\} = \alpha||x||_B$$

To show the triangle inequality property is true, it is suffient to show that

$$\frac{x+y}{||x||_B + ||y||_B} \in B$$

This is true because by definition of $||x||_B$ and $||y||_B$, $\frac{x}{||x||_B} \in B$ and $\frac{y}{||y||_B} \in B$. Since B is a convex set,

$$\frac{x+y}{||x||_B + ||y||_B} = \frac{||x||_B}{||x||_B + ||y||_B} \cdot \frac{x}{||x||_B} + \frac{||y||_B}{||x||_B + ||y||_B} \cdot \frac{y}{||y||_B} \in B$$

Moreover,

$$\begin{aligned} B_{||\cdot||_B} &= \{x \in R^n : ||x||_B \leq 1\} \\ &= \{x \in R^n : \inf\{t > 0 : \frac{x}{t} \in B\} \leq 1\} \\ &= \{x \in R^n : x \in B\} \end{aligned}$$

The last step come from the fact that if $x \notin B$, then $\inf\{t > 0 : \frac{x}{t} \in B\} > 1$, which is a contradiction.

Exercise 1.4 (Neighborhood of Convex Sets) Let $X \subset \mathbf{R}^n$ be a convex set and let $\epsilon > 0$. The ϵ-*neighborhood* of the set X under norm $\|\cdot\|$ is defined as

$$X^\epsilon = \{x \in \mathbf{R}^n : \inf_{y \in X} \|x - y\| \leq \epsilon\}.$$

Prove that X^ϵ is a convex set.

Solution Let x and y be two arbitary elements in X^ϵ. Then we have for any $\epsilon' > \epsilon$,

$$\exists u, v \in X, \text{s.t. } \|x - u\| \leq \epsilon', \|y - v\| \leq \epsilon'.$$

For any $\lambda \in [0, 1]$, $\exists \lambda u + (1 - \lambda)v \in X$, such that

$$\|[\lambda x + (1 - \lambda)y] - [\lambda u + (1 - \lambda)v]\| \leq \lambda\|x - u\| + (1 - \lambda)\|y - b\| \leq \epsilon'.$$

This implies that $\lambda x + (1 - \lambda)y \in X^\epsilon$.

Problem 2: Strong Duality

Recall that in class we have shown the

[Farkas' Lemma] Let $A \in \mathbf{R}^{n\times m}$ and $b \in \mathbf{R}^m$, exactly one of the two sets must be empty:

$$\begin{aligned} (i): &\quad \{x \in \mathbf{R}^n : Ax = b, x \geq 0\} \\ (ii): &\quad \{y \in \mathbf{R}^m : A^T y \leq 0, b^T y > 0\} \end{aligned}$$

Exercise 2.1 (Variant of Farkas' Lemma) Prove the following variant of Farkas' Lemma: exactly one of the following sets must be empty:

$$\begin{aligned} (i): &\quad \{x \in \mathbf{R}^n : Ax \leq b\} \\ (ii): &\quad \{y \in \mathbf{R}^m : y \geq 0, A^T y = 0, b^T y < 0\} \end{aligned}$$

Solution We first show that system (i) is feasible $\rightarrow$ *system* (ii) infeasible Suppose there is a $x \in R^n$ such that $Ax \leq b$, then for any $y \geq 0$, we have

$$x^T(A^T y) = (Ax)^T y \leq b^T y$$

If system (ii) is feasible, then there exists $y \geq 0, A^T y = 0, b^T y < 0$, which implies

$$0 = x^T(Ay) \leq b^T y < 0$$

leading to a contradiction. Therefore, (ii) is infeasible.

We now show that system (i) infeasible $\rightarrow$ system (ii) feasible We can rewrite $Ax \leq b$ as $A(x^+ - x^-) + s = b$, where $(x^+, x^-, s) \geq 0$. Hence,

$$\{x : Ax \leq b\} \Leftrightarrow \left\{ \begin{bmatrix} x^+ \\ x^- \\ s \end{bmatrix} : \begin{bmatrix} A & -A & I \end{bmatrix} \begin{bmatrix} x^+ \\ x^- \\ s \end{bmatrix} = b, \; \begin{bmatrix} x^+ \\ x^- \\ s \end{bmatrix} \geq 0 \right\}$$

By Farkas' Lemma, the infeasibility of this system implies that there exists y such that

$$b^T y < 0, \; \begin{bmatrix} A & -A & I \end{bmatrix}^T y \geq 0$$

This implies that $y \geq 0, A^T y = 0, b^T y < 0$. Namely, system (ii) is feasible.

