Introduction to Convexity

Recitation Course Notes

---

by

Hongyu Cheng

Department of Applied Mathematics and Statistics
Johns Hopkins University

hongyucheng@jhu.edu

Compiled on December 13, 2023

Contents

Part I

Structure of Convex Sets

Recitation 1 September 1, 2023

1.1 Review

Example 1.1 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 $A \cap B = \varnothing$. □

Example 1.2 Show that $\text{conv}(A \cap B) \subseteq \text{conv}(A) \cap \text{conv}(B)$, and find an example that the containment is strict.

Proof

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

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 $\text{conv}(A \cap B) = \varnothing$. □

The following proposition shows that the convexity of a set is invariant under some operations.

Proposition 1.1 (Operations 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. Let $\mathbf{x}, \mathbf{y} \in \bigcap_{i \in \mathcal{I}} X_i$. This implies that $\mathbf{x}, \mathbf{y} \in X_i$ for every $i \in \mathcal{I}$. Since each X_i is convex, for every $\lambda \in [0, 1]$, $\lambda\mathbf{x} + (1 - \lambda)\mathbf{y} \in X_i$ for all $i \in \mathcal{I}$. Therefore, $\lambda\mathbf{x} + (1 - \lambda)\mathbf{y} \in \bigcap_{i \in \mathcal{I}} X_i$.

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 $\mathbf{x}^1 = A\mathbf{x}^2 + \mathbf{b}$ and $\mathbf{y}^1 = A\mathbf{y}^2 + \mathbf{b}$ for some $A \in \mathbb{R}^{m \times d}$ and $\mathbf{b} \in \mathbb{R}^m$. Therefore,
$$\begin{aligned}\lambda \mathbf{x}^1 + (1-\lambda)\mathbf{y}^1 &= \lambda(A\mathbf{x}^2 + \mathbf{b}) + (1-\lambda)(A\mathbf{y}^2 + \mathbf{b}) \\ &= \lambda A\mathbf{x}^2 + (1-\lambda)A\mathbf{y}^2 + \lambda \mathbf{b} + (1-\lambda)\mathbf{b} \\ &= A(\underbrace{\lambda \mathbf{x}^2 + (1-\lambda)\mathbf{y}^2}_{\in X}) + \mathbf{b} \in T(X).\end{aligned}$$
   (b). Note that $T^{-1}(Y) = \{\mathbf{x} \in \mathbb{R}^d : T(\mathbf{x}) \in Y\}$, 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 $\lambda \mathbf{x}^1 + (1-\lambda)\mathbf{x}^2 \in T^{-1}(Y)$. □

1.2 Exercises

In the first exercise we will see how to use part 4. of Proposition 1.1 to prove the convexity of ellipsoids.

Theorem 1.1 (Eigendecomposition)

Let $A \in \mathbb{R}^{d \times d}$ be a symmetric matrix, then there exists a matrix $U \in \mathbb{R}^{d \times d}$ such that $U^\intercal U = I$ and
$$A = U \Lambda U^\intercal,$$
where Λ is the diagonal matrix with the diagonal entries equal to the eigenvalues of A. ♡

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

✍ **Exercise 1.1** Let $A \in \mathbb{R}^{d \times d}$ be a positive definite matrix and $\mathbf{c} \in \mathbb{R}^d$. Show that the ellipsoid
$$E(A, c) := \{\mathbf{x} \in \mathbb{R}^d : (\mathbf{x} - \mathbf{c})^\intercal A^{-1} (\mathbf{x} - \mathbf{c}) \le 1\}$$

is a convex set.

Proof By Theorem 1.1, we define the affine transformation $T(\mathbf{x}) = A^{-\frac{1}{2}}(\mathbf{x} - \mathbf{c})$. Observe that the image of the set $E(A, c)$ under this transformation, denoted $T(E(A, c))$, is the unit ball in $\mathbb{R}^d$. As the unit ball in $\mathbb{R}^d$ is a convex set, the conclusion immediately follows from part 4(b) of Exercise 1.1. □

The following exercise study the convexity of the solution set of a quadratic inequality.

Exercise 1.2 Let $C := \{\mathbf{x} \in \mathbb{R}^d : \mathbf{x}^\mathsf{T} A\mathbf{x} + \mathbf{b}^\mathsf{T}\mathbf{x} + c \le 0\}$.

1 Show that C is a convex set if A is positive semi-definite.

2 Show that the intersection of C and the hyperplane $\{\mathbf{x} \in \mathbb{R}^d : \langle \alpha, x\rangle = \beta\}$ (where $\alpha \in \mathbb{R}^d \backslash \{0\}$, $\beta \in \mathbb{R}$) is convex if the matrix $A + \lambda\alpha\alpha^\mathsf{T}$ is positive semi-definite for some $\lambda \in \mathbb{R}$.

Proof

1 Consider arbitrary vectors $\mathbf{x}, \mathbf{y} \in C$ and any scalar $\lambda \in [0, 1]$. First, observe that the inequality

$$2\mathbf{x}^\mathsf{T} A\mathbf{y} \le \mathbf{x}^\mathsf{T} A\mathbf{x} + \mathbf{y}^\mathsf{T} A\mathbf{y} \tag{1.1}$$

follows from $(\mathbf{x} - \mathbf{y})^\mathsf{T} A(\mathbf{x} - \mathbf{y}) \ge 0$, which is true since $A \succeq 0$. Therefore, we have

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

thus proving that C is convex.

2 We just need to show $C' := \left\{\mathbf{x} \in \mathbb{R}^d : \mathbf{x}^\mathsf{T} A\mathbf{x} + \mathbf{b}^\mathsf{T}\mathbf{x} + c \le 0,\ \alpha^\mathsf{T}\mathbf{x} = \beta\right\}$ is convex. Observe that $\alpha^\mathsf{T}\mathbf{x} = \beta \implies \lambda\mathbf{x}^\mathsf{T}\alpha\alpha^\mathsf{T}\mathbf{x} - \lambda\beta^2 = 0$, adding this to the first inequality one obtains

$$\begin{aligned}
C' &= \{\mathbf{x} \in \mathbb{R}^d : \mathbf{x}^\mathsf{T}(A + \lambda\alpha\alpha^\mathsf{T})\mathbf{x} + \mathbf{b}^\mathsf{T}\mathbf{x} + c - \lambda\beta^2 \le 0,\ \alpha^\mathsf{T}\mathbf{x} = \beta\} \\
&= \{\mathbf{x} \in \mathbb{R}^d : \mathbf{x}^\mathsf{T}(A + \lambda\alpha\alpha^\mathsf{T})\mathbf{x} + \mathbf{b}^\mathsf{T}\mathbf{x} + c - \lambda\beta^2 \le 0\} \cap \{\mathbf{x} \in \mathbb{R}^d : \alpha^\mathsf{T}\mathbf{x} = \beta\},
\end{aligned}$$

which is the intersection of two convex sets, thus convex. □

Recitation 2 September 8, 2023

2.1 Review

Theorem 2.1 (Characterization of closed sets in Euclidean space)

A set $C \subseteq \mathbb{R}^d$ is closed $\iff$ every convergent sequence in C has its limit point in C. ♡

Definition 2.1 (Compact set)

A set $C \subseteq \mathbb{R}^d$ is compact if for every family $\{U_i\}_{i \in I}$ of sets with $U_i \subseteq \mathbb{R}^d$ open and $C \subseteq \bigcup_{i \in I} U_i$ there exists a finite subset $I' \subseteq I$ with $C \subseteq \bigcup_{i \in I'} U_i$. ♣

Theorem 2.2 (Heine-Borel Theorem)

A set $C \subseteq \mathbb{R}^d$ is compact $\iff$ C is closed and bounded. ♡

Theorem 2.3 (Bolzano-Weierstrass Theorem)

Every bounded sequence in $\mathbb{R}^d$ has a convergent subsequence. ♡

Proof [Proof Sketch] By the Completeness Axiom for real numbers we know that every bounded sequence in $\mathbb{R}$ has a convergent subsequence. Then use this argument iteratively on coordinates to prove that every bounded sequence in $\mathbb{R}^d$ has a convergent subsequence. □

Theorem 2.4

Let $f : \mathbb{R}^d \to \mathbb{R}^n$ be a continuous function, and C be a compact set in $\mathbb{R}^d$. Then $f(C)$ is a compact set in $\mathbb{R}^n$. ♡

Proof Let $\{U_i\}_{i \in I}$ be a family of sets with $U_i \subseteq \mathbb{R}^n$ open and $f(C) \subseteq \bigcup_{i \in I} U_i$. Then $f^{-1}(U_i) \subseteq \mathbb{R}^d$ is open for every $i \in I$ and $C \subseteq \bigcup_{i \in I} f^{-1}(U_i)$. Since C is compact, there exists a finite subset $I' \subseteq I$ such that $C \subseteq \bigcup_{i \in I'} f^{-1}(U_i)$. Then $f(C) \subseteq \bigcup_{i \in I'} U_i$, which implies that $f(C)$ is compact. □

Theorem 2.5 (Weierstrass' Theorem)

Let $C \subseteq \mathbb{R}^d$ be a compact set, and let $f : \mathbb{R}^d \mapsto \mathbb{R}$ be a continuous function. Then there exist $\mathbf{x}^{\min}, \mathbf{x}^{\max} \in C$ such that $f(\mathbf{x}^{\min}) \leq f(\mathbf{x}) \leq f(\mathbf{x}^{\max})$ for all $\mathbf{x} \in C$. ♡

Proof Since f is continuous and C is compact, $f(C)$ is compact by Theorem 2.4. Then by Theorem 2.2 $f(C)$ is closed and bounded. The boundedness implies that $\inf f(C)$ and $\sup f(C)$ are finite, and by definition there exist $\{\mathbf{a}_n\}_{n\in\mathbb{N}}$ and $\{\mathbf{b}_n\}_{n\in\mathbb{N}}$ such that $f(\mathbf{a}_n) \to \inf f(C)$ and $f(\mathbf{b}_n) \to \sup f(C)$. Then the closedness of $f(C)$ implies that $\inf f(C), \sup f(C) \in f(C)$ and there exist $\mathbf{x}^{\min} \in f^{-1}(\inf f(C))$ and $\mathbf{x}^{\max} \in f^{-1}(\sup f(C))$ such that $f(\mathbf{x}^{\min}) \leq f(\mathbf{x}) \leq f(\mathbf{x}^{\max})$ for all $\mathbf{x} \in C$. □

Example 2.1 $C \subseteq \mathbb{R}^d$ is a compact convex set $\not\Rightarrow$ $\text{cone}(C)$ is closed.

Proof An example is provided in Figure 2.1. □

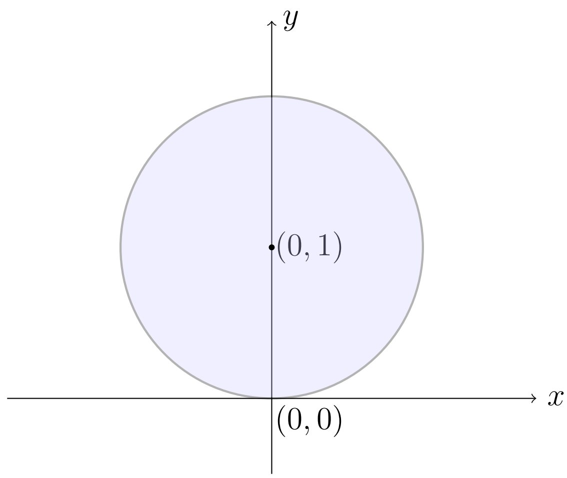


Figure 2.1: $X = \{(x, y) : x^2 + (y-1)^2 \leq 1\}$, but $\text{cone}(X) = \{(x, y) : y > 0\} \cup \{(0, 0)\}$ is not closed.

Example 2.2 $X \subseteq \mathbb{R}^d$ is a closed set $\not\Rightarrow$ $\text{conv}(X)$ is a closed set.

Proof Consider $X = (\{0\} \times [0, 1]) \cup ([0, +\infty) \times \{0\})$, which is a closed set in $\mathbb{R}^2$. However, $\text{conv}(X) = [0, +\infty) \times [0, 1)$ is not closed. □

Example 2.3 $X,\ Y \subseteq \mathbb{R}^d$ are closed sets $\not\Rightarrow$ $X + Y$ is a closed set.

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

Example 2.4 $X,\ Y \subseteq \mathbb{R}^d$ are closed convex sets $\not\Rightarrow X + Y$ is a closed set.

Proof Consider $X = \{(x, y) \in \mathbb{R}^2 : y \geq e^x\}$ and $Y = \{(x, 0) : x \geq 0\}$, which are both closed convex sets in $\mathbb{R}^2$. However, $X + Y = \{(x, y) \in \mathbb{R}^2 : y > 0\}$ is not closed. □

2.2 Exercises

✍ **Exercise 2.1** Show that if X is compact and Y is closed, then $X + Y$ is closed.

Proof Let's consider an arbitrary convergent sequence $\mathbf{x}^i + \mathbf{y}^i \to \mathbf{z}$ where $\mathbf{x}^i \in X$ and $\mathbf{y}^i \in Y$ for each $i \in \mathbb{N}_+$. Our goal is to establish that $\mathbf{z} \in X + Y$. Due to the compactness of X, there exists a convergent subsequence $\mathbf{x}^{i_k} \to \mathbf{x}$ for some $\mathbf{x} \in X$. Along this subsequence, we find that

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

implying that $\mathbf{y}^{i_k} \to \mathbf{z} - \mathbf{x}$ as $k \to +\infty$. Given that Y is closed, it follows that $\mathbf{z} - \mathbf{x} \in Y$. Consequently, $\mathbf{z} = \mathbf{x} + (\mathbf{z} - \mathbf{x})$ belongs to $X + Y$. □

✍ **Exercise 2.2** Let $\mathbf{a}^1, \ldots, \mathbf{a}^n \in \mathbb{R}^d$. Then cone $(\{\mathbf{a}^1, \ldots, \mathbf{a}^n\})$ is closed.

Proof Consider a convergent sequence $\{\mathbf{x}^i\}_{i\in\mathbb{N}} \subseteq \text{cone}\,(\{\mathbf{a}^1, \ldots, \mathbf{a}^n\})$ converging to $\mathbf{x} \in \mathbb{R}^d$. By Carathéodory's Theorem (Theorem 2.2.13), every $\mathbf{x}^i$ is in the conical hull of some linearly independent subset of $\{\mathbf{a}^1, \ldots, \mathbf{a}^n\}$. Since there are only finitely many linearly independent subsets of $\{\mathbf{a}^1, \ldots, \mathbf{a}^n\}$, the conical hull of one of these subsets contains infinitely many elements of the sequence $\{\mathbf{x}^i\}_{i\in\mathbb{N}}$. Thus, after passing to that subsequence, we may assume that $\{\mathbf{x}^i\}_{i\in\mathbb{N}} \subseteq \text{cone}\left(\{\bar{\mathbf{a}}^1, \ldots, \bar{\mathbf{a}}^k\}\right)$ where $\bar{\mathbf{a}}^1, \ldots, \bar{\mathbf{a}}^k$ are linearly independent. For each $\mathbf{x}^i$, there exists $\boldsymbol{\lambda}^i \in \mathbb{R}^k_+$ such that $\mathbf{x}^i = \boldsymbol{\lambda}^i_1\bar{\mathbf{a}}^1 + \ldots + \boldsymbol{\lambda}^i_k\bar{\mathbf{a}}^k$. If we denote by $A \in \mathbb{R}^{d\times k}$ the matrix whose columns are $\bar{\mathbf{a}}^1, \ldots, \bar{\mathbf{a}}^k$, then $\mathbf{x}^i = A\boldsymbol{\lambda}^i$. Since the columns of A are linearly independent, there exists a matrix $B = (A^\mathsf{T}A)^{-1}A^\mathsf{T} \in \mathbb{R}^{k\times d}$ (since in Euclidean space $\text{rank}(A^\mathsf{T}A) = \text{rank}(A)$, and thus $A^\mathsf{T}A$ is invertible) such that BA is the identity matrix. Thus, $B\mathbf{x}^i = BA\boldsymbol{\lambda}^i = \boldsymbol{\lambda}^i$ for every $i \in \mathbb{N}$. Since $\{\mathbf{x}^i\}_{i\in\mathbb{N}}$ is a convergent sequence, it is also a bounded set. This implies that $\{\boldsymbol{\lambda}^i\}_{i\in\mathbb{N}}$ is a bounded set in $\mathbb{R}^k_+$ because it is the image of a bounded set under the linear (and therefore continuous) map given by the matrix B. Thus, by Theorem 2.3 there is a convergent subsequence $\boldsymbol{\lambda}^{i_\ell} \to \boldsymbol{\lambda} \in \mathbb{R}^k_+$. Taking limits,

$$\mathbf{x} = \lim_{\ell\to\infty} \mathbf{x}^{i_\ell} = \lim_{\ell\to\infty} A\boldsymbol{\lambda}^{i_\ell} = A\boldsymbol{\lambda}.$$

Since $\boldsymbol{\lambda} \in \mathbb{R}^k_+$, we find that $\mathbf{x} \in \text{cone}\left(\{\bar{\mathbf{a}}^1, \dots, \bar{\mathbf{a}}^k\}\right) \subseteq \text{cone}\left(\{\mathbf{a}^1, \dots, \mathbf{a}^n\}\right)$. □

✍ **Exercise 2.3** If X is a set of affinely independent points, then $\dim(\text{aff}(X)) = |X| - 1$.

Proof Let $X = \{\mathbf{x}^1, \dots, \mathbf{x}^d, \mathbf{x}\}$ for some $d \in \mathbb{N}_+$, and now $|X| = d + 1$. By Theorem 2.2.6, $L = \text{aff}(X) - \mathbf{x}$ is a linear subspace. We claim that $(X\backslash\{\mathbf{x}\}) - \mathbf{x} = \{\mathbf{x}^1 - \mathbf{x}, \dots, \mathbf{x}^d - \mathbf{x}\} := \mathcal{B}$ is a basis for L. By Proposition 2.2.5, $\mathcal{B}$ is linearly independent, so we just need to show that $\text{span}(\mathcal{B}) = L$. Notice that

$$\begin{aligned}
\mathbf{z} \in L \iff & \mathbf{z} = \lambda_1\mathbf{x}^1 + \cdots + \lambda_d\mathbf{x}^d + \lambda_{d+1}\mathbf{x} - \mathbf{x}, \text{ where } \sum_{i=1}^{d+1} \lambda_i = 1 \\
\iff & \mathbf{z} = \lambda_1\mathbf{x}^1 + \cdots \lambda_d\mathbf{x}^d + \lambda_{d+1}\mathbf{x} - \sum_{i=1}^{d+1} \lambda_i\mathbf{x}, \text{ where } \sum_{i=1}^{d+1} \lambda_i = 1 \\
\iff & \mathbf{z} = \sum_{i=1}^{d} \lambda_i(\mathbf{x}^i - \mathbf{x}) \\
\iff & \mathbf{z} \in \text{span}(\mathcal{B}).
\end{aligned}$$

Therefore, $\dim(\text{aff}(X)) = \dim(L) = d = |X| - 1$. □

✍ **Exercise 2.4** Let $X \subseteq \mathbb{R}^d$ be a set. Then X is a linear subspace if and only if X is both a cone and an affine subset.

Proof It is direct to show the forward direction. For the reverse direction, suppose X is both a cone and an affine subset. $\forall \mathbf{x}, \mathbf{y} \in X$, $\lambda, \gamma \in \mathbb{R}$, then

$$\lambda\mathbf{x} + \gamma\mathbf{y} = \lambda\mathbf{x} + \gamma\mathbf{y} + (1 - \lambda - \gamma)\mathbf{0} \in X,$$

where the first equality holds since X is a cone. □

Exercise 2.5 Let $C \subseteq \mathbb{R}^d$, then $\operatorname{span}(C - C) = \operatorname{aff}(C) - \bar{\mathbf{x}}$ for any $\bar{\mathbf{x}} \in C$.

