Chapter I
Background

March 11, 2023

1 Euclidean Spaces

1.1.1

Prove the intersection of an arbitrary collection of convex sets is convex. Deduce that the convex hull of a set $D \subseteq \mathbb{E}$ is well-defined as the intersection of all convex sets containing D.

Proof: Let C_i, $i \in \mathcal{I}$ be a collection of convex sets. Then for all $x, y \in \cap_{i \in \mathcal{I}} C_i$, and all $\lambda \in [0,1]$, we have

$$\lambda x + (1-\lambda) y \in C_i \quad \forall i \in \mathcal{I}, \text{ since } C_i \text{ is convex and } x, y \in C_i.$$

Thus $\cap_{i \in \mathcal{I}} C_i$ is convex. The rest is clear.

1.1.2

- Prove that if the set $C \subseteq \mathbb{E}$ is convex and if

$$x^1, \cdots, x^m \in C, 0 \le \lambda_!, \cdots, \lambda_m \in \mathbb{R},$$

and $\sum \lambda_i = 1$ then $\sum \lambda_i x^i \in C$. Prove, furthermore, that if $f : C \to \mathbb{R}$ is a convex function then $f(\sum \lambda_i x^i) \le \sum \lambda_i f(x^i)$.

- We know that $-\log$ is convex. Deduce, for any strictly positive reals $x^1, \cdots, x^m$, and any nonnegative reals $\lambda_1, \cdots, \lambda_m$ with sum 1, the *arithmetic-geometric* mean inequality

$$\prod_i (x^i)^{\lambda_i} \le \sum_i \lambda_i x^i.$$

- Prove that for any set $D \subseteq \mathbb{E}$, *conv*D is the set of all convex combinations of elements of D.

Proof:

- Obvious induction.
- Since $-\log$ is convex, we have

$$-\log(\sum_i \lambda_i x^i) \le \sum \lambda_i(-\log(x^i)) \Rightarrow \sum \lambda_i(\log(x^i)) \le \log(\sum_i \lambda_i x^i).$$

So,

$$\log(\prod_i (x^i)^{\lambda_i}) \le \log(\sum_i \lambda_i x^i) \Rightarrow \prod_i (x^i)^{\lambda_i} \le \sum_i \lambda_i x^i..$$

- Easy.

1.1.3

Prove that a convex set $D \subseteq \mathbb{E}$ has convex closure, and deduce that $cl(\operatorname{conv} D)$ is the smallest closed convex set containing D.

Proof:

Let $x, y \in cl(D)$ and suppose $x_i \to x$ and $y_i \to y$ with $x_i, y_i \in D$. Then for $\lambda \in [0,1]$,

$$\lambda x + (1-\lambda) y = \lim_i \lambda x_i + (1-\lambda) y_i.$$

Thus $\lambda x + (1-\lambda)y$ belongs to the closure of D, and thus $cl(D)$ is convex.

Now for $D \subseteq \mathbb{E}$, if C is the smallest closed convex set containing it, then $cl(conv(D)) \subseteq C$ as first C is convex and contains D and so contains $conv(D)$, also C is closed and hence contains $cl(conv(D))$. On the other hand, since $conv(D)$ is convex, $cl(conv(D))$ is also convex. However, C is the smallest closed convex set containing D, and therefore $C = cl(conv(D))$.

1.1.4. Randstorm cancellation
Suppose sets $A, B, C \subseteq \mathbb{E}$ satisfy

$$A + C \subseteq B + C.$$

If A, B are convex, B is closed, and C is bounded, prove

$$A \subseteq B.$$

Show this result can fail if B is not convex.

Proof: Since A is convex we have $2A = A + A$. In fact, $A \subseteq \frac{1}{2}(A + A)$ as $a = \frac{1}{2}(a + a)$. On the other hand, for $a, b \in A$, $\frac{1}{2}(a + b) \in A$, by definition of convexity. Similarly, $2B = B + B$. Thus, we have

$$2A + C = A + A + C = A + (A + C) \subseteq A + (B + C) = (A + C) + B \subseteq (B + C) + B = 2B + C$$

.

By induction,

$$nA + C \subseteq nB + C \;\forall n \in \mathbb{N}.$$

Now, suppose $a \in A$. Then there exist $b_n \in B$ and $c_n \in C$ such that $na = nb_n + c_n$. Since, C is bounded, we can assume there exists a subsequence c_{n_k} of c_n such that c_{n_k} converges. Now since $a = b_{n_k} + \frac{1}{n_k} c_{n_k}$. Since, c_{n_k} is convergent and so bounded, we deduce, $\lim_k \frac{1}{n_k} c_{n_k} = 0$. So, $a = \lim_k b_{n_k}$. But, B is closed and so $\lim_k b_{n_k}$, if it exists, belongs to B. Hence, $a \in B$ and so $A \subseteq B$.

Now let $A = \{\frac{1}{2}\}, B = \{0, 1\}, C = [0, 1]$. Then $A + C = [\frac{1}{2}, \frac{3}{2}], B + C = [0, 2]$. So, $A + C \subseteq B + C$ and also $A \not\subseteq B$.

1.1.5 Strong separation
Suppose that the set $C \subseteq \mathbb{E}$ is closed and convex, and that the set $D \subseteq \mathbb{E}$ is compact and convex.

- Prove the set $D - C$ is closed and convex.
- Deduce that if in addition D and C are disjoint then there exists a nonzero element a in $\mathbb{E}$ with $\inf_{x \in D} \langle a, x\rangle > \sup_{y \in C} \langle a, y\rangle$. Interpret geometrically.
- Show part (b) fails for the closed convex sets in $\mathbb{R}^2$,

$$D = \{x : x_1 > 0, x_1 x_2 \geq 1\}$$
$$C = \{x : x_2 = 0\}.$$

Proof:

- Note that for $d_1, d_2 \in D$ and $c_1, c_2 \in C$, and $\lambda \in [0, 1]$,

  $$\lambda(d_1 - c_1) + (1 - \lambda)(d_2 - c_2) = (\lambda d_1 + (1 - \lambda)d_2) - (\lambda c_1 + (1 - \lambda)c_2) \in D - C.$$

  Thus, $D - C$ is convex.

  Now, let $d_i \in D$ and $c_i \in C$ such that $d_i - c_i \to x$. We wish to prove that $x \in D - C$. Since D is compact, we may assume $d_i \to d \in D$. So c_i converges to some $c \in \mathbb{E}$. Now since C is closed we have $c \in C$. Thus $x = d - c$ belongs to $D - C$.

- Since $D \cap C \neq \emptyset$, we have $0 \notin D - C$ and so, due to the basic separation, there exists $a \in \mathbb{E}$ such that $\langle a, x\rangle > b > 0$ for all $x \in D - C$ and some fixed $b > 0$. So $\inf_{x \in D} \langle a, x\rangle > \sup_{y \in C} \langle a, y\rangle$.

  Geometrically, it means two disjoint closed, convex sets one of which is compact, can be separated via a hyperplane.

- Let $a \in \mathbb{R}^2$, such that $\inf_{x\in D}\langle a, x\rangle > \sup_{y\in C}\langle a, y\rangle$.. Then $\sup_{y\in C}\langle a, y\rangle = \sup_{y\in C} a_1y_1$, which since it is finite, must be equal to zero and thus $a_1 = 0$. Now,

$$\inf_{x\in D}\langle a, x\rangle = a_2x_2 > 0,$$

which is a contradiction as $x_2 \to 0$ implies $a_2x_2 \to 0$.

1.1.6. Recession cones

Consider a nonempty closed convex set $C \subseteq \mathbb{E}$. We define the *recession cone* of C by

$$0^+(C) = \{d \in \mathbb{E} : C + \mathbb{R}_+d \subseteq C\}.$$

- Prove $0^+(C)$ is a closed convex cone.
- Prove $d \in 0^+(C)$ if and only if $x + \mathbb{R}_+d \subseteq C$ for some point x in C. Show this equivalence can fail if C is not closed.
- Consider a family of closed convex sets C_γ ($\gamma \in \Gamma$) with nonempty intersection. Prove $0^+(\cap C_\gamma) = \cap 0^+(C_\gamma)$.
- For a unit vector u in $\mathbb{E}$, prove $u \in 0^+(C)$ if and only if there is a sequence x^r in C satisfying $||x^r|| \to \infty$ and $||x^r||^{-1}x^r \to u$. Deduce C is unbounded if and only if $0^+(C)$ is nontrivial.
- If Y is a Euclidean space, the map $A : \mathbb{E} \to Y$ is linear, and $N(A) \cap 0^+(C)$ is a linear subspace, prove AC is closed. Show this result can fail without the last assumption.
- Consider another nonempty closed convex set $D \subseteq \mathbb{E}$ such that $0^+(C) \cap 0^+(D)$ is a linear subspace. Prove $C - D$ is closed.

Proof:

- Let $d_1, d_2 \in 0^+(C)$ and $\lambda > 0$, then $C + \mathbb{R}_+(\lambda d_1) = C + \mathbb{R}_+d_1 \subseteq C$. Also, $C + \mathbb{R}_+(d_1 + d_2) \subseteq C + \mathbb{R}_+d_1 + \mathbb{R}_+d_2 \subseteq C + \mathbb{R}_+d_2 \subseteq C$.
- Let $C_\infty(x) = \{d \in \mathbb{E} : x + td \in C, \forall t > 0\}$. Now let $d \in \mathbb{C}_\infty(x)$ and also fix $\bar{y} \in C$. We wish to show that $\bar{y} + d \in C$. Since $d \in C_\infty(x)$ for every $\bar{t} > 0$ we have $x + \bar{t}d \in C$. Thus, for $\lambda > 0$ we have

$$\bar{y}_\lambda = \lambda\bar{y} + (1-\lambda)(x + \frac{1}{1-\lambda}d) \in C,$$

  as C is convex. But $\bar{y}_\lambda = \lambda\bar{y} + (1-\lambda)x + d \in C$. Clearly, $\lim_{\lambda\to 1^-}\bar{y}_\lambda = \bar{y} + d$ and since C is closed we conclude that $\bar{y} + d \in C$. Thus, $C_\infty(x) \subseteq 0^+(C)$.

  Conversely, let $d \in 0^+(C)$. Then, by definition, $x \in C_\infty(x)$.

  Example: Take $C = \{(x, y) : y > 0\} \cup \{(0, 0)\}$.

