553.665 Introduction to Convexity, Fall 2022
Section Notes

Hongyu Cheng

December 11, 2022

Contents

1 Section-Sept-2-2022

1.1 Recall the definitions

Let $X \subseteq \mathbb{R}^d$.

1. X is a convex set if $\forall \mathbf{x}, \mathbf{y} \in X,\ \forall \lambda \in [0,1], \lambda \mathbf{x} + (1-\lambda)\mathbf{y} \in X$.
2. $\operatorname{conv}(X)$ is the smallest convex set containing X.
3. An affine transformation is a function $T : \mathbb{R}^d \to \mathbb{R}^m$ of the form $T(x) = A\mathbf{x} + \mathbf{b}$, where A is an $m \times d$ matrix and $\mathbf{b} \in \mathbb{R}^m$.
4. X is a cone if $\forall \mathbf{x}, \mathbf{y} \in X, \forall \lambda, \gamma \geq 0,\ \lambda \mathbf{x} + \gamma \mathbf{y} \in X$.
5. X is an affine set if $\forall \lambda, \gamma \in \mathbb{R}$ s.t. $\lambda + \gamma = 1,\ \lambda \mathbf{x} + \gamma \mathbf{y} \in X$.
6. X is a linear subspace if $\forall \lambda, \gamma \in \mathbb{R},\ \lambda \mathbf{x} + \gamma \mathbf{y} \in X$.

1.2 A brief review of real analysis and basic topology

Exercise 1. Let A_i, $i \in \mathcal{I}$ be an arbitrary family of sets with $\mathcal{I} \neq \varnothing$, and X is a set. Then,

1. $\forall \alpha \in \mathcal{I}$,
$$\bigcap_{i \in \mathcal{I}} A_i \subseteq A_\alpha \subseteq \bigcup_{i \in \mathcal{I}} A_i$$
2. (Distributive property)
$$X \cap \left(\bigcup_{i \in \mathcal{I}} A_i \right) = \bigcup_{i \in \mathcal{I}} (X \cap A_i)$$
$$X \cup \left(\bigcap_{i \in \mathcal{I}} A_i \right) = \bigcap_{i \in \mathcal{I}} (X \cup A_i)$$
3. (De Morgan's law)
$$X \backslash \left(\bigcap_{i \in \mathcal{I}} A_i \right) = \bigcup_{i \in \mathcal{I}} (X \backslash A_i)$$
$$X \backslash \left(\bigcup_{i \in \mathcal{I}} A_i \right) = \bigcap_{i \in \mathcal{I}} (X \backslash A_i)$$

Proof. The first equality in 2 holds since

$$\begin{aligned} x \in X \cap \left(\bigcup_{i \in \mathcal{I}} A_i \right) &\iff x \in X \text{ and } x \in \bigcup_{i \in \mathcal{I}} A_i \\ &\iff x \in X \text{ and } \exists \alpha \in \mathcal{I} \text{ such that } x \in A_\alpha \\ &\iff \exists \alpha \in \mathcal{I} \text{ such that } x \in X \cap A_\alpha \\ &\iff x \in \bigcup_{i \in \mathcal{I}} (X \cap A_i). \end{aligned}$$

The first equality in 3 holds since

$$\begin{aligned} x \in X \backslash \left(\bigcap_{i \in \mathcal{I}} A_i \right) &\iff x \in X \text{ and } x \notin \bigcap_{i \in \mathcal{I}} A_i \\ &\iff x \in X \text{ and } \exists \alpha \in \mathcal{I} \text{ such that } x \notin A_\alpha \\ &\iff \exists \alpha \in \mathcal{I} \text{ such that } x \in X \backslash A_\alpha \\ &\iff x \in \bigcup_{i \in \mathcal{I}} (X \backslash A_i). \end{aligned}$$

The proofs of the remaining arguments are very similar. □

Definition 1. For any set $X \subseteq \mathbb{R}^d$,

1. $\mathrm{cl}(X)$ is the smallest closed set containing X (the intersection of all closed sets of $\mathbb{R}^d$ containing X).
2. $\mathrm{int}(X)$ is the largest open set contained inside X (the union of all open sets of $\mathbb{R}^d$ contained in X).
3. $\mathrm{bd}(X) := \mathrm{cl}(X) \backslash \mathrm{int}(X)$.

Exercise 2. Let $X \subseteq \mathbb{R}^d$, then $x \in \mathrm{bd}(X) \implies \forall r > 0, B(x,r) \cap X \neq \varnothing$ and $B(x,r) \cap (\mathbb{R}^d \backslash X) \neq \varnothing$.

Proof. Note that $\mathrm{bd}(X)$ is defined to be $\mathrm{cl}(X) \backslash \mathrm{int}(X)$. So if there exists $r > 0$ such that $B(x,r) \subseteq X$, then $\mathrm{int}(B(x,r)$ is an open set contained in X that contains x, which contradicts with the definition.

Similarly, if there exists $r > 0$ such that $B(x,r) \subseteq \mathbb{R}^d \backslash X$, then $\mathbb{R}^d \backslash \mathrm{int}(B(x,r))$ is a closed set contains X, but not x, again contradicting the definition.

Thus, for every $r > 0$, $B(x,r)$ must contain a point in X and a point in $\mathbb{R}^d \backslash X$. □

Remark 1. The reverse direction is also true.

Exercise 3. Let $X \subseteq \mathbb{R}^d$, then $\mathrm{int}(A) = \mathbb{R}^d \backslash \mathrm{cl}(\mathbb{R}^d \backslash A)$.

Proof. $\forall x \in \mathrm{int}(A) \implies x \notin \mathrm{cl}(\mathbb{R}^d \backslash A) \implies x \in \mathbb{R}^d \backslash \mathrm{cl}(\mathbb{R}^d \backslash A)$.

For the reverse inclusion,

$$\begin{aligned} \forall x \in \mathbb{R}^d \backslash \mathrm{cl}(\mathbb{R}^d \backslash A) &\implies \exists r > 0 \text{ s.t. } \mathrm{int}(B(x,r)) \subseteq \mathbb{R}^d \backslash \mathrm{cl}(\mathbb{R}^d \backslash A) \\ &\implies \mathrm{int}(B(x,r)) \cap \mathbb{R}^d \backslash A = \varnothing \\ &\implies \mathrm{int}(B(x,r)) \subseteq A \\ &\implies x \in \mathrm{int}(A). \end{aligned}$$

□

Remark 2. Similarly, $\mathrm{cl}(A) = \mathbb{R}^d \backslash \mathrm{int}(\mathbb{R}^d \backslash A)$.

1.3 A brief review of linear algebra

Definition 2. Let $A \in R^{m \times n}$, then there are several subspaces naturally associated to it:

- $\mathrm{null}(A) = \{x \in \mathbb{R}^n : Ax = 0\}$,
- $\mathrm{range}(A) = \{y \in \mathbb{R}^m : \exists x \in \mathbb{R}^n \text{ such that } Ax = y\}$.

Exercise 4. Let $A \in \mathbb{R}^{m \times n}$, then $\mathrm{null}(A)$ is a subspace of $\mathbb{R}^n$.

Proof. 1. $A0 = 0 \implies 0 \in \text{null}(A)$ and $\text{null}(A)$ is nonempty.

2. $\forall x, y \in \text{null}(A)$, $A(x+y) = Ax + Ay = 0 + 0 = 0 \implies x + y \in \text{null}(A)$.

3. $\forall x \in \text{null}(A)$, $\forall \lambda \in \mathbb{R}$, $A(\lambda x) = \lambda Ax = \lambda 0 = 0 \implies \lambda x \in \text{null}(A)$.

□

Exercise 5. Let $A \in \mathbb{R}^{m\times n}$, then $\text{null}(A)$ is orthogonal to $\text{range}(A^\top)$.

Proof. We just need to show that $\forall x \in \text{null}(A)$, $\forall y \in \text{range}(A^\top)$, $\langle x, y\rangle = 0$. By definition of $\text{range}(A^\top)$, there exists $z \in \mathbb{R}^n$ such that $A^\top z = y$. Therefore,

$$\langle x, y\rangle = \langle x, A^\top z\rangle = \langle Ax, z\rangle = \langle 0, z\rangle = 0.$$

□

Theorem 1 (Rank-nullity theorem)**.** *Let $A \in \mathbb{R}^{m\times n}$, then dim range$(A)$ + dim null$(A) = n$.*

Theorem 2. *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix, then there exists a matrix $S \in \mathbb{R}^{n\times n}$ such that $S^\top S = I$ and*

$$A = S\Lambda S^\top,$$

where Λ is the diagonal matrix with the diagonal entries equal to the eigenvalues of A.

Remark 3. Both directions would hold if A is *normal*, i.e., $A^\top A = AA^\top$.

Exercise 6. Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. The rank of A is defined to be the dimension of the range of A. Prove that this is equal to the number of nonzero eigenvalues of A.

1.4 Convex sets

Theorem 3 (Operations that preserve convexity)**.** *The following are all true.*

1. *Let $X_i, i \in \mathcal{I}$ be an arbitrary family of convex sets. Then $\bigcap_{i\in\mathcal{I}} X_i$ is a convex set.*
2. *Let X be a convex set and $\alpha \in \mathbb{R}$, then αX is a convex set.*
3. *Let X, Y be convex sets, then $X + Y$ is convex.*
4. *Let $T : \mathbb{R}^d \to \mathbb{R}^m$ be any affine transformation.*
   (a) *If $X \subseteq \mathbb{R}^d$ is convex, then $T(X)$ is a convex set.*
   (b) *If $Y \subseteq \mathbb{R}^m$ is convex, then $T^{-1}(Y)$ is convex.*

Proof. 1. See lecture notes.

2. $\forall \mathbf{x}^1, \mathbf{y}^1 \in \alpha X$, $\forall \lambda \in [0, 1]$, there exist $\mathbf{x}^2, \mathbf{y}^2 \in X$ such that $\mathbf{x}^1 = \alpha\mathbf{x}^2, \mathbf{y}^1 = \alpha\mathbf{y}^2$. Therefore, $\lambda\mathbf{x}^1 + (1-\lambda)\mathbf{y}^1 = \lambda(\alpha\mathbf{x}^2) + (1-\lambda)(\alpha\mathbf{y}^2) = \alpha(\underbrace{\lambda\mathbf{x}^2 + (1-\lambda)\mathbf{y}^2}_{\in X}) \in \alpha X$.

3. $\forall \mathbf{a}, \mathbf{b} \in X + Y$, $\forall \lambda \in [0, 1]$, there exist $\mathbf{x}^1, \mathbf{x}^2 \in X$ and $\mathbf{y}^1, \mathbf{y}^2 \in Y$ such that $\mathbf{a} = \mathbf{x}^1 + \mathbf{y}^1$ and $\mathbf{b} = \mathbf{x}^2 + \mathbf{y}^2$. Therefore,

$$\begin{aligned}\lambda\mathbf{a} + (1-\lambda)\mathbf{b} &= \lambda(\mathbf{x}^1 + \mathbf{y}^1) + (1-\lambda)(\mathbf{x}^2 + \mathbf{y}^2)\\ &= \underbrace{\lambda\mathbf{x}^1 + (1-\lambda)\mathbf{x}^2}_{\in X} + \underbrace{\lambda\mathbf{y}^1 + (1-\lambda)\mathbf{y}^2}_{\in Y} \in X + Y.\end{aligned}$$

4. (a) $\forall \mathbf{x}^1, \mathbf{y}^1 \in T(X),\ \forall \lambda \in [0,1]$, there exists $\mathbf{x}^2, \mathbf{y}^2 \in X$ such that $\mathbf{x}^1 = A\mathbf{x}^2 + b$ and $\mathbf{y}^1 = A\mathbf{y}^2 + b$, where $A \in \mathbb{R}^{m\times d},\ b \in \mathbb{R}^m$. Therefore,

$$\begin{aligned}\lambda \mathbf{x}^1 + (1-\lambda)\mathbf{y}^1 &= \lambda(A\mathbf{x}^2 + b) + (1-\lambda)(A\mathbf{y}^2 + b)\\ &= \lambda A\mathbf{x}^2 + (1-\lambda)A\mathbf{y}^2 + \lambda b + (1-\lambda)b\\ &= A(\underbrace{\lambda \mathbf{x}^2 + (1-\lambda)\mathbf{y}^2}_{\in X}) + b \in T(X).\end{aligned}$$

(b) Note that $T^{-1}(Y) = \{\mathbf{x} \in \mathbb{R}^d : T(x) \in Y\}$, then $\forall \mathbf{x}^1, \mathbf{x}^2 \in T^{-1}(Y),\ T(\mathbf{x}^1) \in Y$ and $T(\mathbf{x}^2) \in Y$. Therefore, for any $\lambda \in [0,1]$,

$$T(\lambda \mathbf{x}^1 + (1-\lambda)\mathbf{x}^2) = \lambda \underbrace{T(\mathbf{x}^1)}_{\in Y} + (1-\lambda)\underbrace{T(\mathbf{x}^2)}_{\in Y} \in Y,$$

this proves $\lambda \mathbf{x}^1 + (1-\lambda)\mathbf{x}^2 \in T^{-1}(Y)$.

□

Exercise 7. Show that if $A \subseteq B$ then $\text{conv}(A) \subseteq \text{conv}(B)$. Does the converse hold?

Proof. Observe that $A \subseteq B \subseteq \text{conv}\,(B)$, therefore $\text{conv}\,(B)$ is a convex set that contains A. By definition of convex hull, $\text{conv}\,(A) \subseteq \text{conv}\,(B)$. However, the converse is not true. Consider $A = \mathbb{Q},\ B = \mathbb{R}\backslash\mathbb{Q}$, then $\text{conv}\,(A) = \text{conv}\,(B) = \mathbb{R}$, but $A \cap B = \varnothing$. □

Exercise 8. Show that $\text{conv}\,(A \cap B) \subseteq \text{conv}\,(A) \cap \text{conv}\,(B)$. Is the containment strict?

Proof.

$$\begin{aligned} A \cap B \subseteq \text{conv}\,(A) \text{ and } A \cap B \subseteq \text{conv}\,(B) &\implies \text{conv}\,(A) \text{ and } \text{conv}\,(B) \text{ are convex sets containing } A \cap B\\ &\implies \text{conv}\,(A \cap B) \subseteq \text{conv}\,(A) \text{ and } \text{conv}\,(A \cap B) \subseteq \text{conv}\,(B)\\ &\implies \text{conv}\,(A \cap B) \subseteq \text{conv}\,(A) \cap \text{conv}\,(B)\,.\end{aligned}$$

The containment is strict. Consider $A = \mathbb{Q},\ B = \mathbb{R}\backslash\mathbb{Q}$, then $\text{conv}\,(A) \cap \text{conv}\,(B) = \mathbb{R} \cap \mathbb{R} = \mathbb{R}$, but $\text{conv}\,(A \cap B) = \text{conv}\,(\varnothing) = \varnothing$. □

Exercise 9. Let $A \in \mathbb{R}^{d\times d}$ be a positive definite matrix and $\mathbf{c} \in \mathbb{R}^d$. Show that the ellipsoid

$$E(A, c) := \{\mathbf{x} \in \mathbb{R}^d : (\mathbf{x} - \mathbf{c})^\top A^{-1}(\mathbf{x} - \mathbf{c}) \le 1\}$$

is a convex set.

