Introduction to Machine Learning

Fall 2023

University of Science and Technology of China

Lecturer: Jie Wang — Homework 3

Posted: Oct. 19, 2023 — Due: Nov. 2, 2023

Notice, to get the full credits, please present your solutions step by step.

Exercise 1: Convex Sets

Let $C \subset \mathbb{R}^n$ be a nonempty convex set. Please show the following statements.

1. Please find the interior and relative interior of the following convex sets (you don't need to prove them).
   (a) $\{\mathbf{x} \in \mathbb{R}^3 : x_1^2 + x_2^2 < 1, x_3 = 0\} \subset \mathbb{R}^3$.
   (b) $\{\mathbf{A} \in S^n_{++} : \mathrm{Tr}(\mathbf{A}) = 1\} \subset \mathbb{R}^{n\times n}$.
   (c) $\{\mathbf{A} \in S^n_{++} : \mathrm{Tr}(\mathbf{A}) = 1\} \subset S^n$.
   (d) (Optional) $\{\mathbf{A} \in S^n_{++} : \mathrm{Tr}(\mathbf{A}) \leq 1\} \subset \mathbb{R}^{n\times n}$.
2. Some operations that preserve convexity.
   (a) Both **cl** C and **int** C are convex.
   (b) The set **relint** C is convex.
   (c) The intersection $\bigcap_{i\in I} C_i$ of any collection $\{C_i : i \in \mathcal{I}\}$ of convex sets is convex.
   (d) The set $\{\mathbf{y} \in \mathbb{R}^m : \mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{a}, \mathbf{x} \in C\}$ is convex, where $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{a} \in \mathbb{R}^m$.
   (e) The set $\{\mathbf{y} \in \mathbb{R}^m : \mathbf{x} = \mathbf{B}\mathbf{y} + \mathbf{b}, \mathbf{x} \in C\}$ is convex, where $\mathbf{B} \in \mathbb{R}^{n\times m}$ and $\mathbf{b} \in \mathbb{R}^n$.

Solution: 1. We denote the following sets as C.

   (a) **int** $C = \emptyset$, and **relint** $C = C = \{\mathbf{x} \in \mathbb{R}^3 : x_1^2 + x_2^2 < 1, x_3 = 0\}$
   (b) **int** $C = \emptyset$, and **relint** $C = C = \{\mathbf{A} \in S^n_{++} : \mathrm{Tr}(\mathbf{A}) = 1\} \subset \mathbb{R}^{n\times n}$
   (c) **int** $C = \emptyset$, and **relint** $C = C = \{\mathbf{A} \in S^n_{++} : \mathrm{Tr}(\mathbf{A}) = 1\} \subset S^n$
   (d) **int** $C = C$, and **relint** $C = C$

2. (a) $\forall \mathbf{x}, \mathbf{y} \in$ **cl** C, there exists sequence $\{\mathbf{x}_k\} \subset C$ and sequence $\{\mathbf{y}_k\} \subset C$ such that $\mathbf{x}_k \to \mathbf{x} \in$ **cl** C and $\mathbf{y}_k \to \mathbf{y} \in$ **cl** C.
   Because C is a convex set,

$$\forall \theta \in [0, 1], \theta\mathbf{x}_k + (1-\theta)\mathbf{y}_k = \mathbf{z}_k \in C$$

   It means we can obtain a new sequence $\{\mathbf{z_k}\} \subset C$. We denote the limit point of $\{\mathbf{z}_k\}$ as $\mathbf{z}$. It is obvious that $\mathbf{z} = \theta\mathbf{x} + (1-\theta)\mathbf{y} \in$ **cl** C. It means **cl** C is a convex set.
   Next we show **int** C is convex. $\forall \mathbf{x}, \mathbf{y} \in \int C$, $\exists \epsilon > 0$ such that $B_\epsilon(\mathbf{x}) \subseteq C$. C is a convex set, thus

$$\forall \hat{\mathbf{x}} \in B_\epsilon(\mathbf{x}), \forall \theta \in [0, 1], \theta\hat{\mathbf{x}} + (1-\theta)\mathbf{y} \in C$$

It means $A = \theta B_\epsilon(\mathbf{x}) + (1-\theta)\mathbf{y} \subseteq C$. We know $B_\epsilon(\mathbf{x}) = \{\mathbf{z} \in \mathbb{R}^n | \|\mathbf{z} - \mathbf{x}\| < \epsilon\}$, thus $A = \{\mathbf{z} \in \mathbb{R}^n | \|\mathbf{z} - \frac{1}{\theta}(\mathbf{x} - (1-\theta)\mathbf{y})\| < \frac{\epsilon}{\theta}\}$.
Let $\mathbf{z}_0 = \theta\mathbf{x} + (1-\theta)\mathbf{y} \in A$, $r_0 = d(\mathbf{z}_0, \frac{1}{\theta}(\mathbf{x} - (1-\theta)\mathbf{y}))$ and $r = \min(r_0, \frac{\epsilon}{\theta} - r_0)$. Thus, $B_r(\mathbf{z}_0) \subseteq A \subseteq C$. It means $\mathbf{z}_0 = \theta\mathbf{x} + (1-\theta)\mathbf{y}$ is an interior point of C, or we can say $\mathbf{z}_0 \in \textbf{int } C$. Therefore, $\textbf{int } C$ is a convex set.