Proof $\forall \mathbf{z} \in \operatorname{span}(C-C)$, $\mathbf{z} = \gamma_1(\mathbf{x}^1 - \mathbf{y}^1) + \cdots + \gamma_t(\mathbf{x}^t - \mathbf{y}^t)$ for some $\mathbf{x}^1, \ldots, \mathbf{x}^t, \mathbf{y}^1, \ldots, \mathbf{y}^t \in C$ and $\gamma_1, \ldots, \gamma_t \in \mathbb{R}$. Then $\mathbf{z} = \gamma_1\mathbf{x}^1 + \cdots + \gamma_t\mathbf{x}^t + (-\gamma_1)\mathbf{y}^1 + \cdots + (-\gamma_t)\mathbf{y}^t + \bar{\mathbf{x}} - \bar{\mathbf{x}} \in \operatorname{aff}(C) - \bar{\mathbf{x}}$ where $\bar{\mathbf{x}}$ is any point in C since $\sum_{i=1}^t \gamma_i + \sum_{i=1}^t(-\gamma_i) + 1 = 1$, this proves $\operatorname{span}(C-C) \subseteq \operatorname{aff}(C) - \bar{\mathbf{x}}$. To show the reverse inclusion, $\forall \mathbf{z} \in \operatorname{aff}(C) - \bar{\mathbf{x}}$ for any $\bar{\mathbf{x}} \in C$, by definition $\mathbf{z} = \sum_{i=1}^t \lambda_i \mathbf{x}^i - \bar{\mathbf{x}}$ for some $\mathbf{x}^1, \ldots, \mathbf{x}^t \in C$ and $\lambda_1, \ldots, \lambda_t \in \mathbb{R}$ with $\sum_{i=1}^t \lambda_i = 1$. Then $\mathbf{z} = \sum_{i=1}^t \lambda_i(\mathbf{x}^i - \bar{\mathbf{x}}) \in \operatorname{span}(C-C)$, this proves $\operatorname{aff}(C) - \bar{\mathbf{x}} \subseteq \operatorname{span}(C-C)$. □

Remark Notice that $\operatorname{aff}(C) = \operatorname{aff}(\operatorname{conv}(C))$ since $C \subseteq \operatorname{conv}(C) \subseteq \operatorname{aff}(C)$. Therefore, Exercise 2.5 implies that $\dim(\operatorname{span}(C-C)) = \dim(\operatorname{aff}(C)) = \dim(\operatorname{conv}(C))$.

Exercise 2.6 Let $X \subseteq \mathbb{R}^d$. Show that X is a hyperplane if and only if X is an affine set of dimension $d-1$. [Recall that a hyperplane is a set of the form $\{\mathbf{x} \in \mathbb{R}^d : \langle \mathbf{a}, \mathbf{x} \rangle = \delta\}$.]

Proof Suppose X is an affine set with a dimension of $d-1$. By definition, such a set X contains d affinely independent points and no more. Consequently, there exists a distinct set of d affinely independent points, denoted X', for which $\operatorname{aff}(X') = X$. By Theorem 2.2.6 discussed in our class concerning the characterization of affine subspaces, we can identify a matrix A belonging to $\mathbb{R}^{(d-(d-1))\times d} = \mathbb{R}^{1\times d}$ and a scalar b. With these, the set X can be expressed as $X = \{\mathbf{x} \in \mathbb{R}^d : A\mathbf{x} = b\}$, indicating that X represents a hyperplane. The inverse argument can be derived similarly from the aforementioned theorem. □

Exercise 2.7 Let $X \subseteq \mathbb{R}^d$ be a set of $d+1$ affinely independent points. Show that $\operatorname{int}(\operatorname{conv}(X)) \neq \varnothing$.

Proof Hint: One can construct a point $\mathbf{x} \in \operatorname{conv}(X)$ (consider choosing some $\lambda \in (0,1)^{d+1}$), and prove that $\mathbf{x} \in \operatorname{int}(\operatorname{conv}(X))$. □

Exercise 2.8 Let $X \subseteq \mathbb{R}^d$, and let $\mathbf{y} \in \operatorname{conv}(X)$. Suppose H is a halfspace such that $\mathbf{y} \in H$. Prove that $H \cap X \neq \varnothing$.

Proof Consider the set $H = \{\mathbf{x} : \langle \mathbf{a}, \mathbf{x} \rangle \leq \delta\}$, where $\mathbf{a} \in \mathbb{R}^d$ and $\delta > 0$. Given that $\mathbf{y} \in \operatorname{conv}(X)$, it follows that $\mathbf{y} = \sum_{i=1}^k \lambda_i \mathbf{x}^i$, with $\mathbf{x}^i \in X$, $\lambda_i \geq 0$ for $i = 1, \ldots, k$, and

$\sum_{i=1}^{k} \lambda_i = 1$. Assume, for the sake of contradiction, that $H \cap X = \emptyset$. In this case, $\langle \mathbf{a}, \mathbf{x}^i \rangle > \delta$ for all $i \in \{1, \ldots, k\}$. However, this leads to $\langle \mathbf{a}, \mathbf{y} \rangle > \delta$, which is a contradiction. □

✍ **Exercise 2.9** Let $X \subseteq \mathbb{R}^d$ be a compact set (not necessarily convex). Then $\text{conv}(X)$ is compact.

Proof By Theorem 2.2.14, every $\mathbf{x} \in \text{conv}(X)$ is the convex combination of some $d+1$ points in X. Define the following function $f : \underbrace{\mathbb{R}^d \times \ldots \times \mathbb{R}^d}_{d+1 \text{ times}} \times \mathbb{R}^{d+1} \to \mathbb{R}^d$ as follows:

$$f\left(\mathbf{y}^1, \ldots, \mathbf{y}^{d+1}, \boldsymbol{\lambda}\right) = \lambda_1 \mathbf{y}^1 + \ldots + \lambda_{d+1} \mathbf{y}^{d+1}.$$

It is easily verified that f is a continuous function (each coordinate of $f(\cdot)$ is a bilinear quadratic function of the input). We now observe that $\text{conv}(X)$ is the image of $\underbrace{X \times \ldots \times X}_{d+1 \text{ times}} \times \Delta^{d+1}$ under f, where

$$\Delta^{d+1} := \left\{\boldsymbol{\lambda} \in \mathbb{R}_+^{d+1} : \lambda_1 + \ldots + \lambda_{d+1} = 1\right\}.$$

Since X and Δ^{d+1} are compact sets, we obtain the result by applying Theorem 1.3.15. □

Recitation 3 September 15, 2023

3.1 Review

Each of the following statements can be used as the definition of relative interior.

Theorem 3.1 (Equivalent definitions of relative interior)

Let $C \subseteq \mathbb{R}^d$ be a convex set and $\mathbf{x} \in C$. The following are equivalent.

(i) There exists $\varepsilon > 0$ such that $B(\mathbf{x}, \varepsilon) \cap \operatorname{aff}(C) \subseteq C$.

(ii) There exists $\varepsilon > 0$ such that $\forall \mathbf{y} \in \operatorname{aff}(C), \mathbf{x} + \varepsilon \left(\frac{\mathbf{y}-\mathbf{x}}{\|\mathbf{y}-\mathbf{x}\|} \right) \in C$.

(iii) $\forall \mathbf{y} \in \operatorname{aff}(C),\ \exists \varepsilon_{\mathbf{y}} > 0$ such that $\mathbf{x} + \varepsilon_{\mathbf{y}}(\mathbf{y} - \mathbf{x}) \in C$.

♡

Proof [Proof Sketch] The equivalence between (i) and (ii) is straightforward, and it is clear from the definition that (ii) implies (iii), so here we discuss the derivation of (ii) from (iii). The key of the proof lies in noting that, when we consider any sequence $\{\mathbf{y}^i\}_{i\in\mathbb{N}}$ within a compact set $\operatorname{aff}(C) \cap \operatorname{bd}(B(\mathbf{x}, 1))$, there must exist a convergent subsequence. This observation provides a unified positive ε^* in (ii). □

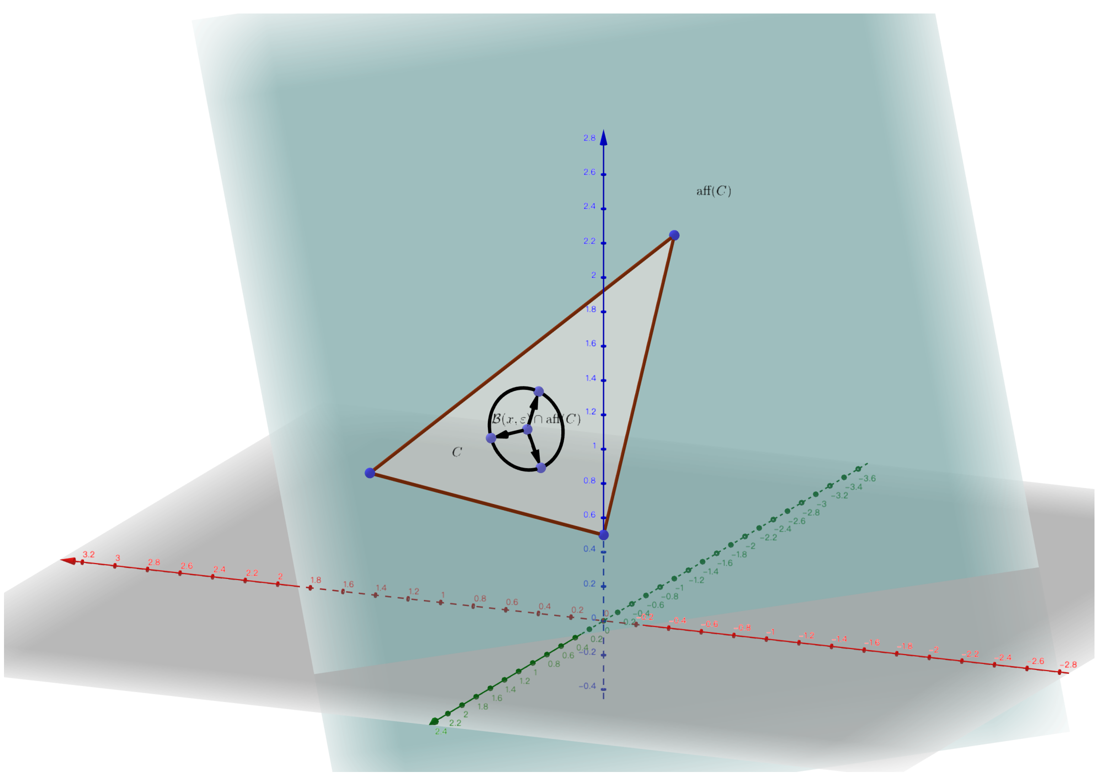


Figure 3.1: Relative interior of a convex set in $\mathbb{R}^3$.

Example 3.1 Let $X, Y \subseteq \mathbb{R}^d$. $X \subseteq Y \not\Rightarrow \text{relint}(X) \subseteq \text{relint}(Y)$.

Proof An example is provided in Figure 3.2. □

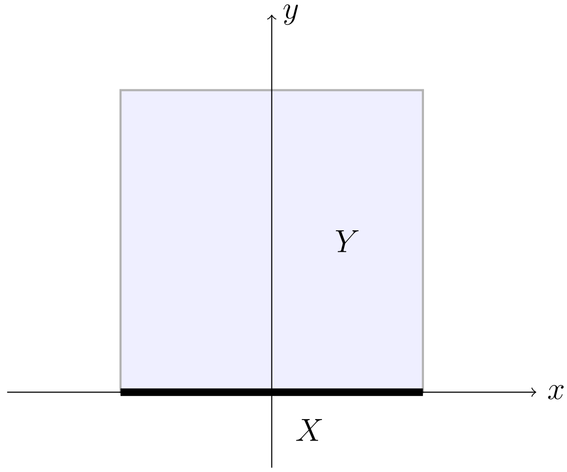


Figure 3.2: $X = \{(x, y) : x \in [-1, 1], y = 0\}$, $Y = \{(x, y) : x \in [-1, 1], y \in [0, 2]\}$, then $X \subseteq Y$ but $\text{relint}(X) \cap \text{relint}(Y) = \varnothing$.

3.2 Revisit Carathéodory's theorem

Lemma 3.1

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

$$\text{conv}\left(H^= \cap X\right) = H^= \cap \text{conv}\left(X\right),$$

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

Proof Consider a half-space H defined by $H = \{\mathbf{x} \in \mathbb{R}^d : \langle \mathbf{a}, \mathbf{x} \rangle \leq \delta\}$, where $\mathbf{a} \in \mathbb{R}^d$ and $\delta > 0$. Let $H^=$ be the boundary of H, defined by $H^= = \{\mathbf{x} : \langle \mathbf{a}, \mathbf{x} \rangle = \delta\}$.

Given any point $\mathbf{x}$ in the intersection $H^= \cap \text{conv}(X)$, we can write $\mathbf{x}$ as a convex combination of points in X: $\mathbf{x} = \sum_{i=1}^k \lambda_i \mathbf{x}^i$, where $\mathbf{x}^i \in X$, $\lambda_i \geq 0$ for $i = 1, \ldots, k$, and $\sum_{i=1}^k \lambda_i = 1$. Then we have:

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

Since $\lambda_i(\langle \mathbf{a}, \mathbf{x}^i \rangle - \delta) \leq 0$ for all i, each λ_i must either be zero, or $\langle \mathbf{a}, \mathbf{x}^i \rangle = \delta$. Therefore, any $\mathbf{x}^i$ with $\lambda_i \neq 0$ must lie in $H^=$. Hence, $\mathbf{x}$ belongs to $\text{conv}(H^= \cap X)$.

To show the converse, for any $\mathbf{x}$ in $\text{conv}(H^= \cap X)$, we can similarly express $\mathbf{x}$ as a convex combination of points in $H^= \cap X$. This implies that $\mathbf{x}$ is in both $H^=$ and $\text{conv}(X)$, confirming that $\mathbf{x} \in H^= \cap \text{conv}(X)$. □

Apply Lemma 3.1, we can provide an alternate proof of the Carathéodory's theorem (Rockafellar 1970).

Theorem 3.2 (Carathéodory's theorem (convex version))

Let $X \subseteq \mathbb{R}^d$ and $\mathbf{x} \in \text{conv}(X)$. Then $\mathbf{x}$ is a convex combination of at most $d+1$ points of X. ♡

Proof
Base Case: For $d = 1$, the statement can be directly verified.
Inductive Step: Assume the statement holds for dimensions smaller than d. Let $\mathbf{x} \in \text{conv}(X)$. By definition, distinct vectors $\mathbf{x}^1, \ldots, \mathbf{x}^k$ exist, such that $\mathbf{x} \in \text{conv}(\{\mathbf{x}^1, \ldots, \mathbf{x}^k\}) := C$.
Case 1: If $\mathbf{x} \in \text{relbd}(C)$, then consider $H^=$, a supporting hyperplane of C that passes through $\mathbf{x}$. It follows that

$$\mathbf{x} \in H^= \cap C = H^= \cap \text{conv}(\{\mathbf{x}^1, \ldots, \mathbf{x}^k\}) \overset{3.1}{=} \text{conv}(H^= \cap \{\mathbf{x}^1, \ldots, \mathbf{x}^k\}).$$

Because $H^=$ is $(d-1)$-dimensional, the induction hypothesis implies that $\mathbf{x}$ is a convex combination of at most d vectors from $H^= \cap \{\mathbf{x}^1, \ldots, \mathbf{x}^k\}$, hence of at most d points from X.
Case 2: If $\mathbf{x} \in \text{relint}(C)$, note that there exists some $i \in \{1, \ldots, k\}$ such that $\mathbf{x} \neq \mathbf{x}^i$. Let $\mathbf{y} \in \{\mathbf{x}^i + \lambda(\mathbf{x} - \mathbf{x}^i) : \lambda > 0\} \cap \text{relbd}(C)$. Hence, $\mathbf{x}$ is a convex combination of at most $d+1$ points from X, as $\mathbf{x}$ can be expressed as a convex combination of $\mathbf{y}$ and $\mathbf{x}^i$, and $\mathbf{y}$ itself is a convex combination of at most d points from X. □

3.3 Exercises

✍ **Exercise 3.1** Prove that the relative interior of a nonempty convex set is nonempty.

Proof An analogy of Exercise 2.7. □

Alternating projection algorithm

Given two closed convex sets C and D in $\mathbb{R}^n$, the *alternating projection* method seeks to find a point in the intersection of these sets by iteratively projecting onto each set. The process is initiated with any point $\mathbf{x}^0 \in C$. In each iteration of the algorithm, the point is first projected onto D followed by a projection onto C. This yields sequences $\mathbf{x}^k$ and $\mathbf{y}^k$ given by:

$$\mathbf{y}^k = \mathrm{Proj}_D(\mathbf{x}^k),\ \mathbf{x}^{k+1} = \mathrm{Proj}_C(\mathbf{y}^k),\ k \in \mathbb{N}. \tag{3.1}$$

Algorithm 1 Alternating projection algorithm

Require: Closed convex sets $C, D \subseteq \mathbb{R}^n$, initial point $\mathbf{x}^0 \in C$

for $k = 0, 1, 2, \dots$ **do**
  $\mathbf{y}^k \leftarrow \mathrm{Proj}_D(\mathbf{x}^k)$
  $\mathbf{x}^{k+1} \leftarrow \mathrm{Proj}_C(\mathbf{y}^k)$
end for

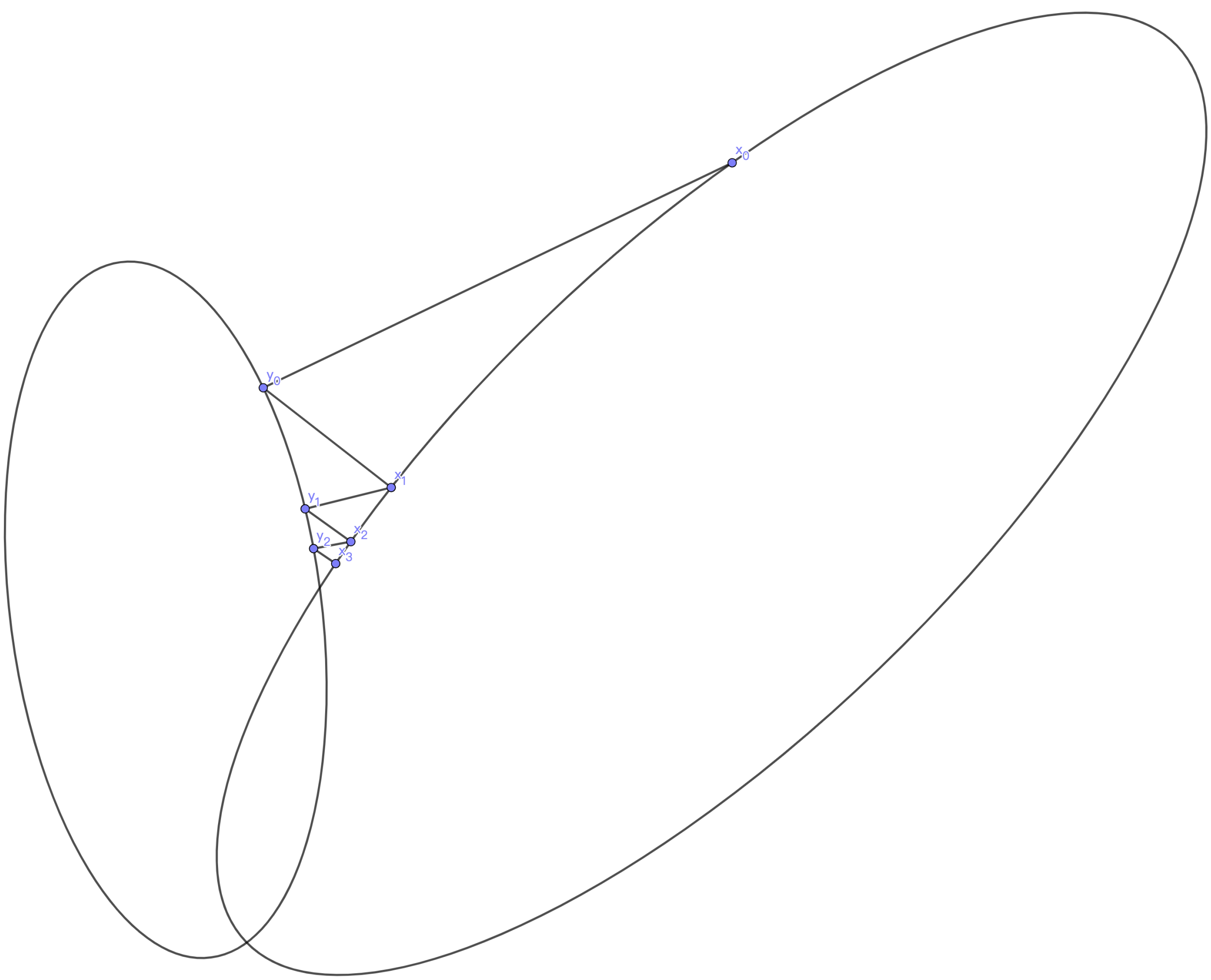


Figure 3.3: Alternating projections algorithm on two closed convex sets with nonempty intersection.

