Introduction to Machine Learning

Fall 2022

University of Science and Technology of China

Lecturer: Jie Wang — Homework 3

Name: Yunqin Zhu — ID: PB20061372

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

Exercise 1: Convex Sets

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

1. Please find the interior and relative interior of the following convex sets (you don't need to prove them).

   (a) $\{\mathbf{x} \in \mathbb{R}^3 : x_1^2 + x_2^2 < 1, x_3 = 0\} \subset \mathbb{R}^3$.

   (b) $\{\mathbf{A} \in S_{++}^n : \mathrm{tr}(\mathbf{A}) = 1\} \subset \mathbb{R}^{n\times n}$.

   (c) $\{\mathbf{A} \in S_{++}^n : \mathrm{tr}(\mathbf{A}) = 1\} \subset S^n$.

   (d) (Optional) $\{\mathbf{A} \in S_{++}^n : \mathrm{tr}(\mathbf{A}) \leq 1\} \subset \mathbb{R}^{n\times n}$.

   (e) $\mathbf{conv}\left(\{x, x^2, x^3\}\right) \subset C[0,1]$ with L^∞ norm, i.e., $\|f\|_\infty = \max_{x\in[0,1]} |f(x)|$ for any $f \in C[0,1]$.

   Solution:

   (a) $\mathbf{int}\ C = \emptyset$. $\mathbf{aff}\ C = \{\mathbf{x} \in \mathbb{R}^3 : x_3 = 0\} \implies \mathbf{relint}\ C = C$.

   (b) $\mathbf{int}\ C = \emptyset$. $\mathbf{relint}\ C = (\mathbf{relint}\ S_{++}^n) \cap (\mathbf{relint}\,\{\mathbf{A} \in S^n : \mathrm{tr}(\mathbf{A}) = 1\}) = C$.

   (c) $\mathbf{int}\ C = \emptyset$. $\mathbf{relint}\ C = C$

   (d) $\mathbf{int}\ C = \emptyset$. $\mathbf{relint}\ C = (\mathbf{relint}\ S_{++}^n) \cap (\mathbf{relint}\,\{\mathbf{A} \in S^n : \mathrm{tr}(\mathbf{A}) \leq 1\}) = \{\mathbf{A} \in S_{++}^n : \mathrm{tr}(\mathbf{A}) < 1\}$.

   (e) $\mathbf{int}\ C = \emptyset$. $\mathbf{relint}\ C = \{\alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 : \alpha_1 + \alpha_2 + \alpha_3 = 1, \alpha_1, \alpha_2, \alpha_3 > 0\}$ ■

2. Some operations that preserve convexity.

   (a) Both $\mathbf{cl}\ C$ and $\mathbf{int}\ C$ are convex.

   (b) The set $\mathbf{relint}\ C$ is convex.

   (c) The intersection $\bigcap_{i\in I} C_i$ of any collection $\{C_i : i \in \mathcal{I}\}$ of convex sets is convex.

   (d) If C_1 and C_2 are convex sets in $\mathbb{R}^n$, then the set

   $$C_1 - C_2 = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \mathbf{x}_1 - \mathbf{x}_2, \mathbf{x}_1 \in C_1, \mathbf{x}_2 \in C_2\}$$

   is convex.

   (e) The set $\{\mathbf{y} \in \mathbb{R}^m : \mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{a}, \mathbf{x} \in C\}$ is convex, where $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{a} \in \mathbb{R}^m$.

   (f) The set $\{\mathbf{y} \in \mathbb{R}^m : \mathbf{x} = \mathbf{B}\mathbf{y} + \mathbf{b}, \mathbf{x} \in C\}$ is convex, where $\mathbf{B} \in \mathbb{R}^{n\times m}$ and $\mathbf{b} \in \mathbb{R}^n$.

   Solution:

   (a) Let $\mathbf{x}, \mathbf{y} \in \mathbf{cl}\ C$. There exist $\{\mathbf{x}_n\}, \{\mathbf{y}_n\} \subset C$ such that $\mathbf{x}_n \to \mathbf{x}$ and $\mathbf{y}_n \to \mathbf{y}$. The convexity of C implies that $\{\theta\mathbf{x}_n + (1-\theta)\mathbf{y}_n\} \subset C$ and $\theta\mathbf{x}_n + (1-\theta)\mathbf{y}_n \to \theta\mathbf{x} + (1-\theta)\mathbf{y}$ for any $\theta \in [0,1]$; that is, $\theta\mathbf{x} + (1-\theta)\mathbf{y} \in \mathbf{cl}\ C$. Hence $\mathbf{cl}\ C$ is convex.

Let $\mathbf{x}, \mathbf{y} \in \mathbf{int}\, C$. There exists $\epsilon > 0$ such that $N_\epsilon(\mathbf{x}), N_\epsilon(\mathbf{y}) \subset C$. The convexity of C implies that $N_\epsilon(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \subset \{\theta\tilde{\mathbf{x}} + (1-\theta)\tilde{\mathbf{y}} : \tilde{\mathbf{x}} \in N_\epsilon(\mathbf{x}), \tilde{\mathbf{y}} \in N_\epsilon(\mathbf{y})\} \subset C$ for any $\theta \in [0,1]$; that is, $\theta\mathbf{x} + (1-\theta)\mathbf{y} \in \mathbf{int}\, C$. Hence $\mathbf{int}\, C$ is convex.

(b) Let $\mathbf{x}, \mathbf{y} \in \mathbf{relint}\, C$. There exists $\epsilon > 0$ such that $N_\epsilon(\mathbf{x}) \cap \mathbf{aff}\, C \subset C$ and $N_\epsilon(\mathbf{y}) \cap \mathbf{aff}\, C \subset C$. The convexity of C implies that $N_\epsilon(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \cap \mathbf{aff}\, C \subset \{\theta\tilde{\mathbf{x}} + (1-\theta)\tilde{\mathbf{y}} : \tilde{\mathbf{x}} \in N_\epsilon(\mathbf{x}) \cap \mathbf{aff}\, C, \tilde{\mathbf{y}} \in N_\epsilon(\mathbf{y}) \cap \mathbf{aff}\, C\} \subset C$ for any $\theta \in [0,1]$; that is, $\theta\mathbf{x} + (1-\theta)\mathbf{y} \in \mathbf{relint}\, C$. Hence $\mathbf{relint}\, C$ is convex.

(c) Let $\mathbf{x}, \mathbf{y} \in \bigcap_{i \in I} C_i$. The convexity of $C_i, \forall i \in I$ implies that $\theta\mathbf{x} + (1-\theta)\mathbf{y} \in C_i$ for any $\theta \in [0,1]$; that is, $\theta\mathbf{x} + (1-\theta)\mathbf{y} \in \bigcap_{i \in I} C_i$. Hence $\bigcap_{i \in I} C_i$ is convex.

(d) Let $\mathbf{x}, \mathbf{y} \in C_1 - C_2$. There exist $\mathbf{x}_1, \mathbf{y}_1 \in C_1$ and $\mathbf{x}_2, \mathbf{y}_2 \in C_2$ such that $\mathbf{x} = \mathbf{x}_1 - \mathbf{x}_2$ and $\mathbf{y} = \mathbf{y}_1 - \mathbf{y}_2$. The convexity of C_1 and C_2 implies that $\theta\mathbf{x} + (1-\theta)\mathbf{y} = \theta\mathbf{x}_1 + (1-\theta)\mathbf{y}_1 - \theta\mathbf{x}_2 - (1-\theta)\mathbf{y}_2 \in C_1 - C_2$ for any $\theta \in [0,1]$. Hence $C_1 - C_2$ is convex.

(e) Let $\mathbf{y}_1, \mathbf{y}_2 \in \{\mathbf{y} \in \mathbb{R}^m : \mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{a}, \mathbf{x} \in C\}$. There exist $\mathbf{x}_1, \mathbf{x}_2 \in C$ such that $\mathbf{y}_1 = \mathbf{A}\mathbf{x}_1 + \mathbf{a}$ and $\mathbf{y}_2 = \mathbf{A}\mathbf{x}_2 + \mathbf{a}$. The convexity of C implies that $\theta\mathbf{y}_1 + (1-\theta)\mathbf{y}_2 = \mathbf{A}(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2) + \mathbf{a} \in \{\mathbf{y} \in \mathbb{R}^m : \mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{a}, \mathbf{x} \in C\}$ for any $\theta \in [0,1]$. Hence $\{\mathbf{y} \in \mathbb{R}^m : \mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{a}, \mathbf{x} \in C\}$ is convex.

(f) Let $\mathbf{y}_1, \mathbf{y}_2 \in \{\mathbf{y} \in \mathbb{R}^m : \mathbf{x} = \mathbf{B}\mathbf{y} + \mathbf{b}, \mathbf{x} \in C\}$. There exist $\mathbf{x}_1, \mathbf{x}_2 \in C$ such that $\mathbf{x}_1 = \mathbf{B}\mathbf{y}_1 + \mathbf{b}$ and $\mathbf{x}_2 = \mathbf{B}\mathbf{y}_2 + \mathbf{b}$. The convexity of C implies that $\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2 = \mathbf{B}(\theta\mathbf{y}_1 + (1-\theta)\mathbf{y}_2) + \mathbf{b} \in C$ for any $\theta \in [0,1]$; that is, $\theta\mathbf{y}_1 + (1-\theta)\mathbf{y}_2 \in \{\mathbf{y} \in \mathbb{R}^m : \mathbf{x} = \mathbf{B}\mathbf{y} + \mathbf{b}, \mathbf{x} \in C\}$. Hence $\{\mathbf{y} \in \mathbb{R}^m : \mathbf{x} = \mathbf{B}\mathbf{y} + \mathbf{b}, \mathbf{x} \in C\}$ is convex. $\blacksquare$

Exercise 2: Affine Sets

Please show the following statements about affine sets.