(b) $\forall \mathbf{x}, \mathbf{y} \in \textbf{relint } C$, $\exists \epsilon_1, \epsilon_2 > 0$ such that

$$B_{\epsilon_1}(\mathbf{x}) \bigcap \textbf{aff} C \subseteq C$$
$$B_{\epsilon_2}(\mathbf{y}) \bigcap \textbf{aff} C \subseteq C$$

Thus there must be $\exists \mathbf{r_1}, \mathbf{r_2} \in \mathbb{R}^n$, $\|\mathbf{r}_1\| < \epsilon_1$, $\|\mathbf{r}_2\| < \epsilon_2$ such that

$$\mathbf{x} + \mathbf{r}_1 \in \textbf{aff} C$$
$$\mathbf{y} + \mathbf{r}_2 \in \textbf{aff} C$$

And

$$\mathbf{x} + \mathbf{r}_1 \in C$$
$$\mathbf{y} + \mathbf{r}_2 \in C$$

Thus we can get $\forall \theta \in [0,1]$, $\theta(\mathbf{x} + \mathbf{r_1}) + (1-\theta)(\mathbf{y} + \mathbf{r_2}) = \theta\mathbf{x} + (1-\theta)\mathbf{y} + \theta\mathbf{r}_1 + (1-\theta)\mathbf{r}_2 \in \textbf{aff} C$(The definition of affine hull).
And because C is convex, we can obtain$\forall \theta \in [0,1]$, $\theta(\mathbf{x} + \mathbf{r_1}) + (1-\theta)(\mathbf{y} + \mathbf{r_2}) = \theta\mathbf{x} + (1-\theta)\mathbf{y} + \theta\mathbf{r}_1 + (1-\theta)\mathbf{r}_2 \in C$. This means that

$$\forall \theta \in [0,1], \theta\mathbf{x} + (1-\theta)\mathbf{y} + \theta\mathbf{r}_1 + (1-\theta)\mathbf{r}_2 \in C$$
$$\theta\mathbf{x} + (1-\theta)\mathbf{y} + \theta\mathbf{r}_1 + (1-\theta)\mathbf{r}_2 \in \textbf{aff} C$$

Besides, $\exists \epsilon > \|\theta\mathbf{r}_1 + (1-\theta)\mathbf{r}_2\|$ such that

$$\theta\mathbf{x} + (1-\theta)\mathbf{y} + \theta\mathbf{r}_1 + (1-\theta)\mathbf{r}_2 \in B_\epsilon(\theta\mathbf{x} + (1-\theta)\mathbf{y})$$

In other words,

$$B_\epsilon(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \bigcap \textbf{aff} C \subseteq C$$

$\theta\mathbf{x} + (1-\theta)\mathbf{y} \in \textbf{relint } C$. Therefore $\textbf{relint } C$ is a convex set.

(c) $\forall \mathbf{x}_1, \mathbf{x}_2 \in \bigcap_{i \in I} C_i$, $\forall i \in \mathcal{I}$, we have $\mathbf{x}_1, \mathbf{x}_2 \in C_i$. C_i is a convex set, so we have $\forall \theta \in [0,1]$,

$$\theta\mathbf{x_1} + (1-\theta)\mathbf{x_2} \in C_i$$

This equation is right for all C_i, thus $\theta\mathbf{x_1} + (1-\theta)\mathbf{x_2} \in \bigcap_{i \in I} C_i$. It tells us that $\bigcap_{i \in I} C_i$ is a convex set.

(d) We denote this set by D. $\forall \mathbf{y}_1, \mathbf{y}_2 \in D$, $\exists \mathbf{x_1}, \mathbf{x_1} \in C$ such that $\mathbf{y_1} = \mathbf{A}\mathbf{x_1} + \mathbf{a}$ and $\mathbf{y_2} = \mathbf{A}\mathbf{x_2} + \mathbf{a}$. $\forall \theta \in [0, 1]$,

$$\theta \mathbf{y}_1 + (1-\theta)\mathbf{y}_2 = \mathbf{A}(\theta \mathbf{x}_1 + (1-\theta \mathbf{x}_2)) + \mathbf{a}$$

C is convex, so $\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in C$. Thus $\theta \mathbf{y}_1 + (1-\theta)\mathbf{y}_2 \in D$. It means D is a convex set.