Proof. By Theorem 2, we can let $\mathbf{y} = A^{-\frac{1}{2}}(\mathbf{x} - \mathbf{c})$, then the argument follows from 4(a) in Theorem 3. □

Exercise 10. Let $A \in \mathbb{R}^{n\times n}$ be a positive definite matrix, and $C := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}^\top A\mathbf{x} + \mathbf{b}^\top\mathbf{x} + \mathbf{c} \le 0\}$. Show that C is a convex set.

Proof. $\forall \mathbf{x}, \mathbf{y} \in C,\ \forall \lambda \in [0,1]$, one can easily obtain the following:

$$\mathbf{x}^\top A\mathbf{x} + \mathbf{b}^\top\mathbf{x} + \mathbf{c} \le 0, \tag{1.1}$$

$$\mathbf{y}^\top A\mathbf{y} + \mathbf{b}^\top\mathbf{y} + \mathbf{c} \le 0, \tag{1.2}$$

$$2\mathbf{x}^\top A\mathbf{y} < \mathbf{x}^\top A\mathbf{x} + \mathbf{y}^\top A\mathbf{y}, \tag{1.3}$$

where the last inequality comes from $(\mathbf{x}-\mathbf{y})^\top A(\mathbf{x}-\mathbf{y}) > 0$ since $A \succ 0$. Therefore,

$$\begin{aligned}
&[\lambda\mathbf{x}+(1-\lambda)\mathbf{y}]^\top A\,[\lambda\mathbf{x}+(1-\lambda)\mathbf{y}] + \mathbf{b}^\top[\lambda\mathbf{x}+(1-\lambda)\mathbf{y}] + \mathbf{c}\\
=&\lambda^2\mathbf{x}^\top A\mathbf{x} + (1-\lambda)^2\mathbf{y}^\top A\mathbf{y} + 2\lambda(1-\lambda)\mathbf{x}^\top A\mathbf{y} + \lambda\mathbf{b}^T\mathbf{x} + (1-\lambda)\mathbf{b}^T\mathbf{y} + \mathbf{c}\\
\overset{(1.3)}{<}&\lambda^2\mathbf{x}^\top A\mathbf{x} + (1-\lambda)^2\mathbf{y}^\top A\mathbf{y} + \lambda(1-\lambda)(\mathbf{x}^\top A\mathbf{x} + \mathbf{y}^\top A\mathbf{y}) + \lambda\mathbf{b}^T\mathbf{x} + (1-\lambda)\mathbf{b}^T\mathbf{y} + \mathbf{c}\\
=&\left[\lambda^2 + \lambda(1-\lambda)\right]\mathbf{x}^\top A\mathbf{x} + \left[(1-\lambda)^2 + \lambda(1-\lambda)\right]\mathbf{y}^\top A\mathbf{y} + \lambda\mathbf{b}^\top\mathbf{x} + (1-\lambda)\mathbf{b}^\top\mathbf{y} + \mathbf{c}\\
=&\lambda\left(\mathbf{x}^\top A\mathbf{x} + \mathbf{b}^\top\mathbf{x} + \mathbf{c}\right) + (1-\lambda)\left(\mathbf{y}^\top A\mathbf{y} + \mathbf{b}^\top\mathbf{y} + \mathbf{c}\right) \overset{(1.1),(1.2)}{\leq} 0.
\end{aligned}$$

□

Exercise 11. Let $X \subseteq \mathbb{R}^n$ be a finite set and let $\mathbf{c} \in \mathbb{R}^n$. Show that $\max\{\mathbf{c}^\top\mathbf{x} : \mathbf{x} \in X\} = \max\{\mathbf{c}^\top\mathbf{x} : x \in \mathrm{conv}\,(X)\}$.

Proof. $X \subseteq \mathrm{conv}(X) \implies \max\{c^T x : x \in X\} \leq \max\{c^T x : x \in \mathrm{conv}(X)\}$, then it suffices to show that $\max\{c^T x : x \in X\} \geq \max\{c^T x : x \in \mathrm{conv}(X)\}$.

X is a finite set $\implies \mathrm{conv}(X)$ is a polytope (Minkowski-Weyl) $\implies \mathrm{conv}(X)$ is compact. Then Weierstrass theorem yields that there exists $x_* \in \mathrm{conv}(X)$ such that

$$c^T x_* = \max\{c^T x : x \in \mathrm{conv}(X)\}.$$

Therefore,

$$\begin{aligned}
\max\{c^T x : x \in \mathrm{conv}(X)\} &= c^T x_*\\
&= c^T(\sum_{j=1}^{t}\lambda_j x^j)\\
&= \sum_{j=1}^{t}\lambda_j c^T x^j\\
&\leq \sum_{j=1}^{t}\lambda_j \max\{c^T x : x \in X\}\\
&= \max\{c^T x : x \in X\}
\end{aligned}$$

□

2 Section-Sept-9-2022

2.1 Review

Theorem 4 (Equivalent definitions of relative interior)**.** *Let $C \subseteq \mathbb{R}^d$ and $\mathbf{x} \in C$. The following are equivalent.*

1. $\mathbf{x} \in$ *relint*(C).
2. *There exists $\varepsilon > 0$ such that $B(\mathbf{x},\varepsilon) \cap \mathit{aff}(C) \subseteq C$.*
3. *There exists $\varepsilon > 0$ such that $\forall \mathbf{y} \in \mathit{aff}(C)$, $\mathbf{x} + \varepsilon\left(\frac{\mathbf{y}-\mathbf{x}}{\|\mathbf{y}-\mathbf{x}\|}\right) \in C$.*
4. $\forall \mathbf{y} \in \mathit{aff}(C)$, $\exists \varepsilon_{\mathbf{y}} > 0$ *such that* $\mathbf{x} + \varepsilon_{\mathbf{y}}(\mathbf{y}-\mathbf{x}) \in C$.

Question 1. Let $X \subseteq \mathbb{R}^d$ be a compact convex set, is it true that cone(X) is closed?
Answer. No, a counterexample is provided in Figure 5.

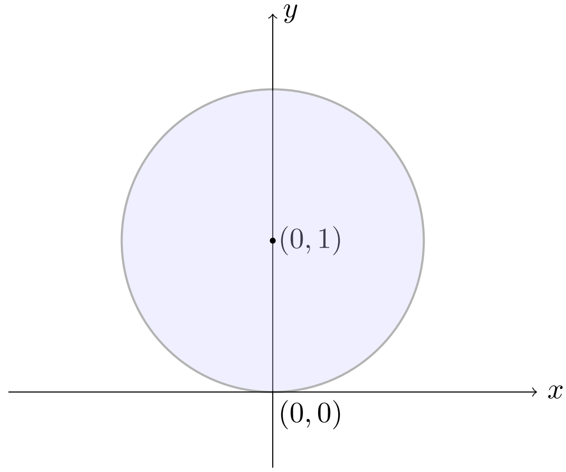


Figure 1: $X = \{(x, y) : x^2 + (y-1)^2 \leq 1\}$, but $\text{cone}(X) = \{(x, y) : y > 0\} \cup \{(0, 0)\}$ is not closed.

Question 2. Let $X, Y \subseteq \mathbb{R}^d$, is it true that $X \subseteq Y$ implies $\text{relint}(X) \subseteq \text{relint}(Y)$?
Answer. No, a counterexample is provided in Figure 6.

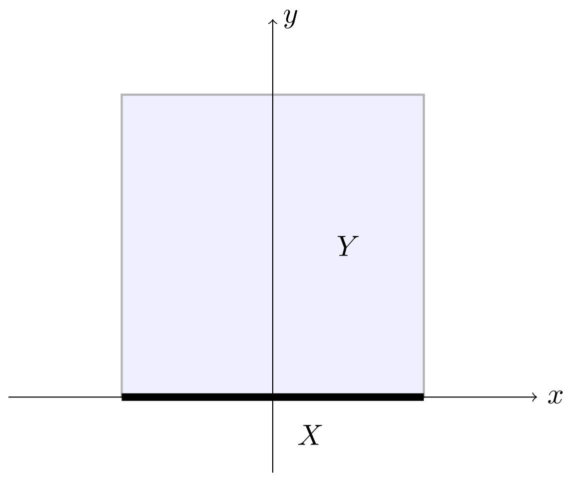


Figure 2: $X = \{(x, y) : x \in [-1, 1], y = 0\}$, $Y = \{(x, y) : x \in [-1, 1], y \in [0, 2]\}$, then $X \subseteq Y$ but $\text{relint}(X) \cap \text{relint}(Y) = \varnothing$.

2.2 Exercises

Exercise 12. Prove that the relative interior of a nonempty convex set is nonempty.

Proof. Hint: Let C be any nonempty convex set with dimension $d \geq 0$, then one can construct a point lie in $\text{relint}(C)$ by some affinely independent points in C. Left as Homework. □

Exercise 13. Let $X \subseteq \mathbb{R}^d$, and let $\mathbf{y} \in \text{conv}(X)$. Suppose H is a halfspace such that $\mathbf{y} \in H$. Prove that $H \cap X \neq \varnothing$.

Proof. Let $H = \{\mathbf{x} : \langle \mathbf{a}, \mathbf{x} \rangle \le \delta\}$ for some $\mathbf{a} \in \mathbb{R}^d,\ \delta > 0$. Since $\mathbf{y} \in \operatorname{conv}(X)$, we have $\mathbf{y} = \sum_{i=1}^k \lambda_i \mathbf{x}^i$ for some $\mathbf{x}^i \in X,\ \lambda_i \ge 0,\ i = 1, \ldots, k$ and $\sum_{i=1}^k \lambda_i = 1$. Now if $H \cap X = \varnothing$, then we would have $\langle \mathbf{a}, \mathbf{x}^i \rangle > \delta,\ \forall i \in \{1, \ldots, k\}$, but this would give $\langle \mathbf{a}, \mathbf{y} \rangle > \delta$, which leads to a contradiction. □

Exercise 14. Let $X \subseteq \mathbb{R}^d$. Suppose H is a halfspace such that $X \subseteq H$. Prove that

$$\operatorname{conv}(H^= \cap X) = H^= \cap \operatorname{conv}(X),$$

where $H^=$ is the hyperplane associated with H.

Note: Talk about the relationship with MILPs, as well as the exercise 11 in session 1.

Proof. Let $H = \{\mathbf{x} : \langle \mathbf{a}, \mathbf{x} \rangle \le \delta\}$ for some $\mathbf{a} \in \mathbb{R}^d,\ \delta > 0$, and $H^= = \{\mathbf{x} : \langle \mathbf{a}, \mathbf{x} \rangle = \delta\}$.

$\forall \mathbf{x} \in H^= \cap \operatorname{conv}(X)$, we have $\langle \mathbf{a}, \mathbf{x} \rangle = \delta,\ \mathbf{x} = \sum_{i=1}^k \lambda_i \mathbf{x}^i$ for some $\mathbf{x}^i \in X,\ \lambda_i \ge 0,\ i = 1, \ldots, k$ and $\sum_{i=1}^k \lambda_i = 1$. Therefore,

$$\sum_{i=1}^k \lambda_i(\langle \mathbf{a}, \mathbf{x}^i \rangle - \delta) = 0.$$

Observe that $\lambda_i(\langle \mathbf{a}, \mathbf{x}^i \rangle - \delta) \le 0,\ \forall i \in \{1, \ldots, k\}$, then for each i, either $\lambda_i = 0$ or $\langle \mathbf{a}, \mathbf{x}^i \rangle = \delta$. Therefore, for those i such that $\lambda_i \neq 0$, $\mathbf{x}^i \in H^=$. Thus, $\mathbf{x} \in \operatorname{conv}(H^= \cap X)$.

To show the reverse inclusion, $\forall \mathbf{x} \in \operatorname{conv}(H^= \cap X)$, $\mathbf{x} = \sum_{i=1}^k \lambda_i \mathbf{x}^i$ for some $\mathbf{x}^i \in H^= \cap X,\ \lambda_i \ge 0,\ i = 1, \ldots, k$ and $\sum_{i=1}^k \lambda_i = 1$. This implies $\mathbf{x} \in \operatorname{conv}(X)$ and $\mathbf{x} \in H^=$. That is, $\mathbf{x} \in H^= \cap \operatorname{conv}(X)$. □

Apply Exercise 14, we can provide an alternate proof of Theorem 5.

Theorem 5 (Carathéodory's theorem (convex version))**.** *Let $X \subseteq \mathbb{R}^d$ and $\mathbf{x} \in$ conv(X). Then $\mathbf{x}$ is a convex combination of at most $d+1$ points of X.*

Proof. Base case: It's easy to check when $d = 1$.
Induction step: Suppose it's true for all dimensions less than d. $\mathbf{x} \in \operatorname{conv}(X)$, then by definition there exist distinct $\mathbf{x}^1, \ldots, \mathbf{x}^k$ such that $\mathbf{x} \in \operatorname{conv}\left(\{\mathbf{x}^1, \ldots, \mathbf{x}^k\}\right) := C$.

Case 1: $\mathbf{x} \in \operatorname{relbd}(C)$, then let $H^=$ be a supporting hyperplane of C through $\mathbf{x}$, then

$$\mathbf{x} \in H^= \cap C = H^= \cap \operatorname{conv}\left(\{\mathbf{x}^1, \ldots, \mathbf{x}^k\}\right) \overset{\textit{Exercise 14}}{=} \operatorname{conv}\left(H^= \cap \{\mathbf{x}^1, \ldots, \mathbf{x}^k\}\right).$$

Since $H^=$ is $(d-1)$-dimensional, by induction hypothesis, $\mathbf{x}$ is a convex combination of at most d points of $H^= \cap \{\mathbf{x}^1, \ldots, \mathbf{x}^k\}$, in particular of at most d points of X.

Case 2: $\mathbf{x} \in \operatorname{relint}(C)$, observe that there exists some $i \in \{1, \ldots, k\}$ such that $\mathbf{x} \neq \mathbf{x}^i$, then let $\mathbf{y} \in \{\mathbf{x}^i + \lambda(\mathbf{x} - \mathbf{x}^i) : \lambda > 0\} \cap \operatorname{relbd}(C)$. Therefore, $\mathbf{x}$ is a convex combination of at most $d+1$ points of X since $\mathbf{x}$ is a convex combination of $\mathbf{y}$ and $\mathbf{x}^i$, and $\mathbf{y}$ is a convex combination of at most d points of X. □

Definition 3. Let $X \subseteq \mathbb{R}^d$ be a linear subspace. We define $X^\perp := \{\mathbf{y} \in \mathbb{R}^d : \langle \mathbf{y}, \mathbf{x} \rangle = 0,\ \forall \mathbf{x} \in X\}$ as the *orthogonal complement* of X.