1. If $U \subset \mathbb{R}^n$ and $\mathbf{0} \in U$, then U is an affine set if and only if it is a subspace.

   Solution:

   ($\Rightarrow$) Since U is an affine set, for any $\mathbf{x}, \mathbf{y} \in U$ and $\alpha, \beta \in \mathbb{R}$, we have $\alpha\mathbf{x} + \beta\mathbf{y} = \alpha\mathbf{x} + \beta\mathbf{y} + (1 - \alpha - \beta)\mathbf{0} \in U$. Thus, U is a subspace.

   ($\Leftarrow$) Since U is a subspace, for any $\mathbf{x}, \mathbf{y} \in U$ and $\theta \in \mathbb{R}$, we have $\theta\mathbf{x} + (1 - \theta)\mathbf{y} \in U$. Thus, U is an affine set. ■

2. If $U \subset \mathbb{R}^n$ is an affine set, there is a unique subspace $V \subset \mathbb{R}^n$ such that $U = \mathbf{u} + V$ for any $\mathbf{u} \in U$.

   Solution:
   Let $\mathbf{u} \in U$ be arbitrary. Then, for any $\mathbf{x}, \mathbf{y} \in U - \mathbf{u}$ and $\alpha, \beta \in \mathbb{R}$, we have $\alpha\mathbf{x} + \beta\mathbf{y} = [\alpha(\mathbf{u} + \mathbf{x}) + \beta(\mathbf{u} + \mathbf{y}) + (1 - \alpha - \beta)\mathbf{u}] - \mathbf{u} \in U - \mathbf{u}$. Thus, $V = U - \mathbf{u}$ is a subspace. For any $\tilde{\mathbf{u}} \in U$, it is clear that $\tilde{\mathbf{u}} - \mathbf{u} \in V$, and hence $U - \tilde{\mathbf{u}} = V - (\tilde{\mathbf{u}} - \mathbf{u}) = V$, from which we conclude that V must be unique. ■

3. Let $U = \mathbf{aff}(\{(1,0,0)^\top, (0,1,0)^\top, (0,0,1)^\top\})$. Please find two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ such that we can represent any vectors $\mathbf{v} \in U$ in the form of $\mathbf{v} = (1,0,0)^\top + \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2$ uniquely, where α_1 and α_2 are two real numbers depending on $\mathbf{v}$. Furthermore, given a point $\mathbf{x}_0 \in U$, find two vectors $\mathbf{w}_1$ and $\mathbf{w}_2$ such that we can represent any vectors $\mathbf{w} \in U$ in the form of $\mathbf{w} = \mathbf{x}_0 + \alpha_1\mathbf{w}_1 + \alpha_2\mathbf{w}_2$ uniquely.

   Solution:
   For any $\mathbf{v} \in U$, there exist $\alpha_1, \alpha_2 \in \mathbb{R}$ depending on $\mathbf{v}$ such that $\mathbf{v} = (1 - \alpha_1 - \alpha_2)(1,0,0)^\top + \alpha_1(0,1,0)^\top + \alpha_2(0,0,1)^\top = (1,0,0)^\top + \alpha_1(-1,1,0)^\top + \alpha_2(-1,0,1)^\top$. Equivalently, $\mathbf{v}_1 = (-1,1,0)^\top$ and $\mathbf{v}_2 = (-1,0,1)^\top$ span the subspace $V = U - (1,0,0)^\top$. By Question 2, $V = U - \mathbf{x}_0$ for any $\mathbf{x}_0 \in U$. So $\mathbf{w} \in U$ can also be represented in the form of $\mathbf{w} = \mathbf{x}_0 + \alpha_1\mathbf{w}_1 + \alpha_2\mathbf{w}_2$, where $\mathbf{w}_1 = \mathbf{v}_1 = (-1,1,0)^\top$ and $\mathbf{w}_2 = \mathbf{v}_2 = (-1,0,1)^\top$. Since (α_1, α_2) serves as the coordinate of $\mathbf{w} - \mathbf{x}_0$ in the basis $\{\mathbf{w}_1, \mathbf{w}_2\}$, we know it must be unique. ■

Exercise 3: Convex Hull and Affine Hull (Optional)

Let A be a subset of $\mathbb{R}^n$.

1. (a) Please show that the convex hull of A is the smallest convex set containing A, i.e., all the convex sets containing A also contain $\mathbf{conv}\ A$.

   (b) Please find the convex hull of the following sets.

   i. $\{\mathbf{A} \in S^n_{++} : \mathrm{tr}(\mathbf{A}) = 1\} \cup \{\mathbf{A} \in S^n_{++} : \mathrm{tr}(\mathbf{A}) \geq 2\} \subset \mathbb{R}^{n\times n}$.

   ii. $\{f \in C[0,1] : \|f\|_\infty = 1\} \cup \{f \in C[0,1] : \|f\|_\infty = 2\} \subset C[0,1]$.

   Solution:

   (a) First, it is clear that $A \subset \mathbf{conv}\ A$. Second, let C be an arbitrary convex set containing A. We prove by induction on k that any $\mathbf{x} = \sum_{i=1}^k \theta_i \mathbf{x}_i \in \mathbf{conv}\ A$ also belongs to C, where $\mathbf{x}_i \in A$, $\theta_i \geq 0$, and $\sum_{i=1}^k \theta_i = 1$. If $k = 1$, we have $\mathbf{x} = \mathbf{x}_1 \in C$ by definition. Now, assume that the statement holds for $k-1$ and consider $\mathbf{x} = \sum_{i=1}^k \theta_i \mathbf{x}_i$, where $\sum_{i=1}^{k-1} \theta_i \neq 0$. Let $\alpha_i = \theta_i / \sum_{i=1}^{k-1} \theta_i$. Then,

   $$\sum_{i=1}^{k-1} \alpha_i \mathbf{x}_i \in C,\ \mathbf{x}_k \in C \implies \mathbf{x} = \left(\sum_{i=1}^{k-1} \theta_i\right)\left(\sum_{i=1}^{k-1} \alpha_i \mathbf{x}_i\right) + \theta_k \mathbf{x}_k \in C,$$

   which completes the proof, i.e., $\mathbf{conv}\ A \subset C$.

   (b) i. $\mathbf{conv}\ A = \{\mathbf{A} \in S^n_{++} : \mathrm{tr}(\mathbf{A}) \geq 1\}$, which is a convex set containing A and any $\mathbf{A}$ belongs to this set can be written as the convex combination

   $$\mathbf{A} = \theta \frac{\mathbf{A}}{\mathrm{tr}(\mathbf{A})} + (1-\theta)\left(\frac{\mathbf{A}}{\mathrm{tr}(\mathbf{A})} + \mathbf{A}\right),\ \text{where } \theta = \frac{1}{\mathrm{tr}(\mathbf{A})}.$$

   ii. $\mathbf{conv}\ A = \{f \in C[0,1] : \|f\|_\infty \leq 2\}$, which is a convex set containing A and any f belongs to this set can be written as the convex combination ... ■

2. (a) Please show that the affine hull of A is the smallest affine set containing A, i.e., all the affine sets containing A also contain $\mathbf{aff}A$.

   (b) Please find the affine hull of the following sets.

   i. $\{\mathbf{A} \in S^n_{++} : \mathrm{tr}(\mathbf{A}) = 1\} \subset \mathbb{R}^{n\times n}$.

   ii. $\{\mathbf{A} \in S^n_{++} : \mathrm{tr}(\mathbf{A}) = 1\} \cup \{\mathbf{A} \in S^n_{++} : \mathrm{tr}(\mathbf{A}) \geq 2\} \subset \mathbb{R}^{n\times n}$.

   Solution:

   (a) First, it is clear that $A \subset \mathbf{aff}\ A$. Second, let C be an arbitrary affine set containing A. We prove by induction on k that any $\mathbf{x} = \sum_{i=1}^k \theta_i \mathbf{x}_i \in \mathbf{aff}\ A$ also belongs to C, where $\mathbf{x}_i \in A$ and $\sum_{i=1}^k \theta_i = 1$. If $k = 1$, we have $\mathbf{x} = \mathbf{x}_1 \in C$ by definition. Now, assume that the statement holds for $k-1$ and consider $\mathbf{x} = \sum_{i=1}^k \theta_i \mathbf{x}_i$. Without loss of generality, we can assume that $\sum_{i=1}^{k-1} \theta_i \neq 0$. Let $\alpha_i = \theta_i / \sum_{i=1}^{k-1} \theta_i$. Then,

   $$\sum_{i=1}^{k-1} \alpha_i \mathbf{x}_i \in C,\ \mathbf{x}_k \in C \implies \mathbf{x} = \left(\sum_{i=1}^{k-1} \theta_i\right)\left(\sum_{i=1}^{k-1} \alpha_i \mathbf{x}_i\right) + \theta_k \mathbf{x}_k \in C,$$

which completes the proof, i.e., $\mathbf{aff}\ A \subset C$.