(e) We denote this set by D. $\forall \mathbf{y}_1, \mathbf{y}_2 \in D$, $\exists \mathbf{x_1}, \mathbf{x_1} \in C$ such that $\mathbf{x_1} = \mathbf{B}\mathbf{y_1} + \mathbf{b}$ and $\mathbf{x_2} = \mathbf{B}\mathbf{y_2} + \mathbf{b}$. $\forall \theta \in [0, 1]$,

$$\theta \mathbf{x_1} + (1-\theta)\mathbf{x_2} = \mathbf{B}(\theta \mathbf{y}_1 + (1-\theta \mathbf{y}_2)) + \mathbf{b}$$

C is convex, so $\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in C$. Thus $\theta \mathbf{y}_1 + (1-\theta)\mathbf{y}_2 \in D$. It means D is a convex set.

$\blacksquare$

Exercise 2: Affine Sets

Please show the following statements about affine sets.

1. If $U \subset \mathbb{R}^n$ and $\mathbf{0} \in U$, then U is an affine set if and only if it is a subspace.
2. If $U \subset \mathbb{R}^n$ is an affine set, there is a unique subspace $V \subset \mathbb{R}^n$ such that $U = \mathbf{u} + V$ for any $\mathbf{u} \in U$.
3. Let $U = \mathbf{aff}(\{(1,0,0)^\top, (0,1,0)^\top, (0,0,1)^\top\})$. Given a point $\mathbf{x}_0 \in U$, find two vectors $\mathbf{w}_1$ and $\mathbf{w}_2$ such that we can represent any vectors $\mathbf{w} \in U$ in the form of $\mathbf{w} = \mathbf{x}_0 + \alpha_1 \mathbf{w}_1 + \alpha_2 \mathbf{w}_2$ uniquely.

Solution:

1. (a) (Sufficiency) If U is a subspace of $\mathbb{R}^n$. $\forall \mathbf{x} \in U$, $\forall \theta \in \mathbb{R}$, we have $\theta \mathbf{x} \in U$. U is closed under addition. Thus we can take $\forall \mathbf{x_1}, \mathbf{x_2} \in U$, $\forall \theta \in \mathbb{R}$, $\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in U$. It means U is an affine set.

   (b) (Necessity) U is an affine set, so we have $\forall \mathbf{x_1}, \mathbf{x_2} \in U$,$\forall \theta \in \mathbb{R}$,

   $$\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in U \tag{1}$$
   $$\theta \mathbf{x}_2 + (1-\theta)\mathbf{x}_1 \in U \tag{2}$$

   Because $\mathbf{0} \in U$, we let $\mathbf{x_2} = \mathbf{0}$ to get

   $$\theta \mathbf{x_1} \in U$$

   It means U is closed under scalar multiplication.
   We let $\theta = \frac{1}{2}$ in (1),

   $$\frac{1}{2}\mathbf{x}_1 + \frac{1}{2}\mathbf{x}_2 \in U \tag{3}$$

   Using scalar multiplication, we can get

   $$\mathbf{x}_1 + \mathbf{x}_2 \in U \tag{4}$$

   It means U is closed under addition. Therefore, U is a subspace of $\mathbb{R}^n$.

2. $\forall \mathbf{u} \in U$, $U - \mathbf{u} = \{\mathbf{x} - \mathbf{u} : \mathbf{x} \in U\}$. And because U is an affine set, we have $\forall \mathbf{x_1}, \mathbf{x_2} \in U$, $\forall \theta \in \mathbb{R}$,

   $$\begin{aligned} \theta \mathbf{x_1} + (1-\theta)\mathbf{x_2} &\in U \\ \theta \mathbf{x_1} - \mathbf{u} + (1-\theta)(\mathbf{x_2} - \mathbf{u}) &= \theta \mathbf{x_1} + (1-\theta)\mathbf{x_2} - \mathbf{u} \\ &\in U - \mathbf{u} \end{aligned}$$

   It means that $U - \mathbf{u}$ is an affine set. And $\mathbf{0} \in U - \mathbf{u}$, because $\mathbf{u} \in U$. We know $V = U - \mathbf{u}$ is a subspace of $\mathbb{R}^n$(the result of 1.).