Definition 4. Let $X \subseteq \mathbb{R}^d$ be any set. We define

$$X^\circ := \{\mathbf{y} \in \mathbb{R}^d : \langle \mathbf{y}, \mathbf{x} \rangle \le 1,\ \forall \mathbf{x} \in X\},$$

$$X^\bullet := \{\mathbf{y} \in \mathbb{R}^d : \langle \mathbf{y}, \mathbf{x} \rangle \le 0,\ \forall \mathbf{x} \in X\}.$$

Exercise 15. Let $X \subseteq \mathbb{R}^d$ be a convex cone, prove that $X^\circ = X^\bullet$.

Proof. It's easy to check that $X^\bullet \subseteq X^\circ$.

To show the reverse inclusion, $\forall \mathbf{y} \in X^\circ$, if $\langle \mathbf{y}, \mathbf{x} \rangle > 0$ for some $\mathbf{x} \in X$, then

$$\langle \mathbf{y}, \underbrace{\frac{2}{\langle \mathbf{y}, \mathbf{x} \rangle}\mathbf{x}}_{\in X} \rangle = 2 > 1$$

contradicts with $\mathbf{y} \in X^\circ$. □

Remark 4. Similarly, one can prove that $\text{cone}(X)^\circ = X^\bullet$ for any $X \subseteq \mathbb{R}^d$.

Exercise 16. Let $X \subseteq \mathbb{R}^d$ be a linear subspace, prove that $X^\circ = X^\perp$.

Proof sketch. Similar to the proof of Exercise 15. $\forall \mathbf{y} \in X^\circ$, if $\langle \mathbf{y}, \mathbf{x} \rangle \neq 0$ for some $\mathbf{x} \in X$, then $\langle \mathbf{y}, \frac{2}{\langle \mathbf{y}, \mathbf{x} \rangle}\mathbf{x} \rangle = 2 > 1$ contradicts with $\mathbf{y} \in X^\circ$. □

Exercise 17. Let $X \subseteq \mathbb{R}^d$ be any set, prove that $(X^\bullet)^\bullet = \text{cl}(\text{cone}(X))$.

Proof. First we show that $X^\bullet$ is a convex cone. $\forall \mathbf{a}, \mathbf{b} \in X^\bullet, \lambda, \gamma \geq 0 \implies \forall \mathbf{x} \in X, \langle \lambda\mathbf{a} + \gamma\mathbf{b}, \mathbf{x} \rangle = \lambda\langle \mathbf{a}, \mathbf{x} \rangle + \gamma\langle \mathbf{b}, \mathbf{x} \rangle \leq 0 \implies \lambda\mathbf{a} + \gamma\mathbf{b} \in X^\bullet \implies X^\bullet$ is a convex cone. Therefore,

$$\begin{aligned}
(X^\bullet)^\bullet &= (X^\bullet)^\circ && (2.1)\\
&= (\text{cone}(X)^\circ)^\circ && (2.2)\\
&= \text{cone}(X)^{\circ\circ} \\
&= \text{cl}(\text{conv}(\text{cone}(X) \cup \{0\})) && (2.3)\\
&= \text{cl}(\text{cone}(X)),
\end{aligned}$$

where (6.1) follows from Exercise 15, (2.2) follows from Remark 4, (2.3) follows from Proposition 2.4.9 (2) in lecture notes. □

3 Section-Sept-16-2022

3.1 Review

Question 1. Let $C \subseteq \mathbb{R}^d$ be a closed convex set, is every face of C an exposed face?
Answer. No. In $\mathbb{R}^2$, let $X = \text{conv}(\{(-2,0),(0,0),(0,2),(-2,2)\})$, $Y = \{(x,y) : x^2 + (y-1)^2 \leq 1\}$, then $X \cup Y$ has a face $\{(0,0)\}$ which is not an exposed face.

Question 2. Let $X,\ Y \subseteq \mathbb{R}^d$ be closed sets, is $X + Y$ a closed set?
Answer. It's not always true. Some counterexamples:

1. $X = \{n + 2^{-n} : n \in \mathbb{N}_+\}$, $Y = \mathbb{Z}$ are closed sets in $\mathbb{R}$. Observe that a convergent sequence $\{2^{-n}\}_{n=1}^{+\infty} \subseteq X + Y$, but $\lim_{n\to+\infty} 2^{-n} = 0 \notin X + Y$.

2. $X = \{(x,y) : xy \geq 1\}$, $Y = \{(0,y) : y \in \mathbb{R}\}$ are closed sets in $\mathbb{R}^2$. However, $X + Y = \{(x,y) : x \neq 0\}$ is not closed.

Question 3. Let $X \subseteq \mathbb{R}^d$ be a convex set, is every face of X a closed set?
Answer. It's not always true. Let X be an open unit ball in $\mathbb{R}^d$, then a trivial face is X itself, which is not a closed set.

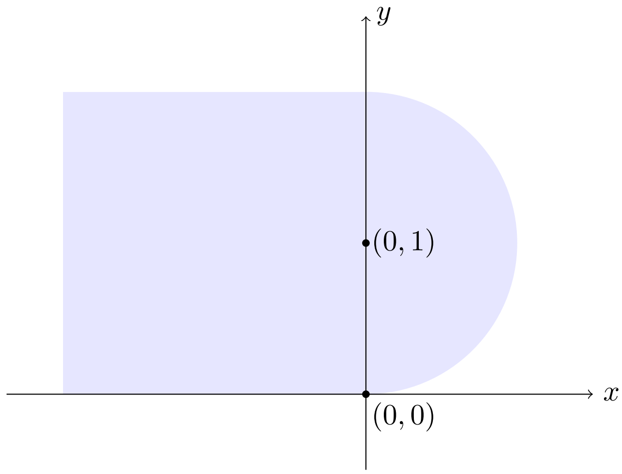


Figure 3: A counterexample for Question 1.

3.2 Exercises

Exercise 18. Show that if X is compact and Y is closed, then $X+Y$ is closed.

Proof. Consider any convergent sequence $\mathbf{x}^i+\mathbf{y}^i \to \mathbf{z}$ such that $\mathbf{x}^i \in X,\ \mathbf{y}^i \in Y$ for all $i \in \mathbb{N}_+$. We need to show that $\mathbf{z} \in X+Y$. Since X is compact, it has a convergent subsequence, and there is some $\mathbf{x} \in X$ such that $\mathbf{x}^{i_k} \to \mathbf{x}$. Along this subsequence, we have that

$$\lim_{k\to+\infty}(\mathbf{x}^{i_k}+\mathbf{y}^{i_k}) = \mathbf{z},$$

so $\mathbf{y}^{i_k} \to \mathbf{z}-\mathbf{x}$ as $k \to +\infty$. Since Y is closed, $\mathbf{z}-\mathbf{x} \in Y$, so we get $\mathbf{z} = \mathbf{x}+(\mathbf{z}-\mathbf{x}) \in X+Y$. □

Exercise 19 (Equivalent definition of face)**.** Let $C \subseteq \mathbb{R}^d$ be a convex set. Then a convex subset $F \subseteq C$ is a face if and only if $\forall \mathbf{z} \in F$, $\mathbf{z} = \lambda\mathbf{x}^1+(1-\lambda)\mathbf{x}^2$ for some $\mathbf{x}^1, \mathbf{x}^2 \in C$ and $\lambda \in (0,1)$ implies $\mathbf{x}^1, \mathbf{x}^2 \in F$.

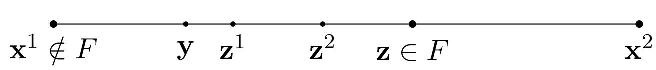
Proof. ($\Longleftarrow$) This is a trivial case by taking $\lambda = \frac{1}{2}$.

($\Longrightarrow$) Consider any $\mathbf{x}^1, \mathbf{x}^2 \in C,\ \lambda \in (0,1)$ with $\mathbf{z} = \lambda\mathbf{x}^1+(1-\lambda)\mathbf{x}^2 \in F$, and WLOG we suppose $\mathbf{x}^1 \in C\backslash F$. Now if $\mathbf{z}^1 = (\mathbf{x}^1+\mathbf{z})/2 \in F$, we would have a contradiction, so we may construct $\mathbf{z}^k = (\mathbf{z}^{k-1}+\mathbf{z})/2 \in C\backslash F$ for all $k \geq 1$. Then if for some $\lambda' \in (\lambda, 1)$, $\mathbf{y} = \lambda'\mathbf{x}^1+(1-\lambda')\mathbf{x}^2 \in F$, the line segment connecting $\mathbf{y}$ and $\mathbf{z}$ would have to lie in F by convexity, and this would contain $\mathbf{z}^k$ for some $k \geq 1$. Thus $\{\lambda'\mathbf{x}^1+(1-\lambda')\mathbf{x}^2 : \lambda' \in (\lambda,1)\} \subseteq C\backslash F$. However,

$$\mathbf{z} = \frac{\lambda_1\mathbf{x}^1+(1-\lambda_1)\,\mathbf{x}^2}{2}+\frac{\lambda_2\mathbf{x}^1+(1-\lambda_2)\,\mathbf{x}^2}{2} \in F,$$

where $\lambda_1 = \lambda-\varepsilon,\ \lambda_2 = \lambda+\varepsilon$ with $\varepsilon = \frac{1}{2}\min\{\lambda, 1-\lambda\}$. This leads to a contradiction since the second vector is not in F.

□



Exercise 20. Let $C \subseteq \mathbb{R}^d$ be a convex set and let $X \subseteq C$ be convex. Let $\mathbf{x} \in \operatorname{relint}(X)$. Suppose F is a face of C such that $\mathbf{x} \in F$. Prove that $X \subseteq F$.

Proof. $\forall \mathbf{y} \in X$, observe that $2\mathbf{x} - \mathbf{y} \in \operatorname{aff}(X)$ since $\mathbf{x},\ \mathbf{y} \in X$ and $2 + (-1) = 1$. Since $\mathbf{x} \in \operatorname{relint}(X)$, there is some $\varepsilon > 0$ such that $\mathbf{z} = \mathbf{x} + \varepsilon((2\mathbf{x} - \mathbf{y}) - \mathbf{x}) \in X$. Therefore,

$$\mathbf{x} = \frac{1}{1+\varepsilon}\mathbf{z} + \frac{\varepsilon}{1+\varepsilon}\mathbf{y},$$

so $\mathbf{y} \in F$, since F is a face. This proves that $X \subseteq F$. □

Exercise 21. Let $C \subseteq \mathbb{R}^d$ be a compact, convex set. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a linear function given by $f(\mathbf{y}) = \langle \mathbf{a}, \mathbf{y} \rangle$ for some $\mathbf{a} \in \mathbb{R}^d$. Show that there exists an extreme point $\mathbf{v} \in C$ such that $f(\mathbf{v}) \le f(\mathbf{x})$ for every $\mathbf{x} \in C$.

Proof. f is a continuous function by definition, then Weierstrass theorem yields that there is some point $\mathbf{u} \in C$ such that $f(\mathbf{u}) \le f(\mathbf{x})$ for all $\mathbf{x} \in C$. Since C is compact and convex, by Krein-Milman $C = \operatorname{conv}(\operatorname{ext}(C))$. Then for some $\mathbf{v}^i \in \operatorname{ext}(C)$ and $\lambda_i \ge 0$ summing to 1, we have $\mathbf{u} = \sum_{i=1}^k \lambda_i \mathbf{v}^i$ and

$$\langle \mathbf{a}, \mathbf{u} \rangle = \sum_{i=1}^{k} \lambda_i \left\langle \mathbf{a}, \mathbf{v}^i \right\rangle \ge \sum_{i=1}^{k} \lambda_i \langle \mathbf{a}, \mathbf{u} \rangle = \langle \mathbf{a}, \mathbf{u} \rangle.$$

Thus the inequality must be satisfied with equality, and $\langle \mathbf{a}, \mathbf{v}^i \rangle = \langle \mathbf{a}, \mathbf{u} \rangle$ for each of these $\mathbf{v}^i$. □

Exercise 22. Show that if D is a closed convex cone, then any face of D is a convex cone.

Proof. Let $F \subseteq D$ be a face. We first want to show that $0 \in F$. Let $\mathbf{x} \in F$ and $\mu > 1$, then

$$\mathbf{x} = \frac{1}{\mu}(\mu \mathbf{x}) + \frac{\mu - 1}{\mu} 0,$$

so $\mu \mathbf{x}, 0 \in D$ since F is a face.

Since F is convex, then we just need to show that $\forall \mathbf{x} \in F,\ \lambda \ge 0$, we have that $\lambda \mathbf{x} \in F$. We have shown the case when $\lambda > 1$ when arguing that $0 \in F$. Now suppose $\lambda \in (0, 1]$, then

$$\lambda \mathbf{x} = \lambda \underbrace{\mathbf{x}}_{\in F} + (1 - \lambda) \underbrace{0}_{\in F} \in F$$

by convexity, which completes the proof. □

4 Section-Sept-23-2022

Exercise 23. Let P_1, P_2 be two polytopes in $\mathbb{R}^d$. Show that $P_1 + P_2$ is a polytope.

Proof. By Minkowski-Weyl theorem, $P_1 = \operatorname{conv}\left(\{\mathbf{u}^1, \dots, \mathbf{u}^m\}\right), P_2 = \operatorname{conv}\left(\{\mathbf{v}^1, \dots, \mathbf{v}^n\}\right)$ for some $\mathbf{u}^i, \mathbf{v}^j \in \mathbb{R}^d$. To prove $P_1 + P_2$ is a polytope, we prove that

$$P_1 + P_2 = \operatorname{conv}\left(\{\mathbf{u}^i + \mathbf{v}^j\}_{(i,j) \in \{1,\dots,m\} \times \{1,\dots,n\}}\right).$$

$\forall \mathbf{x} \in P_1 + P_2$, $\mathbf{x} = \mathbf{u} + \mathbf{v}$ for some $\mathbf{u} \in P_1, \mathbf{v} \in P_2$. Then by convexity, there exist some $\alpha_i, \beta_j \ge 0$ with

$\sum_{i=1}^m \alpha_i = 1, \sum_{j=1}^n \beta_j = 1$ such that $\mathbf{u} = \sum_{i=1}^m \alpha_i \mathbf{u}^i,\ \mathbf{v} = \sum_{j=1}^n \beta_j \mathbf{v}^j$. Therefore,

$$\begin{aligned}
\mathbf{x} = \mathbf{u} + \mathbf{v} &= \sum_{i=1}^m \alpha_i \mathbf{u}^i + \sum_{j=1}^n \beta_j \mathbf{v}^j \\
&= \sum_{i=1}^m \alpha_i \mathbf{u}^i \underbrace{\sum_{j=1}^n \beta_j}_{=1} + \underbrace{\sum_{i=1}^m \alpha_i}_{=1} \sum_{j=1}^n \beta_j \mathbf{v}^j \\
&= \sum_{i=1}^m \left(\alpha_i \mathbf{u}^i \sum_{j=1}^n \beta_j + \alpha_i \sum_{j=1}^n \beta \mathbf{v}^j \right) \\
&= \sum_{i=1}^m \sum_{j=1}^n \alpha_i \beta_j \left(\mathbf{u}^i + \mathbf{v}^j \right).
\end{aligned}$$

Observe that $\sum_{i=1}^m \sum_{j=1}^n \alpha_i\beta_j = \sum_{i=1}^m \alpha_i \sum_{j=1}^n \beta_j = 1$ with $\alpha_i\beta_j \geq 0$, then $\mathbf{x}$ is a convex combination of points in $\{\mathbf{u}^i + \mathbf{v}^j\}_{(i,j)\in\{1,\dots,m\}\times\{1,\dots,n\}}$, so $\mathbf{x} \in \operatorname{conv}\left(\{\mathbf{u}^i + \mathbf{v}^j\}_{(i,j)\in\{1,\dots,m\}\times\{1,\dots,n\}}\right)$.