- Let $x \in \cap C_\gamma$. Then $d \in 0^+(\cap C_\gamma)$ if and only if $x + \mathbb{R}_+d \subseteq \cap C_\gamma$, and this holds, if and only if $x + \mathbb{R}_+d \subseteq C_\gamma$ for all $\gamma \in \Gamma$ or equivalently $d \in 0^+(C_\gamma)$ for all $\gamma \in \Gamma$.
- Let $x \in C$ and $u \in 0^+(C)$. Then $x^r := x + ru \in C$ for $r \in \mathbb{N}$. Note that $\langle x^r, u\rangle = \langle x, u\rangle + r$ and thus $||x^r|| \to +\infty$. We have

$$\lim_r \frac{x^r}{||x^r||} = \lim_r \frac{x^r||x^r||}{||x^r||^2} = \lim_r \frac{x^r||x^r||}{r^2 + 2r\langle x, u\rangle + ||x||^2} = \lim_r \frac{x^r||x^r||}{r^2} = \lim_r (x/r + u)||x/r + u|| = u.$$

  Conversely, suppose $u^r := ||x^r||^{-1}x^r \to u$ for some $||x^r|| \to +\infty$. Now fix $t \geq 0$,

$$x + tu = x + t\lim_r ||x^r||^{-1}x^r = \lim_r (1 - t||x^r||^{-1})x + \lim_r t||x^r||^{-1}x^r = \lim_r [(1 - t||x^r||^{-1})x + t||x^r||^{-1}x^r].$$

  But, $(1 - t||x^r||^{-1})x + t||x^r||^{-1}x^r \in C$ and thus the above limit lies in C as C is closed.

  Now if C is bounded then there is no such sequence x^r in C and hence $0^+(C)$. Now suppose that C is unbounded and thus there is a sequence $x^r \in C$ with $||x^r|| \to +\infty$. By passing to a subsequence, we can suppose $||x^r||^{-1}x^r \to u$ for some $||u|| = 1$.

- Define $L = N(A) \cap 0^+(C)$. Let $c_i \in C$ and $y_i := Ac_i \to y$. Note that if $||c_i||$ is bounded then, passing to a subsequence, we can assume $c_i \to c$ and since C is closed we have $c \in C$. Hence, $Ac_i \to Ac \in AC$. Thus, suppose that $||c_i|| \to +\infty$ and also $||c_i||^{-1}c_i \to u$ for some $||u|| = 1$. Then, according to the above part, $u \in 0^+(C)$. Now note that

$$Au = \lim_i ||c_i||^{-1} Ac_i = 0 \text{ as } Ac_i \text{ is bounded and } ||c_i|| \to +\infty.$$

  Thus, $u \in N(A) \cap 0^+(C) = L$. Hence, if $L = \{0\}$, AC will be closed.

  Note that $C + L \subseteq C + 0^+(C) \subseteq C$. Define $\tilde{C} := C \cap L^\perp$. Then $\tilde{C}$ is nonempty as for $c \in C$, write $c = c' + c''$ wherein $c' \in L, c'' \in L^\perp$. Then, $c'' = c - c' \in C + L \subseteq C$. Hence, $c'' \in C \cap L^\perp \neq \emptyset$. Note that we also proved that $C \subseteq \tilde{C} + L$. Note that also $C = \tilde{C} + L$ as in fact, $\tilde{C} + L \subseteq C + L \subseteq C$, thus $C = \tilde{C} + L$. However, $AC = A\tilde{C} + AL = A\tilde{C}$.

  Further, $\tilde{C}$ is closed and convex as it is the intersection of two closed convex sets in $\mathbb{E}$. Note that $0^+(\tilde{C}) \subseteq 0^+(C)$ as $\tilde{C} \cap C = \tilde{C}$ and thus $0^+(\tilde{C}) = 0^+(\tilde{C}) \cap 0^+(C)$ and so $0^+(\tilde{C}) \subseteq 0^+(C)$.

  Now, we claim that $0^+(\tilde{C}) \cap N(A) = \{0\}$. In fact, let $d \in 0^+(\tilde{C}) \cap N(A)$, then $d \in 0^+(\tilde{C}) \cap N(A) \subseteq 0^+(C) \cap N(A) = L$. Let $c \in \tilde{C}$, then $c + d \in \tilde{C} \subseteq L^\perp$ and so $d \in L^\perp - c \subseteq L^\perp - L^\perp$. Hence, $d \in L^\perp$ and therefore $d \in L \cap L^\perp = \{0\}$. Now according to the above discussion $A\tilde{C}$ is closed. However, we have $AC = A\tilde{C}$.

- Let $A : \mathbb{E} \times \mathbb{E} \to \mathbb{E}$ with $A(x, y) = x - y$. Then $A(C \times D) = C - D$. However, $N(A) = \{(x, x) : x \in \mathbb{E}\}$ and since $0^+(C \times D) = 0^+(C) \times 0^+(D)$, we have $N(A) \cap 0^+(C \times D) = 0^+(C) \cap 0^+(D) = \{0\}$. Thus, based on the previous part, $A(C \times D) = C - D$ is closed.

1.1.7

For any set of vectors $a^1, \cdots, a^m$ in $\mathbb{E}$, prove the function $f(x) = \max_i \langle a^i, x\rangle$ is convex on $\mathbb{E}$.

Proof: We prove if $f : \mathbb{E} \to (-\infty, +\infty]$ are convex then $f(x) = \max f_i(x)$ is convex. Then for $x, y \in \mathbb{E}$ and $\lambda \in [0, 1]$ we have

$$f_i(\lambda x + (1 - \lambda)y) \leq \lambda f_i(x) + (1 - \lambda)f_i(y) \leq \lambda f(x) + (1 - \lambda)f(y)\ \forall i.$$

Thus $f(x) = \max_i f_i(x) \leq \lambda f(x) + (1 - \lambda)f(y)$.

Note that $f_i(x) = \langle a^i, x\rangle$ is obviously linear.

1.1.8

Prove the Weiestrass theorem: Suppose that the set $D \subseteq \mathbb{E}$ is nonempty and closed and that all the level sets of the continuous function $f : D \to \mathbb{R}$ are bounded. Then f has a global minimizer.

Proof:

Let $\alpha \in \mathbb{R}$ such that $\{x \in D : f(x) \leq \alpha\}$ is nonempty. Then there exists $r > 0$ such that $\{x \in D : f(x) \leq \alpha\} \subseteq B_r$. However, $\{x \in D : f(x) \leq \alpha\} = D \cap f^{-1}(-\infty, \alpha]$ is closed and thus it is compact as well. Now if f is unbounded below then there exists $x_i \in D$ such that $f(x_i) \to -\infty$. Then $x_i \in \{x \in D : f(x) \leq \alpha\} = D \cap f^{-1}(-\infty, \alpha]$ and thus $x_i \to x$ and hence $f(x_i) \to f(x)$ and so $f(x) = \infty$. which is a contradiction. So f is bounded below and so if $c = \inf_{x \in \mathbb{E}}$ then there exists $f(x_i) \to c$. Let $x_i \to x$ and hence $f(x) = c$. x is a global minimizer.

1.1.10. Convex growth conditions

- Find a function with bounded level sets which does not satisfy the growth condition:

$$\liminf_{||x|| \to +\infty} \frac{f(x)}{||x||} > 0.$$

- Prove that any function satisfying the above condition has bounded level sets.

- Suppose the convex function $f : C \to \mathbb{R}$ has bounded level sets but the growth condition fails. Deduce the existence of a sequence (x^m) in C with $f(x^m) \leq ||x^m||/m \to +\infty$. For a fixed point $\bar{x}$ in C, derive a contradiction by considering the sequence

$$\bar{x} + \frac{m}{||x^m||}(x^m - \bar{x}).$$

Hence, prove that for a convex function f, it has bounded level sets if and only if it satisfies the growth condition.

Proof:

- Let $f(x) = x^3$ for $f : \mathbb{R} \to \mathbb{R}$.
- Assume that f satisfies the growth condition and does not have bounded level sets. Then there exists $x_1, x_2, \cdots$ such that $f(x_i) \leq M$ for some $M > 0$ and $||x_i|| > i$. But then,

$$0 < \liminf_{||x|| \to +\infty} \frac{f(x)}{||x||} \leq \lim_i \frac{f(x_i)}{||x_i||} \leq 0.$$

- If f does not satisfy the growth condition then

$$\liminf_{||x|| \to +\infty} \frac{f(x)}{||x||} \leq 0,$$

and hence there exists x^m with $||x^m|| \geq m^2$ such that $f(x^m) \leq ||x^m||/m$. Hence, $f(x^m) \leq ||x^m||/m \to +\infty$. We have

$$f(\bar{x} + \frac{m}{||x^m||}(x^m - \bar{x})) \leq \frac{m}{||x^m||} f(x^m) + (1 - \frac{m}{||x^m||}) f(\bar{x}) \leq 1 + (1 - \frac{m}{||x^m||}) f(\bar{x}) \leq 1 + |f(\bar{x})|.$$

However, $\bar{x} + \frac{m}{||x^m||}(x^m - \bar{x})$ is not bounded as in fact $||\frac{m}{||x^m||}(x^m - \bar{x})|| = m$. Hence, we have proved that if f has bounded level sets then f satisfies the growth condition. We proved the opposite for general functions.

---

2 Symmetric Matrices

1.2.1

Prove $\mathbb{S}^n_+$ is a closed convex cone with interior $\mathbb{S}^n_{++}$.

Proof: It is clear that $\mathbb{S}^n_+$ is convex and a cone. However, let $X \notin \mathbb{S}^n_+$ then $\lambda_{\min}(X) < 0$ with $x^T X x \leq -\delta$ for some $||x|| = 1$. Then let $A \in \mathbb{S}^n$ with $||A|| \leq \frac{1}{2}\delta$ then

$$x^T(X + A)x \leq x^T X x + ||x|| ||Ax|| = x^T X x + ||Ax|| \leq x^T X x + ||A|| \leq -\delta + \frac{1}{2}\delta < 0.$$

Thus, $A + X$ can't be positive semidefinite.

On the other hand, let $X \in \mathbb{S}^n_{++}$, then $S^n \to \mathbb{R}$ defined by $x \mapsto x^T X x$ attains its minimum and hence there exists $\delta > 0$ such that $x^T X x \geq \delta$.

Now let $A \in \mathbb{S}^n$ such that $||A|| \leq \frac{1}{2}\delta$ then for all x with $||x|| = 1$ we have

$$x^T(X + A)x \leq \delta - ||x|| ||A|| ||x|| \geq \delta - \frac{1}{2}\delta \geq \frac{1}{2}\delta.$$

Hence, $X + A \succ 0$.

1.2.1

Explain why $\mathbb{S}^2_+$ is not a polyhedron.

Proof: Suppose that $\mathbb{S}^2_+ = \{x \in \mathbb{R}^3 : Ax \geq b\}$. Then since $\mathbb{S}^2_+$ is a cone we have $b = 0$. Let $A = [A_{11} : A_{12} : A_{21} : A_{22}]$, then $A_{11}, A_{22} \geq 0$. But $A_{12} + A_{21}$ is not less than 0 as $e_{12} + e_{21}$ does not belong to $\mathbb{S}^2_+$.

1.2.4. A nonlattice ordering

Suppose the matrix Z in $\mathbb{S}^2$ satisfies

$$W \succeq \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \text{ and } W \succeq \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \iff W \succeq Z.$$

- By considering diagonal W, prove
$$Z = \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix}$$
for some real a.
- By considering $W = I$, prove $Z = I$.
- Derive a contradiction by considering
$$W = \frac{2}{3}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Proof:

- Let
$$W = \begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix}.$$
Then $x, y \geq 1$ if and only if $W \succeq Z$. Thus, $W \succeq I$ if and only if $W \succeq Z$. Hence, $I \succeq Z$ as well as $Z \succeq I$. Hence, $Z = I$.
- However,
$$W = \frac{2}{3}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \succeq \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, W = \frac{2}{3}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \succeq \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$
Hence, $W \succeq I$ which is incorrect.