Overall, we know exactly one of them is empty.

Exercise 2.2 (Duality of Linear Program) The above lemma can be used to derive strong duality of linear programs. Consider the primal and dual pair of linear programs:

$$\min_x \; c^T x \quad \text{s.t.} \;\; Ax = b, x \geq 0 \tag{P}$$

$$\max_y \; b^T y \quad \text{s.t.} \;\; A^T y \leq c \tag{D}$$

Assume that the problem (P) has an optimal solution x_* with finite optimal value p_*.

(a) Let y be a feasible solution to (D), show that $b^T y \leq p_*$.

(b) Apply the above variant Farkas' lemma to show that the following system

$$\{y : A^T y \leq c, y^T b \geq p_*\}$$

is nonempty, namely, there exists y that satisfies

$$\begin{bmatrix} A^T \\ -b^T \end{bmatrix} y \leq \begin{bmatrix} c \\ -p_* \end{bmatrix}.$$

Solution (a) Let y be a feasible solution to (D), then $A^T y \leq c$. Hence,

$$b^T y = (Ax_*)^T y = x_*^T (A^T y) \leq x_*^T c = p_*.$$

(b) Suppose the above system is not feasible. By the variant of Farkas' Lemma, $\exists \tilde{\lambda} = \begin{bmatrix} \lambda \\ \lambda_0 \end{bmatrix} \geq 0$, such that

$$\begin{cases} A\lambda - \lambda_0 b = 0 \\ \lambda^T c - \lambda_0 p^* < 0 \end{cases}$$

If $\lambda_0 > 0$, then $A(\frac{\lambda}{\lambda_0}) = b$ and $c^T(\frac{\lambda}{\lambda_0}) < p_*$, contradicts with p_* being optimal
If $\lambda_0 = 0$, then $A\lambda = 0, \lambda^T c < 0$. then we can construct a feasible solution $x = x_* + \lambda$ that leads to small objective,

$$c^T x = c^T(x_* + \lambda) < c^T x_* = p_*$$

which contradicts with the optimality of x_*. Therefore, the assumption is not true, i.e. above system must be feasible.

Combining (a) and (b), we can see that if problem (P) is solvable, then the dual problem (D) is also solvable, and the optimal values are the same.

Problem 3: Convex Functions

A function f is called *log-convex* if $f(x) > 0, \forall x \in \mathbf{dom}(f)$ and $\ln(f(x))$ is convex.

Exercise 3.1 (Basic Properties) Show that

(a) If f is log-convex, then f is also convex.

(b) f is log-convex if and only if $\forall \lambda \in [0, 1], \forall x, y \in \mathbf{dom}(f)$, we have $f(\lambda x + (1 - \lambda)y) \leq f(x)^\lambda f(y)^{1-\lambda}$.

Solution (a) Let $g(x) = \ln(f(x))$. Since f is log-convex, then $g(x)$ is convex. Note that $f(x) = e^{g(x)}$ is the composition of exponential function (convex and monotonically increasing) and a convex function, then f is also convex.

(b) This is because

$$\begin{aligned}
f(x) \text{ is log-convex } &\Leftrightarrow \ln(f(x)) \text{ is convex} \\
&\Leftrightarrow \forall x, y \in \mathbf{dom}(f), \forall \lambda \in [0,1],\ \ln\big(f(\lambda x + (1-\lambda)y)\big) \leq \lambda \ln\big(f(x)\big) + (1-\lambda)\ln\big(f(y)\big) \\
&\Leftrightarrow \forall x, y \in \mathbf{dom}(f), \forall \lambda \in [0,1],\ e^{\ln\big(f(\lambda x + (1-\lambda)y)\big)} \leq e^{\lambda \ln\big(f(x)\big) + (1-\lambda)\ln\big(f(y)\big)} \\
&\Leftrightarrow \forall x, y \in \mathbf{dom}(f), \forall \lambda \in [0,1],\ f(\lambda x + (1-\lambda)y) \leq f(x)^\lambda f(y)^{1-\lambda}
\end{aligned}$$

Exercise 3.2 (Examples) Show that the following functions are log-convex.

(a) Exponentials: $f(x) = e^{ax}$ is log-convex on $\mathbf{R}$;

(b) Sum of exponentials: $f(u, v) = e^u + e^v$ is log-convex on $\mathbf{R}^2$.