3. It is easy to know $U = \{(x, y, z) \in \mathbb{R}^3 | x + y + z = 1\}$. And $V = U - \mathbf{x_0} = \{(x, y, z) \in \mathbb{R}^3 | x+y+z = 0\}$ is a subspace of $\mathbb{R}^3$. $\mathbf{w_1} = \begin{pmatrix} 1, & 0, & -1 \end{pmatrix}^\top$ and $\mathbf{w_2} = \begin{pmatrix} 0, & 1, & -1 \end{pmatrix}^\top$ is a group of basis of V. It means $\forall \mathbf{v} \in V$, $\mathbf{v}$ can be represented by $\mathbf{w_1}$ and $\mathbf{w_2}$ uniquely. In other words, $\forall \mathbf{w} \in U$, we have $\mathbf{w} - \mathbf{x_0} = \alpha_1 \mathbf{w_1} + \alpha_2 \mathbf{w_2}$ uniquely.

$\blacksquare$

Exercise 3: Relative Interior and Interior

Let $C \subset \mathbb{R}^n$ be a nonempty convex set.

1. Let $\mathbf{x}_0 \in C$. Please show the following statements. The point $\mathbf{x}_0 \in \mathbf{relint}\ C$ if and only if there exists $r > 0$ such that $\mathbf{x}_0 + r\mathbf{v} \in C$ for any $\mathbf{v} \in \mathbf{aff}\, C - \mathbf{x}_0$ and $\|\mathbf{v}\|_2 \leq 1$.

2. (a) Please show that $\mathbf{x} \in \mathbf{relint}\ C$ if and only if for any $\mathbf{y} \in C$, there exists $\gamma > 0$ such that $\mathbf{x} + \gamma(\mathbf{x} - \mathbf{y}) \in C$.
   Hint: the result in Question 1 may be useful.

   (b) Please show that if $\mathbf{x} \in \mathbf{relint}\ C$, $\mathbf{y} \in \mathbf{cl}\ C$, then $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in \mathbf{relint}\ C$ for $\lambda \in (0, 1]$.
   Hint: there exists $r > 0$, such that $B(\mathbf{x}, r) \cap \mathbf{aff}\ C \subset \mathbf{relint}\ C$. Then consider the convex hull of $(B(\mathbf{x}, r) \cap \mathbf{aff}\ C) \cup \{\mathbf{y}\}$.

3. (Optional) Please show the following statements.

   (a) Suppose $\mathbf{int}\ C$ is nonempty, then $\mathbf{int}\ C = \mathbf{int}\,(\mathbf{cl}\ C)$.
   Hint: notice that $\mathbf{relint}\ C = \mathbf{int}\ C$ if $\mathbf{int}\ C$ is nonempty, then apply **Ex** 3.2(b). (in fact, the result still holds when $C = \emptyset$.)

   (b) $\mathbf{cl}\,(\mathbf{relint}\ C) = \mathbf{cl}\ C$.
   Hint: you can use **Ex** 3.2(b).

   (c) $\mathbf{relint}\,(\mathbf{cl}\ C) = \mathbf{relint}\ C$.

Solution:

1. ($\Longrightarrow$) $\mathbf{x}_0 \in \mathbf{relint}\ C$, so $\exists \epsilon > 0$ such that $B_\epsilon(\mathbf{x}_0) \bigcap \mathbf{aff} C \subseteq C$. We can find $\exists 0 < r < \epsilon$, s.t. $\|\mathbf{x_0} + r\mathbf{v} - \mathbf{x_0}\| = r\|\mathbf{v}\| \leq r < \epsilon$. It means that $\mathbf{x_0} + r\mathbf{v} \in B_\epsilon(\mathbf{x}_0)$. Besides, $\mathbf{x_0} + \mathbf{v} \in \mathbf{aff} C$ and $\mathbf{x_0} \in \mathbf{aff} C$, we know $(1-r)\mathbf{x_0} + r(\mathbf{x_0} + \mathbf{v}) = \mathbf{x_0} + r\mathbf{v} \in \mathbf{aff} C$.
   In other words, $\mathbf{x_0} + r\mathbf{v} \in B_\epsilon(\mathbf{x_0}) \bigcap \mathbf{aff} C$. $\mathbf{x_0} + r\mathbf{v} \in C$.
   ($\Longleftarrow$) We can write $\mathbf{aff} C$ as $\mathbf{aff} C = \{\mathbf{v} \in \mathbf{aff} C - \mathbf{x_0} | \mathbf{v} + \mathbf{x_0}\}$. $\mathbf{x_0} + \mathbf{v} \in \mathbf{aff} C$ and $\mathbf{x_0} \in \mathbf{aff} C$, thus $(1-r)\mathbf{x_0} + r(\mathbf{x_0} + \mathbf{v}) = \mathbf{x_0} + r\mathbf{v} \in \mathbf{aff} C$. So we can rewrite the $\mathbf{aff} C$: $\mathbf{aff} C = \{\mathbf{v} \in \mathbf{aff} C - \mathbf{x_0} | r\mathbf{v} + \mathbf{x_0}\}$.
   $\exists \epsilon > r > 0$, such that $\mathbf{x_0} + r\mathbf{v} \in B_\epsilon(\mathbf{x_0})$. Because $\mathbf{x}_0 + r\mathbf{v} \in C$, $B_\epsilon(\mathbf{x}_0) \bigcap \mathbf{aff} C \subseteq C$. It means $\mathbf{x_0} \in \mathbf{relint}\ C$.