1.2.5 Order preservation

- Prove any matrix X in $\mathbb{S}^n$ satisfies $(X^2)^{\frac{1}{2}} \succeq X$.
- Find matrices $X \succeq Y$ in $\mathbb{S}^2_+$ such that $X^2 \not\succeq Y^2$.
- For matrices $X \succeq Y$ in $\mathbb{S}^n_+$, prove $X^{\frac{1}{2}} \succeq Y^{\frac{1}{2}}$.

Proof:

- Let $X = \sum_i \lambda_i u_i u_i^T$ with u_i form an orthogonal basis for $\mathbb{R}^n$. $X^2 = \sum_i \lambda_i^2 u_i u_i^T$, and hence $(X^2)^{\frac{1}{2}} = \sum_i |\lambda_i| u_i u_i^T$. However, since $|\lambda_i| \geq \lambda_i$, we have $\sum_i |\lambda_i| u_i u_i^T \geq \sum_i \lambda_i u_i u_i^T$.
- Let
$$X = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \succeq \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = Y.$$
However,
$$X^2 = \begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix} \succeq \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = Y^2.$$

- Let v be an eigenvector of $X^{\frac{1}{2}} - Y^{\frac{1}{2}}$ with $(X^{\frac{1}{2}} - Y^{\frac{1}{2}})v = \lambda v$. Then
$$\langle (X^{\frac{1}{2}} + Y^{\frac{1}{2}})v, (X^{\frac{1}{2}} - Y^{\frac{1}{2}})v\rangle = \langle (X-Y)v, v\rangle \geq 0.$$
However, $\langle (X^{\frac{1}{2}} + Y^{\frac{1}{2}})v, (X^{\frac{1}{2}} - Y^{\frac{1}{2}})v\rangle = \lambda v^T(X^{\frac{1}{2}} + Y^{\frac{1}{2}})v$. If $(X^{\frac{1}{2}} + Y^{\frac{1}{2}})v = 0$ then $2X^{\frac{1}{2}}v = (X^{\frac{1}{2}} + Y^{\frac{1}{2}})v + (X^{\frac{1}{2}} - Y^{\frac{1}{2}})v = \lambda v$. Hence, $\lambda \geq 0$. Thus, $v^T(X^{\frac{1}{2}} + Y^{\frac{1}{2}})v > 0$.

1.2.6. Square-root iteration
Suppose a matrix A in $\mathbb{S}^n_+$ satisfies $I \succeq A \succeq 0$. Prove that the iteration
$$Y_0 = 0, Y_{n+1} = \frac{1}{2}(A + Y_n^2) \ (n = 0, 1, \cdots)$$
is nondecreasing and converges to the matrix $I - (I - A)^{\frac{1}{2}}$.

Proof: Note that A, Y_0 commute and Y_1 is a polynomial in A and so so forth. Hence, there exists $Q \in O(n)$ such that $Q^T Y_i Q = D_i$ is a diagonal matrix for $i = 0, 1, \cdots, T+1$. Now the below argument complete the proof.

Consider $x_0 = 0$ and $x_{n+1} = \frac{1}{2}(a + x_n^2)$ with $1 \geq a \geq 0$. First, note that $0 \leq x_n \leq 1$ for all n obviously, simple induction.

Then $x_{n+1} - x_n = \frac{1}{2}(a + x_n^2 - 2x_n) = \frac{1}{2}(a + (x_n - 1)^2 - 1)$. Thus, $x_{n+1} \geq x_n$ if and only if $(1 - x_n)^2 \geq 1 - a$ which holds if and only if $(1 - x_n) \geq \sqrt{1-a}$. So assume $1 - \sqrt{1-a} \geq x_t \geq 0$, them $x_{t+1} \leq \frac{1}{2}(a + (1 - \sqrt{1-a})^2) = \frac{1}{2}(a + 1 + 1 - a - 2\sqrt{1-a}) = 1 - \sqrt{1-a}$.

Now let $Y_n \to Y$, then $2Y = A + Y^2$. Hence, $I - A = (Y - I)^2$ and thus $\sqrt{I - A} = I - Y$.

1.2.14 Level sets of perturbed log barriers

- For δ in $\mathbb{R}_{++}$, prove the function
$$t \in \mathbb{R}^n_{++} \to \delta t - \log t$$
has compact level sets.

- For c in R^n_{++}, prove the function
$$x \in \mathbb{R}^n_{++} \mapsto c^T x - \sum_{i=1}^n \log x_i$$
has compact level sets.

- For C in S^n_{++}, prove the function
$$X \in \mathbb{S}^n_{++} \mapsto \langle C, X\rangle - \log\det X$$
has compact level sets.

Proof:

- Since $\delta t - \log t = \delta t - \log \delta t + \log \delta$, without loss of generality suppose, $\delta = 1$. We need to show $\{t \in \mathbb{R}_{++} : t - \log t \leq c\}$ is bounded. If not, $\exists t_n \to +\infty$, s.t. $t_n - \log t_n \leq c$ for some constant c. $\frac{t_n}{\log t_n} - 1 \leq \frac{c}{\log t_n}$. However, $\lim_{t \to +\infty} \frac{t}{\log t} = +\infty$. This contradiction completes the proof.

- Note that $\sum c_i x_i - \sum \log x_i = \sum c_i x_i - \sum \log c_i x_i + \sum \log c_i$, without loss of generality, suppose that $c_i = 1$. But $t - \log t \geq 0$ for all $t > 0$. In fact, for $0 \leq t \leq 1$, $t - \log t \geq t > 0$. Also for $t \geq 1$, $(t -] \log t)' = 1 - \frac{1}{t} \leq 0$ and thus $t - \log t$ is nondecreasing on $t \geq 1$. Note that $t - \log t|_{t=1} = 0$. Now, $t_1 = \log t_1 \leq \sum t_i - \log t_i \leq c$. Thus, to the previous part, t_1 is bounded above. This completes the proof.

- Let $\mu(C)_i = \lambda(C)_{n+1-i}$, then $\mu(C)^T \lambda(X) \leq \langle C, X\rangle$. Thus, $\mu(C)^T\lambda(X) - \sum \log \lambda_i(X) \leq c$.Hence, $\lambda(X)$ is upperbounded and so is $||X|| = \sqrt{\lambda_i(X)^{\frac{1}{2}}}$.

Chapter II
Inequality Constraints

March 11, 2023

1 Optimality Conditions

2.1.1

Prove the normal cone is a closed convex cone.

Proof: Let $C \subseteq \mathbb{E}$ be a convex set and $\bar{x} \in C$. Then

$$N_C(\bar{x}) = \{d \in \mathbb{E} : \langle d, x - \bar{x} \rangle \leq 0\}.$$

If $d \in N_C(\bar{x})$, then obviously $\alpha d \in N_C(\bar{x})$ for $\alpha \geq 0$. Also, if $d_i \in N_C(\bar{x})$ and $d_i \to d \in \mathbb{E}$, then for a fixed $x \in C$,

$$\langle d_i, x - \bar{x} \rangle \to \langle d, x - \bar{x} \rangle,$$

and since $\langle d_i, x - \bar{x} \rangle \leq 0$, $\langle d, x - \bar{x} \rangle \leq 0$. Thus $d \in N_C(\bar{x})$ and so $N_C(\bar{x})$ is closed. Convexity is also clear.

2.1.3 Self-dual cones

Prove that each of the following cones K satisfy the relationship $N_K(0) = -K$.

- $\mathbb{R}^n_+$

  Proof: Recall that

  $$N_K(\bar{x}) = \{d \in \mathbb{E} : \langle d, x - \bar{x} \rangle \leq 0\}.$$

  Thus, $d \in N_{\mathbb{R}^n_+}(0)$ if and only if

  $$\langle d, x \rangle \leq 0 \quad \forall x \in \mathbb{R}^n_+.$$

  So, for $x = e_i$, we realize that $x_i \leq 0$ and hence $x \in -\mathbb{R}^n_+$. Conversely, for all $x \in -\mathbb{R}^n_+$ and all $d \in \mathbb{R}^n_+$ we clearly have $\langle d, x \rangle \leq 0$. Thus $N_{\mathbb{R}^n_+}(0) = -\mathbb{R}^n_+$.

- $\mathbb{S}^n_+$

  Proof: Recall that

  $$N_{\mathbb{S}^n_+}(0) = \{X \in \mathbb{S}^n : \langle X, A \rangle \leq 0\} \quad \forall A \in \mathbb{S}^n_+.$$

  So clearly if $-X \in \mathbb{S}^n_+$ then $X \in N_{\mathbb{S}^n_+}(0)$. Conversely, suppose $X \in N_{\mathbb{S}^n_+}(0)$, then

  $$a^T X a \leq 0 \quad \forall a \in \mathbb{R}^n \text{ or equivalently } \ a^T(-X)a \geq 0 \quad \forall a \in \mathbb{R}^n.$$

  Thus by definition $-X \in \mathbb{S}^n_+$.

- $K = \{x \in \mathbb{R}^n : x_1 \geq 0, x_1^2 \geq x_2^2 + \cdots + x_n^2\}$

  Proof: Let $y \in N_K(0)$. Then

  $$\begin{gathered} \inf \langle -y, x \rangle \\ x_1^2 \geq x_2^2 + \cdots + x_n^2 \end{gathered}$$

  has nonnegative optimum value. Note that if $\bar{x}$ is a local minimum of the function $f(x) := \langle -y, x \rangle$, then if $\bar{x} \neq 0$ then $\bar{x}_1 \neq 0$ as otherwise since $\bar{x}_1^2 \geq \bar{x}_2^2 + \cdots + \bar{x}_n^2$, $\bar{x} = 0$. So $y \in N_K(0)$ if and only if the following problem has optimum value at least 0,

  $$\begin{gathered} \inf -y_1 - \sum_{i=2}^n y_i x_i \\ 1 \geq x_2^2 + \cdots + x_n^2. \end{gathered}$$

However,

$$\sum_{i=2}^{n} y_i x_i \le (x_2^2 + \cdots + x_n^2)^{\frac{1}{2}} (y_2^2 + \cdots + y_n^2)^{\frac{1}{2}} \le (y_2^2 + \cdots + y_n^2)^{\frac{1}{2}}.$$

Now let $\bar{x} = (1, 0, 0, \cdots, 0)$ and so $-y_1 \ge 0$. Now let $x_i = \frac{y_i}{\sqrt{\sum_{i=2}^{n} y_i^2}}$ for $2 \le i \le n$, the above inequality holds with equality. Thus,

$$\sup \sum_{i=2}^{n} x_i y_i$$
$$1 \ge x_2^2 + \cdots + x_n^2,$$

has optimum value $(y_2^2 + \cdots + y_n^2)^{\frac{1}{2}}$. Thus, $y \in N_K(0)$ if and only if $(y_2^2 + \cdots + y_n^2)^{\frac{1}{2}} \le -y_1$ or equivalently $-y \in N_K(0)$.