Exercise 3.2 Let $C, D \subseteq \mathbb{R}^d$ be nonempty closed convex sets with $C \cap D \neq \varnothing$. Show that the alternating projection algorithm starts at any $\mathbf{x}^0 \in C$ converges to a point $\mathbf{x}^* \in C \cap D$.

Proof Consider any $\bar{\mathbf{x}} \in C \cap D$. We first notice that $\forall n \in \mathbb{N}$, the following holds:

$$\langle \mathbf{x}^n - \mathbf{y}^n, \mathbf{z} - \mathbf{y}^n \rangle \leq 0, \ \forall \mathbf{z} \in D. \tag{3.2}$$

From this, we deduce:

$$\begin{aligned} \|\mathbf{x}^n - \bar{\mathbf{x}}\|_2^2 &= \|\mathbf{x}^n - \mathbf{y}^n + \mathbf{y}^n - \bar{\mathbf{x}}\|_2^2 \\ &= \|\mathbf{x}^n - \mathbf{y}^n\|_2^2 + \|\mathbf{y}^n - \bar{\mathbf{x}}\|_2^2 + 2\langle \mathbf{x}^n - \mathbf{y}^n, \mathbf{y}^n - \bar{\mathbf{x}} \rangle \\ &\geq \|\mathbf{x}^n - \mathbf{y}^n\|_2^2 + \|\mathbf{y}^n - \bar{\mathbf{x}}\|_2^2, \end{aligned}$$

given that $\bar{\mathbf{x}}$ lies in $C \cap D$ and based on the inequality in (3.2). Rearranging the terms, we have:

$$\|\mathbf{y}^n - \bar{\mathbf{x}}\|_2^2 \leq \|\mathbf{x}^n - \bar{\mathbf{x}}\|_2^2 - \|\mathbf{x}^n - \mathbf{y}^n\|_2^2 \leq \|\mathbf{x}^n - \bar{\mathbf{x}}\|_2^2 \tag{3.3}$$

and subsequently:

$$\left\|\mathbf{x}^{n+1} - \bar{\mathbf{x}}\right\|_2^2 \leq \|\mathbf{y}^n - \bar{\mathbf{x}}\|_2^2 - \left\|\mathbf{x}^{n+1} - \mathbf{y}^n\right\|_2^2 \leq \|\mathbf{y}^n - \bar{\mathbf{x}}\|_2^2 \leq \|\mathbf{x}^n - \bar{\mathbf{x}}\|_2^2. \tag{3.4}$$

By (3.4) we know the sequence $\{\|\mathbf{x}^n - \bar{\mathbf{x}}\|_2\}_{n \in \mathbb{N}}$ is bounded in $\mathbb{R}^d$ by $\|\mathbf{x}^0 - \bar{\mathbf{x}}\|_2$, ensuring the existence of a convergent subsequence $\{\mathbf{x}^{n_i}\}_{i \in \mathbb{N}}$ to a point $\mathbf{x}^*$ in $\mathbb{R}^d$, with $\mathbf{x}^* \in C$ since C is closed. A similar argument assures the existence of a convergent subsequence $\{\mathbf{y}^{n_j}\}_{j \in \mathbb{N}}$ that converges to a point $\mathbf{y}^*$ in D. Our goal is to show that $\mathbf{x}^*$ equals $\mathbf{y}^*$. Given the inequality (3.3), the sequence $\{\|\mathbf{x}^0 - \bar{\mathbf{x}}\|_2, \|\mathbf{y}^0 - \bar{\mathbf{x}}\|_2, \|\mathbf{x}^1 - \bar{\mathbf{x}}\|_2, \ldots\}$ is monotone decreasing and lower bounded, thus a convergent sequence. As a result, the norm difference $\|\mathbf{y}^n - \mathbf{x}^n\|_2 \to 0$ as $n \to +\infty$, which completes the proof. □

3.4 Linear separability of boolean functions

Consider any finite set $F = \{\mathbf{x}^1, \ldots, \mathbf{x}^t\} \subseteq \mathbb{R}^d$ with $t \geq d + 1$, we say a subset $X \subseteq F$ is linearly separable if there exists some $\mathbf{a} \in \mathbb{R}^d$, $\delta \in \mathbb{R}$ such that $X \subseteq H^{\leq}(\mathbf{a}, \delta)$ and $X^c \subseteq H^{>}(\mathbf{a}, \delta)$, where $X^c := F \backslash X$ is the complement of X in F. Let $K_d = \{0, 1\}^d$ be the set of vertices of a d-dimensional hypercube.

Exercise 3.3 Show that a subset $X \subseteq F$ is linearly separable $\iff$ $\text{conv}(X) \cap \text{conv}(X^c) = \varnothing$.

Proof By definition. □

Lemma 3.2 (Linearly separable sets)

Consider any finite subset $F = \{\mathbf{x}^1, ..., \mathbf{x}^t\} \subseteq \mathbb{R}^d$ with $t \geq d+1$, then there are at most $2\sum_{i=0}^{d}\binom{t-1}{i}$ linearly separable subsets. The equality could be achieved if $\forall S \subseteq F$ with $|S| \leq d+1$, S is affinely independent. ♡

Proof The number of linearly separable subsets of F is equal to the cardinality of the following set:

$$\Omega = \{(\mathbf{a}, b) \in \mathbb{R}^{d+1} : (\mathrm{sgn}(\langle \mathbf{a}, \mathbf{x}^1\rangle - b), ..., \mathrm{sgn}(\langle \mathbf{a}, \mathbf{x}^1\rangle - b))\}.$$

By Lemma 3.3 in Anthony and Bartlett 1999,

$$|\Omega| \leq 2\sum_{i=0}^{d}\binom{t-1}{i},$$

and the equality is achieved if any $d+1$ points of $\{(\mathbf{x}^1, -1), ..., (\mathbf{x}^t, -1)\}$ are linearly independent, that is, any $d+1$ points of F is affinely independent. □

✍ **Exercise 3.4** Let $\mathcal{S}_d$ be the collection of all the linearly separable sets in K_d, prove that $|\mathcal{S}_d| \leq 2^{d^2+1}/d!$.

Proof By a direct application of Lemma 3.2, we have:

$$|\mathcal{S}_d| \leq 2\sum_{i=0}^{d}\binom{2^d-1}{i} \leq \frac{2^{d^2+1}}{d!}.$$

□

✍ **Exercise 3.5** Let $U = \left\{\mathbf{x} \in \{0,1\}^d : \sum_{i=1}^{d}\mathbf{x}_i \text{ is even}\right\} \subseteq K_d$. Prove that U is linear separable $\iff |U| \leq 1$.

Proof [Proof Sketch] For $|U| \geq 2$, take any two distinct points within U, then construct two points in U^c that share the same midpoint. □

Recitation 4 September 22, 2023

4.1 Review

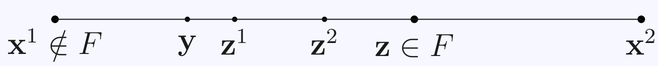
Proposition 4.1 (Equivalent definition of a face)

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 ($\Longleftarrow$) For this direction, the case is trivial by setting $\lambda = \frac{1}{2}$.
($\Longrightarrow$) Let $\mathbf{x}^1, \mathbf{x}^2 \in C$ and $\lambda \in (0,1)$ be given, such that $\mathbf{z} = \lambda \mathbf{x}^1 + (1-\lambda)\mathbf{x}^2 \in F$. Without loss of generality, assume $\mathbf{x}^1 \in C \backslash F$. If $\mathbf{z}^1 = \frac{\mathbf{x}^1 + \mathbf{z}}{2} \in F$, it would lead to a contradiction. Hence, we can iteratively construct a sequence $\mathbf{z}^k = \frac{\mathbf{z}^{k-1} + \mathbf{z}}{2} \in C \backslash F$ for $k \geq 1$.
Now, assume that for some $\lambda' \in (\lambda, 1)$, $\mathbf{y} = \lambda' \mathbf{x}^1 + (1-\lambda')\mathbf{x}^2 \in F$. By the convexity of F, the line segment connecting $\mathbf{y}$ and $\mathbf{z}$ should be entirely in F, which would imply that it contains $\mathbf{z}^k$ for some $k \geq 1$. Therefore, $\{\lambda' \mathbf{x}^1 + (1-\lambda')\mathbf{x}^2 : \lambda' \in (\lambda, 1)\} \subseteq C \backslash F$.
However, consider

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

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



□

Example 4.1 $C \subseteq \mathbb{R}^d$ is a closed convex set $\not\Rightarrow$ every face of C is an exposed face.

Proof An example is provided in Figure 4.1. □

Example 4.2 $X \subseteq \mathbb{R}^d$ is convex $\not\Rightarrow$ every face of X is a closed set.

Proof Let X be any open convex sets in $\mathbb{R}^d$, then a trivial face is X itself. □

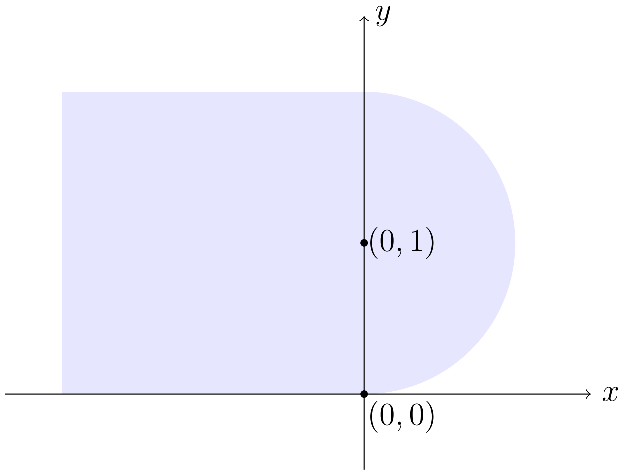


Figure 4.1: In $\mathbb{R}^2$, let $X = \operatorname{conv}(\{(-2,0),(0,0),(0,2),(-2,2)\})$, $Y = \{(x,y) : x^2 + (y-1)^2 \le 1\}$, then $X \cup Y$ has a face $\{(0,0)\}$ which is not an exposed face.

Definition 4.1 (Dual sets)

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

1. The *polar* of X is defined as $X^\circ = \{\mathbf{y} \in \mathbb{R}^d : \langle \mathbf{y}, \mathbf{x} \rangle \le 1, \forall \mathbf{x} \in X\}$.
2. The *polar cone* of X is defined as $X^\bullet = \{\mathbf{y} \in \mathbb{R}^d : \langle \mathbf{y}, \mathbf{x} \rangle \le 0, \forall \mathbf{x} \in X\}$.
3. The *dual cone* of X is defined as $X^* = \{\mathbf{y} \in \mathbb{R}^d : \langle \mathbf{y}, \mathbf{x} \rangle \ge 0, \forall \mathbf{x} \in X\}$.

♣

Remark $X^\bullet = -X^*$.

Definition 4.2 (Orthogonal complement)

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.

♣

4.2 Exercises

✍ **Exercise 4.1** Let $X \subseteq \mathbb{R}^d$ be a convex cone, prove that $X^\circ = X^\bullet$.

Proof It's direct to check that $X^\bullet \subseteq X^\circ$. For the reverse inclusion, consider an arbitrary vector $\mathbf{y} \in X^\circ$. Assume, for the sake of contradiction, that $\langle \mathbf{y}, \mathbf{x} \rangle > 0$ for some $\mathbf{x} \in X$. Then, we have

$$\langle \mathbf{y}, \frac{2}{\langle \mathbf{y}, \mathbf{x} \rangle} \mathbf{x} \rangle = 2 > 1,$$

which is incompatible with $\mathbf{y} \in X^\circ$, thereby establishing the desired inclusion. □

Remark Similarly, it can be shown that $\operatorname{cone}(X)^\circ = X^\bullet$ for any set $X \subseteq \mathbb{R}^d$.

✍ **Exercise 4.2** Let $X \subseteq \mathbb{R}^d$ be a linear subspace, prove that $X^\circ = X^\perp$.

Proof [Proof sketch] Similar to the proof of Exercise 4.1. $\forall \mathbf{y} \in X^\circ$, 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 4.3** Let $X \subseteq \mathbb{R}^d$ be any set, prove that $(X^\bullet)^\bullet = \operatorname{cl}(\operatorname{cone}(X))$.

Proof We first establish that $X^\bullet$ is a convex cone. Given arbitrary vectors $\mathbf{a}, \mathbf{b} \in X^\bullet$ and scalars $\lambda, \gamma \geq 0$, for every vector $\mathbf{x} \in X$, we have

$$\langle \lambda \mathbf{a} + \gamma \mathbf{b}, \mathbf{x} \rangle = \lambda \langle \mathbf{a}, \mathbf{x} \rangle + \gamma \langle \mathbf{b}, \mathbf{x} \rangle \leq 0.$$

This implies $\lambda \mathbf{a} + \gamma \mathbf{b} \in X^\bullet$, establishing that $X^\bullet$ is a convex cone.
Consequently, we can deduce the following:

$$\begin{aligned}
(X^\bullet)^\bullet &= (X^\bullet)^\circ && (4.1)\\
&= (\operatorname{cone}(X)^\circ)^\circ && (4.2)\\
&= \operatorname{cl}(\operatorname{conv}(\operatorname{cone}(X) \cup \{0\})) && (4.3)\\
&= \operatorname{cl}(\operatorname{cone}(X)),
\end{aligned}$$

where Equation (4.1) is a result of Exercise 4.1, Equation (4.2) follows from Remark 4.2, and Equation (4.3) is corroborated by Proposition 2.4.9 (2) in the Basu 2023. □

✍ **Exercise 4.4** 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 $\mathbf{w} \in \operatorname{rec}(P)$. This proves $\{\mathbf{x} \in \mathbb{R}^d : A\mathbf{x} \leq 0\} \subseteq \operatorname{rec}(P)$.
To show reverse inclusion, $\forall \mathbf{y} \notin \{\mathbf{x} \in \mathbb{R}^d : A\mathbf{x} \leq 0\}$, there exists $i \in \{1, \ldots, m\}$ 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 $\beta \leq \mathbf{b}_i$.

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 4.5 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 $\mathrm{cone}(\{\mathbf{a}^1, \ldots, \mathbf{a}^m\}) = \mathbb{R}^d$, 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 Basu 2023 yields that

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

so we just need to show

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

Let $X = \{\mathbf{a}^1, \ldots, \mathbf{a}^m\}$, by Exercise 4.3, we have

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

where the second equality follows from Proposition 2.2.15 in Basu 2023, 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 4.4 shows $\mathrm{rec}(P) = \{\mathbf{x} \in \mathbb{R}^d : A\mathbf{x} \le 0\} = \{\mathbf{x} \in \mathbb{R}^d : \langle \mathbf{a}^i, \mathbf{x}\rangle \le 0,\ i = 1, \ldots, m\}$, this completes the proof.

□

Exercise 4.6 Let $C \subseteq \mathbb{R}^d$ be a convex set and let $X \subseteq C$ be nonempty and convex. Let F be any face of C such that $\mathrm{relint}(X) \cap F \neq \varnothing$, then $X \subseteq F$.

Proof Consider any $\mathbf{x} \in \operatorname{relint}(X) \cap F$ and let $\mathbf{y} \in X$. Note that $2\mathbf{x} - \mathbf{y} \in \operatorname{aff}(X)$ follows directly from the fact that both $\mathbf{x}$ and $\mathbf{y}$ are elements of X, and $2 + (-1) = 1$. Given that $\mathbf{x} \in \operatorname{relint}(X)$, we can find some $\varepsilon > 0$ such that $\mathbf{z} = \mathbf{x} + \varepsilon((2\mathbf{x} - \mathbf{y}) - \mathbf{x})$ is also an element of X. Hence, we can express $\mathbf{x}$ as

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

confirming that $\mathbf{y} \in F$ since F is a face of X. This establishes that $X \subseteq F$. □

Corollary 4.1 (Face of a closed convex set is closed)

Let $C \subseteq \mathbb{R}^d$ be a nonempty, closed, convex set. Let F be any face of C, then F is closed. ♡

Proof If $F = \varnothing$, then F is closed, so we suppose $F \neq \varnothing$ now. Consider $X = C \cap \operatorname{cl} F = \operatorname{cl} F$, then Exercise 4.6 implies that $\operatorname{cl} F \subseteq F$, which proves that F is closed. □

Corollary 4.2 (Proper face is a subset of relative boundary)

Let $C \subseteq \mathbb{R}^d$ be a nonempty, closed, convex set. Let F be a proper face of C, then $F \subseteq \operatorname{relbd}(C)$. ♡

Proof Suppose $\operatorname{relint}(C) \cap F \neq \varnothing$, then by Exercise 4.6 we have $F \subseteq C$, which contradicts the fact that F is a proper face of C. □

✍ **Exercise 4.7** Show that if D is a closed convex cone, then any face of D is a closed convex cone.

Proof Let $F \subseteq D$ be a face. 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. Now take any $\lambda \in (0, 1]$, then

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

by convexity, and Corollary 4.1 yields that F is closed, which completes the proof. □

Recitation 5 September 29, 2023

5.1 Polyhedra and linear programming

In this section, we consider the linear programming problem

$$\begin{array}{ll} \text{minimize} & \mathbf{c}^\mathsf{T}\mathbf{x} \\ \text{subject to} & A\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq 0, \end{array} \tag{5.1}$$

for some $A \in \mathbb{R}^{m\times d}$, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{c} \in \mathbb{R}^d$, and we say $P := \{\mathbf{x} \in \mathbb{R}^d : A\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq 0\}$ is a polyhedron in standard form.

✍ **Exercise 5.1** 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 $f(\mathbf{x}) = \langle \mathbf{c}, \mathbf{x}\rangle$ for some $\mathbf{c} \in \mathbb{R}^d$. Show that there exists an extreme point $\mathbf{v} \in C$ such that $f(\mathbf{v}) \leq f(\mathbf{x})$ for every $\mathbf{x} \in C$.