To show the reverse inclusion, $\forall \mathbf{x} \in \operatorname{conv}\left(\{\mathbf{u}^i + \mathbf{v}^j\}_{(i,j)\in\{1,\dots,m\}\times\{1,\dots,n\}}\right)$, there exist $\lambda_{ij} \geq 0$ with $\sum_{i=1}^m \sum_{j=1}^n = 1$ such that $\mathbf{x} = \sum_{i=1}^m \sum_{j=1}^n \lambda_{ij} \left(\mathbf{u}^i + \mathbf{v}^j\right)$.

Then

$$\begin{aligned}
\mathbf{x} &= \sum_{i=1}^m \sum_{j=1}^n \lambda_{ij} \left(\mathbf{u}^i + \mathbf{v}^j\right) \\
&= \sum_{i=1}^m \sum_{j=1}^n \lambda_{ij} \mathbf{u}^i + \sum_{j=1}^n \sum_{i=1}^m \lambda_{ij} \mathbf{v}^j \\
&= \underbrace{\sum_{i=1}^m \alpha_i \mathbf{u}^i}_{\in P_1} + \underbrace{\sum_{j=1}^n \beta_j \mathbf{v}^j}_{\in P_2} \in P_1 + P_2,
\end{aligned}$$

where $\alpha_i = \sum_{j=1}^n \lambda_{ij} \geq 0,\ i = 1, \dots, m$ and $\beta_j = \sum_{j=1}^m \lambda_{ij} \geq 0,\ j = 1, \dots, n$ with

$$\sum_{i=1}^m \alpha_i = \sum_{j=1}^n \beta_j = \sum_{i=1}^m \sum_{j=1}^n \lambda_{ij} = 1.$$

Minkowski-Weyl theorem implies that $P_1 + P_2$ is a polytope since it's a convex hull of finitely many points. □

Exercise 24. Let $A \in \mathbb{R}^{m\times d}, \mathbf{b} \in \mathbb{R}^m$. Consider the polyhedron $P = \{\mathbf{x} \in \mathbb{R}^d : A\mathbf{x} \leq \mathbf{b}\}$. Show that

$$\operatorname{rec}(P) = \{\mathbf{x} \in \mathbb{R}^d : A\mathbf{x} \leq 0\}, \quad \operatorname{lin}(P) = \{\mathbf{x} \in \mathbb{R}^d : A\mathbf{x} = 0\}.$$

Proof. $\forall \mathbf{w} \in \{\mathbf{x} \in \mathbb{R}^d : A\mathbf{x} \leq 0\}$. For any $\lambda \geq 0$ and $\mathbf{x} \in P$,

$$A(\mathbf{x} + \lambda\mathbf{w}) = A\mathbf{x} + \lambda A\mathbf{w} \leq \mathbf{b} + \lambda 0 = \mathbf{b},$$

so $\mathbf{w} \in \mathrm{rec}(P)$. This proves $\{\mathbf{x} \in \mathbb{R}^d : A\mathbf{x} \le 0\} \subseteq \mathrm{rec}(P)$.

To show reverse inclusion, $\forall \mathbf{y} \notin \{\mathbf{x} \in \mathbb{R}^d : A\mathbf{x} \le 0\}$, there exists $i \in \{1, \dots, m\}$ such that

$$\alpha = (A\mathbf{y})_i > 0.$$

Suppose that $\mathbf{x} \in P$, and let $\beta = (A\mathbf{x})_i$. Consider $\lambda = (\mathbf{b}_i + 1 - \beta)/\alpha$, then $\lambda > 0$ since $\beta \le \mathbf{b}_i$. Therefore,

$$\begin{aligned}(A(\mathbf{x} + \lambda\mathbf{y}))_i &= (A\mathbf{x})_i + \lambda(A\mathbf{y})_i \\ &= \beta + \lambda\alpha \\ &= \beta + (\mathbf{b}_i + 1 - \beta) \\ &= \mathbf{b}_i + 1 > \mathbf{b}_i,\end{aligned}$$

so $\mathbf{x} + \lambda\mathbf{y} \notin P$, thus $\mathbf{y} \notin \mathrm{rec}(P)$.

To show $\mathrm{lin}(P) = \{\mathbf{x} \in \mathbb{R}^d : A\mathbf{x} = 0\}$, by definition of lineality space,

$$\mathrm{lin}(P) = \mathrm{rec}(P) \cap -\mathrm{rec}(P) = \{\mathbf{x} \in \mathbb{R}^d : A\mathbf{x} \le 0, A\mathbf{x} \ge 0\} = \{\mathbf{x} \in \mathbb{R}^d : A\mathbf{x} = 0\}.$$

□

Exercise 25. Let $A \in \mathbb{R}^{m \times d}, \mathbf{b} \in \mathbb{R}^m$. Consider the polyhedron $P = \{\mathbf{x} \in \mathbb{R}^d : A\mathbf{x} \le \mathbf{b}\}$. Show that P is bounded if and only if $\mathrm{cone}(\{\mathbf{a}^1, \dots, \mathbf{a}^m\}) = \mathbb{R}^d$, where $\mathbf{a}^i$ is the ith row of A.

Proof. Notice that P is closed and convex since P is the intersection of halfspaces, then Theorem 2.4.21 in lecture notes yields that

$$P \text{ is bounded} \iff \mathrm{rec}(P) = \{0\},$$

so we just need to show

$$\mathrm{rec}(P) = \{0\} \iff \mathrm{cone}(\{\mathbf{a}^1, \dots, \mathbf{a}^m\}) = \mathbb{R}^d.$$

Let $X = \{\mathbf{a}^1, \dots, \mathbf{a}^m\}$, by Exercise 6 in session 2, we have

$$(X^\bullet)^\bullet = \mathrm{cl}(\mathrm{cone}(X)) = \mathrm{cone}(X),$$

where the second equality follows from Proposition 2.2.15 in lecture notes, and one can easily derive that

$$(X^\bullet)^\bullet = \mathbb{R}^d \iff X^\bullet = \{0\}.$$

Therefore, we just need to show

$$\mathrm{rec}(P) = \{0\} \iff X^\bullet = \{0\}.$$

Observe that $\mathrm{rec}(P) = X^\bullet$ since Exercise 24 shows $\mathrm{rec}(P) = \{\mathbf{x} \in \mathbb{R}^d : A\mathbf{x} \le 0\} = \{\mathbf{x} \in \mathbb{R}^d : \langle \mathbf{a}^i, \mathbf{x} \rangle \le 0,\ i = 1, \dots, m\}$, this completes the proof. □

Exercise 26. Let $C \subseteq \mathbb{R}^d$ be a compact, convex set. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a linear function given by $f(\mathbf{y}) = \langle \mathbf{a}, \mathbf{y} \rangle$ for some $\mathbf{a} \in \mathbb{R}^d$. Show that there exists an extreme point $\mathbf{v} \in C$ such that $f(\mathbf{v}) \le f(\mathbf{x})$ for every $\mathbf{x} \in C$.

Proof. f is a continuous function by definition, then Weierstrass theorem yields that there is some point $\mathbf{u} \in C$ such that $f(\mathbf{u}) \le f(\mathbf{x})$ for all $\mathbf{x} \in C$. Since C is compact and convex, by Krein-Milman $C = \mathrm{conv}\,(\mathrm{ext}(C))$. Then for some $\mathbf{v}^i \in \mathrm{ext}(C)$ and $\lambda_i \ge 0$ summing to 1, we have $\mathbf{u} = \sum_{i=1}^k \lambda_i \mathbf{v}^i$ and

$$\langle \mathbf{a}, \mathbf{u} \rangle = \sum_{i=1}^k \lambda_i \left\langle \mathbf{a}, \mathbf{v}^i \right\rangle \ge \sum_{i=1}^k \lambda_i \langle \mathbf{a}, \mathbf{u} \rangle = \langle \mathbf{a}, \mathbf{u} \rangle.$$

Thus the inequality must be satisfied with equality, and $\langle \mathbf{a}, \mathbf{v}^i \rangle = \langle \mathbf{a}, \mathbf{u} \rangle$ for each of these $\mathbf{v}^i$. □

5 Section-Sept-30-2022

Recall Farkas' lemma:

Theorem 6 (Farkas' Lemma)**.** *Let $A \in \mathbb{R}^{d\times n}$ and $\mathbf{b} \in \mathbb{R}^d$. Exactly one of the following is true.*

1. *$A\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq 0$ has a solution.*
2. *$\exists \mathbf{u} \in \mathbb{R}^d$ such that $\mathbf{u}^T A \leq 0$ and $\mathbf{u}^T\mathbf{b} > 0$.*

Exercise 27. Let $A \in \mathbb{R}^{d\times n}, \mathbf{b} \in \mathbb{R}^d$. Exactly one of the following is true.

1. $A\mathbf{x} \leq \mathbf{b}$ has a solution.
2. $\exists \mathbf{u} \geq 0$ such that $\mathbf{u}^T A = 0$ and $\mathbf{u}^T\mathbf{b} < 0$.

Proof. Let $\mathbf{x}_i^+ = \max\{0, \mathbf{x}_i\} \geq 0,\ \mathbf{x}_i^- = -\min\{0, \mathbf{x}_i\} \geq 0,\ i = 1, \ldots, n$. Then $\mathbf{x} = \mathbf{x}^+ - \mathbf{x}^-$, and

$$\begin{aligned} A\mathbf{x} \leq \mathbf{b} \text{ has no solution} &\iff A\mathbf{x} + \mathbf{s} = \mathbf{b},\ \mathbf{s} \geq 0 \text{ has no solution} \\ &\iff A(\mathbf{x}^+ - \mathbf{x}^-) + \mathbf{s} = \mathbf{b},\ \mathbf{s}, \mathbf{x}^+, \mathbf{x}^- \geq 0 \text{ has no solution} \\ &\iff \begin{bmatrix} A & -A & \mathbf{I} \end{bmatrix} \mathbf{y} = \mathbf{b},\ \mathbf{y} \geq 0 \text{ has no solution} \\ &\iff \exists \tilde{\mathbf{u}} \in \mathbb{R}^d \text{ such that } \tilde{\mathbf{u}}^T \begin{bmatrix} A & -A & \mathbf{I} \end{bmatrix} \leq 0 \text{ and } \tilde{\mathbf{u}}^T\mathbf{b} > 0 \end{aligned}$$

that is, $\tilde{\mathbf{u}}^T A \leq 0,\ -\tilde{\mathbf{u}}^T A \leq 0,\ \tilde{\mathbf{u}}^T \mathbf{I} \leq 0,\ \tilde{\mathbf{u}}^T\mathbf{b} > 0$, which gives $\tilde{\mathbf{u}}^T A = 0,\ \tilde{\mathbf{u}} \leq 0,\ \tilde{\mathbf{u}}^T\mathbf{b} > 0$. Then $\mathbf{u} = -\tilde{\mathbf{u}}$ is the desired vector. □

Exercise 28. Let $A \in \mathbb{R}^{d\times n}$. Exactly one of the following is true.

1. $A\mathbf{x} = 0,\ \mathbf{x} \geq 0$ has a nontrivial solution.
2. $\mathbf{y}^T A > 0$ has a solution.

Proof. We just need to prove

$$\exists \mathbf{x} \neq 0, \mathbf{x} \geq 0 \text{ such that } A\mathbf{x} = 0 \iff \nexists \mathbf{y} \in \mathbb{R}^d \text{ satisfying } \mathbf{y}^T A > 0.$$

Consider any fixed vector $\mathbf{b} > 0 \in \mathbb{R}^n$, then

$$\begin{aligned} \nexists \mathbf{y} \in \mathbb{R}^d \text{ satisfying } \mathbf{y}^T A > 0 &\iff \nexists \mathbf{y} \in \mathbb{R}^d \text{ such that } \mathbf{y}^T A \geq \mathbf{b} \\ &\iff A^T(-\mathbf{y}) \leq -\mathbf{b} \text{ has no solution} \\ &\overset{Ex.32}{\iff} \exists \mathbf{x} \geq 0 \text{ such that } \mathbf{x}^T A^T = 0 \text{ and } \mathbf{x}^T(-\mathbf{b}) < 0 \\ &\iff \exists \mathbf{x} \geq 0 \text{ such that } A\mathbf{x} = 0 \text{ and } \mathbf{x}^T\mathbf{b} > 0 \\ &\iff \exists \mathbf{x} \neq 0,\ \mathbf{x} \geq 0 \text{ such that } A\mathbf{x} = 0. \end{aligned}$$

□

Exercise 29. Let $C \subseteq \mathbb{R}^d$ be a nonempty closed convex set. Then C has at least one extreme point if and only if C is pointed.

Proof. Let $\mathbf{x}$ be an extreme point of C. Suppose C is not pointed, then for any $\mathbf{r} \in \text{lin}(C)\backslash\{0\}$, notice that $\mathbf{x}+\mathbf{r},\ \mathbf{x}-\mathbf{r} \in C$ and $\frac{1}{2}((\mathbf{x}+\mathbf{r})+(\mathbf{x}-\mathbf{r})) = \mathbf{x}$, contradicting that $\mathbf{x}$ is an extreme point.