2.1.7

Suppose a convex function $g : [0, 1] \to \mathbb{R}$ satisfies $g(0) = 0$. Prove the function $t \in (0, 1] \mapsto g(t)/t$ is nondecreasing. Hence prove that for a convex function $f : C \to \mathbb{R}$ and points $\bar{x}, x \in C \subseteq \mathbb{E}$, the quotient $(f(\bar{x} + t(x - \bar{x}) - f(\bar{x}))/t$ is nondecreasing as a function of t in $(0, 1]$, and complete the proof of Proposition 2.1.2.

Proof: Note that g is convex and thus $g(ts) \le tg(s)$ for all $t, s \in [0, 1]$. Thus if $t, s \ne 0$ then

$$\frac{g(ts)}{ts} \le \frac{g(s)}{s},$$

and this means that g is nondecreasing. The rest is clear as $t \in (0, 1] \mapsto (f(\bar{x} + t(x - \bar{x})) - f(\bar{x})/t$ is convex. Hence, since $f'(\bar{x}, x - \bar{x}) = \lim_{t \to 0}(f(\bar{x} + t(x - \bar{x})) - f(\bar{x})/t \ge 0$, we conclude that $(f(\bar{x} + 1(x - \bar{x})) - f(\bar{x})/1 \ge 0$ or equivalently $f(x) \ge f(\bar{x})$.

2.1.10

- Prove the function $f : \mathbb{S}^n_{++} \to \mathbb{R}$ defined by $f(X) = \mathrm{Tr}(X^{-1})$ is differentiable on $\mathbb{S}^n_{++}$.
- Define a function $f : \mathbb{S}^n_{++} \to \mathbb{R}$ by $f(X) = \log\det(X)$. Prove $\nabla f(I) = I$. Deduce $\nabla f(X) = X^{-1}$ for any X in $\mathbb{S}^n_{++}$.

Proof:

- Let $f(X) = \mathrm{Tr}(X^{-1})$. Note that for $H \in \mathbb{S}^n$ and small enough $|t|$,

$$\mathrm{Tr}((X + tH)^{-1}) = \mathrm{Tr}(X^{-\frac{1}{2}}(I + tX^{-\frac{1}{2}}HX^{-\frac{1}{2}})^{-1}X^{-\frac{1}{2}}) =$$
$$\mathrm{Tr}(X^{-\frac{1}{2}}(I - tX^{-\frac{1}{2}}HX^{-\frac{1}{2}} + O(t^2))X^{-\frac{1}{2}}) = \mathrm{Tr}(X^{-1}) - t\,\mathrm{Tr}(X^{-2}H) + O(t^2).$$

  Hence, $\nabla f(X)[H] = \mathrm{Tr}(-X^{-2}H)$. Thus, $\nabla f(X) = -X^{-2}$.

- Note that for $H \in \mathbb{S}^n$,

$$\log\det(X + tH) = \log(1 + t\,\mathrm{Tr}(X^{-\frac{1}{2}}HX^{-\frac{1}{2}}) + O(t^2)) + \log\det(X)$$

  So,

$$\nabla f(X)(H) = \lim_{t \to 0} \frac{1}{t} \log(1 + t\,\mathrm{Tr}(X^{-\frac{1}{2}}HX^{-\frac{1}{2}}) + O(t^2))$$
$$= \lim_{t \to 0} \frac{\mathrm{Tr}(X^{-\frac{1}{2}}HX^{-\frac{1}{2}}) + O(t)}{1 + t\,\mathrm{Tr}(X^{-\frac{1}{2}}HX^{-\frac{1}{2}}) + O(t^2)} = \mathrm{Tr}(X^{-\frac{1}{2}}HX^{-\frac{1}{2}}) = \mathrm{Tr}(X^{-1}H).$$

  Thus $\nabla f(X) = X^{-1}$.

  Side: This is also an immediate consequence of chain rule via using the fact $\nabla \det(X) = adj(X)$.

2.1.10 Matrix completion
For a set $\Delta \subseteq \{(i,j) : 1 \leq i \leq j \leq n\}$, suppose the subspace $L \subseteq \mathbb{S}^n$ of matrices with (i,j)th entry of zero for all (i,j) in Δ satisfies $L \cap \mathbb{S}^n_{++} \neq \emptyset$. By considering the problem (for $C \in \mathbb{S}^n_{++}$)

$$\inf\{\langle C, X\rangle - \log\det X : X \in L \cap \mathbb{S}^n_{++}\}, \tag{1}$$

prove there exists a matrix X in $L \cap \mathbb{S}^n_{++}$ with $C - X^{-1}$ having (i,j)th entry of zero for all (i,j) not in Δ.

We now the function $X \in \mathbb{S}^n_{++} \mapsto \langle C, X\rangle - \log\det(X)$ has compact level sets. Now let $A_{i,j} \in \mathbb{S}^n$ be symmetric matrices with (i,j)th and (j,i)th entry equals to 1 and 0 elsewhere. Then

$$L = \{X \in \mathbb{S}^n : \langle A_{i,j}, X\rangle = 0 \ \forall\ (i,j) \in \Delta\},$$

which is a closed subspace of $\mathbb{S}^n$. So the level sets of 1 are also compact and thus there exists a global minimizer $\bar{X}$ in $L \cap \mathbb{S}^n_{++}$. So due to *"First order conditions for linear constraints"* there exists $y_{i,j}$ for all $(i,j) \in \Delta$ such that

$$C - \bar{X}^{-1} = \sum_{(i,j)\in\Delta} y_{i,j} A_{i,j}.$$

$\bar{X}$ satisfies the desired property.

2.1.13. BFGS update Given a matrix C in $\mathbb{S}^n_{++}$ and vectors s and y in $\mathbb{R}^n$ satisfying $\langle s, y\rangle > 0$, consider the problem

$$\inf\{\langle C, X\rangle - \log\det(X) : Xs = y, X \in \mathbb{S}^n_{++}\}.$$

- Prove that for the problem above, the point

$$X = \frac{(y-\delta s)(y-\delta s)^T}{\langle s, y-\delta s\rangle} + \delta I$$

  is feasible for small $\delta > 0$.

- Prove problem has an optimal solution.

- Use *"First order conditions for linear constraints"* to find the solution. (The solution is called BFGS update of C^{-1} under the secant condition $Xs = y$.)

Proof:

- Note that

$$Xs = \frac{(y-\delta s)(y-\delta s)^T s}{\langle s, y-\delta s\rangle} + \delta X = y - \delta X + \delta X = y.$$

  Also, for $\delta > 0$ small enough, $\langle s, y-\delta s\rangle > 0$ and thus X will be the sum of a positive semi-definite matrix with δI and thus positive definite. Therefore, X is feasible for small $\delta > 0$.

- We know the map $X \in \mathbb{S}^n_{++} \mapsto \langle C, X\rangle - \log\det(X)$ has compact level sets and also $\{X \in \mathbb{S}^n : Xs = y\}$ is a closed affine subspace of $\mathbb{S}^n$. Thus the map $X \in \mathbb{S}^n_{++} \cap \{X \in \mathbb{S}^n : Xs = y\} \mapsto \langle C, X\rangle - \log\det(X)$ has compact level sets.

- From *"First order conditions for linear constraints"* for a local minimum $\bar{X}$ we know there exists w such that

$$C - \bar{X}^{-1} = sw^T + ws^T \Rightarrow (C - (sw^T + ws^T))^{-1} = \bar{X}.$$

  Therefore,

$$(C - (sw^T + ws^T))y = s \Rightarrow (sw^T + ws^T)y = Cy - s.$$

  So,

$$y^T(sw^T + ws^T)y = y^TCy - y^Ts \Rightarrow 2\langle y, s\rangle\langle y, w\rangle = y^TCy - y^Ts.$$

  Now,

$$\langle w, y\rangle = \frac{y^TCy - \langle s, y\rangle}{2\langle s, y\rangle}.$$

2.1.15. Nearest polynomial with a given root

Consider the Euclidean space of complex polynomials of degree no more than n, with inner product

$$\left\langle \sum_{j=0}^{n} x_j z^j, \sum_{j=0}^{n} y_j z^j \right\rangle = \sum_{j=0}^{n} \overline{x_j} y_j.$$

Given a polynomial p in this space, calculate the nearest polynomial with a given complex root α, and prove the distance to this polynomial is

$$\left(\sum_{j=0}^{n} |\alpha|^{2j} \right)^{-\frac{1}{2}} ||p(\alpha)||.$$

Proof: The problem translates into the following optimization problem

$$\inf\{\sum_{i=0}^{n} ||a_i - b_i||^2 : \langle b, \overline{\tilde{\alpha}} \rangle = 0\},$$

where $p(x) = \sum_{j=0}^{n} a_i x^i$ and $\tilde{\alpha} = (1, \alpha, \cdots, \alpha^n)^T \in \mathbb{R}^{2n+2}$. Note that $\langle ., . \rangle : \mathbb{R}^{2n+2} \to \mathbb{R}^2$ is a linear map. Then for $b \in \mathbb{R}^{2n+2}$ a local minimizer to the above problem, we have

$$b - a = z\tilde{\alpha} \text{ for some } z \in \mathbb{C}$$

Now note that

$$b_i - a_i = z\alpha^i \Rightarrow 0 = \sum_{i=0}^{n} b_i \alpha^i = p(\alpha) + z \sum_{i=0}^{n} ||\alpha||^{2i}$$

So,

$$\sum_{i=0}^{n} ||a_i - b_i||^2 ||z||^2 \sum_{i=0}^{n} |\alpha|^{2i} \Rightarrow \sqrt{\sum_{i=0}^{n} ||a_i - b_i||^2} = \left(\sum_{j=0}^{n} |\alpha|^{2j} \right)^{-\frac{1}{2}} ||p(\alpha)||.$$

2 Max Functions

2.3.3

Prove by induction that if the functions $g_1, \cdots, g_m : \mathbb{E} \to \mathbb{R}$ are all continous at the point $\bar{x}$ then so is the max-function $g(x) = \max_i \{g_i(x)\}$.

Proof: It is clearly enough to prove the question for $m = 2$. If $g_1(\bar{x}) \neq g_2(\bar{x})$ then for instance if $g_1(\bar{x}) < g_2(\bar{x})$, then due to continuity for a small neighborhood about $\bar{x}$, g_1 is smaller than g_2 and thus g is equal to g_2. Since g_2 is continuous at $\bar{x}$ so is g.