(b) i. $\mathbf{aff}\ A = \{\mathbf{A} \in S^n : \operatorname{tr}(\mathbf{A}) = 1\}$, which is an affine set containing A and any $\mathbf{A}$ belongs to this set can be written as the affine combination

$$\mathbf{A} = \theta \frac{\mathbf{A} - \lambda \mathbf{I}}{1 - \lambda n} + (1 - \theta)\frac{\mathbf{I}}{n},$$

where $\lambda < \lambda_{\min}(\mathbf{A})$ and $\theta = 1 - \lambda n$.

ii. $\mathbf{aff}\ A = S^n$, which is an affine set containing A and any $\mathbf{A}$ belongs to this set can be written as the affine combination

$$\mathbf{A} = \theta(\mathbf{A} + \alpha \mathbf{I}) + (1 - \theta)(\mathbf{A} + 2\alpha \mathbf{I}),$$

where $\alpha > \max\{-\lambda_{\min}(\mathbf{A}), \frac{2-\operatorname{tr}(\mathbf{A})}{n}\}$ and $\theta = 2$. ■

Exercise 4: Relative Interior and Interior

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

1. Let $\mathbf{x}_0 \in C$. Please show the following statements.

   (a) The point $\mathbf{x}_0 \in \textbf{relint } C$ if and only if there exists $r > 0$ such that $\mathbf{x}_0 + r\mathbf{v} \in C$ for any $\mathbf{v} \in \textbf{aff } C - \mathbf{x}_0$ and $\|\mathbf{v}\|_2 \leq 1$.

   (b) Let $\{\mathbf{v}_i\}_{i=1}^m$ be a basis of $\textbf{aff } C - \mathbf{x}_0$. Then $\mathbf{x}_0 \in \textbf{relint } C$ if and only if there exists $r > 0$ such that $\mathbf{x}_0 + r\sum_i \alpha_i \mathbf{v}_i \in C$ for any $\{\alpha_i\}_{i=1}^m$ with $\sum_i \alpha_i^2 \leq 1$.

   Solution:

   (a) By definition, $\mathbf{x}_0 \in \textbf{relint } C$ if and only if there exists $r > 0$ such that $B(\mathbf{x}_0, r) \cap \textbf{aff } C \subset C$. Since $B(\mathbf{x}_0, r) = \{\mathbf{x} : \|\mathbf{x} - \mathbf{x}_0\|_2 \leq r\} = \{\mathbf{x}_0 + r\mathbf{v} : \|\mathbf{v}\|_2 \leq 1\}$, we have $B(\mathbf{x}_0, r) \cap \textbf{aff } C = \{\mathbf{x}_0 + r\mathbf{v} : r\mathbf{v} \in \textbf{aff } C - \mathbf{x}_0 \text{ and } \|\mathbf{v}\|_2 \leq 1\}$. Moreover, the fact that $\textbf{aff } C - \mathbf{x}_0$ is a subspace implies that $r\mathbf{v} \in \textbf{aff } C - \mathbf{x}_0$ if and only if $\mathbf{v} \in \textbf{aff } C - \mathbf{x}_0$. So $B(\mathbf{x}_0, r) \cap \textbf{aff } C = \{\mathbf{x}_0 + r\mathbf{v} : \mathbf{v} \in \textbf{aff } C - \mathbf{x}_0 \text{ and } \|\mathbf{v}\|_2 \leq 1\}$. This completes the proof.

   (b) Let $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_m)$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)$. Then any $\mathbf{v} = \sum_{i=1}^m \alpha_i \mathbf{v}_i \in \textbf{aff } C - \mathbf{x}_0$ can be written as $\mathbf{v} = \mathbf{V}\boldsymbol{\alpha}$. So $\|\boldsymbol{\alpha}\|_2 \leq 1 \implies \|\mathbf{v}\|_2 \leq \|\mathbf{V}\|_2 \|\boldsymbol{\alpha}\|_2 \leq \|\mathbf{V}\|_2$. On the other hand, $\|\mathbf{v}\|_2 \leq \|\mathbf{V}\|_2 \implies \|\alpha\|_2 \leq \|\mathbf{V}^{-1}\|_2 \|\mathbf{v}\|_2 \leq 1$. Therefore, $\mathbf{x}_0 \in \textbf{relint } C$ if and only if there exists $\rho = r\|\mathbf{V}\|_2 > 0$ such that $C \supset B(\mathbf{x}_0, \rho) \cap \textbf{aff } C = \{\mathbf{x}_0 + r\mathbf{v} : \mathbf{v} \in \textbf{aff } C - \mathbf{x}_0 \text{ and } \|\mathbf{v}\|_2 \leq \|\mathbf{V}\|_2\} = \{\mathbf{x}_0 + r\mathbf{V}\boldsymbol{\alpha} : \|\boldsymbol{\alpha}\|_2 \leq 1\}$, which is equivalent to the statement. ■

2. (a) We let $\mathbf{x}_0 \in \textbf{int } C$, $\mathbf{x}_1 \in \textbf{bd } C$ and $\mathbf{x}_2 = \lambda(\mathbf{x}_1 - \mathbf{x}_0) + \mathbf{x}_0$.

      i. Please show that if $\lambda > 1$, then $\mathbf{x}_2 \notin C$.

      ii. Please show that if $\lambda \in (0, 1)$, then $\mathbf{x}_2 \in \textbf{int } C$.

   (b) i. Please show that $\mathbf{x} \in \textbf{relint } C$ if and only if for any $\mathbf{y} \in C$, there exists $\gamma > 0$ such that $\mathbf{x} + \gamma(\mathbf{x} - \mathbf{y}) \in C$.

      ii. Please show that if $\mathbf{x} \in \textbf{relint } C$, $\mathbf{y} \in \textbf{cl } C$, then $\lambda\mathbf{x} + (1 - \lambda)\mathbf{y} \in \textbf{relint } C$ for $\lambda \in (0, 1]$.

   Solution:

   (a) i. Assume that $\mathbf{x}_2 \in C$. Since $\mathbf{x}_0 \in \textbf{int } C$, there exists $r > 0$ such that $B(\mathbf{x}_0, r) \in C$. By convexity of C, we have $(1 - \frac{1}{\lambda})B(\mathbf{x}_0, r) + \frac{1}{\lambda}\mathbf{x}_2 = B(\mathbf{x}_1, \frac{(\lambda-1)r}{\lambda}) \subset C$, which contradicts $\mathbf{x}_1 \in \textbf{bd } C$. So $\mathbf{x}_2 \notin C$.

      ii. Since $\mathbf{x}_0 \in \textbf{int } C$, there exists $r > 0$ such that $B(\mathbf{x}_0, r) \in C$. Since $\mathbf{x}_1 \in \textbf{bd } C$, there exists $0 < \epsilon < \frac{1-\lambda}{\lambda} r$ such that $\mathbf{x}_1 \in C + B(0, \epsilon)$. Then

$$
\begin{aligned}
& B(\mathbf{x}_2, (1 - \lambda)r - \lambda\epsilon) \\
= & B((1 - \lambda)\mathbf{x}_0, (1 - \lambda)r - \lambda\epsilon) + \lambda\mathbf{x}_1 \\
\subset & B((1 - \lambda)\mathbf{x}_0, (1 - \lambda)r - \lambda\epsilon) + \lambda C + \lambda B(0, \epsilon) \\
= & (1 - \lambda)B(\mathbf{x}_0, r) + \lambda C \\
\subset & (1 - \lambda)C + \lambda C = C,
\end{aligned}
$$

      implying that $\mathbf{x}_2 \in \textbf{int } C$.

(b) i.($\Rightarrow$) Clearly, $\mathbf{x}-\mathbf{y} \in \mathbf{aff}\ C - \mathbf{x}$. According to Question 1(a), there exists $r = \gamma\|\mathbf{x}-\mathbf{y}\|_2 > 0$ such that $\mathbf{x} + r\frac{(\mathbf{x}-\mathbf{y})}{\|\mathbf{x}-\mathbf{y}\|_2} = \mathbf{x} + \gamma(\mathbf{x}-\mathbf{y}) \in C$.

($\Leftarrow$) Let $\mathbf{y} \in \mathbf{relint}\ C$ and $\mathbf{x}+\gamma(\mathbf{x}-\mathbf{y}) \in \mathbf{cl}\ C$. Then $\mathbf{x} = \frac{1}{\gamma+1}(\mathbf{x}+\gamma(\mathbf{x}-\mathbf{y})) + \frac{\gamma}{\gamma+1}\mathbf{y}$ implies that $\mathbf{x} \in \mathbf{relint}\ C$, according to (b)ii.

ii. If $\lambda = 1$, it is clear that $\mathbf{x} \in \mathbf{relint}\ C$. Consider $0 < \lambda < 1$. Since $\mathbf{x} \in \mathbf{relint}\ C$, there exists $r > 0$ such that $B(\mathbf{x}, r) \cap \mathbf{aff}\ C \subset C$. Since $\mathbf{y} \in \mathbf{cl}\ C$, there exists $0 < \epsilon < \frac{\lambda}{1-\lambda}r$ such that $\mathbf{y} \in C + B(0,\epsilon) \cap \mathbf{aff}\ C$. Then

$$\begin{aligned}
&B(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}, \lambda r - (1-\lambda)\epsilon) \cap \mathbf{aff}\ C \\
&= B(\lambda\mathbf{x}, \lambda r - (1-\lambda)\epsilon) \cap \mathbf{aff}\ C + (1-\lambda)\mathbf{y} \\
&\subset B(\lambda\mathbf{x}, \lambda r - (1-\lambda)\epsilon) \cap \mathbf{aff}\ C + (1-\lambda)C + (1-\lambda)B(0,\epsilon) \cap \mathbf{aff}\ C \\
&= \lambda B(\mathbf{x}, r) \cap \mathbf{aff}\ C + (1-\lambda)C \\
&\subset \lambda C + (1-\lambda)C = C,
\end{aligned}$$

implying that $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in \mathbf{relint}\ C$. ■

3. (a) Please show the following statements.

i. Suppose $\mathbf{int}\ C$ is nonempty, then $\mathbf{int}\ C = \mathbf{int}\,(\mathbf{cl}\ C)$ (in fact, the result still holds when $C = \emptyset$).