Conversely, we prove by induction on the dimension of the space to show that if C does not contain a line, then it must have an extreme point. It is trivial for the case when $d = 1$, so assume it is true for $d-1$ with $d \geq 2$. Then for any nonempty closed convex set $C \subseteq \mathbb{R}^d$, there must exist points $\mathbf{x} \in C$ and $\mathbf{y} \notin C$ since $\text{lin}(C) = \{0\}$. The line segment connecting $\mathbf{x}$ and $\mathbf{y}$ intersects the relative boundary of C at some point $\bar{\mathbf{x}}$ (see Figure 5). Consider a supporting hyperplane H of C passing through $\bar{\mathbf{x}}$, then $C \cap H$ lies in a $(d-1)$-dimensional space and does not contain a line. Hence, by induction hypothesis, $C \cap H$ must have an extreme point, and this extreme point must also be an extreme point of C. ☐

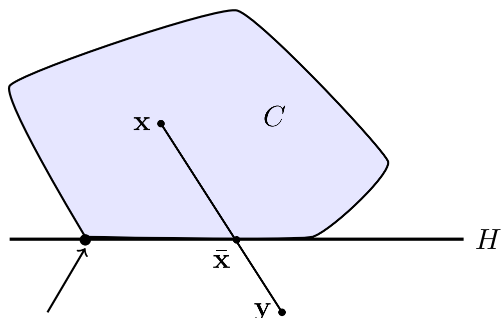


Figure 4: Illustration of the proof in Exercise 29

Corollary 1. *Every nonempty polyhedron* $P = \{\mathbf{x} \in \mathbb{R}^d : A\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq 0\}$ *has at least one extreme point.*

Proof. We just need to show that P is pointed. Observe that

$$P = \{\mathbf{x} \in \mathbb{R}^d : A\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq 0\} = \left\{\mathbf{x} \in \mathbb{R}^d : \begin{bmatrix} A \\ -A \\ -\mathbf{I} \end{bmatrix} \mathbf{x} \leq \begin{bmatrix} \mathbf{b} \\ -\mathbf{b} \\ \mathbf{0} \end{bmatrix} \right\},$$

then by Exercise 2 in Section 4,

$$\text{lin}(P) = \left\{\mathbf{x} \in \mathbb{R}^d : \begin{bmatrix} A \\ -A \\ -\mathbf{I} \end{bmatrix} \mathbf{x} = 0 \right\} = \{0\}.$$

☐

Exercise 30. Let $A \in \mathbb{R}^{m \times d}, \mathbf{b} \in \mathbb{R}^m$. Consider the polyhedron $P = \{\mathbf{x} \in \mathbb{R}^d : A\mathbf{x} \leq \mathbf{b}\}$. Suppose there is some $\bar{\mathbf{x}} \in \mathbb{R}^d$ such that $A\bar{\mathbf{x}} < b$, show that $\dim(P) = d$.

Proof. For any $i \in \{1, \ldots, d\}$, there exists $\varepsilon_i > 0$ such that $A(\bar{\mathbf{x}} + \varepsilon_i \mathbf{e}^i) \leq \mathbf{b}$, where $\mathbf{e}^i$ is the ith standard unit vector. Then

$$\{\bar{\mathbf{x}}, \bar{\mathbf{x}} + \varepsilon_1 \mathbf{e}^1, \ldots, \bar{\mathbf{x}} + \varepsilon_d \mathbf{e}^d\}$$

is a set of $d+1$ affinely independent points in P, so $\dim(P) = d$. ☐

6 Section-Oct-7-2022

Definition 5 (Lower semicontinuous)**.** A function $f : D \to \mathbb{R} \cup \{\pm\infty\}$ is called *lower semicontinuous at* $\mathbf{x} \in D$ if

$$f(\mathbf{x}) \le \liminf_{k\to\infty} f(\mathbf{x}_k)$$

for every sequence $\{\mathbf{x}_k\} \subseteq D$ with $\mathbf{x}_k \to \mathbf{x}$. We say that f is *lower semicontinuous* if it is lower semicontinuous at each point $\mathbf{x}$ in its domain D. We say that f is *upper semicontinuous* if $-f$ is lower semicontinuous.

Exercise 31. For a function $f : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$, the following are equivalent:

1. $f_\gamma = \{\mathbf{x} \in \mathbb{R}^d : f(\mathbf{x}) \le \gamma\}$ is closed $\forall \gamma \in \mathbb{R}$.
2. f is lower semicontinuous.
3. $\mathrm{epi}(f) = \{(\mathbf{x}, t) \in \mathbb{R}^d \times \mathbb{R} : f(\mathbf{x}) \le t\}$ is closed.

Proof. If $f(\mathbf{x}) = \infty$ for all $\mathbf{x}$, the result trivially holds. We thus assume that $f(\mathbf{x}) < \infty$ for at least one $\mathbf{x} \in \mathbb{R}^d$, so that $\mathrm{epi}(f)$ is nonempty and there exist level sets of f that are nonempty.
<u>(i) $\implies$ (ii)</u>. Assume that the level set f_γ is closed for every scalar γ. Suppose to the contrary that

$$f(\bar{\mathbf{x}}) > \liminf_{k\to\infty} f(\mathbf{x}_k)$$

for some $\bar{\mathbf{x}}$ and sequence $\{\mathbf{x}_k\}$ converging to $\bar{\mathbf{x}}$, and let $\bar{\gamma}$ be a scalar such that

$$f(\bar{\mathbf{x}}) > \bar{\gamma} > \liminf_{k\to\infty} f(\mathbf{x}_k).$$

Then there exists a subsequence $\{\mathbf{x}_{k_i}\}$ such that $f(\mathbf{x}_{k_i}) \le \bar{\gamma}$ for all $i \in \mathbb{N}_+$, so that $\{\mathbf{x}_{k_i}\} \subseteq f_{\bar{\gamma}}$. Since $f_{\bar{\gamma}}$ is closed, $\bar{\mathbf{x}}$ must also belong to $f_{\bar{\gamma}}$, so $f(\bar{\mathbf{x}}) \le \bar{\gamma}$, which leads to a contradiction.
<u>(ii) $\implies$ (iii)</u>. Assume that f is lower semicontinuous over $\mathbb{R}^d$, and let $\bar{\mathbf{x}}, \bar{t}$ be the limit of a sequence

$$\{(\mathbf{x}_k, t_k)\} \subseteq \mathrm{epi}(f).$$

Then we have $f(\mathbf{x}_k) \le t_k$, and by taking the limit as $k \to \infty$ and by using the lower semicontinuity of f at $\bar{\mathbf{x}}$, we obtain

$$f(\bar{\mathbf{x}}) \le \liminf_{k\to\infty} f(\mathbf{x}_k) \le \bar{t}.$$

Hence, $(\bar{\mathbf{x}}, \bar{t}) \in \mathrm{epi}(f)$ and $\mathrm{epi}(f)$ is closed.
<u>(iii) $\implies$ (i)</u>. Assume that $\mathrm{epi}(f)$ is closed and let $\{\mathbf{x}_k\}$ be a sequence that converges to some $\bar{\mathbf{x}}$ and belongs to f_γ for some $\gamma \in \mathbb{R}$. Then $(\mathbf{x}_k, \gamma) \in \mathrm{epi}(f)$ for all k and $(\mathbf{x}_k, \gamma) \to (\bar{\mathbf{x}}, \gamma)$, so since $\mathrm{epi}(f)$ is closed, we have $(\bar{\mathbf{x}}, \gamma) \in \mathrm{epi}(f)$. Hence, $\bar{\mathbf{x}}$ belongs to f_γ, implying that this set is closed. ☐

Exercise 32. Consider any convex function $f : \mathbb{R}^d \to \mathbb{R}$. Fix any $\bar{\mathbf{x}} \in \mathbb{R}^d$ and suppose $\mathbf{x}^* \in \mathbb{R}^d$ globally minimizes

$$\min_{\mathbf{x}\in\mathbb{R}^d} f(\mathbf{x}) + \frac{\rho}{2}\|\mathbf{x} - \bar{\mathbf{x}}\|_2^2$$

for some $\rho > 0$. Prove that $\rho(\bar{\mathbf{x}} - \mathbf{x}^*) \in \partial f(\mathbf{x}^*)$ is a subgradient of f at $\mathbf{x}^*$.

Proof. Let $\mathbf{g} = \rho(\bar{\mathbf{x}} - \mathbf{x}^*)$, then by definition we just need to prove that f is lower bounded by the following linear function for all $\mathbf{x} \in \mathbb{R}^d$

$$f(\mathbf{x}) \ge f(\mathbf{x}^*) + \mathbf{g}^T(\mathbf{x} - \mathbf{x}^*).$$

We can prove this directly. For any $\mathbf{x} \in \mathbb{R}^d$, we have

$$\begin{aligned} f(\mathbf{x}^*) + \frac{\rho}{2}\|\mathbf{x}^* - \bar{\mathbf{x}}\|_2^2 &\leq f(\mathbf{x}) + \frac{\rho}{2}\|\mathbf{x} - \bar{\mathbf{x}}\|_2^2 \\ &= f(\mathbf{x}) + \frac{\rho}{2}\|\mathbf{x} - \mathbf{x}^* + \mathbf{x}^* - \bar{\mathbf{x}}\|_2^2 \\ &= f(\mathbf{x}) + \underbrace{\rho(\mathbf{x}^* - \bar{\mathbf{x}})^T}_{\mathbf{g}^T}(\mathbf{x} - \mathbf{x}^*) + \frac{\rho}{2}\|\mathbf{x} - \mathbf{x}^*\|_2^2 + \frac{\rho}{2}\|\mathbf{x}^* - \bar{\mathbf{x}}\|_2^2 \end{aligned}$$

Cancelling the last quadratic from both sides gives a weaker result than needed:

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \mathbf{g}^T(\mathbf{x} - \mathbf{x}^*) - \frac{\rho}{2}\|\mathbf{x} - \mathbf{x}^*\|_2^2, \quad \forall \mathbf{x} \in \mathbb{R}^d. \tag{6.1}$$

For any $\lambda \in (0, 1]$, applying Equation (6.1) at $\lambda\mathbf{x} + (1-\lambda)\mathbf{x}^*$, one can obtain that

$$\begin{aligned} f(\lambda\mathbf{x} + (1-\lambda)\mathbf{x}^*) &\geq f(\mathbf{x}^*) + \mathbf{g}^T(\lambda\mathbf{x} + (1-\lambda)\mathbf{x}^* - \mathbf{x}^*) - \frac{\rho}{2}\|\lambda\mathbf{x} + (1-\lambda)\mathbf{x}^* - \mathbf{x}^*\|_2^2 \\ &= f(\mathbf{x}^*) + \mathbf{g}^T(\lambda\mathbf{x} - \lambda\mathbf{x}^*) - \frac{\rho}{2}\|\lambda\mathbf{x} - \lambda\mathbf{x}^*\|_2^2 \\ &= f(\mathbf{x}^*) + \lambda\mathbf{g}^T(\mathbf{x} - \mathbf{x}^*) - \lambda^2\frac{\rho}{2}\|\mathbf{x} - \mathbf{x}^*\|_2^2. \end{aligned}$$

Using convexity of f, we can strengthen this since

$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{x}^*) \leq \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{x}^*),$$

therefore,

$$\lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{x}^*) \geq f(\mathbf{x}^*) + \lambda\mathbf{g}^T(\mathbf{x} - \mathbf{x}^*) - \lambda^2\frac{\rho}{2}\|\mathbf{x} - \mathbf{x}^*\|_2^2,$$

that is,

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \mathbf{g}^T(\mathbf{x} - \mathbf{x}^*) - \lambda\frac{\rho}{2}\|\mathbf{x} - \mathbf{x}^*\|_2^2.$$

Taking the limit as $\lambda \to 0$ gives the claim. □

Exercise 33. For any μ-strongly convex f with global minimizer $\mathbf{x}^*$, show that $\forall \mathbf{x} \in \mathbb{R}^d$ has

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \frac{\mu}{2}\|\mathbf{x} - \mathbf{x}^*\|_2^2.$$

Proof. Let $h(\mathbf{x}) := f(\mathbf{x}) - \frac{\mu}{2}\|\mathbf{x}\|_2^2$, which is convex by assumption. Then $\mathbf{x}^*$ minimizes $f(\mathbf{x}) = h(\mathbf{x}) + \frac{\mu}{2}\|\mathbf{x} - 0\|_2^2$. This is precisely the shape of the problem considered in Exercise 32. Letting $\rho = \mu$, $\bar{\mathbf{x}} = 0$, we then know for any $\mathbf{x} \in \mathbb{R}^d$,

$$\begin{aligned} h(\mathbf{x}) &\geq h(\mathbf{x}^*) + \rho(0 - \mathbf{x}^*)^T(\mathbf{x} - \mathbf{x}^*) \\ &= h(\mathbf{x}^*) + \frac{\rho}{2}\|\mathbf{x}^*\|_2^2 + \frac{\rho}{2}\|\mathbf{x} - \mathbf{x}^*\|_2^2 - \frac{\rho}{2}\|\mathbf{x}\|_2^2. \end{aligned}$$

Therefore,

$$\underbrace{h(\mathbf{x}) + \frac{\rho}{2}\|\mathbf{x}\|_2^2}_{f(\mathbf{x})} \geq \underbrace{h(\mathbf{x}^*) + \frac{\rho}{2}\|\mathbf{x}^*\|_2^2}_{f(\mathbf{x}^*)} + \frac{\rho}{2}\|\mathbf{x} - \mathbf{x}^*\|_2^2.$$

□

Remark 5. Exercise 32 and Exercise 33 did not require differentiability of f.

7 Section-Oct-28-2022

In lecture notes (Theorem 3.4.5) we presented an important result that $\partial(f+g)(\mathbf{x}) = \partial f(\mathbf{x}) + \partial g(\mathbf{x})$ for convex functions $f, g : \mathbb{R}^d \to \mathbb{R}$. We will prove this in today's section. More precisely, let $E, Y \subseteq \mathbb{R}^d$, for any convex functions $f\colon E \to \mathbb{R} \cup \{+\infty\}$ and $g\colon Y \to \mathbb{R} \cup \{+\infty\}$ and linear map $A\colon E \to Y$ with $0 \in \operatorname{int}(\operatorname{dom} g - A\operatorname{dom} f)$, then $\partial(f + g \circ A)(\mathbf{x}) = \partial f(\mathbf{x}) + A^* \partial g(A\mathbf{x})$.

Exercise 34 (Subdifferential Calculus)**.**

(a) Let $E, Y \subseteq \mathbb{R}^d$, then for any convex functions $f\colon E \to \mathbb{R} \cup \{+\infty\}$ and $g\colon Y \to \mathbb{R} \cup \{+\infty\}$ and linear map $A\colon E \to Y$, show that the following perturbed value function is convex:

$$h(\mathbf{u}) = \inf_{\mathbf{x} \in E} \{f(\mathbf{x}) + g(A\mathbf{x} + \mathbf{u})\}\,.$$

(b) For any f and g, use the definition of subgradients to show the following partial calculus rule

$$\partial(f + g \circ A)(\mathbf{x}) \supseteq \partial f(\mathbf{x}) + A^* \partial g(A\mathbf{x})\ .$$

(Note: In our case, here $A^* = A^\top$.)

(c) Show that equality holds above whenever f and g are convex with $0 \in \operatorname{int}(\operatorname{dom} g - A\operatorname{dom} f)$.