Proof By definition, f is a continuous function. Applying the Weierstrass theorem, we conclude that there exists a point $\mathbf{u} \in C$ such that $f(\mathbf{u}) \leq f(\mathbf{x})$ for all $\mathbf{x} \in C$. Invoking the Krein-Milman theorem due to the compactness and convexity of C, we can express C as the convex hull of its extreme points, i.e., $C = \text{conv}(\text{ext}(C))$. Therefore, there exist $\mathbf{v}^i \in \text{ext}(C)$ and $\lambda_i \geq 0$ (with $\sum_{i=1}^k \lambda_i = 1$) such that $\mathbf{u} = \sum_{i=1}^k \lambda_i \mathbf{v}^i$. It follows that

$$\langle \mathbf{c}, \mathbf{u}\rangle = \sum_{i=1}^k \lambda_i \langle \mathbf{c}, \mathbf{v}^i\rangle \geq \langle \mathbf{c}, \mathbf{u}\rangle.$$

Thus, the inequality holds with equality, implying that $\langle \mathbf{c}, \mathbf{v}^i\rangle = \langle \mathbf{c}, \mathbf{u}\rangle$ for each $\mathbf{v}^i$. □

✍ **Exercise 5.2** 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 $\mathbf{r} \in \text{lin}(C)\setminus\{0\}$, 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 5.1). 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 (Exercise: why?). □

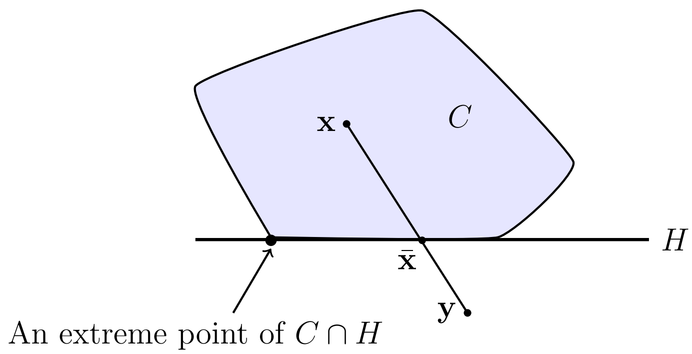


Figure 5.1: Illustration of the proof in Exercise 5.2

Corollary 5.1

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 (which is direct since $P \subseteq \mathbb{R}^d_+$, but as an example to translate a polyhedron in standard form to a regular form), 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,

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

□

✍ **Exercise 5.3** If the linear programming problem 5.1 has a minimizer, then some extreme point of $P = \{\mathbf{x} \in \mathbb{R}^d : A\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq 0\}$ is a minimizer.

Proof Let $\mathbf{x}^*$ be a minimizer of $\mathbf{c}^\mathsf{T}\mathbf{x}$ over $P = \{\mathbf{x} \in \mathbb{R}^d : A\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq 0\}$. Define

$Q = \{\mathbf{x} \in \mathbb{R}^d : A\mathbf{x} = \mathbf{b}, \mathbf{c}^\mathsf{T}\mathbf{x} = \mathbf{c}^\mathsf{T}\mathbf{x}^*,\ \mathbf{x} \geq 0\} = \left\{\mathbf{x} \in \mathbb{R}^d : \begin{bmatrix} A \\ \mathbf{c}^\mathsf{T} \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{b} \\ \mathbf{c}^\mathsf{T}\mathbf{x}^* \end{bmatrix},\ \mathbf{x} \geq 0\right\}$, which is a polyhedron in standard form as well. Therefore, by Corollary 5.1, Q must have an extreme point $\bar{\mathbf{x}}$.

We claim that $\bar{\mathbf{x}}$ is an extreme point of P as well. Suppose there exists $\mathbf{y}, \mathbf{z} \in P \backslash \{\bar{\mathbf{x}}\}$ such that $\frac{1}{2}(\mathbf{y} + \mathbf{z}) = \bar{\mathbf{x}}$. If both of them are in $Q \backslash \{\bar{\mathbf{x}}\}$, then we are done since $\bar{\mathbf{x}}$ is an extreme point of Q, so this case is impossible. If at least one of them is in $P \backslash Q$, say $\mathbf{y}$, then there is a contradiction since $\left\langle \mathbf{c}, \frac{1}{2}(\mathbf{y} + \mathbf{z}) \right\rangle = \frac{1}{2}\left(\langle \mathbf{c}, \mathbf{y} \rangle + \langle \mathbf{c}, \mathbf{z} \rangle\right) < \langle \mathbf{c}, \mathbf{x}^* \rangle$ but $\langle \mathbf{c}, \bar{\mathbf{x}} \rangle = \langle \mathbf{c}, \mathbf{x}^* \rangle$. Thus, $\bar{\mathbf{x}}$ is an extreme point of P and satisfies $\mathbf{c}^\mathsf{T}\bar{\mathbf{x}} = \mathbf{c}^\mathsf{T}\mathbf{x}^*$, which completes the proof. □

Exercise 5.4 Let $X \subseteq \mathbb{R}^d$ be a finite set, and let $\mathbf{c} \in \mathbb{R}^d$. Show that $\max\{\mathbf{c}^\mathsf{T}\mathbf{x} : \mathbf{x} \in X\} = \max\{\mathbf{c}^\mathsf{T}\mathbf{x} : \mathbf{x} \in \operatorname{conv}(X)\}$.

Proof To prove this statement, we first note that $X \subseteq \operatorname{conv}(X)$, which directly implies that $\max\{\mathbf{c}^\mathsf{T}\mathbf{x} : \mathbf{x} \in X\} \leq \max\{\mathbf{c}^\mathsf{T}\mathbf{x} : \mathbf{x} \in \operatorname{conv}(X)\}$. Thus, it is sufficient to show $\max\{\mathbf{c}^\mathsf{T}\mathbf{x} : \mathbf{x} \in X\} \geq \max\{\mathbf{c}^\mathsf{T}\mathbf{x} : \mathbf{x} \in \operatorname{conv}(X)\}$.

Given that X is a finite set, we deduce $\operatorname{conv}(X)$ is a polytope based on the Minkowski-Weyl theorem. Consequently, $\operatorname{conv}(X)$ is compact. Then the Weierstrass theorem yields that there exists a vector $\mathbf{x}_* \in \operatorname{conv}(X)$ such that

$$\mathbf{c}^\mathsf{T}\mathbf{x}_* = \max\{\mathbf{c}^\mathsf{T}\mathbf{x} : \mathbf{x} \in \operatorname{conv}(X)\}.$$

Since $\mathbf{x}_* \in \operatorname{conv}(X)$, it can be expressed as a convex combination of vectors $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^t \in X$, with convex coefficients $\lambda_1, \lambda_2, \ldots, \lambda_t \geq 0$ that sum to 1. Therefore, we have

$$\begin{aligned} \max\{\mathbf{c}^\mathsf{T}\mathbf{x} : \mathbf{x} \in \operatorname{conv}(X)\} &= \mathbf{c}^\mathsf{T}\mathbf{x}_* \\ &= \mathbf{c}^\mathsf{T}\left(\sum_{j=1}^t \lambda_j \mathbf{x}^j\right) \\ &= \sum_{j=1}^t \lambda_j \mathbf{c}^\mathsf{T}\mathbf{x}^j \\ &\leq \sum_{j=1}^t \lambda_j \max\{\mathbf{c}^\mathsf{T}\mathbf{x} : \mathbf{x} \in X\} \\ &= \max\{\mathbf{c}^\mathsf{T}\mathbf{x} : \mathbf{x} \in X\}. \end{aligned}$$

□

5.2 Exercises

Exercise 5.5 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 $A\bar{\mathbf{x}} < \mathbf{b}$, 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 $\dim(P) = d$. □

Exercise 5.6 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, $P_1 = \operatorname{conv}(\{\mathbf{u}^1, \ldots, \mathbf{u}^m\}), P_2 = \operatorname{conv}(\{\mathbf{v}^1, \ldots, \mathbf{v}^n\})$ for some $\mathbf{u}^i, \mathbf{v}^j \in \mathbb{R}^d$. To prove $P_1 + P_2$ is a polytope, we prove that

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

$\forall \mathbf{x} \in P_1 + P_2$, $\mathbf{x} = \mathbf{u} + \mathbf{v}$ for some $\mathbf{u} \in P_1, \mathbf{v} \in P_2$. Then by convexity, there exist some $\alpha_i, \beta_j \geq 0$ 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 $\{\mathbf{u}^i + \mathbf{v}^j\}_{(i,j)\in\{1,\ldots,m\}\times\{1,\ldots,n\}}$, so $\mathbf{x} \in \operatorname{conv}\left(\{\mathbf{u}^i + \mathbf{v}^j\}_{(i,j)\in\{1,\ldots,m\}\times\{1,\ldots,n\}}\right)$. To show the reverse inclusion, $\forall \mathbf{x} \in \operatorname{conv}\left(\{\mathbf{u}^i + \mathbf{v}^j\}_{(i,j)\in\{1,\ldots,m\}\times\{1,\ldots,n\}}\right)$, 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 $\alpha_i = \sum_{j=1}^{n}\lambda_{ij} \geq 0,\ i = 1, \ldots, m$ and $\beta_j = \sum_{j=1}^{m}\lambda_{ij} \geq 0,\ j = 1, \ldots, n$ 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. □

Recitation 6 October 6, 2023

6.1 Theorem of alternatives

Recall Definition 4.1, we give the following 'alternative' type result.

Theorem 6.1 (Generalized Farkas' Lemma)

Let $A \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$, S be a closed cone in $\mathbb{R}^n$. If $\text{cone}_S(A) := \{A\mathbf{x} \mid \mathbf{x} \in S\}$ is closed, then exactly one of the following two statements is true:

1. There exists a vector $\mathbf{x} \in \mathbb{R}^n$ such that $A\mathbf{x} = \mathbf{b}$ and $\mathbf{x} \in S$.
2. There exists a vector $\mathbf{y} \in \mathbb{R}^m$ such that $A^\mathsf{T}\mathbf{y} \in S^*$ and $\mathbf{b}^\mathsf{T}\mathbf{y} < 0$.

♡

Proof Let $\mathbf{b} \in \mathbb{R}^m$, then either $\mathbf{b} \in \text{cone}_S(A)$ or $\mathbf{b} \notin \text{cone}_S(A)$. For the first case, by definition, we have $\mathbf{b} = A\mathbf{x}$ for some $\mathbf{x} \in S$. Thus, $\mathbf{x}$ is the desired vector, and the proof for this case is complete.

For the second case, $\mathbf{b} \notin \text{cone}_S(A)$. Using separating hyperplane theorem, there exists a vector $\mathbf{y} \in \mathbb{R}^m$ and a scalar $\delta \in \mathbb{R}$ such that $\langle \mathbf{y}, \mathbf{b}\rangle < \delta$ and $\langle \mathbf{y}, A\mathbf{s}\rangle = \langle A^\mathsf{T}\mathbf{y}, \mathbf{s}\rangle \geq \delta$ for all $\mathbf{s} \in S$. Note that $\mathbf{0} \in S$, which implies $0 = \langle \mathbf{y}, \mathbf{0}\rangle \geq \delta > \langle \mathbf{y}, \mathbf{b}\rangle$.

To complete the proof, assume for the sake of contradiction that there exists a vector $\tilde{\mathbf{s}}$ such that $\langle A^\mathsf{T}\alpha, \tilde{\mathbf{s}}\rangle < 0$. Let $\lambda = -\frac{|\delta|+1}{\langle A^\mathsf{T}\alpha, \tilde{\mathbf{s}}\rangle} > 0$, then this leads to $\langle A^\mathsf{T}\alpha, \lambda\tilde{\mathbf{s}}\rangle < \delta$, contradicting the inequality $\langle A^\mathsf{T}\alpha, \mathbf{s}\rangle \geq \delta$ for all $\mathbf{s} \in S$.

Therefore, we must have $\langle A^\mathsf{T}\alpha, \mathbf{s}\rangle \geq 0$ for all $\mathbf{s} \in S$, which further implies $A^\mathsf{T}\alpha \in S^*$. □

Remark In convex optimization, various kinds of constraint qualification, e.g. Slater's condition, are responsible for closedness of $\text{cone}_S(A)$.

Let $S = \mathbb{R}^n_+$ in Theorem 6.1, we obtain the Farkas' 'Lemma'.

Theorem 6.2 (Farkas' Lemma)

Let $A \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Exactly one of the following is true:

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

♡

Let $S = \mathbb{R}^n$ and $S^* = \{0\}$ in Theorem 6.1, we have the following corollary.

Corollary 6.1 (Theorem of the alternative)

Let $A \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Exactly one of the following is true:

1. There exists a vector $\mathbf{x} \in \mathbb{R}^m$ such that $A\mathbf{x} = \mathbf{b}$.
2. There exists a vector $\mathbf{y} \in \mathbb{R}^m$ such that $A^\mathsf{T}\mathbf{y} = \mathbf{0}$ and $\mathbf{b}^\mathsf{T}\mathbf{y} \neq 0$.

♡

The following theorem is a consequence of Theorem 6.2. The underlying idea is straightforward: if the set $\{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} \le \mathbf{b}\}$ is non-empty, then no multipliers $\mathbf{u} \ge 0$ exist that yield the contradictory inequality $0 = \mathbf{u}^\mathsf{T} A\mathbf{x} \le \mathbf{u}^\mathsf{T}\mathbf{b} < 0$.

Theorem 6.3 (Farkas' Lemma (Inequality Version))

Let $A \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Exactly one of the following must be true:

1. There exists a vector $\mathbf{x}$ that satisfies $A\mathbf{x} \le \mathbf{b}$.
2. There exists a vector $\mathbf{u} \ge \mathbf{0}$ that satisfies $\mathbf{u}^\mathsf{T} A = \mathbf{0}$ and $\mathbf{u}^\mathsf{T}\mathbf{b} < 0$.

♡

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

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

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

6.2 Application to Polyhedra Complexes

Definition 6.1

A *polyhedral complex* $\mathcal{P}$ is a collection of polyhedra having the following properties:

(A) For every $P, P' \in \mathcal{P}, P \cap P'$ is a common face of P and P'.

(B) every face of a polyhedron in $\mathcal{P}$ belongs to $\mathcal{P}$.

♣

Exercise 6.1 Let $\mathcal{P}$ be a finite polyhedral complex in $\mathbb{R}^n$ with distinct full dimensional polyhedra $\{P_1, \dots, P_m\}$, where $m \in \mathbb{N}_+$. If the union of all polyhedra in $\mathcal{P}$ equals $\mathbb{R}^n$, then the following statements are all true.

1. $\bigcup_{i=1}^m P_i = \mathbb{R}^n$.

2. For any k dimensional polyhedron $F \in \mathcal{P}$ with $0 \leq k \leq n$, there exist $n-k+1$ distinct full-dimensional polyhedra in the complex whose common intersection equals F.
3. $m \leq |\mathcal{P}| < \left(\frac{em}{n+1}\right)^{n+1}$, where $e \approx 2.71828$ is Euler's number.

Proof

Suppose $\bigcup_{i=1}^m P_i \neq \mathbb{R}^n$, and consider $\mathbf{x} \in \mathbb{R}^n \setminus (\bigcup_{i=1}^m P_i)$, then there exists some $\varepsilon > 0$ such that $\mathcal{B}(\mathbf{x}, \varepsilon) \subseteq \mathbb{R}^n \setminus (\bigcup_{i=1}^m P_i)$ since $\bigcup_{i=1}^m P_i$ is closed as it is a finite union of polyhedra. This leads to a contradiction since $\mathcal{P}$ covers $\mathbb{R}^n$ but a finite union of polyhedra with dimension at most $n-1$ cannot cover $\mathcal{B}(\mathbf{x}, \varepsilon)$. This proves part 1.

For part 2., we first observe that $F \in \mathcal{P}$ if and only if F is a face of some full-dimensional polyhedra in $\mathcal{P}$. One direction follows from the definition of a polyhedral complex. For the other direction, consider any $F \in \mathcal{P}$. Using part 1.,

$$F = \mathbb{R}^n \cap F = \left(\bigcup_{i=1}^m P_i\right) \cap F = \bigcup_{i=1}^m (P_i \cap F)\,.$$

By definition of a polyhedral complex, $P_i \cap F$ is a face of F, $\forall i \in [m]$. The above equality thus implies that F is a finite union of some faces of F. This implies that one of these faces must be F itself, i.e., there exists $i \in [m]$ such that $P_i \cap F = F$. Also, by definition $F = P_i \cap F$ is a face of P_i, which proves that F is a face of some full-dimensional polyhedra in $\mathcal{P}$.

Next consider any k dimensional polyhedron $F \in \mathcal{P}$. By the argument above, there exists $i \in [m]$ such that F is a face of P_i. When $k = n$, the result is trivial with $F = P_i$. We now show the result for $k = n-1$. Let $\langle \mathbf{a}, \mathbf{x} \rangle \leq b$ be a facet defining inequality for P_i corresponding to F. Let $\mathbf{x}_0$ be a point in the relative interior of F. Consider the sequence $\mathbf{x}_0 + \frac{1}{n}\mathbf{a}$ and observe that no point in this sequence is contained in P_i. Since this is an infinite sequence, there must exist $j \in [m]$ with $j \neq i$ such that P_j contains infinitely many points from this sequence. Taking limits over this subsequence and using the fact that P_j is closed, we obtain that $\mathbf{x}_0 \in P_j$. Thus, $\mathbf{x}_0 \in P_i \cap P_j$ and $P_i \cap P_j$ is a common face of P_i and P_j. However, since $\mathbf{x}_0$ is in the relative interior of the facet F, this common face must be F. Thus we are done for the case $k = n-1$. For any $k \leq n-2$, the face F must be the intersection of $n-k$ distinct facets of P_i. By the argument above, each of these $n-k$ facets is given by the intersection of P_i with another full-dimensional polyhedron in the complex. Since these are distinct facets, the corresponding full-dimensional polyhedra must be distinct. Including P_i, the intersection of these $n-k+1$ polyhedra equals the intersection of these $n-k$ facets of P_i, which is precisely F. This finishes the proof of part 2.

The first inequality of 3. follows from the fact that $P_1, \ldots, P_m \in \mathcal{P}$. From 2., every polyhedron in the complex of dimension k must be the intersection of $n-k+1$ distinct full-dimensional polyhedra.

Therefore, $\binom{m}{n-k+1}$ gives an upper bound for the number of all the k dimensional polyhedra in $\mathcal{P}$. Now we can give an upper bound for $|\mathcal{P}|$:

$$|\mathcal{P}| \leq \sum_{k=0}^{n} \binom{m}{n-k+1} < \left(\frac{em}{n+1}\right)^{n+1},$$

where the second inequality comes from using Stirling's approximation. □

Part II

Structure of Convex Functions

Recitation 7   October 27, 2023

7.1 Strongly convex functions

Definition 7.1 (Strictly convex function)

A function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is called *strictly convex* if for all $\mathbf{x} \neq \mathbf{y} \in \mathbb{R}^d$ and all $\lambda \in (0,1)$,

$$f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) < \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}).$$

♣

Definition 7.2 (Strongly convex function)

A function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is called *c-strongly convex* if $\mathbf{x} \mapsto f(\mathbf{x}) - \frac{1}{2}c \|\mathbf{x}\|_2^2$ is convex. ♣

Proposition 7.1 (Equivalent definition of strongly convex function)

A function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is c-strongly convex if and only if

$$f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \le \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}) - \frac{1}{2}c\lambda(1-\lambda)\|\mathbf{x}-\mathbf{y}\|_2^2$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ and $\lambda \in (0,1)$.