2. (a) ($\Longrightarrow$) $\mathbf{x} \in \mathbf{relint}\ C$. Using the result of question 1, we have : there exists $r > 0$ such that $\mathbf{x} + r\mathbf{v} \in C$ for any $\mathbf{v} \in \mathbf{aff}\, C - \mathbf{x}$ and $\|\mathbf{v}\|_2 \leq 1$.
   $\forall \mathbf{y} \in C$, $\frac{\mathbf{x}-\mathbf{y}}{\|\mathbf{x}-\mathbf{y}\|} + \mathbf{x} \in \mathbf{aff} C$. Thus $\mathbf{v} = \frac{\mathbf{x}-\mathbf{y}}{\|\mathbf{x}-\mathbf{y}\|} + \mathbf{x} - \mathbf{x} = \frac{\mathbf{x}-\mathbf{y}}{\|\mathbf{x}-\mathbf{y}\|} \in \mathbf{aff} C - \mathbf{x}$ and $\|\mathbf{v}\| = 1$. It satisfies the conditions of question 1. Therefore, $\mathbf{x} + r\frac{\mathbf{x}-\mathbf{y}}{\|\mathbf{x}-\mathbf{y}\|} = \mathbf{x} + \gamma(\mathbf{x} - \mathbf{y}) \in C$, $\gamma = \frac{r}{\|\mathbf{x}-\mathbf{y}\|}$.
   ($\Longleftarrow$) $\forall \mathbf{y} \in C, \exists \gamma > 0$ such that $\mathbf{x} + \gamma(\mathbf{x} - \mathbf{y}) \in C$. $\exists \epsilon > \gamma\|\mathbf{x} - \mathbf{y}\| > 0$, $\|\mathbf{x} + \gamma(\mathbf{x} - \mathbf{y}) - \mathbf{x}\| = \gamma\|\mathbf{x} - \mathbf{y}\| < \epsilon$. It means $\mathbf{x} + \gamma(\mathbf{x} - \mathbf{y}) \in B_\epsilon(\mathbf{x})$.
   Besides, $\mathbf{x} + \gamma(\mathbf{x} - \mathbf{y}) = (1+\gamma)\mathbf{x} - \gamma\mathbf{y} \in \mathbf{aff} C$. $B_\epsilon(\mathbf{x}) \bigcap \mathbf{aff} C \subseteq C$. It means $\mathbf{x} \in \mathbf{relint}\ C$.

■

Exercise 4: Relative Boundary

The relative boundary of a set $S \subset \mathbb{R}^n$ is defined as **relbd** $S =$ **cl** $S \setminus$ **relint** S. Please show the following statements **or give counter-examples**.

1. For a set $S \subset \mathbb{R}^n$, **relbd** $S \subset$ **bd** S.
2. For a set $S \subset \mathbb{R}^n$, **relbd** $S =$ **bd** S.
3. For a set $S \subset \mathbb{R}^n$, **relbd** $S =$ **relbd cl** S.
4. For a set $S \subset \mathbb{R}^n$ and $\mathbf{x}_0 \in$ **cl** S, we can find a sequence $\{\mathbf{x}_k\} \subset \mathbb{R}^n \setminus$ **cl** S such that $\mathbf{x}_k \to \mathbf{x}_0$ as $k \to \infty$.