So suppose that $g_1(\bar{x}) = g_2(\bar{x})$ and let $x^k \in \mathbb{E}$ be a converging sequence to $\bar{x}$. We aim to show $g(x^k) \to g(\bar{x})$. However, let R_1, R_2 be two subsequences of $\mathbb{N}$ such that for all $r \in R_i$, $g(x^r) = g_i(x^r)$, note that $R_1 \cap R_2$ is not necessarily empty, but $R_1 \cup R_2 = \mathbb{N}$. Now, suppose R_1 and R_2 are both infinite sized, then $g(x^k)$ is divided into two subsequences which both converge to $g(\bar{x}) = g_1(\bar{x}) = g_2(\bar{x})$. If only one of R_1 and R_2 are infinite, say for instance R_1, then $\lim g(x^k) = \lim g_1(x^k) = g_1(\bar{x})$ which equals to $g(\bar{x})$. The proof is complete.

2.3.5. Cauchy-Schwarz and steepest descent

For a nonzero vector y in $\mathbb{E}$, use Karush-Kuhn-Tucker conditions to solve the problem

$$\inf\{\langle y, x \rangle : ||x||^2 \leq 1\}$$

Deduce the Cauchy-Schwarz inequality.

Proof: Note that the feasible region is compact and the objective function is linear and thus continous. So there exists an optimal solution, not necessary unique, detoned by $\bar{x}$. Suppose $\bar{x} \neq 0$, then $\langle \nabla g(\bar{x}), -\bar{x} \rangle = 2\langle \bar{x}, -\bar{x} \rangle < 0$ and thus Mangasarian-Fromovitz constraint qualification holds at $\bar{x}$. Hence, there exists $\lambda \in \mathbb{R}_+$ such that

$$y + \lambda \bar{x} = 0.$$

So, since $y \neq 0$, λ is also nonzero and thus $||\bar{x}|| = 1$ and also $\bar{x} = -\frac{1}{\lambda} y$. Thus $\lambda = ||y||$. Hence, the objective value at $\bar{x}$ equals to

$$\langle y, \frac{-y}{||y||} \rangle = -||y|| < 0,$$

which is negative and thus the assumption that $\bar{x} \neq 0$ is justified. Finally we have for $x \neq 0$

$$\langle y, \frac{x}{||x||} \rangle \geq -||y|| \Rightarrow \langle y, x \rangle \geq -||y|| ||x||.$$

Intechanging x with $-x$ results in, which also holds for $x \neq 0$.

$$-||y|| ||x|| \leq \langle y, x \rangle \leq ||y|| ||x||.$$

2.3.7.
Consider a matrix $A \in \mathbb{S}^n_{++}$ and a real $b > 0$.

- Assuming the problem
$$\inf\{-\log\det X : \operatorname{Tr} AX \leq b, X \in \mathbb{S}^n_{++}\}$$
has a solution, find it.
- Repeat using the objective function $\operatorname{Tr}(X^{-1})$.

Proof:

- Note that $X \mapsto A^{\frac{1}{2}} X A^{\frac{1}{2}}$ is a homeomorphism and thus the following problem has a solution
$$\inf\{-\log\det(A^{\frac{1}{2}} X A^{\frac{1}{2}}) : \operatorname{Tr} A^{\frac{1}{2}} X A^{\frac{1}{2}} \leq b, A^{\frac{1}{2}} X A^{\frac{1}{2}} \in \mathbb{S}^n_{++}\}.$$
Thus without loss of generality suppose that $A = I$. Thus we know the problem
$$\inf\{-\log\det X : \operatorname{Tr} X \leq b, X \in \mathbb{S}^n_{++}\}$$
has a solution. Now, for any feasible matrix X, we have
$$\sqrt[1/n]{\det(X)} \leq \frac{1}{n} \operatorname{Tr}(X) \leq \frac{b}{n} \Rightarrow \det(X) \leq (\frac{b}{n})^{1/n} \Rightarrow -\log\det(X) \geq -\log(\frac{b}{n})^{1/n}.$$
However, in the above equation, equality happens if and only if $\lambda_1(X) = \cdots = \lambda_n(X)$ and also $\operatorname{Tr}(X) = b$. So the optimal solution equals to $\bar{X} = \frac{b}{n} I$.
Now for the original problem if X is the optimal solution $\bar{X}$, then $A^{\frac{1}{2}} X A^{\frac{1}{2}} = \frac{b}{n} I$ and so $\bar{X} = \frac{b}{n} A^{-1}$.
- Note that $\langle A, -A \rangle < 0$ and thus MFCQ holds at any $X \in \mathbb{S}^n_{++}$. Now note that if $\bar{X}$ is a local minimizer for the problem then since MFCQ holds at $\bar{X}$ there exists $\lambda \geq 0$ such that
$$-\bar{X}^{-1} + \lambda A = 0 \Rightarrow \bar{X}^{-1} = \lambda A$$
Note that λ can't be zero and so $\operatorname{Tr}(A\bar{X}) = b$. Hence, $\operatorname{Tr}(\frac{1}{\lambda} I) = b$ and so $\lambda = \frac{n}{b}$. Finally, $\bar{X} = \frac{b}{n} A^{-1}$.

2.3.8. Minimum volume ellipsoid

- For a $y \in \mathbb{R}^n$ and the function $g : \mathbb{S}^n \to \mathbb{R}$ defined by $g(X) = ||Xy||2$, prove $\nabla g(X) = Xyy^T + yy^T X$ for all the matrices X in $\mathbb{S}^n$.

- Consider a set $\{y^1, \cdots, y^m\} \subseteq \mathbb{R}^n$. Prove this set spans $\mathbb{R}^n$ if and only if the matrix $\sum_i y^i(y^i)^T$ is positive definite.
- Prove the problem

$$\begin{gathered}\inf -\log\det X \\ \text{subject to } ||Xy^i||^2 - 1 \leq 0 \text{ for } i = 1, 2, \cdots, m \\ X \in \mathbb{S}^n_{++}\end{gathered}$$

  has an optimal solution.
- Show that the Mangasarian-Fromovitz constraint qualification holds at $\bar{X}$ by considering the direction $d = -\bar{X}$.
- Write down the KKT conditions that $\bar{X}$ must satisfy.
- When $\{y^1, \cdots, y^n\}$ is the standard basis of $\mathbb{R}^n$, the optimal solution of the problem in part (c) is $\bar{X} = I$. Find the corresponding Lagrange multiplier vector.

  Proof:

- Let $A = yy^T \in \mathbb{S}^n$, then $g(X) = \mathrm{Tr}(XAX)$ and thus

$$\lim_{t\to 0}\frac{1}{t}(g(X+tY) - g(X)) = \lim_{t\to 0}\frac{1}{t}(t\,\mathrm{Tr}(XAY) + \mathrm{Tr}(AXY) + t^2\,\mathrm{Tr}(YAY)) = \mathrm{Tr}((XA + AX)Y).$$

  Thus $\nabla g(X) = XA + AX$. Note that in the above equation we are using the fact that $\mathrm{Tr}(YAX) = \mathrm{Tr}(AXY)$.

- Clearly, $\sum_i y^i(y^i)^T \succeq 0$, also note that

$$x^T\left[\sum_i y^i(y^i)^T\right]x = \sum_i \langle x, y^i\rangle^2 \Rightarrow \left[\sum_i y^i(y^i)^T\right]x = 0 \iff \langle x, y^i\rangle = 0\ \forall i$$

  So $\mathrm{Ker}(\sum_i y^i(y^i)^T) = 0$ if and only if it doesn't exist a vector x such that $\langle x, y^i\rangle = 0$ for all i and this holds if and only if the set $\{y^1, \cdots, y^m\} \subseteq \mathbb{R}^n$ spans $\mathbb{R}^n$.

  Now suppose the vector $y^1, \cdots, y^m$ span $\mathbb{R}^n$.

- Denote the feasible region of the above problem by Ω. Let $A = \sum_i y^i(y^i)^T$. Then as $y^1, \cdots, y^m$ span $\mathbb{R}^n$, we have $A \succ 0$. Also for $X \in \Omega$, we have $\langle X^2, A\rangle \leq n$. Thus if for $X \in \Omega$, $-\log\det X \leq c$ for some $c \in \mathbb{R}$, then

$$\langle A, X^2\rangle - \log\det X^2 \leq n - \frac{1}{2}\log\det X \leq n - \frac{1}{2}c.$$

  But we know that the level sets of $\langle C, X\rangle - \log\det X$ are compact for any $C \in \mathbb{S}^n_{++}$ from section 1.2, Question 14. However, $X \mapsto X^2$ is a homeomorphism from $\mathbb{S}^n_{++}$ to $\mathbb{S}^n_{++}$. Thus the set of $x \in \Omega$ which satisfies $-\log\det X \leq c$ lie in a compact set. Thus the optimum is obtained.

  Now suppose that $\bar{X}$ is an optimal solution for the problem in part (c).

- Note that for all i

$$\langle \bar{X}y^i(y^i)^T + y^i(y^i)^T\bar{X}, -\bar{X}\rangle = -2\,\mathrm{Tr}(\bar{X}y^i(y_i)^T\bar{X}) = -2 < 0.$$

- The KKT conditions are as the followings

$$\begin{gathered}\bar{X} \in \mathbb{S}^n_{++} \\ ||\bar{X}y^i||^2 - 1 \leq 0 \text{ for } i = 1, 2, \cdots, m \\ \exists\lambda \in \mathbb{R}^m_+,\ s.t.\ -\bar{X}^{-1} + \sum_i \lambda_i(\bar{X}y^i(y^i)^T + y^i(y^i)^T\bar{X}) = 0 \\ \lambda_i(||Xy^i||^2 - 1) = 0\ \forall i\end{gathered}$$

- Note that $I \in \Omega$ and $\log\det I = 0$. So we need to show that for $X \in \Omega$ we have $\log\det X \leq 0$ or equivalently $\det X \leq 1$. However, as we mentioned before, $\langle X, \sum_i y^i (y^i)^T\rangle \leq n$, but $\sum_i y^i (y^i)^T = I$ and thus $\langle X, I\rangle = \mathrm{Tr}(X) \leq n$. Now if $\lambda_1(X), \cdots, \lambda_n(X) > 0$ are the eigenvalues of X, the we have

$$\sqrt[1/n]{\prod_i \lambda_i(X)} \leq \frac{1}{n}(\sum_i \lambda_i(X)) = \frac{1}{n}\,\mathrm{Tr}(X) \leq 1 \Rightarrow \det(X) = \prod_i \lambda_i(X) \leq 1.$$

So I is the optimal solution. Now if $\lambda \in \mathbb{R}^n_+$ is a Lagrange multiplier then

$$-I + 2\,\mathrm{diag}(\lambda) = 0 \Rightarrow \lambda_i = \frac{1}{2}\ \forall i.$$

Chapter III
Fenchel Duality

March 11, 2023

1 Subgradients and Convex Functions

3.1.1

A function $f : \mathbb{E} \to (+\infty, +\infty]$ is sublinear if and only if it is positively homogeneous and subadditive. For a sublinear function f, the lineality space $\operatorname{lin} f$ is the largest subspace of $\mathbb{E}$ on which f is linear. Recall that

$$\operatorname{lin} f = \{x \in \mathbb{E} : -f(x) = f(-x)\}.$$

Proof: First suppose that f is sublinear. Then $f(\lambda x + \mu y) \leq \lambda f(x) + \mu f(y)$ for all $x, y \in \mathbb{E}$ and $\lambda, \mu \in \mathbb{R}_+$. Now let $x = y = 0$, then

$$f(0) \leq f(0) + f(0) \Rightarrow 0 \leq f(0).$$

Now let $\lambda = \mu = 0$ and so

$$f(0) \leq 0.$$

So, we have $f(0) = 0$. Now let $y = 0$ and conclude that

$$f(\lambda x) \leq \lambda f(x) \; \forall \lambda \in \mathbb{R}_+.$$

Let $\lambda > 0$ and thus

$$f(x) \leq \frac{1}{\lambda} f(\lambda x) \Rightarrow \lambda f(x) \leq f(\lambda x).$$

So $f(x) = \lambda f(x)$ for all $\lambda \in \mathbb{R}_+$ and thus f is positively homogeneous. Finally let $\lambda = \mu = 1$ and thus $f(x + y) \leq f(x) + f(y)$ and thus f is subadditive.