♠

Proof

$$\begin{aligned}
& f \text{ is } c\text{-strongly convex} \\
\iff & \mathbf{x} \mapsto f(\mathbf{x}) - \frac{1}{2}c\|\mathbf{x}\|_2^2 \text{ is convex} \\
\iff & f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) - \frac{1}{2}c\|\lambda \mathbf{x} + (1-\lambda)\mathbf{y}\|_2^2 \\
& \le \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}) - \frac{1}{2}c\lambda\|\mathbf{x}\|_2^2 - \frac{1}{2}c(1-\lambda)\|\mathbf{y}\|_2^2,\ \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d,\ \lambda \in (0,1) \\
\iff & f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) - \frac{1}{2}c\lambda^2\|\mathbf{x}\|_2^2 - c\langle \lambda \mathbf{x}, (1-\lambda)\mathbf{y}\rangle - \frac{1}{2}c(1-\lambda)^2\|\mathbf{y}\|_2^2 \\
& \le \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}) - \frac{1}{2}c\lambda\|\mathbf{x}\|_2^2 - \frac{1}{2}c(1-\lambda)\|\mathbf{y}\|_2^2,\ \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d,\ \lambda \in (0,1) \\
\iff & f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \le \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}) - \frac{1}{2}c\lambda(1-\lambda)\|\mathbf{x}-\mathbf{y}\|_2^2,\ \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, \lambda \in (0,1).
\end{aligned}$$

□

Proposition 7.2

Let $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be differentiable over an open set $X \subseteq \mathrm{dom}(f)$, and let $C \subseteq X$ be convex. Then the following are all equivalent.

- f is c-strongly convex over C.
- $f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y}-\mathbf{x} \rangle + \frac{1}{2}c \|\mathbf{y}-\mathbf{x}\|_2^2$ for all $\mathbf{x}, \mathbf{y} \in C$.
- $\langle \nabla f(\mathbf{y}) - \nabla f(\mathbf{x}), \mathbf{y}-\mathbf{x} \rangle \geq c \|\mathbf{y}-\mathbf{x}\|_2^2$ for all $\mathbf{x}, \mathbf{y} \in C$.

Proof To see this, just notice that $g(\mathbf{x}) := f(\mathbf{x}) - \frac{1}{2}c\|\mathbf{x}\|_2^2$ is a convex function, and $\nabla g(\mathbf{x}) = \nabla f(\mathbf{x}) - c\mathbf{x}$. Then, by Theorem 3.2.8 for convex functions in Basu 2023, we have

$$\begin{aligned} f(\mathbf{y}) - \frac{1}{2}c\|\mathbf{y}\|_2^2 &\geq f(\mathbf{x}) - \frac{1}{2}c\|\mathbf{x}\|_2^2 + \langle \nabla f(\mathbf{x}) - c\mathbf{x}, \mathbf{y}-\mathbf{x} \rangle, \quad \forall \mathbf{x}, \mathbf{y} \in C, \\ \iff f(\mathbf{y}) &\geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y}-\mathbf{x} \rangle + \frac{1}{2}c\|\mathbf{y}\|_2^2 - \frac{1}{2}c\|\mathbf{x}\|_2^2 - c\langle \mathbf{x}, \mathbf{y}-\mathbf{x} \rangle \quad \forall \mathbf{x}, \mathbf{y} \in C, \\ \iff f(\mathbf{y}) &\geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y}-\mathbf{x} \rangle + \frac{1}{2}c\|\mathbf{y}-\mathbf{x}\|_2^2, \quad \forall \mathbf{x}, \mathbf{y} \in C, \end{aligned}$$

and

$$\begin{aligned} &\langle (\nabla f(\mathbf{y}) - c\mathbf{y}) - (\nabla f(\mathbf{x}) - c\mathbf{x}), \mathbf{y}-\mathbf{x} \rangle \geq 0, \quad \forall \mathbf{x}, \mathbf{y} \in C, \\ \iff &\langle \nabla f(\mathbf{y}) - \nabla f(\mathbf{x}), \mathbf{y}-\mathbf{x} \rangle \geq c\|\mathbf{y}-\mathbf{x}\|_2^2, \quad \forall \mathbf{x}, \mathbf{y} \in C. \end{aligned}$$

□

Proposition 7.3

Let $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be a twice differentiable function over an open set $X \subseteq \operatorname{dom}(f)$ and let $C \subseteq X$ be convex, then f is c-strongly convex over C if and only if $\nabla^2 f(\mathbf{x}) - c\mathbf{I}_d$ is positive semidefinite for all $\mathbf{x} \in C$.

Proof By definition $f(\mathbf{x}) - \frac{1}{2}c\|\mathbf{x}\|_2^2$ is convex, and notice that $\nabla^2 \left(f(\mathbf{x}) - \frac{1}{2}c\|\mathbf{x}\|_2^2 \right) = \nabla^2 f(\mathbf{x}) - c\mathbf{I}$. Then part 1 in Theorem 3.2.11 in Basu 2023 gives the result. □

Exercise 7.1 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

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

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

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

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \mathbf{s}^\mathsf{T}(\mathbf{x} - \mathbf{x}^*).$$

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

$$\begin{aligned} f(\mathbf{x}^*) + \frac{c}{2}\left\|\mathbf{x}^* - \bar{\mathbf{x}}\right\|_2^2 &\le f(\mathbf{x}) + \frac{c}{2}\left\|\mathbf{x} - \bar{\mathbf{x}}\right\|_2^2 \\ &= f(\mathbf{x}) + \frac{c}{2}\left\|\mathbf{x} - \mathbf{x}^* + \mathbf{x}^* - \bar{\mathbf{x}}\right\|_2^2 \\ &= f(\mathbf{x}) + \underbrace{c(\mathbf{x}^* - \bar{\mathbf{x}})^\mathsf{T}}_{\mathbf{s}^\mathsf{T}}(\mathbf{x} - \mathbf{x}^*) + \frac{c}{2}\left\|\mathbf{x} - \mathbf{x}^*\right\|_2^2 + \frac{c}{2}\left\|\mathbf{x}^* - \bar{\mathbf{x}}\right\|_2^2 \end{aligned}$$

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

$$f(\mathbf{x}) \ge f(\mathbf{x}^*) + \mathbf{s}^\mathsf{T}(\mathbf{x} - \mathbf{x}^*) - \frac{c}{2}\left\|\mathbf{x} - \mathbf{x}^*\right\|_2^2, \quad \forall \mathbf{x} \in \mathbb{R}^d. \tag{7.1}$$

For any $\lambda \in (0, 1]$, applying (7.1) at $\lambda\mathbf{x} + (1-\lambda)\mathbf{x}^*$, one can obtain that

$$\begin{aligned} f(\lambda\mathbf{x} + (1-\lambda)\mathbf{x}^*) &\ge f(\mathbf{x}^*) + \mathbf{s}^\mathsf{T}(\lambda\mathbf{x} + (1-\lambda)\mathbf{x}^* - \mathbf{x}^*) - \frac{c}{2}\left\|\lambda\mathbf{x} + (1-\lambda)\mathbf{x}^* - \mathbf{x}^*\right\|_2^2 \\ &= f(\mathbf{x}^*) + \mathbf{s}^\mathsf{T}(\lambda\mathbf{x} - \lambda\mathbf{x}^*) - \frac{c}{2}\left\|\lambda\mathbf{x} - \lambda\mathbf{x}^*\right\|_2^2 \\ &= f(\mathbf{x}^*) + \lambda\mathbf{s}^\mathsf{T}(\mathbf{x} - \mathbf{x}^*) - \lambda^2\frac{c}{2}\left\|\mathbf{x} - \mathbf{x}^*\right\|_2^2. \end{aligned}$$

Using convexity of f, we can strengthen this since

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

therefore,

$$\lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{x}^*) \ge f(\mathbf{x}^*) + \lambda\mathbf{s}^\mathsf{T}(\mathbf{x} - \mathbf{x}^*) - \lambda^2\frac{c}{2}\left\|\mathbf{x} - \mathbf{x}^*\right\|_2^2,$$

that is,

$$f(\mathbf{x}) \ge f(\mathbf{x}^*) + \mathbf{s}^\mathsf{T}(\mathbf{x} - \mathbf{x}^*) - \lambda\frac{c}{2}\left\|\mathbf{x} - \mathbf{x}^*\right\|_2^2.$$

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

Recitation 8 November 3, 2023

8.1 Review

Example 8.1 A convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ may not be closed.

Proof Consider a convex function $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ given by

$$f(x) = \begin{cases} +\infty & \text{if } x < 0 \\ 1 & \text{if } x = 0 \\ x^2 & \text{if } x > 0 \end{cases},$$

then $\text{epi}(f)$ is not closed. □

The example also implies the following:

Example 8.2 A convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ may not be continuous on a compact set $C \subseteq \mathbb{R}^n$.

Below is an if and only if condition for closedness of a function's epigraph.

Definition 8.1 (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. ♣

Proposition 8.1

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. $\text{epi}(f) = \{(\mathbf{x}, t) \in \mathbb{R}^d \times \mathbb{R} : f(\mathbf{x}) \le t\}$ 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 $\text{epi}(f)$ is nonempty and there exist level sets of f that are nonempty.
(i) $\Longrightarrow$ (ii). 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 $f(\mathbf{x}_{k_i}) \leq \bar{\gamma}$ 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 $f(\bar{\mathbf{x}}) \leq \bar{\gamma}$, which leads to a contradiction.
(ii) $\implies$ (iii). Assume that f is lower semicontinuous over $\mathbb{R}^d$, and let $\bar{\mathbf{x}}, \bar{t}$ be the limit of a sequence

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

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

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

Hence, $(\bar{\mathbf{x}}, \bar{t}) \in \text{epi}(f)$ and $\text{epi}(f)$ is closed.
(iii) $\implies$ (i). Assume that $\text{epi}(f)$ 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 $(\mathbf{x}_k, \gamma) \in \text{epi}(f)$ for all k and $(\mathbf{x}_k, \gamma) \to (\bar{\mathbf{x}}, \gamma)$, so since $\text{epi}(f)$ is closed, we have $(\bar{\mathbf{x}}, \gamma) \in \text{epi}(f)$. Hence, $\bar{\mathbf{x}}$ belongs to f_γ, implying that this set is closed. □

Example 8.3 A strictly convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ may not be strongly convex.

Proof Consider $f(x) = e^x$, then for any $c > 0$ there exists $\bar{x} = \log(c/2)$ such that $\nabla^2 \left(f(\cdot) - \frac{c}{2}\|\cdot\|_2^2\right)(\bar{x}) = e^{\bar{x}} - c = -\frac{1}{2}c < 0$. Therefore, $f(x) - \frac{c}{2}x^2$ is not convex. □

Example 8.4 A strictly convex twice differentiable function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ may not have its Hessian positive definite everywhere.

Proof Consider $f(x) = x^4$. f is strictly convex, but $\nabla^2 f(0) = 0$. □

Example 8.5 A strictly convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ may not always have a unique global minimizer.

Proof Consider $f(x) = e^x$. □

However, for strongly convex functions, indeed there is always a unique global minimizer (based on the result from Exercise 7.1). Moreover, we have the following:

✍ **Exercise 8.1** For any c-strongly convex f with global minimizer $\mathbf{x}^*$, show that $\forall \mathbf{x} \in \mathbb{R}^d$ has

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

Proof Let $h(\mathbf{x}) := f(\mathbf{x}) - \frac{c}{2}\|\mathbf{x}\|_2^2$, which is convex by assumption. Then $\mathbf{x}^*$ minimizes $f(\mathbf{x}) = h(\mathbf{x}) + \frac{c}{2}\|\mathbf{x}-0\|_2^2$. This is precisely the shape of the problem considered in Exercise 7.1. Letting $c = c$, $\bar{\mathbf{x}} = 0$, we then know for any $\mathbf{x} \in \mathbb{R}^d$,

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

Therefore,

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

□

Remark Exercise 7.1, Exercise 8.1, and the existence of the global minimizer did not require the strongly convex function f to be differentiable.

8.2 Exercises

✍ **Exercise 8.2 Theorem 3.3.14 in Basu 2023**

Let $N : \mathbb{R}^d \to \mathbb{R}$ be a norm. Then $B_N(\mathbf{0}, 1) = \{\mathbf{x} \in \mathbb{R}^d : N(\mathbf{x}) \leq 1\}$ is a $\mathbf{0}$-symmetric, compact convex set with $\mathbf{0}$ in its interior. Moreover, $\gamma_{B_N(\mathbf{0},1)} = N$.

Conversely, let B be a $\mathbf{0}$-symmetric, compact convex set containing $\mathbf{0}$ in its interior. Then γ_B is a norm on $\mathbb{R}^d$ and $B = B_{\gamma_B}(\mathbf{0}, 1)$.

Proof For the first part, since N is sublinear, it is convex (Proposition 3.3.2). By definition, $B_N(\mathbf{0}, 1) = \{\mathbf{x} \in \mathbb{R}^d : N(\mathbf{x}) \leq 1\}$ is a sublevel set for N, and is thus a convex set (Proposition 3.1.10). It is closed, since N is continuous by Theorem 3.2.3. Since $N(\mathbf{x}) = N(-\mathbf{x})$, this also shows that $B_N(\mathbf{0}, 1)$ is $\mathbf{0}$-symmetric. We now show that $\operatorname{rec}(B_N(\mathbf{0}, 1)) = \{\mathbf{0}\}$; this will imply that it is compact by Theorem 2.4.22. Consider any nonzero vector $\mathbf{r}$, and let $N(\mathbf{r}) = M > 0$. Then, $\frac{2}{M}\mathbf{r} = \mathbf{0} + \frac{2}{M}\mathbf{r}$, but $N\left(\frac{2}{M}\mathbf{r}\right) = 2$. Thus, $\frac{2}{M}\mathbf{r} \notin B_N(\mathbf{0}, 1)$, and so $\mathbf{r}$ cannot be a recession direction for $B_N(\mathbf{0}, 1)$.
We next verify that $\mathbf{0} \in \operatorname{int}(B_N(\mathbf{0}, 1))$. If not, then by the Supporting Hyperplane Theorem 2.4.5, there exists $\mathbf{a} \in \mathbb{R}^d \backslash \{\mathbf{0}\}$ and $\delta \in \mathbb{R}$ such that $B_N(\mathbf{0}, 1) \subseteq H \leq (\mathbf{a}, \delta)$ and $\langle \mathbf{a}, \mathbf{0} \rangle = \delta$. Thus,

$\delta = 0$. Now, since $\mathbf{a} \neq \mathbf{0}$, $N(\mathbf{a}) > 0$. Thus, $N\left(\frac{\mathbf{a}}{N(\mathbf{a})}\right) = 1$ and by definition, $\frac{\mathbf{a}}{N(\mathbf{a})} \in B_N(\mathbf{0}, 1)$. However, $\left\langle \mathbf{a}, \frac{\mathbf{a}}{N(\mathbf{a})} \right\rangle = \frac{\|\mathbf{a}\|^2}{N(\mathbf{a})} > 0$ which contradicts the fact that $B_N(\mathbf{0}, 1) \subseteq H \leq (\mathbf{a}, 0)$. Finally, from Corollary 3.3.13, we obtain that $N = \gamma_{B_N(\mathbf{0},1)}$ since N is a nonnegative, sublinear function taking value 0 at the origin.

For the second part, we know that γ_B is nonnegative and sublinear, and since B is compact, $\gamma_B(\mathbf{r}) > 0$ for all $\mathbf{r} \neq \mathbf{0}$ by Corollary 3.3.12. Since $\mathbf{0} \in \mathrm{int}(B)$, Exercise 2 from Section 3.3.5 below implies that γ_C is finite valued everywhere. To confirm that γ_B is a norm, all that remains to be checked is that $\gamma_B(\mathbf{x}) = \gamma_B(-\mathbf{x})$ for all $\mathbf{x} \neq \mathbf{0}$. Suppose to the contrary that $\gamma_B(\mathbf{x}) > \gamma_B(-\mathbf{x})$ (note that this is without loss of generality). This implies that $\gamma_B\left(\frac{1}{\gamma_B(-\mathbf{x}))}\mathbf{x}\right) > 1$. Therefore, $\frac{1}{\gamma_B(-\mathbf{x}))}\mathbf{x} \notin B$ by Theorem 3.3.11, part 2 . However, $\gamma_B\left(-\frac{1}{\gamma_B(-\mathbf{x}))}\mathbf{x}\right) = \frac{1}{\gamma_B(-\mathbf{x})}\gamma_B(-\mathbf{x}) = 1$ showing that $-\frac{1}{\gamma_B(-\mathbf{x}))}\mathbf{x} \in B$ by Theorem 3.3.11, part 2. This contradicts the fact that B is $\mathbf{0}$-symmetric. Thus, γ_B is a norm on $\mathbb{R}^d$. Moreover, by Theorem 3.3.11, part 2., $B = \left\{\mathbf{x} \in \mathbb{R}^d : \gamma_B(\mathbf{x}) \leq 1\right\} = B_{\gamma_B}(\mathbf{0}, 1)$.

□

Part III

Convex Optimization

Recitation 9 Subdifferential Calculus

Introduction

- *Support function calculus*
- *Subdifferential calculus*
- *Chain rule of directional derivatives*
- *Fenchel duality*

9.1 Calculus of support functions

Proposition 9.1 (Calculus of support functions)

The following are all true.

1. Let $A, B \subseteq \mathbb{R}^d$ be closed, convex sets. Show that $A \subseteq B$ if and only if $\sigma_A \leq \sigma_B$.
2. Let $A, B \subseteq \mathbb{R}^d$ be closed, convex sets, and $\lambda_1, \lambda_2 \geq 0$. Then $\sigma_{\lambda_1 A + \lambda_2 B} = \lambda_1 \sigma_A + \lambda_2 \sigma_B$.
3. Let $C_i, i \in I$ be a family of closed, convex sets, and let $C = \operatorname{cl}(\operatorname{conv}(\cup_{i \in I} C_i))$. Then $\sigma_C = \sup_{i \in I} \sigma_{C_i}$.
4. Let $T : \mathbb{R}^d \to \mathbb{R}^m$ be a linear transformation, and let $T^* : \mathbb{R}^m \to \mathbb{R}^d$ be its adjoint transformation, i.e., for all $\mathbf{x} \in \mathbb{R}^d$ and $\mathbf{y} \in \mathbb{R}^m$, we have $\langle \mathbf{y}, T\mathbf{x} \rangle = \langle T^* \mathbf{y}, \mathbf{x} \rangle$ (in matrix language, T^* is represented by the transpose of the matrix representing T). Show that for any set $S, \sigma_{T(S)}(\mathbf{r}) = \sigma_S(T^* \mathbf{r})$ for all $\mathbf{r} \in \mathbb{R}^m$.