(a) *Proof.* $\forall \mathbf{u}_1, \mathbf{u}_2 \in \operatorname{dom} g - A\operatorname{dom} f$, $\forall \lambda \in [0,1]$. Then $\forall \varepsilon > 0$, by the definition of $h(\mathbf{u}_1), h(\mathbf{u}_2)$, there exists $\mathbf{x}_1, \mathbf{x}_2 \in E$ such that

$$f(\mathbf{x}_1) + g(A\mathbf{x}_1 + \mathbf{u}_1) < h(\mathbf{u}_1) + \varepsilon \tag{7.1}$$
$$f(\mathbf{x}_2) + g(A\mathbf{x}_2 + \mathbf{u}_2) < h(\mathbf{u}_2) + \varepsilon \tag{7.2}$$

Then we consider $\mathbf{x}_0 = \lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$:

$$\begin{aligned}
h(\lambda \mathbf{u}_1 + (1-\lambda)\mathbf{u}_2) &= \inf_{x \in E} \{f(\mathbf{x}) + g(A\mathbf{x} + \lambda \mathbf{u}_1 + (1-\lambda)\mathbf{u}_2)\} \\
&\le f(\mathbf{x}_0) + g(A\mathbf{x}_0 + \lambda \mathbf{u}_1 + (1-\lambda)\mathbf{u}_2) \\
&= f(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) + g(\lambda(A\mathbf{x}_1 + \mathbf{u}_1) + (1-\lambda)(A\mathbf{x}_2 + \mathbf{u}_2)) \\
&\le \lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_2) + \lambda g(A\mathbf{x}_1 + \mathbf{u}_1) + (1-\lambda) g(A\mathbf{x}_2 + \mathbf{u}_2) \\
&\overset{(7.1),(7.2)}{<} \lambda h(\mathbf{u}_1) + \lambda \varepsilon + (1-\lambda) h(\mathbf{u}_2) + (1-\lambda)\varepsilon \\
&= \lambda h(\mathbf{u}_1) + (1-\lambda) h(\mathbf{u}_2) + \varepsilon
\end{aligned}$$

let $\varepsilon \to 0$, we have

$$h(\lambda \mathbf{u}_1 + (1-\lambda)\mathbf{u}_2) \le \lambda h(\mathbf{u}_1) + (1-\lambda) h(\mathbf{u}_2),$$

which implies $h(\mathbf{u})$ is convex. □

(b) *Proof.* $\forall \Phi \in \partial f(\mathbf{x}) + A^* \partial g(A\mathbf{x}) \implies \Phi = \phi + A^* \psi$, where $\phi \in \partial f(\mathbf{x}), \psi \in \partial g(A\mathbf{x})$. Then, $\forall \mathbf{y} \in$

$\text{dom}(g \circ A) \cap \text{dom} f$, we have

$$g(A\mathbf{y}) \geq g(A\mathbf{x}) + \langle A^*\psi, \mathbf{y} - \mathbf{x} \rangle$$
$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \phi, \mathbf{y} - \mathbf{x} \rangle$$

hence,

$$f(\mathbf{y}) + g(A\mathbf{y}) \geq g(A\mathbf{x}) + f(\mathbf{x}) + \langle \phi + A^*\psi, \mathbf{y} - \mathbf{x} \rangle, \ \forall \mathbf{y} \in \text{dom}(g \circ A) \cap \text{dom} f.$$

Thus, by definition of subgradient, $\mathbf{y} = A^*\psi + \phi \in \partial(f + g \circ A)(\mathbf{x})$. This proves $\partial(f + g \circ A)(\mathbf{x}) \supseteq \partial f(\mathbf{x}) + A^*\partial g(A\mathbf{x})$. ☐

(c) *Proof.* $\forall \Phi \in \partial(f+g\circ A)(\mathbf{x}) \implies 0 \in \partial(f+g\circ A - \langle \Phi, \cdot \rangle)(\mathbf{x})$, that is, $\mathbf{x}$ minimizes $f(\mathbf{y}) + g(A\mathbf{y}) - \langle \Phi, \mathbf{y} \rangle$. Let

$$H(\mathbf{u}) = \inf_{\mathbf{y} \in E} \{ f(\mathbf{y}) + g(A\mathbf{y} + \mathbf{u}) - \langle \Phi, \mathbf{y} \rangle \}.$$

Exercise 34 (a) yields that $H(\mathbf{u})$ is convex. Also, $0 \in \text{int dom} H$ since $0 \in \text{int}(\text{dom} g - A\text{dom } f)$, so by Max Formula $-\Psi \in \partial H(0)$ exists. Then by definition of subgradient,

$$H(0) \leq H(\mathbf{u}) + \langle \Psi, \mathbf{u} \rangle. \tag{7.3}$$

Recall that $H(\mathbf{u}) = \inf_{\mathbf{y} \in E}\{f(\mathbf{y}) + g(A\mathbf{y} + \mathbf{u}) - \langle \Phi, \mathbf{y} \rangle\}$ and $\mathbf{x}$ minimizes $f(\mathbf{y}) + g(A\mathbf{y}) - \langle \Phi, \mathbf{y} \rangle$, hence $\forall \mathbf{y}, \forall \mathbf{u}$ we have,

$$\begin{aligned} \underbrace{f(\mathbf{x}) + g(A\mathbf{x}) - \langle \Phi, \mathbf{x} \rangle}_{H(0)} &\overset{(7.3)}{\leq} \underbrace{\inf_{\mathbf{y} \in E} \{ f(\mathbf{y}) + g(A\mathbf{y} + \mathbf{u}) - \langle \Phi, \mathbf{y} \rangle \}}_{H(\mathbf{u})} + \langle \Psi, \mathbf{u} \rangle \\ &\leq f(\mathbf{y}) + g(A\mathbf{y} + \mathbf{u}) - \langle \Phi, \mathbf{y} \rangle + \langle \Psi, \mathbf{u} \rangle \end{aligned} \tag{7.4}$$

Take $\mathbf{y} = \mathbf{x}$ in Equation (7.4), we have

$$\begin{aligned} g(A\mathbf{x}) &\leq g(A\mathbf{x} + \mathbf{u}) + \langle \Psi, \mathbf{u} \rangle, \ \forall \mathbf{u}, \\ \implies g(A\mathbf{x} + \mathbf{u}) &\geq g(A\mathbf{x}) + \langle -\Psi, (A\mathbf{x} + \mathbf{u}) - A\mathbf{x} \rangle, \ \forall \mathbf{u}, \\ \implies g(\mathbf{z}) &\geq g(A\mathbf{x}) + \langle -\Psi, \mathbf{z} - A\mathbf{x} \rangle, \ \forall \mathbf{z}. \end{aligned}$$

This proves $-\Psi \in \partial g(A\mathbf{x})$.

Take $\mathbf{u} = A(\mathbf{x} - \mathbf{y}) \in \text{dom } g - A\text{dom } f$ in Equation (7.4), we have

$$\begin{aligned} f(\mathbf{x}) + g(A\mathbf{x}) - \langle \Phi, \mathbf{x} \rangle &\leq f(\mathbf{y}) + g(A\mathbf{x}) - \langle \Phi, \mathbf{y} \rangle + \langle \Psi, A(\mathbf{x} - \mathbf{y}) \rangle, \ \forall \mathbf{y} \\ \implies f(\mathbf{x}) &\leq f(\mathbf{y}) + \langle \Phi, \mathbf{x} - \mathbf{y} \rangle + \langle A^*\Psi, \mathbf{x} - \mathbf{y} \rangle, \ \forall \mathbf{y} \\ \implies f(\mathbf{y}) &\geq f(\mathbf{x}) + \langle \Phi + A^*\Psi, \mathbf{y} - \mathbf{x} \rangle, \ \forall \mathbf{y} \end{aligned}$$

Therefore, by the definition of subgradient, $\Phi + A^*\Psi \in \partial f(\mathbf{x})$.

Thus,

$$\Phi = \Phi + A^*\Psi - A^*\Psi = \underbrace{\Phi + A^*\Psi}_{\in \partial f(\mathbf{x})} + A^* \underbrace{(-\Psi)}_{\in \partial g(A\mathbf{x})} \in \partial f(\mathbf{x}) + A^*\partial g(A\mathbf{x}),$$

which implies $\partial(f + g \circ A)(\mathbf{x}) \subseteq \partial f(\mathbf{x}) + A^*\partial g(A\mathbf{x})$, then completes the proof. ☐

Exercise 35 (Max Formula.). Let $E \subseteq \mathbb{R}^d$, then for any convex function $f : E \to \mathbb{R} \cup \{+\infty\}$ and $\bar{\mathbf{x}} \in \text{int dom } f$, $\mathbf{r} \in E$, the following exists (is finite):

$$f'(\bar{\mathbf{x}}; \mathbf{r}) = \sup\{\langle \phi, \mathbf{r} \rangle : \phi \in \partial f(\bar{\mathbf{x}})\}.$$

In particular, $\partial f(\bar{\mathbf{x}}) \neq \varnothing$.

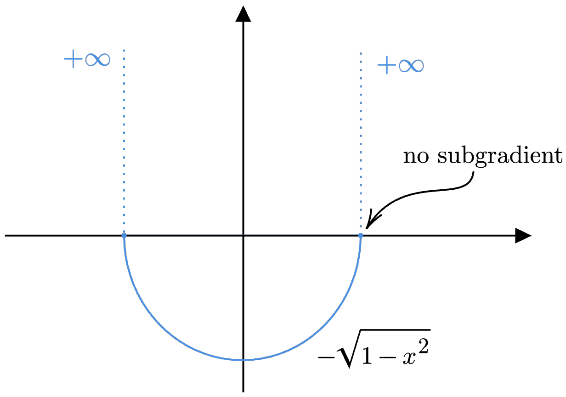


Figure 5: On the boundary of dom f, there may not be any $\phi \in \partial f(x)$.

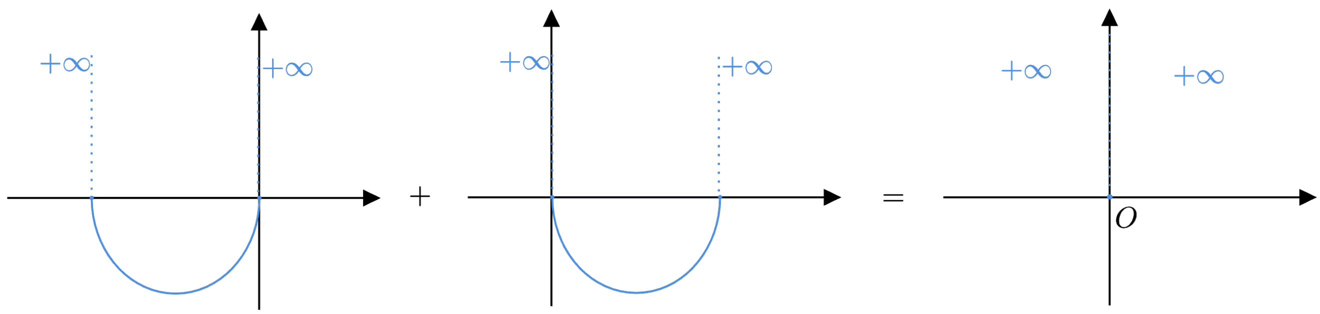


Figure 6: No sum rule in general.

8 Section-Nov-11-2022

Discussions on **Theorem 2.7.6** from lecture notes.

Theorem 7 (Theorem 2.7.1 in lecture notes). *The following are equivalent.*

1. *There exists a positive definite matrix* $A \in \mathbb{R}^{d \times d}$ *and* $\mathbf{c} \in \mathbb{R}^d$ *such that*

$$E = \{\mathbf{x} \in \mathbb{R}^d : (\mathbf{x} - \mathbf{c})^T A^{-1} (\mathbf{x} - \mathbf{c}) \leq 1\}.$$

2. *There exist orthonormal vectors* $\mathbf{b}^1, \dots, \mathbf{b}^d \in \mathbb{R}^d$, $\sigma_1, \dots, \sigma_d > 0$ *and* $\mathbf{c} \in \mathbb{R}^d$ *such that*

$$E = \{\mathbf{c} + \lambda_1 \mathbf{b}^1 + \cdots \lambda_d \mathbf{b}^d : \frac{\lambda_1^2}{\sigma_1^2} + \cdots + \frac{\lambda_d^2}{\sigma_d^2} \leq 1\}.$$

Proof. By eigendecomposition of A,

$$A = \begin{bmatrix} \mathbf{b}^1 & \cdots & \mathbf{b}^d \end{bmatrix} \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_d^2 \end{bmatrix} \begin{bmatrix} (\mathbf{b}^1)^T \\ \vdots \\ (\mathbf{b}^d)^T \end{bmatrix} = \underbrace{\begin{bmatrix} \sigma_1 \mathbf{b}^1 & \cdots & \sigma_d \mathbf{b}^d \end{bmatrix}}_{:=Q} \underbrace{\begin{bmatrix} (\sigma_1 \mathbf{b}^1)^T \\ \vdots \\ (\sigma_d \mathbf{b}^d)^T \end{bmatrix}}_{Q^T},$$

where $\mathbf{b}^1, \ldots, \mathbf{b}^d$ are orthonormal, $\sigma_1, \ldots, \sigma_d > 0$ since A is positive definite. Then we have

$$\begin{aligned} E &= \{\mathbf{x} \in \mathbb{R}^d : (\mathbf{x} - \mathbf{c})^T A^{-1} (\mathbf{x} - \mathbf{c}) \leq 1\} \\ &= \{\mathbf{x} \in \mathbb{R}^d : (\mathbf{x} - \mathbf{c})^T Q^{-T} Q^{-1} (\mathbf{x} - \mathbf{c}) \leq 1\} \\ &= \{\mathbf{x} \in \mathbb{R}^d : \|Q^{-1}(\mathbf{x} - \mathbf{c})\|_2^2 \leq 1\} \\ &= \left\{ \mathbf{c} + \mathbf{x} : \left\| \begin{bmatrix} (\frac{1}{\sigma_1} \mathbf{b}^1)^T \\ \vdots \\ (\frac{1}{\sigma_d} \mathbf{b}^d)^T \end{bmatrix} \mathbf{x} \right\|_2^2 \leq 1 \right\} \\ &= \left\{ \mathbf{c} + \mathbf{x} : \frac{((\mathbf{b}^1)^T \mathbf{x})^2}{\sigma_1^2} + \cdots + \frac{((\mathbf{b}^d)^T \mathbf{x})^2}{\sigma_d^2} \leq 1 \right\} \\ &= \left\{ \mathbf{c} + \lambda_1 \mathbf{b}^1 + \cdots \lambda_d \mathbf{b}^d : \frac{\lambda_1^2}{\sigma_1^2} + \cdots + \frac{\lambda_d^2}{\sigma_d^2} \leq 1 \right\}. \end{aligned}$$

To see the reverse direction, just let $A = QQ^T$, where $Q = \begin{bmatrix} \sigma_1 \mathbf{b}^1 & \cdots & \sigma_d \mathbf{b}^d \end{bmatrix}$, and then start from the last equality above to the first. ☐

Definition 6 (Ellipsoid)**.** Let $A \in \mathbb{R}^{d \times d}$ be a positive definite matrix and $\mathbf{c} \in \mathbb{R}^d$. The set

$$E(A, \mathbf{c}) := \{\mathbf{x} \in \mathbb{R}^d : (\mathbf{x} - \mathbf{c})^T A^{-1} (\mathbf{x} - \mathbf{c}) \leq 1\}$$

is called an ellipsoid centered at $\mathbf{c}$.