Conversely, suppose f is subadditive and positively homogeneous. Then $f(\lambda x + \mu y) \leq f(\lambda x) + f(\mu y) = \lambda f(x) + \mu f(y)$. Thus f is sublinear.

Now let $x, y \in \operatorname{lin} f$, then if $\lambda, \mu \in \mathbb{R}_-$, then

$$f(\lambda x + \mu y) = f(-\lambda(-x) - \mu(-y)) \leq -\lambda f(-x) - \mu f(-y) = \lambda f(x) + \mu f(y)$$

3.1.6.

If the function $f : \mathbb{E} \to (+\infty, +\infty]$ is convex and the point $\bar{x}$ lies in $\operatorname{dom} f$, then an element ϕ of $\mathbb{E}$ is a subgradient of f at $\bar{x}$ if and only if it satisfies $\langle \phi, . \rangle \leq f'(\bar{x}; .)$.

Proof:

$" \Rightarrow "$.

For $d \in \mathbb{E}$ let $\epsilon > 0$ be small enough such that $x_t := \bar{x} + td \in \operatorname{dom} f$ for all $t \in [0, \epsilon]$. Then,

$$\langle \phi, x_t - \bar{x} \rangle \leq f(x_t) - f(\bar{x}) \Rightarrow \langle \phi, d \rangle \leq \frac{f(x + td) - f(\bar{x})}{t} \qquad \forall t \in (0, \epsilon].$$

Now, taking $t \downarrow 0$ completes the proof.

Conversely, suppose that $\langle \phi, . \rangle \leq f'(\bar{x}; .)$ and let $x \in \mathbb{E}$. Then since $t \mapsto \frac{f(\bar{x}+t(x-\bar{x}))-f(\bar{x})}{t}$ is nondecreasing we have $f'(x, x - \bar{x}) \leq f(x) - f(\bar{x})$

$$\langle \phi, x - \bar{x} \rangle \leq f'(x; x - \bar{x}) \leq f(x) - f(\bar{x}).$$

This completes the proof.

3.1.7.

Suppose that the function $p : \mathbb{E} \to (+\infty, +\infty]$ is sublinear and that the point $\bar{x}$ lies in core(dom p). Then the function $q(.) = p'(\bar{x}; .)$ satisfies the conditions

- $q(\lambda\bar{x}) = p(\lambda\bar{x})$.
- $q \leq p$.
- $\operatorname{lin} p + span\{\bar{x}\} \subseteq \operatorname{lin} q$.

Proof:

- $$q(\lambda\bar{x}) = \lim_{t\downarrow 0} \frac{p(\bar{x} + t\lambda\bar{x}) - p(\bar{x})}{t} = \lim_{t\downarrow 0} \frac{(1 + t\lambda)p(\bar{x}) - p(\bar{x})}{t} = \lambda p(\bar{x}).$$

  Note that for small enough t, $1 + t\lambda > 0$.

- For $d \in \mathbb{E}$,

  $$q(d) = \lim_{t\downarrow 0} \frac{p(\bar{x} + td) - p(\bar{x})}{t} \leq \lim_{t\downarrow 0} \frac{p(\bar{x}) + p(td) - p(\bar{x})}{t} = \lim_{t\downarrow 0} \frac{p(td)}{t} = p(d).$$

- Note that $\operatorname{lin} q$ is a linear subspace and thus due to part 1, it suffices to prove $\operatorname{lin} p \subseteq \operatorname{lin} q$. Suppose $d \in \operatorname{lin} p$ and hence $p(-d) + p(d) = 0$. Now we have

  $$q(d) \leq p(d) \Rightarrow -q(d) \geq -p(d) = p(-d) \geq q(-d) \Rightarrow q(-d) \leq -q(d) \leq q(-d) \Rightarrow q(d)+q(-d) = 0.$$

  Note that $-q(d) \leq q(-d)$ holds true since q is sublinear.

3.1.9. Subgradients of maximum eigenvalue

Prove

$$\partial\lambda_1(0) = \{Y \in \mathbb{S}^n_+ : \operatorname{Tr}(Y) = 1\}.$$

Proof: $Y \in \partial\lambda_1(0)$ if and only if $\operatorname{Tr}(XY) \leq \lambda_1(X)$ for all $X \in \mathbb{S}^n$. Let $X = I$ and $X = -I$ respectively to conclude that $\operatorname{Tr}(Y) = 1$. Now from Fan inequality we know

$$\operatorname{Tr}(XY) \leq \lambda(X)^T\lambda(Y) = \sum \lambda_i(X)\lambda_i(Y) \leq \lambda_1(X) \sum \lambda_i(Y) = \lambda_1(X).$$

3.1.6. (Bregman distances)

For a function $\phi : \mathbb{E} \to (\infty, +\infty]$ that is strictly convex and differentiable on $int(\operatorname{dom}\phi))$, define the *Bregman distance* $d_\phi : \operatorname{dom}\phi \times int(\operatorname{dom}\phi) \to \mathbb{R}$ by

$$d_\phi(x, y) = \phi(x) - \phi(y) - \phi'(y)(x - y).$$

- Prove $d_\phi(x, y) \geq 0$ with equality if and only if $x = y$.

- Compute d_ϕ when $\phi(t) = \frac{t^2}{2}$ and when ϕ is the function p defined in Exercise 27.
- Suppose ϕ is three times differentiable. Prove d_ϕ is convex if and only if $-1/\phi''$ is convex on $int(\text{dom}\,\phi)$.

Proof:

- By definition of strictly convex.
- Let $\phi = t^2/2$, then $d_\phi(x,y) = \frac{(x-y)^2}{2}$. Also for the function p from Exercise 27, if $u, v > 0$, then

$$u \log u - u - v \log v + v - \log v(u - v) = u(\log u - \log v) - (u - v).$$

  Note that
$$u(\log u - \log v) - (u - v) \geq 0 \iff u/v \geq e^{1-\frac{v}{u}}.$$
  On the other hand, $e^x - xe$ is a convex function with minimum occurs at $x = 1$, and so $e^x - xe \geq 0$.

  However, if $u = 0$, then

$$d_\phi(0, v) = -v \log v + v - \log v(-v) = v.$$

- Note that the second derivative of d_ϕ can be calculated as the following:

$$\nabla^2 d_\phi(x,y) = \begin{bmatrix} \phi''(x) & -\phi''(y) \\ -\phi''(y) & \phi''(y) + \phi'''(y)(y-x) \end{bmatrix}$$

  Now, due to Schur complement criterion, $\nabla^2 d_\phi(x,y)$ is positive semi-definite if and only if
$$\phi''(x) > 0, \quad \phi''(y) + \phi'''(y)(y-x) - \phi''(y)^2/\phi''(x) > 0$$

3.1.20. Monotonicity of gradients

Suppose that $S \subseteq \mathbb{R}^n$ is open and convex and the function $f : S \subseteq \mathbb{R}$ is differentiable. Prove f is convex if and only if

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0 \qquad \text{for all } x, y \in S,$$

and f is strictly convex if and only if the above inequality holds strictly whenever $x \neq y$.

Proof: First suppose that f is convex and let $x, y \in S$, then

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle\,,$$
$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle\,.$$

Summing the above two inequalities we obtain

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0,$$

as desired.

Conversely, suppose that for all $x, y \in S$, $\langle \nabla f(y) - \nabla f(x), y - x\rangle \geq 0$ holds true. We wish to show that f is convex. It suffices to show that $g : t \in [0,1] \to f((1-t)x + ty)$ is convex. For that, we just need to show that $g'(t)$ is non-decreasing. But, $g'(t) = \langle \nabla f((1-t)x + ty), y - x\rangle$. Now let $0 \leq t_1 < t_2 \leq 1$ and notice

$$(t_2 - t_1)(g'(t_2) - g'(t_1)) = \langle \nabla f((1-t_2)x + t_2 y) - \nabla f((1-t_1)x + t_1 y), y - x\rangle$$
$$= \langle \nabla f((1-t_2)x + t_2 y) - \nabla f((1-t_1)x + t_1 y), ((1-t_2)x + t_2 y) - ((1-t_1)x + t_1 y)\rangle \geq 0,$$

thus $g'(t)$ is non-decreasing as desired.

Now suppose f is strictly convex and then

$$f(y) > f(x) + \langle \nabla f(x), y - x\rangle\,,$$
$$f(x) > f(y) + \langle \nabla f(y), x - y\rangle\,.$$

Summing the above two inequalities we obtain

$$\langle \nabla f(y) - \nabla f(x), y - x\rangle > 0,$$

as desired.

Conversely, suppose that for all $x, y \in S$, $\langle \nabla f(y) - \nabla f(x), y - x\rangle > 0$ holds true. We wish to show that f is strictly convex. It suffices to show that $g : t \in [0,1] \to f((1-t)x+ty)$ is strictly convex. For that, we just need to show that $g'(t)$ strictly increasing. But, $g'(t) = \langle \nabla f((1-t)x + ty), y - x\rangle$. Now let $0 \leq t_1 < t_2 \leq 1$ and notice

$$(t_2 - t_1)(g'(t_2) - g'(t_1)) = \langle \nabla f((1-t_2)x + t_2 y) - \nabla f((1-t_1)x + t_1 y), y - x\rangle$$
$$= \langle \nabla f((1-t_2)x + t_2 y) - \nabla f((1-t_1)x + t_1 y), ((1-t_2)x + t_2 y) - ((1-t_1)x + t_1 y)\rangle > 0,$$

thus $g'(t)$ is strictly increasing as desired.

3.1.21. The log barrier

Use Exercise 20 (Monotonicity of gradients), Exercise 10 in Section 2.1. and Exercise 8 in Section 1.2 to prove that the function $f : \mathbb{S}^n_{++} \to \mathbb{R}$ defined by $f(X) = -\log\det(X)$ is strictly convex. Deduce the uniqueness of the minimum volume ellipsoid in Section 2.3, Exercise 8, and the matrix completion in Section 2.1, Exercise 12.