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

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

[Hint: if C contains more than one point, then $\mathbf{relint}\ C$ is nonempty. You may also use the results in Question 2.]

(b) Using the results in Question 3(a), please prove the following statement.

For a convex set $C \subset \mathbb{R}^n$ and $\mathbf{x}_0 \in \mathbf{bd}\ C$, we can find a sequence $\{\mathbf{x}_k\} \subset \mathbb{R}^n \setminus \mathbf{cl}\ C$ such that $\mathbf{x}_k \to \mathbf{x}_0$ as $k \to \infty$.

Solution:

(a) i. It is clear that $\mathbf{int}\ C \subset \mathbf{int}\,(\mathbf{cl}\ C)$ as $C \subset \mathbf{cl}\ C$. Consider $\mathbf{x} \in \mathbf{int}\,(\mathbf{cl}\ C)$ and $\mathbf{y} \in \mathbf{int}\ C$, where $\mathbf{x} \neq \mathbf{y}$. There exist $r = \gamma\|\mathbf{x}-\mathbf{y}\|_2 > 0$ such that $\mathbf{z} = \mathbf{x} + \gamma(\mathbf{x}-\mathbf{y}) \in B(\mathbf{x}, r) \subset \mathbf{cl}\ C$. Then, by Question 2(a)ii, $\mathbf{x} \in \mathbf{relint}\ C$ follows from $\mathbf{x} = \frac{1}{\gamma+1}\mathbf{z} + \frac{\gamma}{\gamma+1}\mathbf{y}$.

ii. It is clear that $\mathbf{cl}\,(\mathbf{relint}\ C) \subset \mathbf{cl}\ C$ as $\mathbf{relint}\ C \subset C$. Consider $\mathbf{x} \in \mathbf{cl}\ C$ and $\mathbf{y} \in \mathbf{relint}\ C$, where $\mathbf{x} \neq \mathbf{y}$. According to Question 2(b)ii, we have $\mathbf{x}_k = (1-\frac{1}{k})\mathbf{x} + \frac{1}{k}\mathbf{y} \in \mathbf{relint}\ C$ for any $k \in \mathbb{N}^+$. Since $\mathbf{x}_k \to \mathbf{x}$, we have $\mathbf{x} \in \mathbf{cl}\,(\mathbf{relint}\ C)$. Hence $\mathbf{cl}\ C \subset \mathbf{cl}\,(\mathbf{relint}\ C)$.

iii. It is clear that $\mathbf{relint}\,(\mathbf{cl}\ C) \supset \mathbf{relint}\ C$ as $\mathbf{cl}\ C \supset C$ and $\mathbf{aff}\,(\mathbf{cl}\ C) = \mathbf{aff}\ C$. Consider $\mathbf{x} \in \mathbf{relint}\,(\mathbf{cl}\ C)$ and $\mathbf{y} \in \mathbf{relint}\ C$, where $\mathbf{x} \neq \mathbf{y}$. By Question 2(b)i, there exists $\mathbf{z} = \mathbf{x} + \gamma(\mathbf{x}-\mathbf{y}) \in \mathbf{cl}\ C$ for some $\gamma > 0$. Then, by Question 2(b)ii, $\mathbf{x} \in \mathbf{relint}\ C$ follows from $\mathbf{x} = \frac{1}{\gamma+1}\mathbf{z} + \frac{\gamma}{\gamma+1}\mathbf{y}$.

(b) $\mathbf{x}_0 \in \mathbf{bd}\ C \implies \mathbf{x}_0 \notin \mathbf{int}\ C \implies \mathbf{x}_0 \notin \mathbf{int}\,(\mathbf{cl}\ C)$. That is, for any $r > 0$, there exists $\mathbf{x} \in B(\mathbf{x}_0, r)$ but $\mathbf{x} \notin \mathbf{cl}\ C$. Let $r = \frac{1}{k}$ and pick $\mathbf{x}_k \in B(\mathbf{x}_0, r)$ such that $\mathbf{x}_k \notin \mathbf{cl}\ C$. Then $\{\mathbf{x}_k\}$ is the desired sequence. ■

Exercise 5: Relative Boundary

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

1. For a set $S \subset \mathbb{R}^n$, $\textbf{relbd}\ S \subset \textbf{bd}\ S$.

   Solution:
   By definition, we know that $\textbf{relint}\ S \supset \textbf{int}\ S$. Hence, $\textbf{relbd}\ S \subset \textbf{cl}\ S \setminus \textbf{int}\ S = \textbf{bd}\ S$. ■

2. For a set $S \subset \mathbb{R}^n$, $\textbf{relbd}\ S = \textbf{bd}\ S$.

   Solution:
   Counter-example: $S = [0,1] \times \{0\} \subset \mathbb{R}^2$.
   $\textbf{relbd}\ S = \{(0,0),(1,0)\}$, $\textbf{bd}\ S = S$, $\textbf{relbd}\ S \neq \textbf{bd}\ S$. ■

3. For a set $S \subset \mathbb{R}^n$, $\textbf{relbd}\ S = \textbf{relbd}\ \textbf{cl}\ S$.

   Solution:
   Counter-example: $S = \mathbb{R} \setminus \{0\} \subset \mathbb{R}$.
   $\textbf{relbd}\ S = \{0\}$, $\textbf{relbd}\,(\textbf{cl}\ S) = \emptyset$, $\textbf{relbd}\ S \neq \textbf{relbd}\,(\textbf{cl}\ S)$. ■

4. (Optional) For a convex set $C \subset \mathbb{R}^n$, $\textbf{relbd}\ C = \textbf{relbd}\ \textbf{cl}\ C$.

   Solution:
   If S is empty, then the statement is clear. If S is nonempty, by Exercise 4 Question 3(a)iii, we have $\textbf{relint}\ S = \textbf{relint}\,(\textbf{cl}\ S)$. So $\textbf{relbd}\ S = \textbf{cl}\ S \setminus \textbf{relint}\ S = \textbf{cl}\,(\textbf{cl}\ S) \setminus \textbf{relint}\,(\textbf{cl}\ S) = \textbf{relbd}\,(\textbf{cl}\ S)$. ■

5. For a set $S \subset \mathbb{R}^n$ and $\mathbf{x}_0 \in \textbf{cl}\ S$, we can find a sequence $\{\mathbf{x}_k\} \subset \mathbb{R}^n \setminus \textbf{cl}\ S$ such that $\mathbf{x}_k \to \mathbf{x}_0$ as $k \to \infty$.

   Solution:
   If $\mathbf{x}_0 \in \textbf{int}\ S$, then the statement is clearly false. If $\mathbf{x}_0 \in \textbf{bd}\ S$, consider $S = \mathbb{R} \setminus \{0\} \subset \mathbb{R}$ for counter-example. It is impossible to find the desired sequence $\{\mathbf{x}_k\} \subset \mathbb{R} \setminus \textbf{cl}\ S = \emptyset$. ■

Exercise 6: Minkowski Summation of Sets (Optional)

The Minkowski sum of two sets S_1 and S_2 is defined by

$$S_1 + S_2 = \{\mathbf{x} + \mathbf{y} : \mathbf{x} \in S_1, \mathbf{y} \in S_2\}.$$

1. Let $S_1 = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\|_2 \leq 1\}$ and $S_2 = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\|_\infty \leq 1\}$.

   (a) Please draw the set $S_1 + S_2$.

   (b) How do you tell if a point $\mathbf{x}$ is in the set $S_1 + S_2$?

   Solution:

   (a) The plot of $S_1 + S_2$ is shown below.

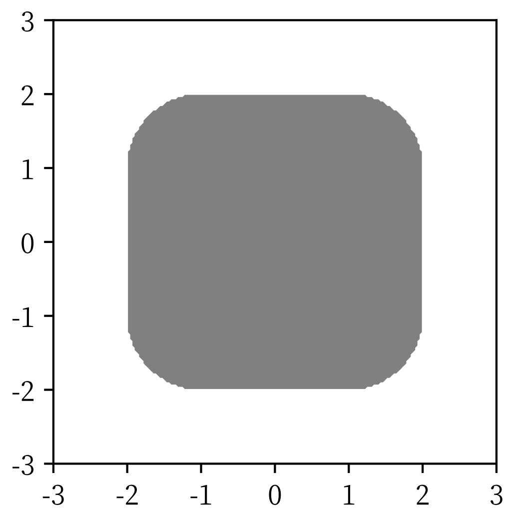

Figure 1: Plot of $S_1 + S_2$

   (b) $S_1 + S_2 = \{\mathbf{x} \in \mathbb{R}^n : S_1 \cap (\mathbf{x} - S_2) \neq \emptyset\} = \{\mathbf{x} \in \mathbb{R}^n : S_2 \cap (\mathbf{x} - S_1) \neq \emptyset\}$. ■

2. Recall that $\mathbb{R}^n$ can be decomposed as $\mathbb{R}^n = S \oplus S^\perp$, i.e., $\mathbb{R}^n = S + S^\perp$ and $S \cap S^\perp = \emptyset$, where $S \subset \mathbb{R}^n$ is a subspace and $S^\perp = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \perp S\}$. Let $C \subset \mathbb{R}^n$ be a convex set. Define $\hat{C} = C + (\mathbf{aff}\ C - \mathbf{x}_0)^\perp$. Please show that:

   (a) $\dim(\mathbf{aff}\ \hat{C}) = n$;

   (b) $\mathbf{relint}\ C + (\mathbf{aff}\ C - \mathbf{x}_0)^\perp = \mathbf{relint}\ \hat{C}$;

   (c) $\mathbf{relbd}\ C + (\mathbf{aff}\ C - \mathbf{x}_0)^\perp = \mathbf{relbd}\ \hat{C}$.