Theorem 8 (Theorem 2.7.3 in lecture notes)**.** *The following are all true.*

1. *Let $A \in \mathbb{R}^{d \times d}$ be a positive definite matrix and $\mathbf{c} \in \mathbb{R}^d$. Then*

$$\mathrm{vol}(E(A, \mathbf{c})) = \sqrt{\det(A)}\ \mathrm{vol}(B(\mathbf{0}, 1)).$$

2. *Let $\mathbf{b}^1, \ldots, \mathbf{b}^d \in \mathbb{R}^d$ be orthonormal vectors, $\sigma_1, \ldots, \sigma_d > 0$ and $\mathbf{c} \in \mathbb{R}^d$. Then*

$$\mathrm{vol}(E) = \left(\prod_{i=1}^d \sigma_i \right) \mathrm{vol}(B(\mathbf{0}, 1)),$$

where $E = \{\mathbf{c} + \lambda_1 \mathbf{b}^1 + \cdots \lambda_d \mathbf{b}^d : \frac{\lambda_1^2}{\sigma_1^2} + \cdots + \frac{\lambda_d^2}{\sigma_d^2} \leq 1\}$.

Proof sketch.

1. Notice that $E(A,\mathbf{c}) = \{\mathbf{x}\in\mathbb{R}^d : \sqrt{(\mathbf{x}-\mathbf{c})^T A^{-1}(\mathbf{x}-\mathbf{c})}\} = \{A^{\frac{1}{2}}\mathbf{y}+\mathbf{c} : \|\mathbf{y}\|_2 \le 1\}$, then

$$\begin{aligned}\mathrm{vol}(E(A,\mathbf{c})) &= \int_{E(A,\mathbf{c})} 1\mathrm{d}x_1\cdots\mathrm{d}x_d \\ &= \int_{B(\mathbf{0},1)} 1\cdot\left|\frac{\partial(x_1,\dots,x_d)}{\partial(y_1,\dots,y_d)}\right|\mathrm{d}y_1\cdots\mathrm{d}y_d \\ &= \int_{B(\mathbf{0},1)} 1\cdot\left|\det(A^{\frac{1}{2}})\right|\mathrm{d}y_1\cdots\mathrm{d}y_d \\ &= \sqrt{\det(A)}\int_{B(\mathbf{0},1)} 1\mathrm{d}y_1\cdots\mathrm{d}y_d \\ &= \sqrt{\det(A)}\mathrm{vol}(B(\mathbf{0},1)).\end{aligned}$$

2. In this case, $A = QQ^T$, where $Q = \begin{bmatrix}\sigma_1\mathbf{b}^1 & \cdots & \sigma_d\mathbf{b}^d\end{bmatrix}$. Notice that $\sqrt{\det(A)} = \sqrt{\det(Q\cdot Q^T)} = \sqrt{\det(Q)\cdot\det(Q^T)} = \sqrt{\det(Q)^2} = |\det(Q)|$ and the fact that

$$\det(Q) = \det(\begin{bmatrix}\sigma_1\mathbf{b}^1 & \cdots & \sigma_d\mathbf{b}^d\end{bmatrix}) = (\sigma_1\cdots\sigma_d)\det(\begin{bmatrix}\mathbf{b}^1 & \cdots & \mathbf{b}^d\end{bmatrix}) = (\sigma_1\cdots\sigma_d)(\pm 1).$$

□

Theorem 9 (Theorem 2.7.4 in lecture notes)**.** *Volume of the unit ball $B(\mathbf{0},1)$ in $\mathbb{R}^d$.*

$$\mathrm{vol}(B(\mathbf{0},1)) = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} \sim \frac{1}{\sqrt{\pi n}}\left(\frac{2e\pi}{n}\right)^{\frac{n}{2}}.$$

Moreover,

$$\left(\frac{2}{\sqrt{d}}\right)^d \le \mathrm{vol}(B(\mathbf{0},1)) \le 2^d.$$

Theorem 10 (Theorem 2.7.6 in lecture notes)**.** *Let $d\ge 2$ and let $E\subseteq\mathbb{R}^d$ be an ellipsoid with center $\mathbf{c}\in\mathbb{R}^d$. Let $H\subseteq\mathbb{R}^d$ be a halfspace and let $0\le\beta<\frac{1}{d}$ such that H does not contain $\mathbf{c}+\beta(E-\mathbf{c})$. Then, there exists another ellipsoid E' such that $E\cap H\subseteq E'$ and*

$$\mathrm{vol}(E') \le e^{-\frac{(1-\beta d)^2}{2(d+1)}}\mathrm{vol}(E).$$

Proof. See the main proof in lecture notes. Some notes on choosing σ and σ' :

Assume $\mathbf{c} = (-t,0,\dots,0)\in\mathbb{R}^d,\ t>0$, and E' is given by

$$E' = \left\{\mathbf{x}\in\mathbb{R}^d : \frac{(x_1+t)^2}{\sigma^2} + \frac{x_2^2}{\sigma'^2} + \cdots + \frac{x_d^2}{\sigma'^2} \le 1\right\}.$$

Then imposing the conditions $-\mathbf{e}^1\in\mathrm{bd}(E'),\ \beta\mathbf{e}^1 \pm \sqrt{1-\beta^2}\mathbf{e}^i\in\mathrm{bd}(E'),\ i=2,\dots,d$:

$$\begin{cases}\frac{(-1+t)^2}{\sigma^2} = 1\\ \frac{(\beta+t)^2}{\sigma^2} + \frac{1-\beta^2}{\sigma'^2} = 1\end{cases} \implies \begin{cases} t = 1-\sigma\\ \frac{(\beta+t)^2}{(1-t)^2} + \frac{1-\beta^2}{\sigma'^2} = 1\end{cases} \implies \sigma'^2 = \frac{(1-t)^2(1-\beta)}{1-\beta-2t}.$$

Then consider the 1-d continuous optimization problem:

$$\min_{t\in(0,1)}\left[(1-t)\left(\frac{(1-t)^2(1-\beta)}{1-\beta-2t}\right)^{\frac{d-1}{2}}=(1-t)^d\frac{(1-\beta)^{\frac{d-1}{2}}}{(1-\beta-2t)^{\frac{d-1}{2}}}\right].$$

Take the logarithm of the objective function and let

$$f(t)=d\log(1-t)-\frac{d-1}{2}\log(1-\beta-2t)+\frac{d-1}{2}(1-\beta).$$

Then

$$\begin{aligned}f'(t)=0 &\iff \frac{-d}{1-t}-\frac{d-1}{2}\frac{-2}{1-\beta-2t}=0\\ &\iff \frac{d-1}{1-\beta-2t}=\frac{d}{1-t}\\ &\iff d-1-dt+t=d-\beta d-2dt\\ &\iff dt+t=1-\beta d\\ &\iff t=\frac{1-\beta d}{1+d}\in(0,1).\end{aligned}$$

[Note: Need to check $t=\frac{1-\beta d}{1+d}$ is indeed a minimizer of $f(t)$ over $(0,1)$]

Therefore,

$$\sigma=1-t=\frac{(1+\beta)d}{d+1}\text{ and }\sigma'^2=\frac{(1-t)^2(1-\beta)}{1-\beta-2t}=\frac{(1-\beta^2)d^2}{d^2-1}.$$

□

9 Section-Nov-18-2022

Definition 7 (Cone of feasible directions)**.** Let $C\subseteq\mathbb{R}^d$ be a convex set, and let $\mathbf{x}\in C$. Define *the cone of feasible directions* as

$$F_C(\mathbf{x})=\{\mathbf{r}\in\mathbb{R}^d:\exists\varepsilon>0\text{ such that }\mathbf{x}+\varepsilon\mathbf{r}\in C\}.$$

Exercise 36. Let $C\subseteq\mathbb{R}^d$ and $\mathbf{x}\in C$, show that $F_C(\mathbf{x})$ is a convex cone.

Proof. $\forall\mathbf{r}^1,\mathbf{r}^2\in F_C(\mathbf{x})$, by definition there exist $\varepsilon_1,\varepsilon_2>0$ such that $\mathbf{x}+\varepsilon_1\mathbf{r}^1,\mathbf{x}+\varepsilon_2\mathbf{r}^2\in C$. For any $\lambda,\gamma\geq 0$, notice that

$$\mathbf{x}+\frac{\varepsilon_1}{2\lambda}(2\lambda\mathbf{r}^1),\ \mathbf{x}+\frac{\varepsilon_2}{2\gamma}(2\gamma\mathbf{r}^2)\in C,$$

so we have that $2\lambda\mathbf{r}^1,2\gamma\mathbf{r}^2\in F_C(\mathbf{x})$. Let $\varepsilon_3=\min\{\frac{\varepsilon_1}{2\lambda},\frac{\varepsilon_2}{2\gamma}\}$, by convexity of C one can obtain that $\mathbf{x}+\varepsilon_3(2\lambda\mathbf{r}^1),\mathbf{x}+\varepsilon_3(2\gamma\mathbf{r}^2)\in C$. Again by convexity, we have

$$\mathbf{x}+\varepsilon_3(\lambda\mathbf{r}^1+\gamma\mathbf{r}^2)=\frac{1}{2}(\mathbf{x}+\varepsilon_3(2\lambda\mathbf{r}^1))+\frac{1}{2}(\mathbf{x}+\varepsilon_3(2\gamma\mathbf{r}^2))\in C,$$

so $\lambda\mathbf{r}^1+\gamma\mathbf{r}^2\in F_C(\mathbf{x})$, which proves $F_C(\mathbf{x})$ is a convex cone. □

Remark 6. $F_C(\mathbf{x})$ may not be closed: consider $C=\{\mathbf{x}\in\mathbb{R}^2:\|\mathbf{x}\|\leq 1\}$, and let $\mathbf{x}=(-1,0)$. Then $F_C(\mathbf{x})=\{\mathbf{r}\in\mathbb{R}^2:\mathbf{r}_1>0\}\cup\{\mathbf{0}\}$.

Exercise 37. Let $P \subseteq \mathbb{R}^d$ be a polyhedron given by $P = \{\mathbf{x} \in \mathbb{R}^d : A\mathbf{x} \leq \mathbf{b}\}$. Let $\mathbf{a}^i,\ i = 1, \ldots, m$ be the rows of A. For any $\bar{\mathbf{x}} \in P$, define $\text{tight}(\bar{\mathbf{x}}) = \{i : \langle \mathbf{a}^i, \mathbf{x} \rangle = \mathbf{b}_i\}$. Show that

$$F_P(\bar{\mathbf{x}}) = \{\mathbf{r} \in \mathbb{R}^d : \langle \mathbf{a}^i, \mathbf{r} \rangle \leq 0 \text{ for all } i \in \text{tight}(\bar{\mathbf{x}})\}.$$

Proof. The set $\{\mathbf{r} \in \mathbb{R}^d : \langle \mathbf{a}^i, \mathbf{r} \rangle \leq 0 \text{ for all } i \in \text{tight}(\bar{\mathbf{x}})\} \subseteq F_P(\bar{\mathbf{x}})$, since
<u>Case 1: $i \in \text{tight}(\bar{\mathbf{x}})$.</u> For all $\varepsilon > 0$, $\langle \mathbf{a}^i, \mathbf{x} + \varepsilon \mathbf{r} \rangle = \langle \mathbf{a}^i, \mathbf{x} \rangle + \varepsilon \langle \mathbf{a}^i, \mathbf{r} \rangle \leq \langle \mathbf{a}^i, \mathbf{x} \rangle = \mathbf{b}_i$.
<u>Case 2: $i \notin \text{tight}(\bar{\mathbf{x}})$.</u> $\langle \mathbf{a}^i, \mathbf{x} \rangle < \mathbf{b}_i$, once $\varepsilon > 0$ is small enough, the inequality $\langle \mathbf{a}^i, \mathbf{x} + \varepsilon \mathbf{r} \rangle \leq \mathbf{b}_i$ will hold.

To show the reverse direction, consider any $\mathbf{r} \in \mathbb{R}^d$ has $\langle \mathbf{a}^i, \mathbf{r} \rangle > 0$ for some $i \in \text{tight}(\bar{\mathbf{x}})$, then for any $\varepsilon > 0$ we have

$$\langle \mathbf{a}^i, \bar{\mathbf{x}} + \varepsilon \mathbf{r} \rangle = \langle \mathbf{a}^i, \bar{\mathbf{x}} \rangle + \varepsilon \langle \mathbf{a}^i, \mathbf{r} \rangle > \mathbf{b}_i,$$

so $\mathbf{r} \notin F_C(\bar{\mathbf{x}})$. □

Exercise 38. Let $C \subseteq \mathbb{R}^d$ be a nonempty, closed, convex set. Then the following are all true:

1. Let $\mathbf{y} \in \mathbb{R}^d \backslash C$, then $\text{Proj}_C(\mathbf{y}) = \mathbf{x}$ if and only if $\mathbf{y} - \mathbf{x} \in N_C(\mathbf{x})$.
2. If $\mathbf{x} \in \text{int}\,(C)$, then $N_C(\mathbf{x}) = \{0\}$.
3. If $\mathbf{x} \in \text{bd}\,(C)$, then $\{0\} \subsetneq N_C(\mathbf{x})$.

Proof.

1. $(\Longrightarrow)$: Notice that

$$\begin{array}{lcr} & \langle \mathbf{y} - \mathbf{x}, \mathbf{z} \rangle \leq \langle \mathbf{y} - \mathbf{x}, \mathbf{x} \rangle \leq \langle \mathbf{y} - \mathbf{x}, \mathbf{y} \rangle, & \forall \mathbf{z} \in C, \\ \Longrightarrow & \langle \mathbf{y} - \mathbf{x}, \mathbf{z} - \mathbf{x} \rangle \leq 0, & \forall \mathbf{z} \in C, \\ \Longrightarrow & \mathbf{y} - \mathbf{x} \in N_C(\mathbf{x}). & \end{array}$$

   $(\Longleftarrow)$: Since $\mathbf{y} - \mathbf{x} \in N_C(\mathbf{x})$, by definition of normal cone,

$$\begin{array}{lcr} & \langle \mathbf{y} - \mathbf{x}, \mathbf{z} - \mathbf{x} \rangle \leq 0, & \forall \mathbf{z} \in C, \\ \Longrightarrow & \langle \mathbf{y} - \mathbf{x}, \mathbf{y} - \mathbf{x} \rangle + \langle \mathbf{y} - \mathbf{x}, \mathbf{z} - \mathbf{y} \rangle \leq 0, & \forall \mathbf{z} \in C, \\ \Longrightarrow & \|\mathbf{y} - \mathbf{x}\|^2 \leq (\mathbf{y} - \mathbf{x})^T(\mathbf{y} - \mathbf{z}) \leq \|\mathbf{y} - \mathbf{x}\| \|\mathbf{y} - \mathbf{z}\|, & \forall \mathbf{z} \in C, \\ \Longrightarrow & \|\mathbf{y} - \mathbf{x}\| \leq \|\mathbf{y} - \mathbf{z}\|, & \forall \mathbf{z} \in C, \\ \Longrightarrow & \mathbf{x} = \text{Proj}_C(\mathbf{y}). & \end{array}$$

2. $\forall \mathbf{r} \in N_C(\mathbf{x})$, there exists $\varepsilon > 0$ such that $\mathbf{x} + \varepsilon \mathbf{r} \in C$ since $\mathbf{x} \in \text{int}\,(C)$. Then

$$\begin{array}{lc} & \langle \mathbf{r}, (\mathbf{x} + \varepsilon \mathbf{r}) - \mathbf{x} \rangle \leq 0, \\ \Longrightarrow & \varepsilon \langle \mathbf{r}, \mathbf{r} \rangle \leq 0, \\ \Longrightarrow & \mathbf{r} = 0, \end{array}$$

   which implies $N_C(\mathbf{x}) = \{0\}$.