Proof

1. For the 'if' direction, we consider any $\mathbf{r} \in \mathbb{R}^d$, then $\sigma_A(\mathbf{r}) = \sup_{\mathbf{x} \in A} \langle \mathbf{x}, \mathbf{r} \rangle \leq \sup_{\mathbf{x} \in B} \langle \mathbf{x}, \mathbf{r} \rangle = \sigma_B(\mathbf{r})$ since $A \subseteq B$. For the reverse direction, just notice that

$$\begin{aligned} A = C_{\sigma_A} &= \{\mathbf{x} \in \mathbb{R}^d : \langle \mathbf{x}, \mathbf{r} \rangle \leq \sigma_A(\mathbf{r}), \forall \mathbf{r} \in \mathbb{R}^d\} \\ &\subseteq \{\mathbf{x} \in \mathbb{R}^d : \langle \mathbf{x}, \mathbf{r} \rangle \leq \sigma_B(\mathbf{r}), \forall \mathbf{r} \in \mathbb{R}^d\} \\ &= C_{\sigma_B} = B. \end{aligned}$$

2. For any $\mathbf{r} \in \mathbb{R}^d$, we have

$$\begin{aligned}
\sigma_{\lambda_1 A+\lambda_2 B}(\mathbf{r}) &= \sup_{\mathbf{x}\in\lambda_1 A+\lambda_2 B} \langle \mathbf{x}, \mathbf{r}\rangle \\
&= \sup_{\mathbf{x}_1\in A,\ \mathbf{x}_2\in B} \langle \lambda_1\mathbf{x}_1 + \lambda_2\mathbf{x}_2, \mathbf{r}\rangle \\
&= \sup_{\mathbf{x}_1\in A,\ \mathbf{x}_2\in B} \left(\lambda_1\langle \mathbf{x}_1, \mathbf{r}\rangle + \lambda_2\langle \mathbf{x}_2, \mathbf{r}\rangle\right) \\
&= \lambda_1 \sup_{\mathbf{x}_1\in A} \langle \mathbf{x}_1, \mathbf{r}\rangle + \lambda_2 \sup_{\mathbf{x}_2\in B} \langle \mathbf{x}_2, \mathbf{r}\rangle \\
&= \lambda_1\sigma_A(\mathbf{r}) + \lambda_2\sigma_B(\mathbf{r}).
\end{aligned}$$

3. For any $\mathbf{r} \in \mathbb{R}^d$, we have

$$\sigma_C(\mathbf{r}) = \sigma_{\mathrm{cl}(\mathrm{conv}(\cup_{i\in I} C_i))}(\mathbf{r}) = \sigma_{\cup_{i\in I} C_i}(\mathbf{r}) = \sup_{\mathbf{x}\in\cup_{i\in I} C_i} \langle \mathbf{x}, \mathbf{r}\rangle = \sup_{i\in I}\sup_{\mathbf{x}\in C_i} \langle \mathbf{x}, \mathbf{r}\rangle = \sup_{i\in I} \sigma_{C_i}(\mathbf{r}).$$

To see second last equality, notice that $\forall \mathbf{x} \in \cup_{i\in I} C_i$, $\mathbf{x} \in C_i$ for some $i \in I$, then we have $\langle \mathbf{x}, \mathbf{r}\rangle \le \sup_{\mathbf{x}\in C_i}\langle \mathbf{x}, \mathbf{r}\rangle \le \sup_{i\in I}\sup_{\mathbf{x}\in C_i}\langle \mathbf{x}, \mathbf{r}\rangle$. Thus, $\sup_{\mathbf{x}\in\cup_{i\in I} C_i}\langle \mathbf{x}, \mathbf{r}\rangle \le \sup_{i\in I}\sup_{\mathbf{x}\in C_i}\langle \mathbf{x}, \mathbf{r}\rangle$. On the other hand, $\forall i \in I$, $\sup_{\mathbf{x}\in C_i}\langle \mathbf{x}, \mathbf{r}\rangle \le \sup_{\mathbf{x}\in\cup_{i\in I} C_i}\langle \mathbf{x}, \mathbf{r}\rangle$, which implies $\sup_{i\in I}\sup_{\mathbf{x}\in C_i}\langle \mathbf{x}, \mathbf{r}\rangle \le \sup_{\mathbf{x}\in\cup_{i\in I} C_i}\langle \mathbf{x}, \mathbf{r}\rangle$. Therefore, $\sup_{\mathbf{x}\in\cup_{i\in I} C_i}\langle \mathbf{x}, \mathbf{r}\rangle = \sup_{i\in I}\sup_{\mathbf{x}\in C_i}\langle \mathbf{x}, \mathbf{r}\rangle$.

4. For any $\mathbf{r} \in \mathbb{R}^m$, we have

$$\sigma_{T(S)}(\mathbf{r}) = \sup_{\mathbf{x}\in T(S)} \langle \mathbf{r}, \mathbf{x}\rangle = \sup_{\mathbf{z}\in S}\langle \mathbf{r}, T\mathbf{z}\rangle = \sup_{\mathbf{z}\in S}\langle T^*\mathbf{r}, \mathbf{z}\rangle = \sigma_S(T^*\mathbf{r}).$$

□

9.1.1 Subdifferential calculus

Theorem 9.1 (Subdifferential calculus)

The following are all true.

1. Let $f_1, f_2 : \mathbb{R}^d \to \mathbb{R}$ be convex functions and let $t_1, t_2 \ge 0$. Then
$$\partial\left(t_1 f_1 + t_2 f_2\right)(\mathbf{x}) = t_1\partial f_1(\mathbf{x}) + t_2\partial f_2(\mathbf{x}) \text{ for all } \mathbf{x} \in \mathbb{R}^d.$$
2. Let $A \in \mathbb{R}^{m\times d}$ and $\mathbf{b} \in \mathbb{R}^m$ and let $T(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$ be the corresponding affine map from $\mathbb{R}^d \to \mathbb{R}^m$ and let $g : \mathbb{R}^m \to \mathbb{R}$ be a convex function. Then
$$\partial(g \circ T)(\mathbf{x}) = A^{\mathsf{T}}\partial g(A\mathbf{x} + \mathbf{b}) \text{ for all } \mathbf{x} \in \mathbb{R}^d.$$
3. Let $f_j : \mathbb{R}^d \to \mathbb{R}, j \in J$ be convex functions for some (possibly infinite) index set J, and let $f = \sup_{j\in J} f_j$. Then
$$\mathrm{cl}\left(\mathrm{conv}\left(\cup_{j\in J(\mathbf{x})}\partial f_j(\mathbf{x})\right)\right) \subseteq \partial f(\mathbf{x}),$$
where $J(\mathbf{x})$ is the set of indices j such that $f_j(\mathbf{x}) = f(\mathbf{x})$.

♡

Proof

1. f_1, f_2 are convex functions with finite values everywhere, so f_1, f_2 are continuous, then by Proposition 8.1 they are closed convex with $\mathrm{dom}(f_1) = \mathrm{dom}(f_2) = \mathbb{R}^d$. Moreover, Exercise 14 from 3.1.1 yields that both sides are closed convex sets. Therefore, by part 1 from Proposition 9.1, we have

$$\begin{aligned}
&\partial(t_1 f_1 + t_2 f_2)(\mathbf{x}) = t_1 \partial f_1(\mathbf{x}) + t_2 \partial f_2(\mathbf{x}), && \forall \mathbf{x} \in \mathbb{R}^d, \\
\overset{9.1,1}{\iff}\ &\sigma_{\partial(t_1 f_1 + t_2 f_2)(\mathbf{x})}(\mathbf{r}) = \sigma_{(t_1 \partial f_1(\mathbf{x}) + t_2 \partial f_2(\mathbf{x}))}(\mathbf{r}), && \forall \mathbf{x} \in \mathbb{R}^d, \forall \mathbf{r} \in \mathbb{R}^d, \\
\overset{9.1,2}{\iff}\ &\sigma_{\partial(t_1 f_1 + t_2 f_2)(\mathbf{x})}(\mathbf{r}) = t_1 \sigma_{\partial f_1(\mathbf{x})}(\mathbf{r}) + t_2 \sigma_{\partial f_2(\mathbf{x})}(\mathbf{r}), && \forall \mathbf{x} \in \mathbb{R}^d, \forall \mathbf{r} \in \mathbb{R}^d, \\
\overset{\text{Thm } 3.4.3}{\iff}\ &(t_1 f_1 + t_2 f_2)'(\mathbf{x}; \mathbf{r}) = t_1 f_1'(\mathbf{x}; \mathbf{r}) + t_2 f_2'(\mathbf{x}; \mathbf{r}), && \forall \mathbf{x} \in \mathbb{R}^d, \forall \mathbf{r} \in \mathbb{R}^d.
\end{aligned}$$

Then we only need to prove the last line. By definition of directional derivative, for any $\mathbf{x}, \mathbf{r} \in \mathbb{R}^d$ and $t_1, t_2 > 0$, we have

$$\begin{aligned}
(t_1 f_1 + t_2 f_2)'(\mathbf{x}; \mathbf{r}) &= \lim_{t \to 0^+} \frac{(t_1 f_1 + t_2 f_2)(\mathbf{x} + t\mathbf{r}) - (t_1 f_1 + t_2 f_2)(\mathbf{x})}{t} \\
&= \lim_{t \to 0^+} \frac{t_1 f_1(\mathbf{x} + t\mathbf{r}) + t_2 f_2(\mathbf{x} + t\mathbf{r}) - (t_1 f_1(\mathbf{x}) + t_2 f_2(\mathbf{x}))}{t} \\
&= \lim_{t \to 0^+} \frac{t_1 f_1(\mathbf{x} + t\mathbf{r}) - t_1 f_1(\mathbf{x})}{t} + \lim_{t \to 0^+} \frac{t_2 f_2(\mathbf{x} + t\mathbf{r}) - t_2 f_2(\mathbf{x})}{t} \\
&= t_1 f_1'(\mathbf{x}; \mathbf{r}) + t_2 f_2'(\mathbf{x}; \mathbf{r}).
\end{aligned}$$

2. Similarly, observe that

$$\begin{aligned}
&\partial(g \circ T)(\mathbf{x}) = A^\mathsf{T} \partial g(A\mathbf{x} + \mathbf{b}), && \forall \mathbf{x} \in \mathbb{R}^d, \\
\overset{9.1,1}{\iff}\ &\sigma_{\partial(g \circ T)(\mathbf{x})}(\mathbf{r}) = \sigma_{A^\mathsf{T} \partial g(A\mathbf{x} + \mathbf{b})}(\mathbf{r}), && \forall \mathbf{x} \in \mathbb{R}^d, \forall \mathbf{r} \in \mathbb{R}^d, \\
\overset{9.1,4}{\iff}\ &\sigma_{\partial(g \circ T)(\mathbf{x})}(\mathbf{r}) = \sigma_{\partial g(A\mathbf{x} + \mathbf{b})}(A\mathbf{r}), && \forall \mathbf{x} \in \mathbb{R}^d, \forall \mathbf{r} \in \mathbb{R}^d, \\
\overset{\text{Thm } 3.4.3}{\iff}\ &(g \circ T)'(\mathbf{x}; \mathbf{r}) = g'(A\mathbf{x} + \mathbf{b}; A\mathbf{r}), && \forall \mathbf{x} \in \mathbb{R}^d, \forall \mathbf{r} \in \mathbb{R}^d.
\end{aligned}$$

Then we only need to verify that the last equality is true. By definition of directional derivative, for any $\mathbf{x}, \mathbf{r} \in \mathbb{R}^d$, we have

$$\begin{aligned}
(g \circ T)'(\mathbf{x}; \mathbf{r}) &= \lim_{t \to 0^+} \frac{(g \circ T)(\mathbf{x} + t\mathbf{r}) - (g \circ T)(\mathbf{x})}{t} \\
&= \lim_{t \to 0^+} \frac{g(A(\mathbf{x} + t\mathbf{r}) + \mathbf{b}) - g(A\mathbf{x} + \mathbf{b})}{t} \\
&= \lim_{t \to 0^+} \frac{g(A\mathbf{x} + \mathbf{b} + t(A\mathbf{r})) - g(A\mathbf{x} + \mathbf{b})}{t} \\
&= g'(A\mathbf{x} + \mathbf{b}; A\mathbf{r}).
\end{aligned}$$

3. Similarly, observe that

$$\begin{aligned}
&\mathrm{cl}(\mathrm{conv}(\cup_{j\in J(\mathbf{x})}\partial f_j(\mathbf{x}))) \subseteq \partial f(\mathbf{x}), && \forall \mathbf{x}\in\mathbb{R}^d,\\
\overset{9.1.1}{\iff}\ & \sigma_{\mathrm{cl}(\mathrm{conv}(\cup_{j\in J(\mathbf{x})}\partial f_j(\mathbf{x})))}(\mathbf{r}) \le \sigma_{\partial f(\mathbf{x})}(\mathbf{r}), && \forall \mathbf{x}\in\mathbb{R}^d, \forall \mathbf{r}\in\mathbb{R}^d,\\
\overset{9.1.3}{\iff}\ & \sup_{j\in J(\mathbf{x})} \sigma_{\partial f_j(\mathbf{x})}(\mathbf{r}) \le \sigma_{\partial f(\mathbf{x})}(\mathbf{r}), && \forall \mathbf{x}\in\mathbb{R}^d, \forall \mathbf{r}\in\mathbb{R}^d,\\
\iff\ & \sigma_{\partial f_j(\mathbf{x})}(\mathbf{r}) \le \sigma_{\partial f(\mathbf{x})}(\mathbf{r}), && \forall \mathbf{x}\in\mathbb{R}^d, \forall j\in J(\mathbf{x}), \forall \mathbf{r}\in\mathbb{R}^d,\\
\overset{9.1.1}{\iff}\ & \partial f_j(\mathbf{x}) \subseteq \partial f(\mathbf{x}), && \forall \mathbf{x}\in\mathbb{R}^d, \forall j\in J(\mathbf{x}).
\end{aligned}$$

Then we only need to prove that the last inclusion is true. For any $\mathbf{x}\in\mathbb{R}^d$ and $j\in J(\mathbf{x})$, we consider any $\mathbf{s}\in\partial f_j(\mathbf{x})$, then by definition of subgradient, we have

$$\begin{aligned}
& f_j(\mathbf{y}) \ge f_j(\mathbf{x}) + \langle \mathbf{s}, \mathbf{y}-\mathbf{x}\rangle, && \forall \mathbf{y}\in\mathbb{R}^d,\\
\implies & f(\mathbf{y}) \ge f_j(\mathbf{y}) \ge f(\mathbf{x}) + \langle \mathbf{s}, \mathbf{y}-\mathbf{x}\rangle, && \forall \mathbf{y}\in\mathbb{R}^d,\\
\implies & \mathbf{s}\in\partial f(\mathbf{x}).
\end{aligned}$$

□

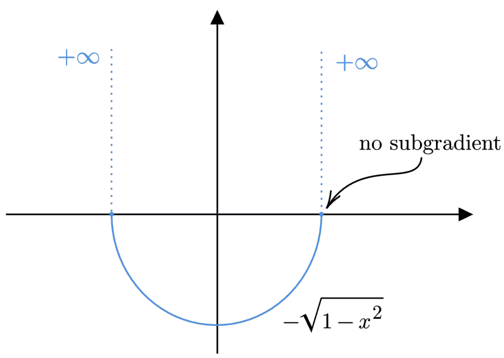


Figure 9.1: On the boundary of dom f, there may not be any $\phi\in\partial f(x)$.

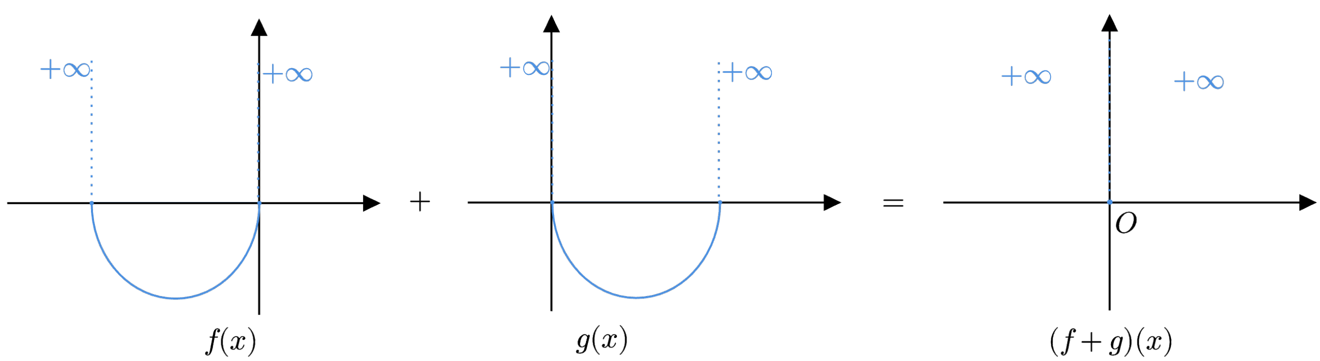


Figure 9.2: No sum rule in general.

9.1.2 Alternative proof using Fenchel duality

Theorem 9.2 (Subdifferential calculus 2)

(a) For any convex functions $f\colon \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ and $g\colon \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, show that the partial calculus rule

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

holds for any $\mathbf{x} \in \operatorname{relint}(\operatorname{dom}(f)) \cap \operatorname{relint}(\operatorname{dom}(g \circ A))$ (Note: In our case $A^* = A^\mathsf{T}$).

(b) For any convex functions $f\colon \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ and $g\colon \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, and linear map $A\colon \mathbb{R}^d \to \mathbb{R}^d$, show that the following perturbed value function is convex for any $\Phi \in \mathbb{R}^d$:

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

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

(a)

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, $\forall \mathbf{y} \in \operatorname{dom}(g \circ A) \cap \operatorname{dom} f$, 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,

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

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

(b)

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 $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^d$ such that

$$f(\mathbf{y}_1) + g(A\mathbf{y}_1 + \mathbf{u}_1) - \langle \Phi, \mathbf{y}_1\rangle < h(\mathbf{u}_1) + \varepsilon \tag{9.1}$$
$$f(\mathbf{y}_2) + g(A\mathbf{y}_2 + \mathbf{u}_2) - \langle \Phi, \mathbf{y}_2\rangle < h(\mathbf{u}_2) + \varepsilon \tag{9.2}$$

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

$$\begin{aligned}
&h(\lambda \mathbf{u}_1 + (1-\lambda)\mathbf{u}_2) \\
=& \inf_{\mathbf{y}\in\mathbb{R}^d} \{f(\mathbf{y}) + g(A\mathbf{y} + \lambda\mathbf{u}_1 + (1-\lambda)\mathbf{u}_2) - \langle \Phi, \mathbf{y}\rangle\} \\
\le& f(\mathbf{y}_0) + g(A\mathbf{y}_0 + \lambda \mathbf{u}_1 + (1-\lambda)\mathbf{u}_2) - \langle \Phi, \mathbf{y}_0\rangle \\
=& f(\lambda\mathbf{y}_1 + (1-\lambda)\mathbf{y}_2) + g(\lambda(A\mathbf{y}_1 + \mathbf{u}_1) + (1-\lambda)(A\mathbf{y}_2 + \mathbf{u}_2)) - \langle \Phi, \lambda\mathbf{y}_1 + (1-\lambda)\mathbf{y}_2\rangle \\
\le& \lambda f(\mathbf{y}_1) + (1-\lambda) f(\mathbf{y}_2) + \lambda g(A\mathbf{y}_1 + \mathbf{u}_1) + (1-\lambda) g(A\mathbf{y}_2 + \mathbf{u}_2) - \langle \Phi, \lambda\mathbf{y}_1 + (1-\lambda)\mathbf{y}_2\rangle \\
\overset{(9.1),(9.2)}{<}& \lambda h(\mathbf{u}_1) + \lambda\varepsilon + (1-\lambda)h(\mathbf{u}_2) + (1-\lambda)\varepsilon \\
=& \lambda h(\mathbf{u}_1) + (1-\lambda)h(\mathbf{u}_2) + \varepsilon
\end{aligned}$$

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. □

(c)

Proof Consider any $\mathbf{x} \in \mathrm{dom}(g\circ A) \cap \mathrm{dom}(f)$. $\forall \Phi \in \partial(f + g\circ A)(\mathbf{x})$, by definition of subgradient we have $0 \in \partial(f + g\circ A - \langle \Phi, \cdot\rangle)(\mathbf{x})$, that is, $\mathbf{x}$ minimizes $f(\mathbf{y}) + g(A\mathbf{y}) - \langle \Phi, \mathbf{y}\rangle$. The above argument (b) yields that $h(\mathbf{u})$ is convex. Also, $\mathbf{0} \in \mathrm{relint}(\mathrm{dom}(h))$ since $\mathbf{0} \in \mathrm{relint}(\mathrm{dom}(g)) - A\,\mathrm{relint}(\mathrm{dom}(f))$ (Verify!), so $-\Psi \in \partial h(0)$ exists. Then by definition of subgradient,

$$h(0) \le h(\mathbf{u}) + \langle \Psi, \mathbf{u}\rangle. \tag{9.3}$$

Recall that $h(\mathbf{u}) = \inf_{\mathbf{y}}\{f(\mathbf{y}) + g(A\mathbf{y}+\mathbf{u}) - \langle\Phi,\mathbf{y}\rangle\}$ 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,

$$\begin{aligned}
\underbrace{f(\mathbf{x}) + g(A\mathbf{x}) - \langle \Phi, \mathbf{x}\rangle}_{h(0)} &\overset{(9.3)}{\le} \underbrace{\inf_{\mathbf{y}}\{f(\mathbf{y}) + g(A\mathbf{y}+\mathbf{u}) - \langle \Phi, \mathbf{y}\rangle\}}_{h(\mathbf{u})} + \langle \Psi, \mathbf{u}\rangle \\
&\le f(\mathbf{y}) + g(A\mathbf{y}+\mathbf{u}) - \langle \Phi, \mathbf{y}\rangle + \langle \Psi, \mathbf{u}\rangle
\end{aligned} \tag{9.4}$$

Take $\mathbf{y} = \mathbf{x}$ in 9.4, we have

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

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 9.4, we have

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

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. □

Recitation 10 Normal Cones

Introduction

- *Normal cone examples*
- *Normal cone calculus*

10.1 Feasible cone and normal cone

Definition 10.1 (Cone of feasible directions and tangent cone)

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

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

and the *tanget cone of* C *at* $\mathbf{x}$ as $T_C(\mathbf{x}) = \mathrm{cl}(F_C(\mathbf{x}))$. ♣

Exercise 10.1 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 $F_C(\mathbf{x})$ may not be closed: consider $C = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\| \leq 1\}$, and let $\mathbf{x} = (-1, 0)$. Then $F_C(\mathbf{x}) = \{\mathbf{r} \in \mathbb{R}^2 : \mathbf{r}_1 > 0\} \cup \{\mathbf{0}\}$.