   Solution:

   (a) We assert that $\mathbf{aff}\ \hat{C} = \mathbf{aff}\ C + \mathbf{aff}\ (\mathbf{aff}\ C - \mathbf{x}_0)^\perp$. Since $(\mathbf{aff}\ C - \mathbf{x}_0)^\perp$ is a subspace, we have $\mathbf{aff}\ (\mathbf{aff}\ C - \mathbf{x}_0)^\perp = (\mathbf{aff}\ C - \mathbf{x}_0)^\perp$. So $\mathbf{aff}\ \hat{C} = \mathbf{x}_0 + (\mathbf{aff}\ C - \mathbf{x}_0) + (\mathbf{aff}\ C - \mathbf{x}_0)^\perp = \mathbf{x}_0 + \mathbb{R}^n = \mathbb{R}^n$. Hence $\dim(\mathbf{aff}\ \hat{C}) = n$.

To complete the proof, we show that $\mathbf{aff}\,(C_1 + C_2) = \mathbf{aff}\,(C_1) + \mathbf{aff}\,(C_2)$ for any sets C_1 and C_2. Let $\mathbf{z} = \sum_i \theta_i(\mathbf{x}_i + \mathbf{y}_i) \in \mathbf{aff}\,(C_1 + C_2)$, where $\mathbf{x}_i \in C_1$, $\mathbf{y}_i \in C_2$ and $\sum_i \theta_i = 1$. Then $\mathbf{z} = \sum_i \theta_i\mathbf{x}_i + \sum_i \theta_i\mathbf{y}_i \in \mathbf{aff}\,(C_1) + \mathbf{aff}\,(C_2)$, and hence $\mathbf{aff}\,(C_1 + C_2) \subset \mathbf{aff}\,(C_1) + \mathbf{aff}\,(C_2)$.

To show the reverse inclusion, let $\mathbf{x} = \sum_i \alpha_i\mathbf{x}_i \in \mathbf{aff}\ C_1$ and $\mathbf{y} = \sum_j \beta_j\mathbf{y}_j \in \mathbf{aff}\ C_2$, where $\mathbf{x}_i \in C_1$, $\mathbf{y}_j \in C_2$ and $\sum_i \alpha_i = \sum_j \beta_j = 1$. Then $\mathbf{x} + \mathbf{y} = \sum_{i,j} \alpha_i\beta_i(\mathbf{x}_i + \mathbf{y}_j) \in \mathbf{aff}\,(C_1 + C_2)$, and hence $\mathbf{aff}\,(C_1) + \mathbf{aff}\,(C_2) \subset \mathbf{aff}\,(C_1 + C_2)$.

(b) We assert that $\mathbf{relint}\ \hat{C} = \mathbf{relint}\ C + \mathbf{relint}\,(\mathbf{aff}\ C - \mathbf{x}_0)^\perp$. Since $(\mathbf{aff}\ C - \mathbf{x}_0)^\perp$ is a subspace, we have $\mathbf{relint}\,(\mathbf{aff}\ C - \mathbf{x}_0)^\perp = (\mathbf{aff}\ C - \mathbf{x}_0)^\perp$. So $\mathbf{relint}\ C + (\mathbf{aff}\ C - \mathbf{x}_0)^\perp = \mathbf{relint}\ \hat{C}$.

To complete the proof, we show that $\mathbf{relint}\,(C_1 + C_2) = \mathbf{relint}\,(C_1) + \mathbf{relint}\,(C_2)$ for any convex sets C_1 and C_2. First, we note that by Exercise 4.2(b), a point $\mathbf{x}$ belongs to the relative interior of a convex set C if and only if $\mathbf{x} = \theta\mathbf{x}_1 + (1 - \theta)\mathbf{x}_2$, where $\mathbf{x}_1, \mathbf{x}_2 \in C$ and $0 < \theta < 1$. For any $\mathbf{z} = \theta\mathbf{z}_1 + (1 - \theta)\mathbf{z}_2 \in \mathbf{relint}\,(C_1 + C_2)$, where $\mathbf{z}_1, \mathbf{z}_2 \in C_1 + C_2$ and $0 < \theta < 1$, there exists $\mathbf{x}_1, \mathbf{x}_2 \in C_1$ and $\mathbf{y}_1, \mathbf{y}_2 \in C_2$ such that $\mathbf{z}_1 = \mathbf{x}_1 + \mathbf{y}_1$ and $\mathbf{z}_2 = \mathbf{x}_2 + \mathbf{y}_2$. Therefore, $\mathbf{z} = \theta\mathbf{x}_1 + (1 - \theta)\mathbf{x}_2 + \theta\mathbf{y}_1 + (1 - \theta)\mathbf{y}_2 \in \mathbf{relint}\ C_1 + \mathbf{relint}\ C_2$, and hence $\mathbf{relint}\,(C_1 + C_2) \subset \mathbf{relint}\,(C_1) + \mathbf{relint}\,(C_2)$.

To show the reverse inclusion, let $\mathbf{x} = \alpha\mathbf{x}_1 + (1 - \alpha)\mathbf{x}_2 \in \mathbf{relint}\ C_1$ and $\mathbf{y} = \beta\mathbf{y}_1 + (1 - \beta)\mathbf{y}_2 \in \mathbf{relint}\ C_2$, where $\mathbf{x}_1, \mathbf{x}_2 \in C_1$ and $\mathbf{y}_1, \mathbf{y}_2 \in C_2$ and $0 < \alpha, \beta < 1$. Actually, by Exercise 4.2(b)ii, we can always find $\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}_1, \mathbf{y}_2$ such that $\alpha = \beta$. Then $\mathbf{x} + \mathbf{y} = \alpha(\mathbf{x}_1 + \mathbf{y}_1) + (1 - \alpha)(\mathbf{x}_2 + \mathbf{y}_2) \in \mathbf{relint}\,(C_1 + C_2)$, and hence $\mathbf{relint}\,(C_1) + \mathbf{relint}\,(C_2) \subset \mathbf{relint}\,(C_1 + C_2)$.

(c) We assert that $\mathbf{relbd}\ \hat{C} = \mathbf{cl}\ \hat{C} \setminus \mathbf{relint}\ \hat{C} = (\mathbf{cl}\ C + (\mathbf{aff}\ C - \mathbf{x}_0)^\perp) \setminus (\mathbf{relint}\ C + (\mathbf{aff}\ C - \mathbf{x}_0)^\perp) = (\mathbf{cl}\ C \setminus \mathbf{relint}\ C) + (\mathbf{aff}\ C - \mathbf{x}_0)^\perp = \mathbf{relbd}\ C + (\mathbf{aff}\ C - \mathbf{x}_0)^\perp$.

To complete the proof, we need to show:

i. $\mathbf{cl}\,(C_1 + C_2) = \mathbf{cl}\,(C_1) + \mathbf{cl}\,(C_2)$ for any sets C_1 and C_2 satisfying $\mathbf{aff}\,(C_1) \perp \mathbf{aff}\,(C_2)$.

ii. $(C_1 + C) \setminus (C_2 + C) = (C_1 \setminus C_2) + C$ for any sets satisfying $C_1 \supset C_2$ and $C_1 \cap C = \{\mathbf{0}\}$.

For the first statement, let $\mathbf{x} + \mathbf{y} \in \mathbf{cl}\,(C_1 + C_2)$ and $\{\mathbf{x}_k + \mathbf{y}_k\} \subset C_1 + C_2$ converges to $\mathbf{x} + \mathbf{y}$, where $\mathbf{x}, \mathbf{x}_k \in C_1$ and $\mathbf{y}, \mathbf{y}_k \in C_2$. Since $\langle \mathbf{x}_k + \mathbf{y}_k, \mathbf{x}_k \rangle = \langle \mathbf{x}_k, \mathbf{x}_k \rangle \to \langle \mathbf{x}, \mathbf{x}_k \rangle$, it follows that $\mathbf{x}_k \to \mathbf{x}$. Analogously, $\mathbf{y}_k \to \mathbf{y}$. Therefore, $\mathbf{x} \in \mathbf{cl}\ C_1$ and $\mathbf{y} \in \mathbf{cl}\ C_2$, and hence $\mathbf{x} + \mathbf{y} \in \mathbf{cl}\,(C_1) + \mathbf{cl}\,(C_2)$, i.e. $\mathbf{cl}\,(C_1 + C_2) \subset \mathbf{cl}\,(C_1) + \mathbf{cl}\,(C_2)$. The reverse inclusion is obvious.

For the second statement, let $\mathbf{x} \in C_1, \mathbf{y} \in C_2$. If $\forall\, \mathbf{z}_1 \in C, \neg\exists\, \mathbf{z}_2 \in C$, s.t. $\mathbf{x} + \mathbf{z}_1 = \mathbf{y} + \mathbf{z}_2$, then it is clear that $\mathbf{x} \neq \mathbf{y}$. This implies that $(C_1 + C) \setminus (C_2 + C) \subset (C_1 \setminus C_2) + C$. Conversely, if $\exists\, \mathbf{x} \in C_1, \mathbf{y} \in C_2$ and $\mathbf{z}_1, \mathbf{z}_2 \in C$, s.t. $\mathbf{x} + \mathbf{z}_1 = \mathbf{y} + \mathbf{z}_2$, then $\mathbf{x} - \mathbf{y} = \mathbf{z}_2 - \mathbf{z}_1 = \mathbf{0} \in C_1 \cap C$. Therefore, $\mathbf{z}_1 = \mathbf{z}_2, \mathbf{x} = \mathbf{y}$. This implies $(C_1 + C) \cap (C_2 + C) \cap ((C_1 \setminus C_2) + C) = \emptyset$.