Solution:

1. It's not right. **relbd** $S \subset$ **bd** $S \Leftrightarrow$ **int** $S \subset$ **relint** S. Consider the case that $S = \{(x, y, z) \in \mathbb{R}^3 | x^2 + y^2 + z^2 < 1\} \subset \mathbb{R}^3$. In this case **int** $S = S =$ **relint** S.
2. It's not right. **relbd** $S =$ **bd** $S \Leftrightarrow$ **int** $S =$ **relint** S. Consider the case that $S = \{\mathbf{x} \in \mathbb{R}^3 : x_1^2 + x_2^2 < 1, x_3 = 0\} \subset \mathbb{R}^3$. **int** $S = \emptyset$ and **relint** $S = S = \{\mathbf{x} \in \mathbb{R}^3 : x_1^2 + x_2^2 < 1, x_3 = 0\}$.
3. It's not right. **cl cl** $S =$ **cl** S, thus **relbd** $S =$ **relbd cl** $S \Leftrightarrow$ **relint cl** $S =$ **relint** S. Consider the case that $S = \{\mathbf{x} \in \mathbb{R}^3 : x_1^2 + x_2^2 < 1, x_3 = 0\} \subset \mathbb{R}^3$. In this case, **relbd** $S = S = \{\mathbf{x} \in \mathbb{R}^3 : x_1^2 + x_2^2 < 1, x_3 = 0\}$ but **relbd cl** $S = \{\mathbf{x} \in \mathbb{R}^3 : x_1^2 + x_2^2 \leq 1, x_3 = 0\}$.
4. It's not right. Consider the case that $S = (0, 1) \subset R$ and $\mathbf{x_0} = \frac{1}{2} \in$ **cl** S. It is obvious that we can't find a sequence $(\mathbf{x}_k) \in R \setminus$ **cl** S such that $\mathbf{x}_k \to \mathbf{x}_0$ as $k \to \infty$.

■

Exercise 5: Supporting Hyperplane

1. From the lecture, we know that there exsits supporting hyperplanes at the boundary point of a convex set. Please solve the following questions.
   (a) Express the closed convex set $\{\mathbf{x} \in \mathbb{R}^2_+ \mid x_1x_2 \geq 1\}$ as an intersection of halfspaces.
   (b) Let $C = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\|_\infty \leq 1\}$, the ∞-norm unit ball in $\mathbb{R}^n$, and let $\hat{\mathbf{x}}$ be a point in the boundary of C. Identify the supporting hyperplanes of C at $\hat{\mathbf{x}}$ explicitly. (The ∞-norm of a point $\mathbf{x} \in \mathbb{R}^n$ is defined as $\max_{1\leq i\leq n} |x_i|$.)

2. On the linear space of symmetric $n \times n$ matrices S^n, we can define the standard inner product $\mathrm{tr}(XY) = \sum_{i,j=1}^n X_{ij}Y_{ij}$. From **Ex7** in **HW2**, we know the positive semi-definite cone S^n_+ isn't a polyhedron. However, please show that we can express S^n_+ as an intersection of halfspaces. Specifically, for $X, Y \in S^n$,

$$\mathrm{tr}(XY) \geq 0 \text{ for all } X \geq 0 \Leftrightarrow Y \geq 0.$$

3. **The set of separating hyperplanes:** Suppose that C and D are disjoint subsets of $\mathbb{R}^n$ (C and D may **not** be the convex sets). Consider the set of $(\mathbf{a}, b) \in \mathbb{R}^{n+1}$ for which $\mathbf{a}^T\mathbf{x} \leq b$ for all $\mathbf{x} \in C$, and $\mathbf{a}^T\mathbf{x} \geq b$ for all $\mathbf{x} \in D$. Show that this set is a convex cone (if there is no hyperplane that separates C and D, the set becomes $\{(\mathbf{0}, 0)\}$).

Solution:

1. (a) $\{\mathbf{x} \in \mathbb{R}^2_+ \mid x_1x_2 \geq 1\} = \bigcap_{\alpha\in\mathbb{R}^2_+} \{\mathbf{x} \in \mathbb{R}^2_+, \alpha = (a,b)^\top \in \mathbb{R}^2_+ | \alpha^\top\mathbf{x} \geq 2\sqrt{ab}\}$.
   (b) Suppose that there are s elements' absolute value of $\hat{\mathbf{x}}$ is 1. Let $\mathbf{a} = (\mathrm{sign}(x_i)) \in \mathbb{R}^n$. $\mathrm{sign}(x_i) = 1$ when $|x_i| = 1$ and $\mathrm{sign}(x_i) = 0$ in other cases. Therefore we have

$$< \mathbf{a}, \hat{\mathbf{x}} >= s$$

   $\forall \mathbf{y} \in C$, we can get

$$< \mathbf{a}, \hat{\mathbf{y}} >= \sum_{i=1}^{s} x_i'y_i'$$

   Because $|x_i'| \leq 1$ and $|y_i'| \leq 1$, we have $< \mathbf{a}, \mathbf{y} >\leq s$. In other words, $< \mathbf{a}, \mathbf{y} >\leq< \mathbf{a}, \hat{\mathbf{x}} >$. The supporting hyperplanes of C at $\hat{\mathbf{x}}$ is $H_{(\mathbf{a},s)} = \{\mathbf{x} \in \mathbb{R}^n | \mathbf{a}^\top\mathbf{x} = s\}$.