Solution (a) First of all, we have $f(x) > 0$. Moreover, we have

$$\ln\big(f(x)\big) = ax$$

is a linear function and is a convex function. Thus, f is log-convex.

(b) First of all, we have $f(u, v) > 0$. It suffices to show that the function $g(u, v) = \ln\big(f(u, v)\big) = \ln\big(e^u + e^v\big)$ is convex. The Hessian of function $g(u, v)$ is given by

$$\nabla^2 g(u, v) = \begin{bmatrix} \frac{e^u e^v}{(e^u+e^v)^2} & -\frac{e^u e^v}{(e^u+e^v)^2} \\ -\frac{e^u e^v}{(e^u+e^v)^2} & \frac{e^u e^v}{(e^u+e^v)^2} \end{bmatrix}$$

Since $\nabla^2 g(u, v) \succeq 0$, $g(u, v)$ is convex.

Exercise 3.3 (Convexity-Preserving Operation) If f and g are log-convex, then $f + g$ is also log-convex.

Solution Let $u(x) = \ln(f(x))$ and $v(x) = \ln(g(x))$. Since f and g are log-convex, we know that $u(x)$ and $v(x)$ are convex. We already show that the function $h(u, v) = \ln(e^u + e^v)$ is convex in (u, v). It is easy to show that this function is also coordinate-wise non-decreasing since the gradient is always positive. Hence, the composition

$$h(x) = \ln(e^{u(x)} + e^{v(x)})$$

is also convex. By definition of $u(x)$ and $v(x)$, we have $h(x) = \ln(f(x) + g(x))$. Therefore, $f + g$ is log-convex.

Problem 4: Convex Sets and Convex Functions

Denote $\mathcal{S}^n$ as the set of real symmetric matrices of size $n \times n$. Denote $\mathcal{S}^n_+$ as the set of positive semi-definite matrices in $\mathcal{S}^n$. Denote $\mathcal{S}^n_{++}$ as the set of positive definite matrices in $\mathcal{S}^n$. The inner product between two matrices in $\mathcal{S}^n$ is defined as $\langle X, Y \rangle = \text{Trace}(X^T Y) = \text{Trace}(XY) = \sum_{k,l=1}^n x_{kl} y_{kl}$.

Exercise 4.1 (Positive Semidefinite Cone) Recall that set C is a *convex cone* if $\forall x, y \in C$, $\lambda_1 x + \lambda_2 y \in C$ for any $\lambda_1, \lambda_2 \geq 0$. Prove that the set $\mathcal{S}^n_+$ is a convex cone.

Solution Let matrix A and B be in S^n_+. Then for any $x \in R^n$, we have $x^T Ax \geq 0, x^T Bx \geq 0$. Then for any $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$, we have

$$x^T(\lambda_1 A + \lambda_2 B)x = \lambda_1 x^T Ax + \lambda_2 x^T Bx \geq 0$$

Therefore, S^n_+ is convex cone.

Exercise 4.2 (Linear Matrix Inequalities) Let $A_1, \ldots, A_k, B \in \mathcal{S}^n$ be symmetric matrices. Show that the following set

$$C = \{x \in \mathbf{R}^k : B - \sum_{i=1}^{k} x_i A_i \in \mathcal{S}_+^n\}$$

is convex.

Solution C can be considered as the inverse affine image of the set $\mathbf{S}^n$ under the affine mapping

$$\mathcal{A}(x) : x \mapsto B - \sum_{i=1}^{k} x_i A_i$$

i.e. $C = \mathcal{A}^{-1}(\mathbf{S}^n)$. Since $\mathbf{S}^n$ is convex, C is also convex.

Exercise 4.3 (Log-Determinant) Follow the steps below to show that the negative log-determinant function $f(X) = -\ln(\det(X)) : \mathcal{S}_{++}^n \to \mathbf{R}$ is convex. Consequently, $\det(X)^{-1}$ is log-convex.

(a) Define $\phi(t) = -\ln(\det(X + tH))$. Verify that $\phi(t)$ is well defined on an open interval $I(X, H)$ that contains the origin.

(b) Prove that $\phi(t) = \phi(0) - \sum_{i=1}^{n} \log(1 + t\lambda_i)$, where $\lambda_1, \ldots, \lambda_n$ are eigenvalues of $X^{-1/2} H X^{-1/2}$.

(c) Combine (a) and (b) to show that $f(X)$ is convex on $\mathcal{S}_{++}^n$.