■

Exercise 7: Convex Sets and Linear Functions

Let $C \subset \mathbb{R}^n$ be a convex set and $l(\mathbf{x}) = \langle \mathbf{a}, \mathbf{x} \rangle$ be a linear function on $\mathbb{R}^n$. The linear function is nontrivial if $\mathbf{a} \neq \mathbf{0}$. Suppose $\mathbf{x}_0 \in C$ and denote

$$B_C(\mathbf{x}_0, r) = B(\mathbf{x}_0, r) \cap \mathbf{aff}\ C.$$

Please show the following statements.

1. If $l(\mathbf{x}) = \alpha, \forall\, \mathbf{x} \in B_C(\mathbf{x}_0, r)$, then $l(\mathbf{x}) = \alpha, \forall\, \mathbf{x} \in C$.

   Solution:
   Let $m = \dim(\mathbf{aff}\ C - \mathbf{x}_0)$ and $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_m) \in \mathbb{R}^{n\times m}$ be a matrix whose columns form an orthonormal basis of $\mathbf{aff}\ C - \mathbf{x}_0$. Then any $\mathbf{x} \in B_C(\mathbf{x}_0, r)$ can be written as $\mathbf{x}_0 + \mathbf{V}\mathbf{y}$, where $\mathbf{y} \in \mathbb{R}^m$ satisfies $\|\mathbf{y}\|_2 \leq r$. If we let $\mathbf{y} = \pm r\mathbf{e}_i$, $i = 1, \ldots, m$, then $l(\mathbf{x}) = \langle \mathbf{a}, \mathbf{x}_0 \rangle \pm r\langle \mathbf{a}, \mathbf{v}_i \rangle = \alpha \implies \langle \mathbf{a}, \mathbf{x}_0 \rangle = \alpha$ and $\langle \mathbf{a}, \mathbf{v}_i \rangle = 0$. Therefore, for any $\mathbf{x}' = \mathbf{x}_0 + \mathbf{V}\mathbf{y}' \in C$, we have $l(\mathbf{x}') = \langle \mathbf{a}, \mathbf{x}_0 \rangle + \langle \mathbf{a}, \mathbf{V}\mathbf{y}' \rangle = \alpha + 0 = \alpha$. ■

2. The linear function $l(\mathbf{x}) = \alpha, \forall\, \mathbf{x} \in B_C(\mathbf{x}_0, r)$ for some constant α if and only if $\mathbf{a} \perp (\mathbf{aff}\ C - \mathbf{x}_0)$.

   Solution:

   (⇒) According to Question 1, $\langle \mathbf{a}, \mathbf{v}_i \rangle = 0 \implies \mathbf{a} \perp \mathbf{v}_i$, $i = 1, \ldots, m$. Since $\mathcal{C}(\mathbf{V}) = \mathbf{aff}\ C - \mathbf{x}_0$, we have $\mathbf{a} \perp (\mathbf{aff}\ C - \mathbf{x}_0)$.

   (⇐) The fact $\mathbf{a} \perp \mathcal{C}(\mathbf{V})$ implies that for any $\mathbf{x} = \mathbf{x}_0 + \mathbf{V}\mathbf{y} \in B_C(\mathbf{x}_0, r)$, we have $l(\mathbf{x}) = \langle \mathbf{a}, \mathbf{x}_0 \rangle + \langle \mathbf{a}, \mathbf{V}\mathbf{y} \rangle = \langle \mathbf{a}, \mathbf{x}_0 \rangle + 0 = \alpha$, where $\alpha = \langle \mathbf{a}, \mathbf{x}_0 \rangle$ is a constant. ■

3. The linear function $l(\mathbf{x}) = \langle \mathbf{a}, \mathbf{x} \rangle$ is not constant if and only if $\Pi_{(\mathbf{aff}\ C - \mathbf{x}_0)}(\mathbf{a}) \neq \mathbf{0}$, where Π means the projection.

   Solution:
   On the basis of Question 2, we only need to prove that $\Pi_{(\mathbf{aff}\ C - \mathbf{x}_0)}(\mathbf{a}) = \mathbf{0}$ if and only if $\mathbf{a} \perp (\mathbf{aff}\ C - \mathbf{x}_0)$.

   (⇒) $\mathbf{V}^\top \Pi_{(\mathbf{aff}\ C - \mathbf{x}_0)}(\mathbf{a}) = \mathbf{V}^\top \mathbf{V}(\mathbf{V}^\top\mathbf{V})^{-1}\mathbf{V}^\top \mathbf{a} = \mathbf{V}^\top\mathbf{a} = \mathbf{0}$. So $\mathbf{a} \perp (\mathbf{aff}\ C - \mathbf{x}_0)$.

   (⇐) $\mathbf{V}^\top\mathbf{a} = \mathbf{0}$. So $\Pi_{(\mathbf{aff}\ C - \mathbf{x}_0)}(\mathbf{a}) = \mathbf{V}(\mathbf{V}^\top\mathbf{V})^{-1}\mathbf{V}^\top\mathbf{a} = \mathbf{V}(\mathbf{V}^\top\mathbf{V})^{-1}\mathbf{0} = \mathbf{0}$. ■

4. If $\mathbf{relbd}\ C \neq \emptyset$, then there exists a nontrivial linear function l, and a constant α such that $l(\mathbf{x}) \leq \alpha$ for $\forall\, \mathbf{x} \in C$.

   Solution:
   If $m = \dim(\mathbf{aff}\ C - \mathbf{x}_0) < n$, then by Question 2, we can find $\mathbf{a} \in (\mathbf{aff}\ C - \mathbf{x}_0)^\perp$ such that $\mathbf{a} \neq \mathbf{0}$ and $l(\mathbf{x}) = l(\mathbf{x}_0) = \alpha$.

   If $m = n$, then $\mathbf{aff}\ C = \mathbb{R}^n$, $\mathbf{relbd}\ C = \mathbf{bd}\ C$. Suppose $\mathbf{x_1} \in \mathbf{bd}\ C$. By the Supporting Hyperplane Theorem, there exists a hyperplane $H_{(\mathbf{a},\alpha)}$ supporting C at $\mathbf{x_1}$ such that $l(\mathbf{x}) \leq l(\mathbf{x}_1) = \alpha$ for $\forall\, \mathbf{x} \in C$. ■

Exercise 8: Separation Theorems

Let $C_1, C_2, C \subset \mathbb{R}^n$ be convex sets. Please show the following statements.

1. If C_1 is compact, C_2 is closed and $C_1 \cap C_2 = \emptyset$, then C_1 and C_2 can be strongly separated.

   Solution:
   Let $C = C_1 - C_2$, which is a nonempty convex closed set because both C_1, C_2 are nonempty, convex and closed. Since $C_1 \cap C_2 = \emptyset$, we know that $\mathbf{0} \notin C$. By Theorem 3 in Lecture 5, C and $\mathbf{0}$ can strongly separated, i.e. there exists $\mathbf{a} \in \mathbb{R}^n$ and $\alpha > \beta$ such that $C \subset H^{+}_{(\mathbf{a},\alpha)}$ and $\mathbf{0} \in H^{--}_{(\mathbf{a},\beta)} \implies \beta > 0$. So $\mathbf{a}^\top(\mathbf{x}_1 - \mathbf{x}_2) > \beta \implies \mathbf{a}^\top\mathbf{x}_1 > \beta + \mathbf{a}^\top\mathbf{x}_2$ for any $\mathbf{x}_1 \in C_1$ and $\mathbf{x}_2 \in C_2$. Note that C_1 is bounded, which implies that there exists $\alpha' = \inf \mathbf{a}^\top\mathbf{x}_1$ and hence exists $\beta' = \sup \mathbf{a}^\top\mathbf{x}_2$. Then we have $\mathbf{a}^\top\mathbf{x}_1 \geq \alpha' \geq \beta + \beta' > \beta' \geq \mathbf{a}^\top\mathbf{x}_2$, i.e. $C_1 \subset H^{+}_{(\mathbf{a},\alpha')}$ and $C_2 \in H^{-}_{(\mathbf{a},\beta')}$. Therefore, C_1 and C_2 can be strongly separated. ■

2. (Optional) The sets C_1 and C_2 can be properly separated if and only if **relint** $C_1 \cap$ **relint** $C_2 = \emptyset$.

   Solution:
   Let $C = C_1 - C_2$, which is a nonempty convex set. Since **relint** $C =$ **relint** $C_1 -$ **relint** C_2, we have **relint** $C_1 \cap$ **relint** $C_2 = \emptyset$ if and only if $\mathbf{0} \notin$ **relint** C, and hence if and only if C and $\mathbf{0}$ can be properly separated, which follows from the Proper Separation Theorem in Lecture 5. That is, there exists $H_{(\mathbf{a},\alpha)}$ such that

   $$C \subset H^{+}_{(\mathbf{a},\alpha)},\ \mathbf{0} \in H^{-}_{(\mathbf{a},\alpha)};$$
   $$\exists\, \mathbf{x} \in C \cup \{\mathbf{0}\},\ \mathbf{x} \notin H_{(\mathbf{a},\alpha)}.$$

   This is equivalent to the conditions that $H_{(\mathbf{a},\alpha)}$ properly separates C_1, C_2:

   $$\forall\, \mathbf{x}_1 \in C_1, \mathbf{x}_2 \in C_2,\ \mathbf{a}^\top\mathbf{x}_1 \geq \mathbf{a}^\top\mathbf{x}_2;$$
   $$\exists\, \mathbf{x}_1 \in C_1, \mathbf{x}_2 \in C_2,\ \mathbf{a}^\top\mathbf{x}_1 > \mathbf{a}^\top\mathbf{x}_2.$$

   To sum up, C_1 and C_2 is properly separated if and only if **relint** $C_1 \cap$ **relint** $C_2 = \emptyset$. ■

3. If $\dim(\textbf{aff}\ C) = n$ and $\mathbf{x} \in \mathbb{R}^n \setminus C$, then $\mathbf{x}$ and C can be properly separated.