2. If $\mathbf{A} \in S^n_+$, $\forall \mathbf{z} \in \mathbb{R}^n, \mathbf{z} \neq \mathbf{0}$, $\mathbf{z}^\top\mathbf{A}\mathbf{z} \geq 0$. Let $\mathbf{B} = \mathbf{z}^\top\mathbf{z} \in S^n$, $\mathbf{z}^\top\mathbf{A}\mathbf{z} = tr(\mathbf{z}^\top\mathbf{A}\mathbf{z}) = tr(\mathbf{AB})$. And$\{\mathbf{A} \in S^n | tr(AB) > 0\}$ is a halfspace. Therefore, we can obtain

$$S^n_+ = \bigcap_{\mathbf{B}\in S^n} \{\mathbf{A} \in S^n | tr(\mathbf{AB}) > 0\}$$

   It is an intersection of halfspaces.
   $\mathrm{tr}(XY) = \mathrm{tr}(X)\mathrm{tr}(Y) \geq 0$ and $\mathrm{tr}(X) \geq 0$. Thus $\mathrm{tr}(Y) \geq 0$. Because $Y \in S^n$, $Y \geq 0$.

3. We denote the set of $(\mathbf{a}, b) \in \mathbb{R}^{n+1}$ by E. $\forall \mathbf{y_1} = (\mathbf{a_1}, b_1)$ and $\mathbf{y_2} = (\mathbf{a_2}, b_2) \in E$,

$$\forall \mathbf{x} \in C, \mathbf{a_1}^\top \mathbf{x} \leq b_1, \mathbf{a_2}^\top \mathbf{x} \leq b_2$$

Next we do multiplication: $\forall \theta_1, \theta_2 \geq 0$,

$$\theta_1 \mathbf{a_1}^\top \mathbf{x} \leq \theta_1 b_1, \theta_2 \mathbf{a_2}^\top \mathbf{x} \leq \theta_2 b_2$$

Adding together we can get

$$(\theta_1 \mathbf{a_1} + \theta_2 \mathbf{a_2})^\top \mathbf{x} \leq (\theta_1 b_1 + \theta_2 b2)$$

In the same way, we can get $\forall \mathbf{x} \in D$, $(\theta_1 \mathbf{a_1} + \theta_2 \mathbf{a_2})^\top \mathbf{x} \geq (\theta_1 b_1 + \theta_2 b2)$

It means $(\theta_1 \mathbf{a_1} + \theta_2 \mathbf{a_2}, \theta_1 b_1 + \theta_2 b2) = \theta_1 \mathbf{y_1} + \theta_2 \mathbf{y_2} \in E$. It means this set is a convex cone. ■

Exercise 6: Farkas' Lemma

Let $\mathbf{A} = (\mathbf{a}_1, \cdots, \mathbf{a}_n) \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Consider a set $A = \{\mathbf{a}_1, \cdots, \mathbf{a}_n\}$. Its conic hall $\mathbf{cone}\, A$ is defined as

$$\mathbf{cone}\, A = \{\sum_{i=1}^{n} \alpha_i \mathbf{a}_i : \alpha_i \geq 0, \mathbf{a}_i \in A\}.$$

1. Please show that $\mathbf{cone}\, A$ is closed and convex.
2. If $\mathbf{b} \in \mathbf{cone}\, A$, please show that there exists $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$.
3. If $\mathbf{b} \notin \mathbf{cone}\, A$, use separation theorems to show that there exists $\mathbf{y} \in \mathbb{R}^m$, such that $\mathbf{A}^\top \mathbf{y} \geq \mathbf{0}$ and $\mathbf{b}^\top \mathbf{y} < 0$.
4. Now you can prove Farkas' Lemma: for given $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$, one and only one of the two statements hold:
   - $\exists \mathbf{x} \in \mathbb{R}^n$, $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$.
   - $\exists \mathbf{y} \in \mathbb{R}^m$, $\mathbf{A}^\top \mathbf{y} \geq \mathbf{0}$ and $\mathbf{b}^\top \mathbf{y} < 0$.

Solution: 1. Let's start with the closure of $\mathbf{cone} A$: Suppose we have a sequence of points in $\mathbf{cone}\, A$, denoted as $\{\mathbf{v}_k\}$, such that $\mathbf{v}_k$ converges to some point $\mathbf{v}$. We need to show that $\mathbf{v}$ is also in $\mathbf{cone}\, A$.

Each point $\mathbf{v}_k$ can be expressed as $\mathbf{v}_k = \sum_{i=1}^{n} \alpha_{ki} \mathbf{a}_i$, where $\alpha_{ki} \geq 0$ for all i.