Solution (a) First of all, we know when $t = 0$, $\phi(t) = -\ln(\det(X))$ is well defined. Moreover, we know in order to make $\phi(t)$ be well defined, we need $det(X + tH)$ to be positive. Since $det(X + tH)$ is a continuous function in terms of t, then we know that an open interval which is related to H and X including 0 have $det(X + tH) > 0$.

(b)

$$\begin{aligned}
\phi(t) &= -\ln\Big(\det\big(X + tH\big)\Big) \\
&= -\ln\Big(\det\big(X^{\frac{1}{2}}(I + tX^{-\frac{1}{2}} H X^{-\frac{1}{2}}) X^{\frac{1}{2}}\big)\Big) \\
&= -\ln\Big(\det\big(X^{\frac{1}{2}}\big)\det\big(I + tX^{-\frac{1}{2}} H X^{-\frac{1}{2}}\big)\det\big(X^{\frac{1}{2}}\big)\Big) \\
&= -\ln\Big(\det\big(I + tX^{-\frac{1}{2}} H X^{-\frac{1}{2}}\big)\det\big(X\big)\Big) \\
&= -\ln\Big(\prod_{i=1}^{n}(1 + t\lambda_i)\Big) - \ln(\det(X)) \\
&= -\sum_{i=1}^{n} \ln(1 + t\lambda_i) + \phi(0)
\end{aligned}$$

(c) Since

$$\phi'(t) = -\sum_{i=1}^{n} \frac{\lambda_i}{1+t\lambda_i}$$
$$\phi''(t) = \sum_{i=1}^{n} \frac{\lambda_i^2}{(1+t\lambda_i)^2} \geq 0$$

Therefore, $\phi(t)$ is a convex function for any X, H. Hence, $f(x)$ is convex.

Problem 5: Convex Optimization

Exercise 5.1 (Convex Reformulation) Show that the following optimization problem

$$\begin{aligned} \min_x \quad & f(x) := \max_{k=1,\dots,n} |\log(a_k^T x) - \log(b_k)| \\ \text{s.t.} \quad & 0 \leq x_i \leq 1, i = 1, \dots, m \end{aligned}$$

where $a_k \in \mathbf{R}^m, b_k \in \mathbf{R}, k = 1, \dots, n$ are given, is equivalent to the following optimization problem

$$\begin{aligned} \min_x \quad & \max_{k=1,\dots,n} h(a_k^T x / b_k) \\ \text{s.t.} \quad & 0 \leq x_i \leq 1, i = 1, \dots, m \end{aligned}$$

where $h(u) = \max(u, 1/u)$ for $u > 0$. And show that the above problem is a convex program.

Solution

$$\begin{aligned} \max_{k=1,\dots,m} \big|\log(a_k^T x) - \log(b_k)\big| &= \max_{k=1,\dots,m} \big|\log(\frac{a_k^T x}{b_k})\big| \\ &= \max_{k=1,\dots,m} \max\big\{\log\big(\frac{a_k^T x}{b_k}\big), \log\big(\frac{b_k}{a_k^T x}\big)\big\} \\ &= \max_{k=1,\dots,m} \log\big(\max\big\{\frac{a_k^T x}{b_k}, \frac{b_k}{a_k^T x}\big\}\big) \\ &= \max_{k=1,\dots,m} \log h\Big(\big(\frac{a_k^T x}{b_k}\big)\Big) \end{aligned}$$

Since the log function is monotonic,

$$\arg\min_{x:0\leq x\leq 1} \max_{k=1,\dots,m} \big|\log(a_k^T x) - \log(b_k)\big| = \arg\min_{x:0\leq x\leq 1} \max_{k=1,\dots,m} \log h\big(\frac{a_k^T x}{b_k}\big) = \arg\min_{x:0\leq x\leq 1} \max_{k=1,\dots,m} h\big(\frac{a_k^T x}{b_k}\big)$$

Therefore, they are equivalent.

Next, since the constraint is obviously a convex set, to show the above program is a convex program, we need to show the function $\max_{k=1,\ldots,m} h\big(\frac{a_k^T x}{b_k}\big)$ is convex.

Since $h(u) = \max(u, \frac{1}{u})$ is convex on $u > 0$ (both u, $1/u$ are convex on $u > 0$). Therefore, $h\big(\frac{a_k^T x}{b_k}\big)$ is a convex function. And we know taking pointwise maximum preserves convexity; therefore, $\max_{k=1,\ldots,m} h\big(\frac{a_k^T x}{b_k}\big)$ remains convex.