   Solution:
   If $\mathbf{x} \notin$ **cl** C, by the Strict Separation Theorem, there exists a hyperplane which strictly separates $\mathbf{x}$ and **cl** C and thus also properly separates $\mathbf{x}$ and C.

   If $\mathbf{x} \in$ **bd** $C =$ **relbd** C, by the Supporting Hyperplane Theorem, there exists a hyperplane $H_{(\mathbf{a},\alpha)}$ such that $C \subset H^{-}_{(\mathbf{a},\alpha)}$. Note that **int** $C =$ **relint** $C \neq \emptyset$. For any $\mathbf{y} \in$ **int** C and $r > 0$ such that $B(\mathbf{y}, r) \subset C$, we have $\mathbf{a}^\top(\mathbf{y} + r\mathbf{a}) \leq \alpha \implies \mathbf{a}^\top\mathbf{y} < \alpha$, so $y \notin H_{(\mathbf{a},\alpha)}$ and thus $C \not\subset H_{(\mathbf{a},\alpha)}$. Therefore, $\mathbf{x}$ and C can be properly separated by $H_{(\mathbf{a},\alpha)}$. ■

Exercise 9: Farkas' Lemma

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

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

1. Please show that **cone** A is closed and convex.

   Solution:
   Without loss of generality, we assume that $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are linearly independent. Let $\{\mathbf{x}_k\} \subset$ **cone** A be an arbitrary sequence that converges to some $\mathbf{x} = \sum_{i=1}^n \alpha_i \mathbf{a}_i \in$ **span** A, where $\mathbf{x}_k = \sum_{i=1}^n \alpha_{ki} \mathbf{a}_i$ for some $\alpha_{ki} \geq 0$. Since A is a basis of **span** A, it follows that $\alpha_{ki} \to \alpha_i$ as $k \to \infty$, and hence $\alpha_i \geq 0$ for all i. Therefore, $\mathbf{x} \in$ **cone** A and **cone** A is closed.

   If $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are linearly dependent, TBD.

   Second, for any $\mathbf{y} = \sum_{i=1}^n \beta_i \mathbf{a}_i \in$ **cone** A and any $0 \leq \theta \leq 1$, we have $\theta\mathbf{x} + (1-\theta)\mathbf{y} = \sum_{i=1}^n (\theta\alpha_i + (1-\theta)\beta_i)\, \mathbf{a}_i \in$ **cone** A, implying that **cone** A is convex. ■

2. If $\mathbf{b} \in$ **cone** A, please show that there exists $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{A}\mathbf{x} = \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$.

   Solution:
   If $\mathbf{b} \in$ **cone** A, then there exists $x_i \geq 0$ such that $\mathbf{b} = \sum_{i=1}^n x_i \mathbf{a}_i$. Let $\mathbf{x} = (x_1, \cdots, x_n)^\top$. Then $\mathbf{A}\mathbf{x} = \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$. ■

3. If $\mathbf{b} \notin$ **cone** A, use separation theorems to show that there exists $\mathbf{y} \in \mathbb{R}^m$, such that $\mathbf{A}^\top \mathbf{y} \geq \mathbf{0}$ and $\mathbf{b}^\top \mathbf{y} < 0$.

   Solution:
   cone A is a nonempty closed convex set. If $\mathbf{b} \notin$ **cone** A, then **cone** A and $\mathbf{b}$ can be strongly separated; that is, there exists $\mathbf{y} \in \mathbb{R}^n$ and $\alpha > \beta$ such that **cone** $A \subset H^+_{(\mathbf{y},\alpha)}$ and $\mathbf{b} \in H^-_{(\mathbf{y},\beta)}$. Since $\mathbf{0} \in$ **cone A**, we have $\alpha \leq 0 \implies \beta < 0$. Thus $\mathbf{b}^\top \mathbf{y} \leq \beta < 0$. Note that $\lambda \mathbf{a}_i \in$ **cone A** for any $\lambda > 0$. Hence $\mathbf{a}_i^\top \mathbf{y} \geq \lim_{\lambda\to\infty} \frac{\alpha}{\lambda} = 0 \implies \mathbf{A}^\top \mathbf{y} \geq \mathbf{0}$. ■

4. Now you can prove Farkas' Lemma: for given $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$, one and only one of the two statements hold:

   - $\exists \mathbf{x} \in \mathbb{R}^n$, $\mathbf{A}\mathbf{x} = \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$.
   - $\exists \mathbf{y} \in \mathbb{R}^m$, $\mathbf{A}^\top \mathbf{y} \geq \mathbf{0}$ and $\mathbf{b}^\top \mathbf{y} < 0$.

   Solution:
   If $\mathbf{b} \in$ **cone** A, then by Question 2, the first statement holds. If $\mathbf{b} \notin$ **cone** A, then by Question 3, the second statement holds.

   Suppose that both the statements hold. Then there exists $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{A}\mathbf{x} = \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$, and there exists $\mathbf{y} \in \mathbb{R}^m$, $\mathbf{A}^\top \mathbf{y} \geq \mathbf{0}$ and $\mathbf{b}^\top \mathbf{y} < 0$. It follows that $\mathbf{b}^\top \mathbf{y} = \mathbf{x}^\top \mathbf{A}^\top \mathbf{y} = \langle \mathbf{x}, \mathbf{A}^\top \mathbf{y} \rangle \geq 0$, which is a contradiction. ■

Exercise 10: Projection to a Polytope

Hint: you may want to read [1, 2].

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

   (a) The projection operator on C, i.e., Π_C, is continuous and firmly nonexpansive. In other words, for any $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{R}^n$, we have

   $$\left\|\Pi_C\left(\mathbf{w}_1\right)-\Pi_C\left(\mathbf{w}_2\right)\right\|_2^2+\left\|\left(\operatorname{Id}-\Pi_C\right)\left(\mathbf{w}_1\right)-\left(\operatorname{Id}-\Pi_C\right)\left(\mathbf{w}_2\right)\right\|_2^2 \leq\left\|\mathbf{w}_1-\mathbf{w}_2\right\|_2^2,$$

   where Id is the identity operator.

   (b) For a point $\mathbf{w} \in \mathbb{R}^n$, let $\mathbf{w}(t)=\Pi_C(\mathbf{w})+t\left(\mathbf{w}-\Pi_C(\mathbf{w})\right)$. Then, the projection of the point $\mathbf{w}(t)$ is $\Pi_C(\mathbf{w})$ for all $t \geq 0$, i.e.,

   $$\Pi_C(\mathbf{w}(t))=\Pi_C(\mathbf{w}), \forall t \geq 0.$$

   Solution:

   (a) Since $\mathbf{w}_{1,2}=\Pi_C(\mathbf{w}_{1,2})+(\mathbf{w}_{1,2}-\Pi_C(\mathbf{w}_{1,2}))=\Pi_C(\mathbf{w}_{1,2})+(\operatorname{Id}-\Pi_C)\mathbf{w}_{1,2}$, we have

   $$\begin{aligned}\|\mathbf{w}_1-\mathbf{w}_2\|_2^2=&\|\Pi_C(\mathbf{w}_1)-\Pi_C(\mathbf{w}_2)\|_2^2+\|(\operatorname{Id}-\Pi_C)(\mathbf{w}_1)-(\operatorname{Id}-\Pi_C)(\mathbf{w}_2)\|_2^2\\&+2\langle\Pi_C(\mathbf{w}_1)-\Pi_C(\mathbf{w}_2),(\operatorname{Id}-\Pi_C)(\mathbf{w}_1)-(\operatorname{Id}-\Pi_C)(\mathbf{w}_2)\rangle.\end{aligned}$$

   We need to show the last term is non-negative, i.e

   $$\langle\Pi_C(\mathbf{w}_1)-\Pi_C(\mathbf{w}_2),\mathbf{w}_1-\mathbf{w}_2-\Pi_C(\mathbf{w}_1)+\Pi_C(\mathbf{w}_2)\rangle\geq 0 \tag{1}$$

   Since $\Pi_C(\mathbf{w}_1),\Pi_C(\mathbf{w}_2)\in C$, by variational inequality, we have

   $$\langle\Pi_C(\mathbf{w}_1)-\Pi_C(\mathbf{w}_2),\mathbf{w}_2-\Pi_C(\mathbf{w}_2)\rangle\leq 0 \tag{2}$$
   $$\langle\Pi_C(\mathbf{w}_2)-\Pi_C(\mathbf{w}_1),\mathbf{w}_1-\Pi_C(\mathbf{w}_1)\rangle\leq 0 \tag{3}$$

   (3) $-$ (2) yields (1). The nonexpansiveness implies $\lim_{\mathbf{w}_2\to\mathbf{w}_1}\|\Pi_C(\mathbf{w}_2)-\Pi_C(\mathbf{w}_1)\|_2=0$ and hence $\lim_{\mathbf{w}_2\to\mathbf{w}_1}\Pi_C(\mathbf{w}_2)=\Pi_C(\mathbf{w}_1)$, i.e. Π_C is continuous.