Since $\mathbf{v}_k$ converges to $\mathbf{v}$, we have:

$$\mathbf{v} = \lim_{k \to \infty} \mathbf{v}_k = \lim_{k \to \infty} \sum_{i=1}^{n} \alpha_{ki} \mathbf{a}_i$$

Now, we can use the properties of limits. The limit of a sum is the sum of the limits:

$$\mathbf{v} = \sum_{i=1}^{n} \lim_{k \to \infty} \alpha_{ki} \mathbf{a}_i$$

Since each $\alpha_{ki} \geq 0$ for all i, and limits preserve inequalities, we have $\lim_{k \to \infty} \alpha_{ki} \geq 0$ for all i. This means that each $\alpha_i = \lim_{k \to \infty} \alpha_{ki}$ is non-negative.

Therefore, we've shown that $\mathbf{v}$ can be expressed as a non-negative combination of the vectors in A, and thus, $\mathbf{v} \in \mathbf{cone}\, A$. This proves that $\mathbf{cone}\, A$ is closed.

Next, let's show that $\mathbf{cone}\, A$ is convex:
Take any points in $\mathbf{cone} A$:

$$\mathbf{u} = \sum_{i=1}^{n} \alpha_i \mathbf{a}_i, \alpha_i \geq 0$$

$$\mathbf{v} = \sum_{i=1}^{n} \beta_i \mathbf{a}_i, \beta_i \geq 0$$

Consider the convex combinations of these 2 points:

$$\begin{aligned}\forall\theta \in [0,1], \mathbf{w} = \theta\mathbf{u} + (1-\theta)\mathbf{v} \\ &= \sum_{i=1}^{n}(\theta\alpha_i + (1-\theta)\beta_i)\mathbf{a}_i \\ &= \sum_{i=1}^{n}\gamma\mathbf{a}_i\end{aligned}$$

Because $\theta \geq 0$, $\alpha_i \geq 0$, $(1-\theta) \geq 0$ and $\beta_i \geq 0$, $\gamma_i \geq 0$. It means $\mathbf{w} \in \mathbf{cone}A$, $\mathbf{cone}A$ is convex.

2. $\mathbf{x} = (x_1, x_2, \cdots, x_n)^\top > \mathbf{0}$ means $x_i > 0$, $i = 1, 2, \cdots, n$. $\mathbf{b} \in \mathbf{cone}A$ means there exists $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_n)^\top > \mathbf{0}$ such that $\mathbf{b} = \sum_{i=1}^{n} \alpha_i \mathbf{a}_i$.
We can just let $\mathbf{x} = \alpha$. At this time, $\mathbf{Ax} = \sum_{i=1}^{n} \alpha_i \mathbf{a}_i = \mathbf{b}$.

3. $\mathbf{cone}A$ is a nonempty closed convex set. According to the separation theorem, we can obtain: $\exists \mathbf{w} \in \mathbb{R}^n$ and $\mathbf{w} \neq \mathbf{0}$, $\exists u > v$ such that $\mathbf{cone}A \subseteq H^+_{(\mathbf{w},u)}$ and $\mathbf{b} \in H^-_{(\mathbf{w},v)}$. It means $\mathbf{b}^\top\mathbf{w} \leq v < u$. Let $\mathbf{y} = \mathbf{w} - u$, we can get $\mathbf{b}^\top\mathbf{y} < 0$.
$\mathbf{cone}A \subseteq H^+_{(\mathbf{w},u)}$. Therefore, $\forall \mathbf{x} \in \mathbf{cone}A$, we have $\mathbf{w}^\top\mathbf{x} \geq u$. Because $\mathbf{a}_1 \in \mathbf{cone}A$, it follows that $\mathbf{a}_1^\top\mathbf{w} \geq u$. Or we can say $\mathbf{a}_1^\top\mathbf{y} \geq 0$. In the same way, we know $\mathbf{a}_i^\top\mathbf{y} \geq 0 (i = 1, 2, \cdots, n)$. It can be written as $\mathbf{A}^\top\mathbf{y} \geq \mathbf{0}$.

4. Using the results of question 2. and question 3. we can prove it easily:

   (a) If $\mathbf{b} \in \mathbf{cone}A$, $\exists \mathbf{x} \in \mathbb{R}^n$, $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$.

   (b) If $\mathbf{b} \notin \mathbf{cone}A$, $\exists \mathbf{y} \in \mathbb{R}^m$, $\mathbf{A}^\top\mathbf{y} \geq \mathbf{0}$ and $\mathbf{b}^\top\mathbf{y} < 0$.

■