Definition 10.2 (Normal cone)

Let $C \subseteq \mathbb{R}^d$ be a convex set, and let $\mathbf{x} \in C$. The *normal cone* of C at $\mathbf{x}$ is

$$N_C(\mathbf{x}) = \{\mathbf{r} \in \mathbb{R}^d : \langle \mathbf{r}, \mathbf{y} - \mathbf{x} \rangle \leq 0,\ \forall \mathbf{y} \in C\}.$$

♣

Proposition 10.1

Let $C \subseteq \mathbb{R}^d$ be a convex set, and let $\mathbf{x} \in C$. Then $N_C(\mathbf{x}) = T_C(\mathbf{x})^\circ$, i.e., the tangent cone and the normal cone are polars of each other. ♠

Proof Direct by definition. □

Proposition 10.2

Consider a closed, convex function $g : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ and a point $\bar{\mathbf{x}} \in \operatorname{int}(\operatorname{dom} g)$, then

$$T_{g_{\leq g(\bar{\mathbf{x}})}}(\bar{\mathbf{x}}) = \{\mathbf{r} \in \mathbb{R}^d : g'(\bar{\mathbf{x}}; \mathbf{r}) \leq 0\}.$$

10.2 Normal cone examples

Proposition 10.3

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 $\operatorname{Proj}_C(\mathbf{y}) = \mathbf{x}$ if and only if $\mathbf{y} - \mathbf{x} \in N_C(\mathbf{x})$.
2. If $\mathbf{x} \in \operatorname{int}(C)$, then $N_C(\mathbf{x}) = \{0\}$.
3. If $\mathbf{x} \in \operatorname{relbd}(C)$, then $\{0\} \subsetneq N_C(\mathbf{x})$.

Proof

1. ($\Longrightarrow$): Notice that by Proposition 2.3.1 in Basu 2023,

$$\operatorname{Proj}_C(\mathbf{y}) = \mathbf{x} \implies \langle \mathbf{y} - \mathbf{x}, \mathbf{z} - \mathbf{x} \rangle \leq 0,\ \forall \mathbf{z} \in C, \implies \mathbf{y} - \mathbf{x} \in N_C(\mathbf{x}).$$

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

$$\begin{aligned}
& & \langle \mathbf{y} - \mathbf{x}, \mathbf{z} - \mathbf{x} \rangle \leq 0, & \quad \forall \mathbf{z} \in C, \\
&\implies & \langle \mathbf{y} - \mathbf{x}, \mathbf{y} - \mathbf{x} \rangle + \langle \mathbf{y} - \mathbf{x}, \mathbf{z} - \mathbf{y} \rangle \leq 0, & \quad \forall \mathbf{z} \in C, \\
&\implies & \|\mathbf{y} - \mathbf{x}\|^2 \leq \langle \mathbf{y} - \mathbf{x}, \mathbf{y} - \mathbf{z} \rangle \leq \|\mathbf{y} - \mathbf{x}\| \|\mathbf{y} - \mathbf{z}\|, & \quad \forall \mathbf{z} \in C, \\
&\implies & \|\mathbf{y} - \mathbf{x}\| \leq \|\mathbf{y} - \mathbf{z}\|, & \quad \forall \mathbf{z} \in C, \\
&\implies & \mathbf{x} = \operatorname{Proj}_C(\mathbf{y}). &
\end{aligned}$$

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

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

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

3. By Lemma 2.3.4 in Basu 2023, there exists $\mathbf{y} \in \operatorname{aff}(C) \backslash C$ such that $\operatorname{Proj}_C(\mathbf{y}) = \mathbf{x}$. Then part 1 implies $\mathbf{y} - \mathbf{x} \in N_C(\mathbf{x})$. Therefore, $\{0\} \subsetneq \{0, \mathbf{y} - \mathbf{x}\} \subseteq N_C(\mathbf{x})$. □

10.3 Normal cone calculus

Definition 10.3 (Indicator function)

For any subset $X \subseteq \mathbb{R}^d$, define

$$I_X(\mathbf{x}) := \begin{cases} 0 & \text{if } \mathbf{x} \in X \\ +\infty & \text{if } \mathbf{x} \notin X \end{cases}.$$

It's not hard to see that I_X is convex if and only if X is convex. ♣

Exercise 10.2 For any subset $X, Y \subseteq \mathbb{R}^d$, we have $I_X + I_Y = I_{X\cap Y}$.

Proof Let $\mathbf{x}$ be any point in $\mathbb{R}^d$, then we have

$$I_X(\mathbf{x}) + I_Y(\mathbf{x}) = 0 \iff \mathbf{x} \in X,\ \mathbf{x} \in Y \iff \mathbf{x} \in X \cap Y \iff I_{X\cap Y}(\mathbf{x}) = 0.$$ □

Proposition 10.4

Let $C \subseteq \mathbb{R}^d$ be a convex set, then $\partial I_C(\mathbf{x}) = N_C(\mathbf{x})$ for any $\mathbf{x} \in C$. ♠

Proof Consider any $\mathbf{x} \in C$.

$$\begin{aligned} \mathbf{s} \in \partial I_C(\mathbf{x}) &\iff I_C(\mathbf{y}) \geq I_C(\mathbf{x}) + \langle \mathbf{s}, \mathbf{y} - \mathbf{x}\rangle,\ \forall \mathbf{y} \in \mathbb{R}^d \\ &\iff \begin{cases} 0 \geq 0 + \langle \mathbf{s}, \mathbf{y} - \mathbf{x}\rangle & \forall \mathbf{y} \in C \\ +\infty \geq \langle \mathbf{s}, \mathbf{y} - \mathbf{x}\rangle & \forall \mathbf{y} \notin C \end{cases} \\ &\iff 0 \geq \langle \mathbf{s}, \mathbf{y} - \mathbf{x}\rangle,\ \forall \mathbf{y} \in C \\ &\iff \mathbf{s} \in N_C(\mathbf{x}). \end{aligned}$$ □

Theorem 10.1 (Normal cone sum rule)

Let $C_1, C_2 \subseteq \mathbb{R}^d$ be two convex sets. If the regularity condition $\operatorname{relint}(C_1) \cap \operatorname{relint}(C_2) \neq \varnothing$ holds, then we have

$$N_{C_1\cap C_2}(\mathbf{x}) = N_{C_1}(\mathbf{x}) + N_{C_2}(\mathbf{x})$$

for any $\mathbf{x} \in C_1 \cap C_2$. ♡

Proof

$$N_{C_1\cap C_2}(\mathbf{x}) \overset{10.4}{=} \partial I_{C_1\cap C_2}(\mathbf{x}) \overset{10.2}{=} \partial(I_{C_1} + I_{C_2})(\mathbf{x}) \overset{9.2}{=} \partial I_{C_1}(\mathbf{x}) + \partial I_{C_2}(\mathbf{x}) \overset{10.4}{=} N_{C_1}(\mathbf{x}) + N_{C_2}(\mathbf{x}).$$ □

Theorem 10.2

Given a closed, convex function $g : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ and a point $\bar{\mathbf{x}} \in \operatorname{int}(\operatorname{dom} g)$ such that $g(\bar{\mathbf{x}}) > \inf_{\mathbf{x} \in \mathbb{R}^d} g(\mathbf{x})$, then

$$N_C(\bar{\mathbf{x}}) = \operatorname{cone}(\partial g(\bar{\mathbf{x}})),$$

where $C := \{\mathbf{x} \in \mathbb{R}^d : g(\mathbf{x}) \leq g(\bar{\mathbf{x}})\}$. Moreover, if g is differentiable, then

$$N_C(\bar{\mathbf{x}}) = \operatorname{cone}(\{\nabla g(\bar{\mathbf{x}})\}).$$

♡

Proof The inclusion $\operatorname{cone}(\partial g(\bar{\mathbf{x}})) \subseteq N_C(\bar{\mathbf{x}})$ is straightforward. To establish the equality, assume there exists a direction $\mathbf{r} \in N_C(\bar{\mathbf{x}}) \backslash \operatorname{cone}(\partial g(\bar{\mathbf{x}}))$. The fact that $\bar{\mathbf{x}}$ is not a minimizer of g implies $\mathbf{0} \notin \partial g(\bar{\mathbf{x}})$. Hence, $\operatorname{cone}(\{\mathbf{r}\}) \cap \operatorname{cone}(\partial g(\bar{\mathbf{x}})) = \{\mathbf{0}\}$ and $\operatorname{cone}(\{\mathbf{r}\}) \cap \partial g(\bar{\mathbf{x}}) = \varnothing$. According to Problem 14 from Exercise 3.1.1, $\partial g(\bar{\mathbf{x}})$ is a compact convex set (as $\bar{\mathbf{x}} \in \operatorname{int}(\operatorname{dom} g)$). Therefore, by Problem 1 from Exercise 2.4.4, there exist $\mathbf{a} \in \mathbb{R}^d \backslash \{\mathbf{0}\}, \delta \in \mathbb{R}, c > 0$ such that $\operatorname{cone}(\{\mathbf{r}\}) \subseteq H^{\leq}(\mathbf{a}, \delta)$ and $\partial g(\bar{\mathbf{x}}) \subseteq H^{\geq}(\mathbf{a}, \delta + c)$. Notice that $\operatorname{cone}(\{\mathbf{r}\}) \subseteq H^{\leq}(\mathbf{a}, \delta)$ implies $\delta \geq \langle \mathbf{a}, \mathbf{0} \rangle = 0$, then it's not hard to verify that $\operatorname{cone}(\{\mathbf{r}\}) \subseteq H^{\leq}(\mathbf{a}, 0)$ and $\partial(g(\bar{\mathbf{x}})) \subseteq H^{\geq}(\mathbf{a}, c)$. It follows that $\langle \mathbf{a}, \mathbf{s} \rangle \geq c > 0$ for all $\mathbf{s} \in \partial g(\bar{\mathbf{x}})$, leading to

$$g'(\bar{\mathbf{x}}; -\mathbf{a}) = \sigma_{\partial g(\bar{\mathbf{x}})}(-\mathbf{a}) = \sup_{\mathbf{s} \in \partial g(\bar{\mathbf{x}})} \langle \mathbf{s}, -\mathbf{a} \rangle \leq -c < 0.$$

Consequently, there exists a sufficiently small $\varepsilon > 0$ such that $g(\bar{\mathbf{x}} - \varepsilon \mathbf{a}) < g(\bar{\mathbf{x}})$, which implies $\bar{\mathbf{x}} - \varepsilon \mathbf{a} \in \operatorname{int}(g(\bar{\mathbf{x}}))$. Since $\mathbf{r} \in N_C(\bar{\mathbf{x}})$, it follows that $-\varepsilon \langle \mathbf{r}, \mathbf{a} \rangle = \langle \mathbf{r}, \bar{\mathbf{x}} - \varepsilon \mathbf{a} - \bar{\mathbf{x}} \rangle \leq 0$, implying $\langle \mathbf{r}, \mathbf{a} \rangle \geq 0$. We claim that $\langle \mathbf{r}, \mathbf{a} \rangle \neq 0$, otherwise consider some $\varepsilon' > 0$ small enough such that $g(\bar{\mathbf{x}} - \varepsilon \mathbf{a} + \varepsilon' \mathbf{r}) \leq g(\bar{\mathbf{x}})$, then $\langle \mathbf{r}, \bar{\mathbf{x}} - \varepsilon \mathbf{a} + \varepsilon' \mathbf{r} - \bar{\mathbf{x}} \rangle = \varepsilon' \|\mathbf{r}\|_2^2 > 0$, contradicting with the fact that $\mathbf{r} \in N_C(\bar{\mathbf{x}})$.
However, $\langle \mathbf{r}, \mathbf{a} \rangle > 0$ would contradict with the fact that $\operatorname{cone}(\{\mathbf{r}\}) \subseteq H^{\leq}(\mathbf{a}, 0)$, thereby proving that such a direction $\mathbf{r}$ cannot exist. □

10.4 Exercises

✍ **Exercise 10.3** 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 $\operatorname{tight}(\bar{\mathbf{x}}) = \{i : \langle \mathbf{a}^i, \mathbf{x} \rangle = \mathbf{b}_i\}$. Show that

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

Proof The set $\{\mathbf{r} \in \mathbb{R}^d : \langle \mathbf{a}^i, \mathbf{r} \rangle \leq 0 \text{ for all } i \in \operatorname{tight}(\bar{\mathbf{x}})\} \subseteq F_P(\bar{\mathbf{x}})$, since
Case 1: $i \in \operatorname{tight}(\bar{\mathbf{x}})$. For all $\varepsilon > 0$, $\langle \mathbf{a}^i, \mathbf{x} + \varepsilon \mathbf{r} \rangle = \langle \mathbf{a}^i, \mathbf{x} \rangle + \varepsilon \langle \mathbf{a}^i, \mathbf{r} \rangle \leq \langle \mathbf{a}^i, \mathbf{x} \rangle = \mathbf{b}_i$.

Case 2: $i \notin \text{tight}(\bar{\mathbf{x}})$. $\langle \mathbf{a}^i, \mathbf{x} \rangle < \mathbf{b}_i$, 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 $\mathbf{r} \notin F_P(\bar{\mathbf{x}})$. □

Exercise 10.4 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 $N_P(\bar{\mathbf{x}}) = \{-(\mathbf{s} + A^\mathsf{T}\mathbf{y}) : (\mathbf{s}, \mathbf{y}) \in \mathbb{R}^d \times \mathbb{R}^m, \mathbf{s} \geq 0, \mathbf{s}_i = 0 \text{ for } i \in I(\bar{\mathbf{x}})\}$, where $I(\bar{\mathbf{x}}) = \{i : \bar{\mathbf{x}}_i > 0\}$.
2. Prove that if $-\mathbf{c} \in \text{int}\,(N_P(\bar{\mathbf{x}}))$ 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

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

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 $\mathbf{c} \in \text{int}(N_P(\bar{\mathbf{x}}))$, for small enough $\varepsilon > 0$, $-\mathbf{c} - \varepsilon(\bar{\mathbf{x}} - \bar{\mathbf{x}}') \in N_P(\bar{\mathbf{x}})$. 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

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

□

Exercise 10.5 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 $\mathrm{tr}(AB) = \sum_{i,j=1}^n A_{ij}B_{ij}$. 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

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

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

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

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

Then we prove that $\mathcal{S}^n_+$ is self-dual, i.e., for $A, B \in \mathcal{S}^n$: $\mathrm{tr}(AB) \geq 0,\ \forall A \in \mathcal{S}^n_+ \iff B \in \mathcal{S}^n_+$.

Suppose $B \notin \mathcal{S}^n_+$, then there exists $\mathbf{x} \in \mathbb{R}^n$ with

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

Hence the positive semidefinite matrix $A = \mathbf{x}\mathbf{x}^\mathsf{T}$ satisfies $\mathrm{tr}(AB) < 0$, 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 $A = \sum_{i=1}^n \lambda_i \mathbf{x}_i\mathbf{x}_i^\mathsf{T}$, where the eigenvalues $\lambda_i \geq 0,\ i = 1, \ldots, n$. Then we have

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

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

Dual of a norm cone (Boyd and Vandenberghe 2004):

Exercise 10.6 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 $\langle \mathbf{u}, \mathbf{x} \rangle + vt \geq 0$. □

Second-order conditions for K-convexity (Boyd and Vandenberghe 2004):

✍ **Exercise 10.7** 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 $\mathbf{c} \in K^*$. For any $\mathbf{x} \in \mathbb{R}^n$, the Hessian of $\langle \mathbf{c}, \mathbf{f}(\mathbf{x}) \rangle$ is

$$\nabla^2(\langle \mathbf{c}, \mathbf{f}(\mathbf{x}) \rangle) = \sum_{k=1}^{n} \mathbf{c}_k \nabla^2 \mathbf{f}_k(\mathbf{x}) = \sum_{k=1}^{n} \mathbf{c}_k \left[\frac{\partial^2 \mathbf{f}_k(\mathbf{x})}{\partial \mathbf{x}_i \partial \mathbf{x}_j} \right]_{(i,j) \in [n] \times [n]}.$$

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

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

which by definition of dual cone is equivalent to

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

□

Appendices

Appendix A Frequently Asked Homework Problems

A.1 Homework 1

Problem A.1 Let $X \subseteq \mathbb{R}^d$ be a set. Then X is a linear subspace if and only if X is both a cone and an affine subset.

Proof It is direct to show the forward direction. For the reverse direction, suppose X is both a cone and an affine subset. $\forall \mathbf{x}, \mathbf{y} \in X$, $\lambda, \gamma \in \mathbb{R}$, then

$$\lambda \mathbf{x} + \gamma \mathbf{y} = \lambda \mathbf{x} + \gamma \mathbf{y} + (1 - \lambda - \gamma)\mathbf{0} \in X,$$

where the first equality holds since X is a cone. $\square$

Problem A.2 Let $C \subseteq \mathbb{R}^d$, then $\operatorname{span}(C - C) = \operatorname{aff}(C) - \bar{\mathbf{x}}$ for any $\bar{\mathbf{x}} \in C$.

Proof $\forall \mathbf{z} \in \operatorname{span}(C - C)$, $\mathbf{z} = \gamma_1(\mathbf{x}^1 - \mathbf{y}^1) + \cdots + \gamma_t(\mathbf{x}^t - \mathbf{y}^t)$ for some $\mathbf{x}^1, \ldots, \mathbf{x}^t, \mathbf{y}^1, \ldots, \mathbf{y}^t \in C$ and $\gamma_1, \ldots, \gamma_t \in \mathbb{R}$. Then $\mathbf{z} = \gamma_1 \mathbf{x}^1 + \cdots + \gamma_t \mathbf{x}^t + (-\gamma_1)\mathbf{y}^1 + \cdots + (-\gamma_t)\mathbf{y}^t + \bar{\mathbf{x}} - \bar{\mathbf{x}} \in \operatorname{aff}(C) - \bar{\mathbf{x}}$ where $\bar{\mathbf{x}}$ is any point in C since $\sum_{i=1}^t \gamma_i + \sum_{i=1}^t (-\gamma_i) + 1 = 1$, this proves $\operatorname{span}(C - C) \subseteq \operatorname{aff}(C) - \bar{\mathbf{x}}$. To show the reverse inclusion, $\forall \mathbf{z} \in \operatorname{aff}(C) - \bar{\mathbf{x}}$ for any $\bar{\mathbf{x}} \in C$, by definition $\mathbf{z} = \sum_{i=1}^t \lambda_i \mathbf{x}^i - \bar{\mathbf{x}}$ for some $\mathbf{x}^1, \ldots, \mathbf{x}^t \in C$ and $\lambda_1, \ldots, \lambda_t \in \mathbb{R}$ with $\sum_{i=1}^t \lambda_i = 1$. Then $\mathbf{z} = \sum_{i=1}^t \lambda_i (\mathbf{x}^i - \bar{\mathbf{x}}) \in \operatorname{span}(C - C)$, this proves $\operatorname{aff}(C) - \bar{\mathbf{x}} \subseteq \operatorname{span}(C - C)$. $\square$

Remark Notice that $\operatorname{aff}(C) = \operatorname{aff}(\operatorname{conv}(C))$ since $C \subseteq \operatorname{conv}(C) \subseteq \operatorname{aff}(C)$. Therefore, Problem A.2 implies that $\dim(\operatorname{span}(C - C)) = \dim(\operatorname{aff}(C)) = \dim(\operatorname{conv}(C))$.