   (b) By variational inequality,

   $$\begin{aligned}0&\geq\langle\Pi_C(\mathbf{w})-\Pi_C(\mathbf{w}(t)),\mathbf{w}(t)-\Pi_C(\mathbf{w}(t))\rangle\\&=\|\Pi_C(\mathbf{w})-\Pi_C(\mathbf{w}(t))\|_2^2-t\langle\Pi_C(\mathbf{w}(t))-\Pi_C(\mathbf{w}),\mathbf{w}-\Pi_C(\mathbf{w})\rangle\\&\geq\|\Pi_C(\mathbf{w})-\Pi_C(\mathbf{w}(t))\|_2^2\end{aligned}$$

   Thus $\|\Pi_C(\mathbf{w})-\Pi_C(\mathbf{w}(t))\|_2=0$, i.e. $\Pi_C(\mathbf{w}(t))=\Pi_C(\mathbf{w})$ for all $t\geq 0$. ■

2. Let $\mathbf{y}$ be an N-dimensional vector ,$\{\mathbf{x}_1,\cdots,\mathbf{x}_p\}$ be N-dimension non-zero vectors and $\lambda\geq 0$ is a regularization parameter. Consider the following optimization problem:

   $$\min_\theta\left\{\left\|\theta-\frac{\mathbf{y}}{\lambda}\right\|_2^2:\left|\mathbf{x}_i^\top\theta\right|\leq 1, i=1,2,\ldots,p\right\}. \tag{4}$$

   For notational convenience, we denote the optimal solution of (4) by $\theta^*(\lambda)$.

(a) We let the feasible set of (4) be F. Please give an interpretation of the geometry of F (you don't need to prove it). Then give a close form of the optimal solution $\theta^*(\lambda)$ in the form of projection.

(b) Let $\lambda, \lambda_0 > 0$ be two regularization parameters. Please show that

$$\theta^*(\lambda) \in B\left(\theta^*\left(\lambda_0\right), \left|\frac{1}{\lambda} - \frac{1}{\lambda_0}\right| \|\mathbf{y}\|_2\right).$$

(c) Let $\lambda, \lambda_0 > 0$ be two regularization parameters. Please show that

$$\theta^*(\lambda) \in B\left(\theta^*\left(\lambda_0\right) + \frac{1}{2}\left(\frac{1}{\lambda} - \frac{1}{\lambda_0}\right)\mathbf{y}, \frac{1}{2}\left|\frac{1}{\lambda} - \frac{1}{\lambda_0}\right| \|\mathbf{y}\|_2\right).$$

(You may use the result in Question 1(a).)

(d) Suppose that $\Pi_F\left(\frac{\mathbf{y}}{\lambda_0}\right) \neq \theta^*\left(\lambda_0\right)$. For any $\lambda \in (0, \lambda_0]$, let us define

$$\begin{aligned}
\mathbf{v}_1\left(\lambda_0\right) &= \frac{\mathbf{y}}{\lambda_0} - \theta^*\left(\lambda_0\right), \\
\mathbf{v}_2\left(\lambda, \lambda_0\right) &= \frac{\mathbf{y}}{\lambda} - \theta^*\left(\lambda_0\right), \\
\mathbf{v}_2^{\perp}\left(\lambda, \lambda_0\right) &= \mathbf{v}_2\left(\lambda, \lambda_0\right) - \frac{\left\langle \mathbf{v}_1\left(\lambda_0\right), \mathbf{v}_2\left(\lambda, \lambda_0\right)\right\rangle}{\left\|\mathbf{v}_1\left(\lambda_0\right)\right\|_2^2}\mathbf{v}_1\left(\lambda_0\right).
\end{aligned}$$

Then, the dual optimal solution $\theta^*(\lambda)$ can be estimated as follows:

$$\theta^*(\lambda) \in B\left(\theta^*\left(\lambda_0\right), \left\|\mathbf{v}_2^{\perp}\left(\lambda, \lambda_0\right)\right\|_2\right) \subseteq B\left(\theta^*\left(\lambda_0\right), \left|\frac{1}{\lambda} - \frac{1}{\lambda_0}\right| \|\mathbf{y}\|_2\right)$$

(You may use the result in Question 1(b).)

Solution:

(a) Define $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_n)$. Then $\mathbf{X}^\top\theta$ is a point in the unit ∞-norm ball; that is, each dimension of $\mathbf{X}^\top\theta$ is bounded within $[-1, 1]$. As F is a nonempty closed convex set, the optimal solution $\theta^*(\lambda) = \Pi_F(\frac{\mathbf{y}}{\lambda})$.

(b) By nonexpansiveness of Π_F, we have

$$\|\theta^*(\lambda) - \theta^*(\lambda_0)\|_2 = \left\|\Pi_F(\frac{\mathbf{y}}{\lambda}) - \Pi_F(\frac{\mathbf{y}}{\lambda_0})\right\|_2 \leq \left\|\frac{\mathbf{y}}{\lambda} - \frac{\mathbf{y}}{\lambda_0}\right\|_2 = \left|\frac{1}{\lambda} - \frac{1}{\lambda_0}\right| \|\mathbf{y}\|_2.$$

(c) By Question 1(a), we have

$$
\begin{aligned}
&\left\|\theta^*(\lambda)-\theta^*(\lambda_0)-\frac{1}{2}\left(\frac{1}{\lambda}-\frac{1}{\lambda_0}\right)\mathbf{y}\right\|_2\\
=&\left\|\Pi_F(\frac{\mathbf{y}}{\lambda})-\Pi_F(\frac{\mathbf{y}}{\lambda_0})-\frac{1}{2}\left(\frac{1}{\lambda}-\frac{1}{\lambda_0}\right)\mathbf{y}\right\|_2\\
=&\frac{1}{2}\left\|\Pi_F(\frac{\mathbf{y}}{\lambda})-\Pi_F(\frac{\mathbf{y}}{\lambda_0})-(\mathrm{Id}-\Pi_F)(\frac{\mathbf{y}}{\lambda})+(\mathrm{Id}-\Pi_F)(\frac{\mathbf{y}}{\lambda_0})\right\|_2\\
\le&\frac{1}{2}\sqrt{\left\|\Pi_F(\frac{\mathbf{y}}{\lambda})-\Pi_F(\frac{\mathbf{y}}{\lambda_0})\right\|_2^2+\left\|(\mathrm{Id}-\Pi_F)(\frac{\mathbf{y}}{\lambda})-(\mathrm{Id}-\Pi_F)(\frac{\mathbf{y}}{\lambda_0})\right\|_2^2}\\
\le&\frac{1}{2}\sqrt{\left\|\frac{\mathbf{y}}{\lambda}-\frac{\mathbf{y}}{\lambda_0}\right\|_2^2}=\frac{1}{2}\left|\frac{1}{\lambda}-\frac{1}{\lambda_0}\right|\|\mathbf{y}\|_2\,,
\end{aligned}
$$

where the first inequality holds as $\left\langle\Pi_F(\frac{\mathbf{y}}{\lambda})-\Pi_F(\frac{\mathbf{y}}{\lambda_0}),(\mathrm{Id}-\Pi_F)(\frac{\mathbf{y}}{\lambda})-(\mathrm{Id}-\Pi_F)(\frac{\mathbf{y}}{\lambda_0})\right\rangle\ge 0$.

(d) Note that $\langle\mathbf{v}_2^\perp(\lambda,\lambda_0),\mathbf{v}_1(\lambda_0)\rangle=0$, so

$$
\begin{aligned}
\|\mathbf{v}_2^\perp(\lambda,\lambda_0)\|_2=&\left\|\frac{\mathbf{y}}{\lambda}-\frac{\mathbf{y}}{\lambda_0}+\left(1-\frac{\langle\mathbf{v}_1(\lambda_0),\mathbf{v}_2(\lambda,\lambda_0)\rangle}{\|\mathbf{v}_1(\lambda_0)\|_2^2}\right)\mathbf{v}_1(\lambda_0)\right\|\\
\le&\left\|\frac{\mathbf{y}}{\lambda}-\frac{\mathbf{y}}{\lambda_0}\right\|_2=\left|\frac{1}{\lambda}-\frac{1}{\lambda_0}\right|\|\mathbf{y}\|_2.
\end{aligned}
$$

By Question 1(b), we know that $\Pi_F\left(\theta^*(\lambda_0)+\frac{\langle\mathbf{v}_1(\lambda_0),\mathbf{v}_2(\lambda,\lambda_0)\rangle}{\|\mathbf{v}_1(\lambda_0)\|_2^2}\mathbf{v}_1(\lambda_0)\right)=\theta^*(\lambda_0)$. Then by nonexpansiveness, we have

$$
\begin{aligned}
\|\theta^*(\lambda)-\theta^*(\lambda_0)\|=&\left\|\Pi_F(\frac{\mathbf{y}}{\lambda})-\Pi_F\left(\theta^*(\lambda_0)+\frac{\langle\mathbf{v}_1(\lambda_0),\mathbf{v}_2(\lambda,\lambda_0)\rangle}{\|\mathbf{v}_1(\lambda_0)\|_2^2}\mathbf{v}_1(\lambda_0)\right)\right\|\\
\le&\left\|\frac{\mathbf{y}}{\lambda}-\theta^*(\lambda_0)-\frac{\langle\mathbf{v}_1(\lambda_0),\mathbf{v}_2(\lambda,\lambda_0)\rangle}{\|\mathbf{v}_1(\lambda_0)\|_2^2}\mathbf{v}_1(\lambda_0)\right\|\\
=&\|\mathbf{v}_2^\perp(\lambda,\lambda_0)\|_2
\end{aligned}
$$

■

References

[1] T. Hastie, R. Tibshirani, and M. Wainwright. Statistical learning with sparsity. *Monographs on statistics and applied probability*, 143:143, 2015.

[2] J. Wang, P. Wonka, and J. Ye. Lasso screening rules via dual polytope projection. *Journal of Machine Learning Research*, 16(32):1063–1101, 2015.