Proof: Recall that $\nabla f(X) = -X^{-1}$ and so we should prove for all $X, Y \in \mathbb{S}^n_{++}$ we have

$$\left\langle -X^{-1} + Y^{-1}, X - Y\right\rangle \geq 0 \iff \operatorname{Tr}(XY^{-1}) + \operatorname{Tr}(YX^{-1}) \geq 2n \iff \sum \lambda_i(A) + \sum \frac{1}{\lambda_i(A)} \geq 2n,$$

where $A = XY^{-1}$ and $\lambda_i(A)$ are eigenvalues of A. Note that eigenvalues of A and the positive definite matrix $Y^{-\frac{1}{2}}XY^{-\frac{1}{2}}$ are identical. So since for any $x > 0$ one has $x + \frac{1}{x} \geq 2$, the proof of convexity of f is complete.

Also, note that $X \mapsto ||Xy||^2 - 1$ is convex for any fixed $y \in \mathbb{R}^n$. In fact, we know that $\nabla g(X) = Xyy^T + yy^T X$. So we need to show that for $X, Y \in \mathbb{S}^n_{++}$ we have

$$\langle \nabla g(X) - \nabla g(Y), X - Y\rangle \geq 0 \iff \left\langle (X-Y)yy^T + yy^T(X-Y), X - Y\right\rangle \geq 0.$$

So, we need to prove for $Z \in \mathbb{S}^n$ one has for the positive semi-definite $A = yy^T$

$$\langle ZA + AZ, Z\rangle \geq 0 \iff 2\operatorname{Tr}(ZAZ) \geq 0 \iff 2\operatorname{Tr}((ZA^{\frac{1}{2}})(A^{\frac{1}{2}}Z)) \geq 0,$$

which is immediate as $(ZA^{\frac{1}{2}})(A^{\frac{1}{2}}Z)$ is positive semi-definite.

3.1.22.
Let $f : \mathbb{R}^n \to \mathbb{R}$ with

$$f(x) = \log(\sum_{i=1}^{m} \exp\langle a^i, x\rangle),$$

where $a^1, \cdots, a^m$ are vectors in $\mathbb{R}^n$. Compute the Hessian of f and prove it is positive semi-definite matrix.

Proof: Note that $e^{f(x)} = \sum_{i=1}^{m} \exp\langle a^i, x\rangle$ and thus

$$e^{f(x)}\nabla f(x) = \sum_{i=1}^{m} \exp(\langle a^i, x\rangle)a^i.$$

So,

$$e^{f(x)}\nabla^2 f(x) + e^{f(x)}\nabla f(x)\nabla^T f(x) = \sum_{i=1}^{m} \exp(\langle a^i, x\rangle)a^i {a^i}^T,$$

and hence,

$$e^{f(x)}\nabla^2 f(x) = \sum_{i=1}^{m} \exp(\langle a^i, x\rangle)a^i {a^i}^T - e^{f(x)}\nabla f(x)\nabla^T f(x).$$

Let $t_i = \exp\langle a^i, x\rangle$, then

$$e^{2f(x)}\nabla^2 f(x) = \left(\sum_{k=1}^{m} t_k\right)\left(\sum_{i=1}^{m} t_i a^i {a^i}^T\right) - \left(\sum_{i=1}^{m} t_i a^i\right)\left(\sum_{i=1}^{m} t_i a^i\right)^T.$$

Now let $\lambda_i = \frac{t_i}{\sum_{i=1}^m t_i}$, then $\lambda_i > 0$ and $\sum_{i=1}^m \lambda_i = 1$. Then

$$\nabla^2 f(x) = \left(\sum_{i=1}^{m} \lambda_i a^i {a^i}^T\right) - \left(\sum_{i=1}^{m} \lambda_i a^i\right)\left(\sum_{i=1}^{m} \lambda_i a^i\right)^T.$$

Now $\nabla^2 f(x) \succeq 0$ if and only if

$$\begin{bmatrix} 1 & \sum_{i=1}^m \lambda_i {a^i}^T \\ \sum_{i=1}^m \lambda_i a^i & \sum_{i=1}^m \lambda_i a^i {a^i}^T \end{bmatrix} \succeq 0.$$

But the above matrix equals to

$$\sum_{i=1}^{m} \lambda_i \begin{bmatrix} 1 & {a^i}^T \\ a^i & a^i {a^i}^T. \end{bmatrix}$$

which is clearly positive semi-definite.

3.1.23

Suppose $f : \mathbb{E} \to (\infty, +\infty]$ is essentially strictly convex, prove all distinct points x and y satisfy $\partial f(x) \cap \partial f(y) = \emptyset$. Deduce that f has at most one minimizer.

Proof: Let $s \in \partial f(x) \cap \partial f(y)$ for some $x, y \in \operatorname{dom} \partial f$. Then $g := f + \langle s, . \rangle$ is an essentially strictly convex that satisfies $0 \in \partial g(x) \cap \partial g(y)$. Thus, without loss of generality, suppose $s = 0$ and thus x and y are minimizer to f. However, since f is convex all the points lying on the line segment $[x, y]$ are also minimizers of f. Thus, $[x, y] \in \operatorname{dom} \partial f$ and this is a contradiction as f is essentially strictly convex on $\operatorname{dom} \partial f$.

3.1.25. Convex matrix functions

Consider a matrix C in $\mathbb{S}^n_+$.

- For matrices $X \in \mathbb{S}^n_{++}$ and D in $\mathbb{S}^n$, use a power series expansion to prove
$$\frac{d^2}{dt^2} \operatorname{Tr}(C(X + tD)^{-1})|_{t=0} \geq 0.$$
- Deduce $X \in \mathbb{S}^n_{++} \mapsto \operatorname{Tr}(CX^{-1})$ is convex.
- Prove similarly the function $X \in \mathbb{S}^n \mapsto \operatorname{Tr}(CX^2)$ and the function $X \in \mathbb{S}^n_+ \mapsto -\operatorname{Tr}(CX^{\frac{1}{2}})$ are convex.
- One version of *Hölder inequality* states, for real $p, q > 1$ satisfying $p^{-1} + q^{-1} = 1$ and functions $u, v : \mathbb{R}_+ \to \mathbb{R}$,
$$\int uv \leq \left(\int |u|^p\right)^{\frac{1}{p}} \left(\int |v|^q\right)^{\frac{1}{q}}$$
when the right hand side is well-defined. Use this to prove the *gamma function* $\Gamma : \mathbb{R}_{++} \to \mathbb{R}$ given by
$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$$
is log-convex.
- Note that
$$(X+tD)^{-1} = X^{-\frac{1}{2}}(I+tX^{-\frac{1}{2}}DX^{-\frac{1}{2}})^{-1}X^{-\frac{1}{2}} = X^{-\frac{1}{2}}(I-tX^{-\frac{1}{2}}DX^{-\frac{1}{2}}+t^2(X^{-\frac{1}{2}}DX^{-\frac{1}{2}})^2+O(t^3))X^{-\frac{1}{2}}.$$
Thus,
$$\operatorname{Tr}(C(X+tD)^{-1}) = \operatorname{Tr}(CX^{-1}) - t\operatorname{Tr}(X^{-1}CX^{-1}D) + t^2 \operatorname{Tr}(CX^{-1}DX^{-1}DX^{-1}) + O(t^3).$$
However,
$$\operatorname{Tr}(CX^{-1}DX^{-1}X^{-1}DX^{-1}) = \operatorname{Tr}(C^{\frac{1}{2}}X^{-1}DX^{-\frac{1}{2}}X^{-\frac{1}{2}}DX^{-1}C^{\frac{1}{2}}) = \operatorname{Tr}(AA^T) \geq 0,$$
where $A := C^{\frac{1}{2}}X^{-1}DX^{-\frac{1}{2}}$. Note that AA^T is positive semidefinite and thus has non-negative trace.

- Let $Y \in \mathbb{S}^n_{++}$ and let $g : t \in [0,1] \to \mathrm{Tr}(C(X + t(Y - X))^{-1})$. Then due to part 1, g is convex and thus so is the function $X \in \mathbb{S}^n_{++} \mapsto \mathrm{Tr}(CX^{-1})$.

- Note that
$$\mathrm{Tr}(C(X + tD)^2) = \mathrm{Tr}(C(X^2 + t(DX + XD) + t^2D^2)).$$
Thus,
$$\frac{d^2}{dt^2}\mathrm{Tr}(C(X + tD)^2)|_{t=0} = 2\,\mathrm{Tr}(CD^2) = 2\,\mathrm{Tr}(C^{\frac{1}{2}}DDC^{\frac{1}{2}}) \geq 0.$$
Also, for $X \in \mathbb{S}^n_{++}$,
$$\mathrm{Tr}(C(X + tY)^{\frac{1}{2}}) =$$

3.1.26. Log-convexity

Given a convex set $C \subseteq \mathbb{E}$, we say that a function $f : C \to \mathbb{R}_{++}$ is *log-convex* if $\log f(.)$ is convex.

- Prove any log-convex function is convex.

- If a polynomial $p : \mathbb{R} \to \mathbb{R}$ has all real roots, prove $1/p$ is log-convex on any interval on which p is strictly positive.

Proof:

- Suppose that $f : C \to \mathbb{R}_{++}$ is log-convex. Then let $x, y \in C$ and $\lambda \in (0,1)$, we need to show

  $f(\lambda x+(1-\lambda)y) \leq \lambda f(x)+(1-\lambda)f(y)$ or equivalently $\log(f(\lambda x+(1-\lambda)y)) \leq \log(\lambda f(x)+(1-\lambda)f(y))$.

  However, since f is log-convex, $\log(f(\lambda x + (1 - \lambda)y)) \leq \lambda \log f(x) + (1 - \lambda)\log f(y)$. Now notice that log is concave and so
$$\lambda \log f(x) + (1 - \lambda)\log f(y) \leq \log(\lambda f(x) + (1 - \lambda)f(y)).$$
  Thus the proof is complete.