Problem A.3 Let $\mathbf{a} \in \mathbb{R}^d$ and $\delta_1 \leq \delta_2$. Show that the distance between the hyperplanes $H^=(\mathbf{a}, \delta_1)$ and $H^=(\mathbf{a}, \delta_2)$ is given by $\frac{\delta_2 - \delta_1}{\|\mathbf{a}\|}$. More precisely, show that

$$\inf\left\{\|\mathbf{x} - \mathbf{y}\| : \mathbf{x} \in H^=(\mathbf{a}, \delta_1), \mathbf{y} \in H^=(\mathbf{a}, \delta_2)\right\} = \frac{\delta_2 - \delta_1}{\|\mathbf{a}\|}$$

Proof $\forall \mathbf{x} \in H^=(\mathbf{a}, \delta_1)$, $\mathbf{y} \in H^=(\mathbf{a}, \delta_2)$, then by B.6 we have

$$\delta_2 - \delta_1 = \langle \mathbf{a}, \mathbf{y} - \mathbf{x} \rangle \leq \|\mathbf{a}\|_2 \|\mathbf{y} - \mathbf{x}\|_2,$$

and the equality holds if and only if $\mathbf{y} - \mathbf{x} = \lambda \mathbf{a}$ for some $\lambda \in \mathbb{R}$. Specifically, one can choose $\mathbf{x} = \frac{\delta_1}{\|a\|_2^2}\mathbf{a}$, $\mathbf{y} = \frac{\delta_2}{\|a\|_2^2}\mathbf{a}$ to achieve the equality. □

A.2 Homework 2

Problem A.4 Let $X \subseteq \mathbb{R}^d$ be a set of $d+1$ affinely independent points. Show that $\mathrm{int}(\mathrm{conv}(X)) \neq \varnothing$.

Proof Let $X = \{\mathbf{x}^1, \ldots, \mathbf{x}^{d+1}\}$ be affinely independent, we can consider proving a point $\bar{\mathbf{x}} = \sum_{i=1}^{d+1} \lambda_i \mathbf{x}^i$ with $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_{d+1}]^\mathsf{T} \in \mathrm{int}(\Delta_{d+1})$ is in the interior of $\mathrm{conv}(X)$. Specifically, one can choose $\bar{\mathbf{x}} = \sum_{i=1}^{d+1} \frac{1}{d+1}\mathbf{x}^i$. To establish that $\bar{\mathbf{x}} \in \mathrm{int}(\mathrm{conv}(X))$, several possible approaches can be employed. One involves invoking Theorem 3.1 to prove that for any direction $\mathbf{r} \in \mathbb{R}^d$, there exists some $\varepsilon_{\mathbf{r}} > 0$ such that $\bar{\mathbf{x}} + \varepsilon_{\mathbf{r}}\mathbf{r} \in \mathrm{conv}(X)$, which is always achievable for sufficiently small $\varepsilon_{\mathbf{r}}$. Alternatively, one can directly construct a ball centered at $\bar{\mathbf{x}}$ that is fully contained in $\mathrm{conv}(X)$. For example, one can consider the optimization problem with a continuous convex objective function over a compact set: $\varepsilon^* := \min\{\|\mathbf{y} - \mathbf{x}\|_2^2 : \mathbf{y} \in \mathrm{bd}(\mathrm{conv}(X))\}$. I claim $\varepsilon^* > 0$ since

$$\bar{\mathbf{x}} \notin \bigcup_{S \subseteq \{\mathbf{x}^1, \ldots, \mathbf{x}^{d+1}\},\ |S| = d} \mathrm{aff}(S) \supseteq \mathrm{bd}(\mathrm{conv}(X)),$$

then $\mathcal{B}\left(\bar{\mathbf{x}}, \frac{\varepsilon^*}{553.665}\right) \subseteq \mathrm{conv}(X)$. □

Problem A.5 Show that $\mathrm{relint}(C)$ is nonempty for any nonempty convex set $C \subseteq \mathbb{R}^d$.

Proof Suppose that $\dim(C) = k \in [0, d] \cap \mathbb{Z}$. Notice that $\mathrm{aff}(C) \cong \mathbb{R}^k$ and use Problem A.4. □

Problem A.6 Let $X, Y \subseteq \mathbb{R}^d$, show that if X is closed convex with $\mathbf{0} \in X$ and Y is a linear subspace, then $(X \cap Y)^\circ = \mathrm{cl}(X^\circ + Y^\perp)$.

Proof It's trivial to see the second set is contained in the first one. To see the reverse inclusion, just notice that

$$(X \cap Y)^\circ = (X^{\circ\circ} \cap Y^{\circ\circ})^\circ = (X^\circ \cup Y^\circ)^{\circ\circ} = \mathrm{cl}(\mathrm{conv}(X^\circ \cup Y^\perp)) \subseteq \mathrm{cl}(X^\circ + Y^\perp).$$

□

A.3 Homework 3

Problem A.7 (Farkas' Type Results in Linear Inequalities)

Let A be a given matrix and $\mathbf{b}$ be a given vector. Then, the following statements hold:

1. There exists $\mathbf{x} \geq 0$ satisfying $A\mathbf{x} \leq \mathbf{b}$ if and only if for each $\mathbf{y} \geq 0, \mathbf{y}^\top A \geq 0 \Rightarrow \mathbf{y}^\top \mathbf{b} \geq 0$.
2. There exists $\mathbf{x} > 0$ satisfying $A\mathbf{x} = \mathbf{0}$ if and only if for each $\mathbf{y} \in \mathbb{R}^m, \mathbf{y}^\top A \geq 0 \Rightarrow \mathbf{y}^\top A = 0$.
3. There exists $\mathbf{x} \neq 0$ satisfying $\mathbf{x} \geq 0$ and $A\mathbf{x} = 0$ if and only if there is no vector $\mathbf{y} \in \mathbb{R}^m$ satisfying $\mathbf{y}^\top A > 0$.

Proof Here we only prove part 3..
To prove part 3., we just need to prove

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

Consider any fixed $\mathbf{b} \in \mathbb{R}^n_{++}$, then

$$\begin{aligned}
\nexists \mathbf{y} \in \mathbb{R}^d \text{ satisfying } \mathbf{y}^\top A > 0 &\iff \nexists \mathbf{y} \in \mathbb{R}^d \text{ such that } \mathbf{y}^\top A \geq \mathbf{b} \\
&\iff A^\top(-\mathbf{y}) \leq -\mathbf{b} \text{ has no solution} \\
&\overset{Thm.6.3}{\iff} \exists \mathbf{x} \geq 0 \text{ such that } \mathbf{x}^\top A^\top = 0 \text{ and } \mathbf{x}^\top(-\mathbf{b}) < 0 \\
&\iff \exists \mathbf{x} \geq 0 \text{ such that } A\mathbf{x} = 0 \text{ and } \mathbf{x}^\top \mathbf{b} > 0 \\
&\iff \exists \mathbf{x} \neq 0,\ \mathbf{x} \geq 0 \text{ such that } A\mathbf{x} = 0.
\end{aligned}$$

□

Appendix B Discrete Inequalities

B.1 Young's Inequality and the AM-GM Inequality

Theorem B.1 (Young's Inequality)

Let $\lambda_1, \ldots, \lambda_m \geq 0$ such that $\lambda_1 + \cdots + \lambda_m = 1$. Then for any real numbers $a_1, \ldots, a_m \geq 0$, we have

$$\lambda_1 a_1 + \cdots + \lambda_m a_m \geq a_1^{\lambda_1} \cdots a_m^{\lambda_m}.$$

♡

Proof The inequality is trivial if there exists any $i \in [m]$ such that $a_i = 0$, so we may assume all the a_i's are positive. Notice that $f(x) = \ln(x)$ is a concave function on $(0, +\infty)$. Therefore,

$$\lambda_1 \ln(a_1) + \cdots + \lambda_m \ln(a_m) \leq \ln(\lambda_1 a_1 + \cdots + \lambda_m a_m),$$

which completes the proof since $\ln(x)$ is monotone increasing on $(0, +\infty)$. □

Theorem B.2 (Weighted AM-GM inequality)

Let $\lambda_1, \ldots, \lambda_m \geq 0$ and $\lambda = \sum_{i=1}^m \lambda_i$. Then for any real numbers $a_1, \ldots, a_m \geq 0$, we have

$$\frac{\lambda_1 a_1 + \cdots + \lambda_m a_m}{\lambda} \geq \sqrt[\lambda]{a_1^{\lambda_1} \cdots a_m^{\lambda_m}}.$$

♡

Proof Notice that $\sum_{i=1}^m \frac{\lambda_i}{\lambda} = 1$, then use Theorem B.1 with λ_i/λ as the weights. □

B.2 Generalized Finite Hölder's Inequality

Theorem B.3 (Generalized Finite Hölder's Inequality)

Let $\lambda_1, \ldots, \lambda_m \geq 0$ such that $\lambda_1 + \cdots + \lambda_m = 1$. Suppose $A = (a_{ij})_{m \times n} > 0$ is an m by n real matrix. Then, we have

$$\prod_{i=1}^m \left(\sum_{j=1}^n a_{ij} \right)^{\lambda_i} \geqslant \sum_{j=1}^n \left(\prod_{i=1}^m a_{ij}^{\lambda_i} \right),$$

or equivalently,

$$(a_{11} + \cdots + a_{1n})^{\lambda_1} (a_{21} + \cdots + a_{2n})^{\lambda_2} \cdots (a_{m1} + \cdots + a_{mn})^{\lambda_m} \geqslant \sum_{j=1}^n \left(a_{1j}^{\lambda_1} a_{2j}^{\lambda_2} \cdots a_{mj}^{\lambda_m} \right). \quad \text{(B.1)}$$

Equality holds if and only if $a_{11} : a_{12} : \cdots : a_{1n} = a_{21} : a_{22} : \cdots : a_{2n} = \cdots = a_{m1} : a_{m2} : \cdots : a_{mn}$, that is, each row of A is parallel to the others. ♡

Proof Let $A_i = \sum_{j=1}^n a_{ij}$ for $i = 1, 2, \ldots, m$ denote the sum of the i-th row of the matrix $A = (a_{ij})_{m\times n}$. By Theorem B.1, we have

$$\lambda_1 \left(\frac{a_{1j}}{A_1}\right) + \lambda_2 \left(\frac{a_{2j}}{A_2}\right) + \cdots \lambda_m \left(\frac{a_{mj}}{A_m}\right) \geq \left(\frac{a_{1j}}{A_1}\right)^{\lambda_1} \left(\frac{a_{2j}}{A_2}\right)^{\lambda_2} \cdots \left(\frac{a_{mj}}{A_m}\right)^{\lambda_m}, \ \forall j \in [n].$$

Summing over j from 1 to n, we obtain

$$1 = \lambda_1 + \lambda_2 + \cdots + \lambda_m \geq \sum_{j=1}^n \left(\frac{a_{1j}}{A_1}\right)^{\lambda_1} \left(\frac{a_{2j}}{A_2}\right)^{\lambda_2} \cdots \left(\frac{a_{mj}}{A_m}\right)^{\lambda_m},$$

which implies

$$A_1^{\lambda_1} A_2^{\lambda_2} \cdots A_m^{\lambda_m} \geq \sum_{j=1}^n a_{1j}^{\lambda_1} a_{2j}^{\lambda_2} \cdots a_{mj}^{\lambda_m},$$

and completes the proof. □

Remark Take $m = 2$, and $a_{1j} = |x_j|^p, \ a_{2j} = |y_j|^p, \ \forall j \in [n]$ for some $p, q > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$ in Theorem B.3, we obtain the Finite Hölder's Inequality:

$$(|x_1|^p + \cdots + |x_n|^p)^{1/p}(|y_1|^q + \cdots + |y_n|^q)^{1/q} \geqslant |x_1y_1| + \cdots + |x_ny_n|, \tag{B.2}$$

or equivalently,

$$|\langle \mathbf{x}, \mathbf{y}\rangle| \leq \|\mathbf{x}\|_p \|\mathbf{y}\|_q \tag{B.3}$$

for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, and equality holds if and only if $|\mathbf{x}|^p$ and $|\mathbf{y}|^q$ are linearly dependent, where $|\mathbf{x}|^p := [|x_1|^p, \ldots, |x_n|^p]^\intercal \in \mathbb{R}^n$.

Remark Let $\lambda_1 = \lambda_2 = \cdots = \lambda_m = \frac{1}{m}$ in Theorem B.3, we obtain the following inequality:

$$\prod_{i=1}^m \left(\sum_{j=1}^n a_{ij}^m\right) \geqslant \left(\sum_{j=1}^n \prod_{i=1}^m a_{ij}\right)^m, \tag{B.4}$$

or equivalently,

$$(a_{11}^m + \cdots + a_{1n}^m) \cdots (a_{m1}^m + \cdots + a_{mn}^m) \geqslant (a_{11} \cdots a_{m1} + \cdots + a_{1n} \cdots a_{mn})^m. \tag{B.5}$$

Example B.1 Show that the following optimization problem has the minimum 5 attained at $a = 1$ and $b = 2$:

$$\begin{aligned} \min \quad & \frac{1}{a} + \frac{8}{b} \\ \text{s.t.} \quad & a^2 + b^2 = 5 \\ & a, b \geq 0. \end{aligned}$$

Proof By (B.5), we have

$$\left(\frac{1}{a}+\frac{8}{b}\right)\left(\frac{1}{a}+\frac{8}{b}\right)\underbrace{(a^2+b^2)}_{5} \geq \left(1+\sqrt[3]{8\times 8}\right)^3 = 5^3,$$

hence

$$\frac{1}{a}+\frac{8}{b} \geq \sqrt{\frac{5^3}{5}} = 5.$$

The equality holds when $[\frac{1}{a}, a^2]$ and $[\frac{8}{b}, b^2]$ are linearly dependent, which implies that $a = 1$ and $b = 2$. □

B.3 Cauchy-Schwarz Inequality

Take $p = q = 2$ in (B.3), we obtain the Cauchy-Schwarz Inequality:

$$|\langle \mathbf{x}, \mathbf{y}\rangle| \leq \|\mathbf{x}\|_2 \|\mathbf{y}\|_2, \tag{B.6}$$

for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, and equality holds if and only if $\mathbf{x}$ and $\mathbf{y}$ are linearly dependent.

Alternatively, the Cauchy-Schwarz inequality can be derived by noting that $\varphi(t) := \langle \mathbf{x} + t\mathbf{y}, \mathbf{x} + t\mathbf{y}\rangle \geq 0$ for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $t \in \mathbb{R}$, or noticing that $\left\|\mathbf{x} - \langle \mathbf{x}, \frac{\mathbf{y}}{\|\mathbf{y}\|_2}\rangle \frac{\mathbf{y}}{\|\mathbf{y}\|_2}\right\|_2 \geq 0$.

B.4 Minkowski's Inequality

Minkowski's Inequality establishes that $\|\cdot\|_p$ is a norm in $\mathbb{R}^n$ for $p \geq 1$, and can be derived by Finite Hölder's Inequality (B.3).

Theorem B.4 (Minkowski's Inequality)

Let $p \in (1, +\infty)$. Then, for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we have

$$\|\mathbf{x}+\mathbf{y}\|_p \leq \|\mathbf{x}\|_p + \|\mathbf{y}\|_p, \tag{B.7}$$

the equality holds if and only if $\mathbf{x} = c\mathbf{y}$ for some $c \geq 0$ or one of the vectors is $\mathbf{0}$. ♡

Proof

$$\begin{aligned}
\|\mathbf{x}+\mathbf{y}\|_p^p &= \sum_{i=1}^n |x_i+y_i|^p \\
&= \sum_{i=1}^n |x_i+y_i||x_i+y_i|^{p-1} \\
&\le \sum_{i=1}^n |x_i||x_i+y_i|^{p-1} + \sum_{i=1}^n |y_i||x_i+y_i|^{p-1} \\
&\overset{(B.3)}{\le} \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} \left(\sum_{i=1}^n |x_i+y_i|^{(p-1)q}\right)^{1/q} + \left(\sum_{i=1}^n |y_i|^p\right)^{1/p} \left(\sum_{i=1}^n |x_i+y_i|^{(p-1)q}\right)^{1/q} \\
&= \left(\|\mathbf{x}\|_p + \|\mathbf{y}\|_p\right) \|\mathbf{x}+\mathbf{y}\|_p^{\frac{p}{q}},
\end{aligned}$$

divide both sides by $\|\mathbf{x}+\mathbf{y}\|_p^{\frac{p}{q}}$ and we obtain the desired result. □

B.5 Power Mean Inequality

Theorem B.5 (Power Mean Inequality)

For any $\mathbf{x} \in \mathbb{R}^n_{\ge 0}$, define

$$M_r(\mathbf{x}) = \left(\frac{1}{n}\sum_{i=1}^n x_i^r\right)^{1/r}, \quad r \in \mathbb{R}\backslash\{0\},$$

with $M_0(\mathbf{x}) = \lim_{r\to 0} M_r(\mathbf{x}) = \sqrt[n]{x_1 \cdots x_n}$, $M_{-\infty}(\mathbf{x}) = \lim_{r\to -\infty} M_r(\mathbf{x}) = \min_{1\le i\le n} x_i$, and $M_{+\infty}(\mathbf{x}) = \lim_{r\to +\infty} M_r(\mathbf{x}) = \max_{1\le i\le n} x_i$. Then, for any $-\infty \le r \le s \le +\infty$, we have

$$M_r(\mathbf{x}) \le M_s(\mathbf{x}).$$

♡

Proof [Proof Sketch] The proof is based on the observation that $f(x) = x^{\frac{s}{r}}$ is a convex function since $r \le s$. Then,

$$f\left(\frac{1}{n}\sum_{i=1}^n x_i^r\right) \le \frac{1}{n}\sum_{i=1}^n f(x_i^r) \Rightarrow \left(\frac{1}{n}\sum_{i=1}^n x_i^r\right)^{\frac{s}{r}} \le \frac{1}{n}\sum_{i=1}^n (x_i^r)^{\frac{s}{r}} \Rightarrow \left(\frac{1}{n}\sum_{i=1}^n x_i^r\right)^{\frac{1}{r}} \le \left(\frac{1}{n}\sum_{i=1}^n x_i^s\right)^{\frac{1}{s}}.$$

□

Bibliography

Anthony, Martin and Peter L Bartlett (1999). *Neural network learning: Theoretical foundations*. Vol. 9. cambridge university press Cambridge.

Basu, Amitabh (2023). *Convexity and its use in discrete and continuous optimization*.

Boyd, Stephen and Lieven Vandenberghe (2004). *Convex optimization*. Cambridge university press.

Rockafellar, R. Tyrrell (1970). *Convex analysis*. en. Princeton mathematical series 28. Princeton, N.J: Princeton University Press. ISBN: 978-0-691-08069-7.