3. By Lemma 2.3.4 in lecture notes, there exists $\mathbf{y} \in \mathbb{R}^d \backslash C$ such that $\text{Proj}_C(\mathbf{y}) = \mathbf{x}$. Then part 1 implies $\mathbf{y} - \mathbf{x} \in N_C(\mathbf{x})$. Therefore, $\{0\} \subsetneq \{0, \mathbf{y} - \mathbf{x}\} \subseteq N_C(\mathbf{x})$.

□

Exercise 39. Consider the following standard form polyhedron in $\mathbb{R}^d$, defined by some $A \in \mathbb{R}^{m \times d}$, $\mathbf{b} \in \mathbb{R}^m$:

$$P = \{\mathbf{x} \in \mathbb{R}^d : A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq 0\}.$$

1. Prove that every $\bar{\mathbf{x}} \in P$ has $N_P(\bar{\mathbf{x}}) = \{-(\mathbf{s} + A^T\mathbf{y}) : (\mathbf{s}, \mathbf{y}) \in \mathbb{R}^d \times \mathbb{R}^m, \mathbf{s} \geq 0, \mathbf{s}_i = 0 \text{ for } i \in I(\bar{\mathbf{x}})\}$, where $I(\bar{\mathbf{x}}) = \{i : \bar{\mathbf{x}}_i > 0\}$.

2. Prove that if $-\mathbf{c} \in \operatorname{int}(N_P(\bar{\mathbf{x}}))$ for some $\bar{\mathbf{x}} \in P$, then $\bar{\mathbf{x}}$ is the unique minimizer of $\langle \mathbf{c}, \cdot \rangle$ over P.

Proof. 1. For any $\mathbf{c} \in \mathbb{R}^d$, the function $f(\mathbf{x}) = \langle \mathbf{c}, \mathbf{x} \rangle$ is convex and differentiable with $\nabla f(\mathbf{x}) = \mathbf{c}$, then

$$
\begin{aligned}
-\mathbf{c} \in N_P(\bar{\mathbf{x}}) &\iff \bar{\mathbf{x}} \text{ minimizes } \langle \mathbf{c}, \mathbf{x} \rangle \text{ over } P, \\
&\iff \exists \mathbf{y} \in \mathbb{R}^m \text{ s.t. } A^T\mathbf{y} \leq \mathbf{c},\ \langle \mathbf{b}, \mathbf{y} \rangle = \langle \mathbf{c}, \bar{\mathbf{x}} \rangle, \\
&\iff \exists \mathbf{y} \in \mathbb{R}^m \text{ s.t. } A^T\mathbf{y} \leq \mathbf{c},\ \langle A\bar{\mathbf{x}}, \mathbf{y} \rangle = \langle \mathbf{c}, \bar{\mathbf{x}} \rangle, \\
&\iff \exists \mathbf{y} \in \mathbb{R}^m \text{ s.t. } A^T\mathbf{y} \leq \mathbf{c},\ \mathbf{x}^T(\mathbf{c} - A^T\mathbf{y}) = 0, \\
&\iff \exists \mathbf{y} \in \mathbb{R}^m, \mathbf{s} \geq 0 \text{ s.t. } \mathbf{s} = \mathbf{c} - A^T\mathbf{y},\ \mathbf{x}^T(\mathbf{c} - A^T\mathbf{y}) = 0, \\
&\iff \exists \mathbf{y} \in \mathbb{R}^m, \mathbf{s} \geq 0 \text{ s.t. } \mathbf{c} = \mathbf{s} + A^T\mathbf{y},\ \mathbf{x}^T\mathbf{s} = 0, \\
&\iff -\mathbf{c} \in \{-\ (\mathbf{s} + A^T\mathbf{y}) : \mathbf{y} \in \mathbb{R}^m, \mathbf{s} \geq 0, \langle \bar{\mathbf{x}}, \mathbf{s} \rangle = 0\}.
\end{aligned}
$$

2. Suppose to the contrary both $\bar{\mathbf{x}}$ and $\bar{\mathbf{x}}'$ minimize the objective $\langle \mathbf{c}, \mathbf{x} \rangle$ but $\bar{\mathbf{x}} \neq \bar{\mathbf{x}}'$. Since $c \in \operatorname{int}(N_P(\bar{\mathbf{x}}))$, for small enough $\varepsilon > 0$, $-c - \varepsilon(\bar{\mathbf{x}} - \bar{\mathbf{x}}') \in N_P(\bar{\mathbf{x}})$. This implies $\bar{\mathbf{x}}$ minimizes $\langle \mathbf{c} + \varepsilon(\bar{\mathbf{x}} - \bar{\mathbf{x}}'), \mathbf{x} \rangle$ over P. However, this is contradicted by $\bar{\mathbf{x}}' \in P$ as it has

$$
\begin{aligned}
\langle \mathbf{c} + \varepsilon(\bar{\mathbf{x}} - \bar{\mathbf{x}}'), \bar{\mathbf{x}}' \rangle &= \langle \mathbf{c} + \varepsilon(\bar{\mathbf{x}} - \bar{\mathbf{x}}'), \bar{\mathbf{x}} \rangle + \langle \mathbf{c} + \varepsilon(\bar{\mathbf{x}} - \bar{\mathbf{x}}'), \bar{\mathbf{x}}' - \bar{\mathbf{x}} \rangle \\
&= \langle \mathbf{c} + \varepsilon(\bar{\mathbf{x}} - \bar{\mathbf{x}}'), \bar{\mathbf{x}} \rangle + \underbrace{\langle \mathbf{c}, \bar{\mathbf{x}}' \rangle - \langle \mathbf{c}, \bar{\mathbf{x}} \rangle}_{=0} - \varepsilon \underbrace{\left\| \bar{\mathbf{x}} - \bar{\mathbf{x}}' \right\|_2^2}_{>0} \\
&< \langle \mathbf{c} + \varepsilon(\bar{\mathbf{x}} - \bar{\mathbf{x}}'), \bar{\mathbf{x}} \rangle.
\end{aligned}
$$

□

10 Section-Dec-2-2022

Definition 8 (Dual cone)**.** Let K be a convex cone, the set $K^* = \{\mathbf{y} : \langle \mathbf{x}, \mathbf{y} \rangle \geq 0 \text{ for all } \mathbf{x} \in K\}$ is called the *dual cone* of K.

Exercise 40 (Positive semidefinite cone)**.** We use $\mathcal{S}^n$ to denote all the symmetric $n \times n$ matrices, and all the $n \times n$ positive semidefinite matrices will be denoted by $\mathcal{S}_+^n$. On the set of $\mathcal{S}^n$, we use the standard inner product $\operatorname{tr}(AB) = \sum_{i,j=1}^n A_{ij}B_{ij}$. Prove that $\mathcal{S}_+^n$ is a closed convex cone and self-dual.

Proof. First we prove that $\mathcal{S}_+^n$ is a closed convex cone. Observe that

$$
\begin{aligned}
\mathcal{S}_+^n &= \{A \in \mathcal{S}^n : \mathbf{x}^T A \mathbf{x} \geq 0 \text{ for all } \mathbf{x} \in \mathbb{R}^n\} \\
&= \bigcap_{\mathbf{x} \in \mathbb{R}^n} \{A \in \mathcal{S}^n : \operatorname{tr}(\mathbf{x}^T A \mathbf{x}) \geq 0\} \\
&= \bigcap_{\mathbf{x} \in \mathbb{R}^n} \{A \in \mathcal{S}^n : \operatorname{tr}(\mathbf{x}\mathbf{x}^T A) \geq 0\} \\
&= \bigcap_{\mathbf{x} \in \mathbb{R}^n} \{A \in \mathcal{S}^n : \langle \mathbf{x}\mathbf{x}^T, A \rangle \geq 0\}
\end{aligned}
$$

is an intersection of closed halfspaces, which implies $\mathcal{S}_+^n$ is a closed convex set. To show it is a cone, we take any $A, B \in \mathcal{S}_+^n$ and $\lambda, \gamma \geq 0$, then for any $\mathbf{x} \in \mathbb{R}^n$ we have

$$\mathbf{x}^T(\lambda A + \gamma B)\mathbf{x} = \lambda(\mathbf{x}^T A\mathbf{x}) + \gamma(\mathbf{x}^T B\mathbf{x}) \geq 0,$$

so $\lambda A + \gamma B \in \mathcal{S}_+^n$.

Then we prove that $\mathcal{S}_+^n$ is self-dual, i.e., for $A, B \in \mathcal{S}^n$, $\operatorname{tr}(AB) \geq 0,\ \forall A \in \mathcal{S}_+^n \iff B \in \mathcal{S}_+^n$. Suppose $B \notin \mathcal{S}_+^n$, then there exists $\mathbf{x} \in \mathbb{R}^n$ with

$$\mathbf{x}^T B\mathbf{x} = \operatorname{tr}(\mathbf{x}\mathbf{x}^T B) < 0.$$

Hence the positive semidefinite matrix $A = \mathbf{x}\mathbf{x}^T$ satisfies $\operatorname{tr}(AB) < 0$, which implies $B \notin (\mathcal{S}_+^n)^*$.

Now suppose $A, B \in \mathcal{S}_+^n$. We can express A in terms of its eigenvalue decomposition as $A = \sum_{i=1}^n \lambda_i \mathbf{x}_i \mathbf{x}_i^T$, where the eigenvalues $\lambda_i \geq 0,\ i = 1, \ldots, n$. Then we have

$$\operatorname{tr}(BA) = \operatorname{tr}\left(B \sum_{i=1}^n \lambda_i \mathbf{x}_i \mathbf{x}_i^T\right) = \sum_{i=1}^n \lambda_i \mathbf{x}_i^T B \mathbf{x}_i \geq 0.$$

This shows that $B \in (\mathcal{S}_+^n)^*$. □

Exercise 41 (Dual of a norm cone [1])**.** Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. The dual of the associated cone $K = \{(\mathbf{x}, t) \in \mathbb{R}^{n+1} : \|\mathbf{x}\| \leq t\}$ is the cone defined by the dual norm, i.e.,

$$K^* = \{(\mathbf{u}, v) \in \mathbb{R}^{n+1} : \|\mathbf{u}\|_* \leq v\},$$

where the dual norm is given by $\|\mathbf{u}\|_* = \sup\{\langle \mathbf{u}, \mathbf{x}\rangle : \|\mathbf{x}\| \leq 1\}$.

Proof. To prove the result we need to show that

$$\langle \mathbf{x}, \mathbf{u}\rangle + tv \geq 0 \text{ whenever } \|\mathbf{x}\| \leq t \iff \|\mathbf{u}\|_* \leq v.$$

" $\Longrightarrow$ " : Suppose to the contrary that $\|\mathbf{u}\|_* > v$, then by the definition of the dual norm, there exists an $\mathbf{x}$ with $\|\mathbf{x}\| \leq 1$ and $\langle \mathbf{x}, \mathbf{u}\rangle > v$. Taking $t = 1$, we have $\langle \mathbf{u}, -\mathbf{x}\rangle + v < 0$, which leads to a contradiction.
" $\Longleftarrow$ " : Suppose $\|\mathbf{u}\|_* \leq v$ and $\|\mathbf{x}\| \leq t$ for some $t > 0$. Applying the definition of the dual norm, and the fact that $\|-\mathbf{x}/t\| \leq 1$, we have $\langle \mathbf{u}, -\mathbf{x}/t\rangle \leq \|\mathbf{u}\|_* \leq v$, and therefore $\langle \mathbf{u}, \mathbf{x}\rangle + vt \geq 0$. □

Exercise 42 (Second-order conditions for K-convexity [1])**.** Let $K \subseteq \mathbb{R}^m$ be a closed, convex, pointed cone, with associated generalized inequality $\preceq_K$. Show that a twice differentiable function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$, with convex domain, is K-convex if and only if for all $\mathbf{x} \in \operatorname{dom} \mathbf{f}$ and all $\mathbf{y} \in \mathbb{R}^n$,

$$0 \preceq_K \sum_{i,j=1}^n \frac{\partial^2 \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}_i \partial \mathbf{x}_j} \mathbf{y}_i \mathbf{y}_j.$$

(Here $\partial^2 \mathbf{f}/\partial \mathbf{x}_i \partial \mathbf{x}_j \in \mathbb{R}^m$, with components $\partial^2 \mathbf{f}_k/\partial \mathbf{x}_i \partial \mathbf{x}_j$, for $k = 1, \ldots, m$.)

Proof. It's not hard to show $\mathbf{f}$ is K-convex if and only if $\langle \mathbf{c}, \mathbf{f}\rangle$ is convex for all $\mathbf{c} \in K^*$ (similar to Proposition 9.2.9 in lecture notes). The Hessian of $\langle \mathbf{c}, \mathbf{f}(\mathbf{x})\rangle$ is

$$\nabla^2(\langle \mathbf{c}, \mathbf{f}(\mathbf{x})\rangle) = \sum_{k=1}^n \mathbf{c}_k \nabla^2 \mathbf{f}_k(\mathbf{x}).$$

This is positive semidefinite if and only if for all $\mathbf{y}$,

$$\mathbf{y}^T \nabla^2(\langle \mathbf{c}, \mathbf{f}(x) \rangle)\mathbf{y} = \sum_{i,j=1}^{n} \sum_{k=1}^{n} \mathbf{c}_k \nabla^2 \mathbf{f}_k(\mathbf{x})\mathbf{y}_i \mathbf{y}_j = \sum_{k=1}^{n} \mathbf{c}_k \left(\sum_{i,j=1}^{n} \nabla^2 \mathbf{f}_k(\mathbf{x})\mathbf{y}_i \mathbf{y}_j \right) \geq 0,$$

which by definition of dual cone is equivalent to

$$\sum_{i,j=1}^{n} \nabla^2 \mathbf{f}_k(\mathbf{x})\mathbf{y}_i \mathbf{y}_j \in K.$$

□

References

[1] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. Convex optimization. Cambridge university press, 2004.