- We show first that $p^{\frac{1}{n}}$ is concave. Let $a, a_1, \cdots, a_n$ be such that $p(x) = a\prod_{i=1}^n (x - a_i)$ and let $I = (\alpha, \beta)$ be any interval on which p is positive. Note that
$$\frac{d}{dt}p(t)^{\frac{1}{n}} = \frac{1}{n}p(t)^{\frac{1}{n}-1}p'(t),$$
  and so,
$$\begin{aligned}\frac{d^2}{dt^2}p(t)^{\frac{1}{n}} &= \frac{1}{n}(\frac{1}{n} - 1)p(t)^{\frac{1}{n}-2}p'(t)^2 + \frac{1}{n}p(t)^{\frac{1}{n}-1}p''(t)\\ &= \frac{1}{n^2}p(t)^{\frac{1}{n}-2}\left[(1 - n)p'(t)^2 + np''(t)p(t)\right].\end{aligned}$$

On the other hand, for $t \neq a_i$,

$$p'(t) = a\left(\sum_{i=1}^{n} \frac{1}{t-a_i}\right) p(t),$$

and,

$$p''(t) = a\left(-\sum_{i=1}^{n} \frac{1}{(t-a_i)^2}\right) p(t) + a\left(\sum_{i=1}^{n} \frac{1}{t-a_i}\right)^2 p(t).$$

Let $r_i := (t-a_i)^{-1}$ and thus we have

$$\frac{n^2}{p(t)^{2-\frac{1}{n}}}\frac{d^2}{dt^2}p(t)^{\frac{1}{n}} = a^2\left[(1-n)(\sum_{i=1}^{n} r_i)^2 + n(-\sum_{i=1}^{n} r_i^2 + (\sum_{i=1}^{n} r_i)^2)\right] p(t)^2.$$

However,

$$(1-n)(\sum_{i=1}^{n} r_i)^2 + n(-\sum_{i=1}^{n} r_i^2 + (\sum_{i=1}^{n} r_i)^2) = (\sum_{i=1}^{n} r_i)^2 - n\sum_{i=1}^{n} r_i^2 \leq 0,$$

due to Cauchy-Schwartz inequality. Thus $p^{\frac{1}{n}}$ is concave and so is $\log p^{\frac{1}{n}}$ since log is non-decreasing and concave. Thus $\log p$ is concave,

- Note that we have the following as a special case of *Hölder inequality*

$$\log\int_0^\infty u^\lambda v^{1-\lambda} dt \leq \lambda \log\int_0^\infty u dt + (1-\lambda)\log\int_0^\infty v dt,$$

where $u, v : \mathbb{R}_{++} \to \mathbb{R}_{++}$ and $\lambda \in (0,1)$. Now suppose $x, y \in \mathbb{R}_{++}$ and let

$$u(t) = t^x e^{-t}, v(t) = t^y e^{-t}.$$

Then the convexity of the Gamma function follows immediately.

---

2 The Value Function

3.2.1 Lagrangian sufficient conditions.

Prove the Lagrangian sufficient conditions:

Suppose $\bar{\lambda}$ is a Lagrangian multiplier for a feasible solution $\bar{x}$ such that $\bar{x}$ minimizes $L(.,\bar{\lambda})$ over $\mathbb{E}$. Then $\bar{x}$ is an optimal solution.

Proof: Note that since $\bar{\lambda}$ is a Lagrangian multiplier we have $L(\bar{x},\bar{\lambda}) = f(\bar{x})$. However,

$$f(\bar{x}) = L(\bar{x},\bar{\lambda}) \leq L(y,\bar{\lambda}) \leq f(y) \quad \forall \text{ feasible solution } y.$$

Thus $\bar{x}$ is an optimal solution.

3.2.2.

Use the Lagrangian sufficient conditions to the following problem s.

- a)

$$\begin{aligned} &\inf x_1^2 + x_2^2 - 6x_1 - 2x_2 + 10 \\ &\text{subject to} \quad 2x_1 + x_2 - 2 \leq 0 \\ &x_2 - 1 \leq 0 \\ &x \in \mathbb{R}^2. \end{aligned}$$

Proof: Let $y_1 = x_1 - 3, y_2 = x_2 - 1$ then

$$\begin{aligned} &\inf y_1^2 + y_2^2 \\ &\text{subject to} \quad 2y_1 + y_2 + 5 \leq 0 \\ &y_2 \leq 0 \\ &y \in \mathbb{R}^2. \end{aligned}$$

If $y_2 = 0$, then the obvious minimum will be $25/4$. So suppose that $y_2 < 0$ and so $\bar{\lambda}_2 = 0$. Now let $\bar{\lambda}_1 = 2$, then

$$L(y, \bar{\lambda}) = y_1^2 + y_2^2 + 4y_1 + 2y_2 + 10 = (y_1 + 2)^2 + (y_2 + 1)^2 + 5.$$

So, $y = -(2, 1)$ minimizes $L(., \bar{\lambda})$. Since y is feasible and its objective value is $5 < 25/4$, the optimal value is 5 with the optimal solution $(1, 0)$.

-

$$\begin{aligned} &\inf -2x_1 + x_2 \\ &\text{subject to} \quad x_1^2 - x_2 \leq 0 \\ &x_2 - 4 \leq 0 \\ &x \in \mathbb{R}^2. \end{aligned}$$

Proof: If $x_2 = 4$ then the obvious minimum will be 0. So suppose that $x_2 < 4$ and then $\bar{\lambda}_2 = 0$,

$$L(x, \bar{\lambda}) = -2x_1 + x_2 + \bar{\lambda}_1(x_1^2 - x_2).$$

Now let $\bar{\lambda}_1 = 1$ and let $\bar{x} = (1, 1)$. Note that since

$$L(x, \bar{\lambda}) = x_1^2 - 2x_1 \geq -1,$$

and since $L(\bar{x}, \bar{\lambda}) = -1$. Thus $\bar{\lambda}$ is a Lagrangian multiplier $\bar{x}$ and since $\bar{x}$ minimizes $L(., \bar{\lambda})$, we conclude the optimum value is -1.

-

$$\begin{aligned} &\inf x_1 + \frac{2}{x_2} \\ &\text{subject to} - x_2 + \frac{1}{2} \leq 0 \\ &-x_1 + x_2^2 \leq 0. \end{aligned}$$

Proof: If $x_2 = \frac{1}{2}$ then the obvious inf will be $4 + \frac{1}{4}$. Now suppose that $x_2 > \frac{1}{2}$. Then $\bar{\lambda}_1 = 0$ and

$$L(x, \bar{\lambda}) = x_1 + \frac{2}{x_2} + \bar{\lambda}_2(-x_1 + x_2^2).$$

Now let $\bar{\lambda}_2 = 1$ then

$$L(x, \bar{\lambda}) = x_1 + \frac{2}{x_2} + \bar{\lambda}_2(-x_1 + x_2^2) = \frac{2}{x_2} + x_2^2 \geq 3.$$

Note that $\frac{2}{x_2} + x_2^2 = \frac{1}{x_2} + \frac{1}{x_2} + x_2^2 \geq 3$. Now let $\bar{x} = (1, 1)$, then the objective value equals to 3 and thus the optimum value is 3.

---

3 The Fenchel Conjugate

3.1.7. Quadratics

For all matrices A in $\mathbb{S}^n_{++}$, prove the function $x \in \mathbb{R}^n \to x^T Ax/2$ is convex and calculate its conjugate. Use the order preserving property to the conjugacy operation to prove

$$A \succeq B \iff B^{-1} \succeq A^{-1} \text{ for all } A \text{ and } B \text{ in } \mathbb{S}^n_{++}.$$

Proof: Note that for $f(x) = \frac{1}{2}x^T Ax$ we have $\nabla^2 f = A$ and thus f is convex. However, $\nabla f(x) = Ax$ and hence $\sup_x \langle x, y\rangle - f(x)$ is realized at $x = A^{-1}y$. Thus, $f^*(x) = \frac{1}{2}x^T A^{-1}x$. Now if $A \succeq B$ then $f_A \geq f_B$ and hence $f^*_B \geq f^*_A$.

3.1.3.
Verify the conjugates of the log barrier Ib and Id claimed in the text.
Proof: Let $f(x) = -\log x$, then

$$f^*(y) = \sup_{x \in \mathbb{E}} \langle x, y\rangle + \log x.$$

However, $(\langle x, y\rangle + \log x)'' = \frac{-1}{x^2} < 0$. Now since $y + \frac{1}{x}$. Hence, $f^*(x) = -1 + f(-x)$.

Now since $Id^*(X) = \sup_{Y \succ 0} \langle X, Y\rangle + \log\det Y$. Then if $X \not\prec 0$ then $\langle X, Y\rangle + \log\det Y$ is unbounded above as for $Xx = \lambda x$ with $\lambda \geq 0$ we have $\langle X, xx^T + I\rangle + \log\det(xx^T + I) = \lambda ||x||^2 + 1 + ||x||^2 + \mathrm{Tr}(X)$. Now let $||x|| \to +\infty$. Hence, $Id^*(X) = \infty$. So suppose that $X \prec 0$. Then since $\langle X, Y\rangle + \log\det(Y)$ is concave with gradient $Y^{-1} + X$ or $Y = -X^{-1}$. Now

$$\langle -X^{-1}, X\rangle + \log\det(-X^{-1}) = -n - \log\det(-X).$$

3.3.4 Self Conjugacy
Consider functions $f : \mathbb{E} \to (\infty, \infty]$.

- Prove $f^* = f$ if and only if $f(x) = ||x||^2/2$ for all points x in $\mathbb{E}$.
- Find two distinct functions f satisfying $f(-x) = f^*(x)$ for all x in $\mathbb{E}$.

Proof:

- Suppose that $f(x) = \frac{1}{2}||x||^2$. Then $f^*(x) = \sup\langle y, x\rangle - \frac{1}{2}||x||^2$. Now note that $\langle y, x\rangle - \frac{1}{2}||x||^2$ is strictly concave and $\nabla(\langle y, x\rangle - \frac{1}{2}||x||^2) = y - x$. Thus, $f^*(y) = \frac{1}{2}||y||^2$. Conversely, suppose that $f = f^*$. Then since f^* is convex, f is convex as well. If $\operatorname{dom} f = \emptyset$, then $f^* \equiv -\infty$ which is a contradiction as $f = f^*$ and f never takes the value $-\infty$. So $\operatorname{dom} f \neq \emptyset$. Let $x \in \operatorname{dom} f$, then $f(x) + f(y) \geq \langle x, y\rangle$ for all $y \in \operatorname{dom} f$. Also, if $y \in \partial f(x)$, then $f(x)+f(y) = \langle x, y\rangle$ and hence $y \in \partial f^*(x)$ and thus $y \in \partial f(x)$. Now since $f(x)+f(y) \ \geq \langle x, y\rangle$ we obtain $f(x) \geq \frac{1}{2}\langle x, x\rangle$. Now let $x \in \mathbb{E}$ and $y \in \partial f(x)$ then $\langle x, y\rangle \geq \frac{1}{2}(||x||^2 + ||y||^2)$. Hence, $f(x) = \frac{1}{2}||x||^2$. This completes the proof.

- Let $f(x) = -\log x$. Note that $f^*(y) = -1\log(-y) = -1 + f(-y)$. Thus, $f^*(x) = -1 + f(-x)$. Let $g(x) = f(x) - \frac{1}{2}$. Then $g^*(x) = g(-x)$.

Question 7. Maximum entropy example

- Prove the function g defined by

$$g(z) = \inf_{x\in\mathbb{R}^{m+1}} \left\{ \sum_i \exp^*(x_i) : \sum_i x_i = 1, \sum_i x_i a^i = z \right\}$$

is convex.

- For any point $y \in \mathbb{R}^{m+1}$, prove

$$g^*(y) = \sup_{x\in\mathbb{R}^{m+1}} \left\{ \sum_i (x_i\langle a^i, y\rangle - \exp^*(x_i)) : \sum_i x_i = 1 \right\}.$$

- Apply Exercise 27 in Section 3.1 to deduce the conjugacy formula 3.3.2.

- Compute the conjugate of the function of $x \in \mathbb{R}^{m+1}$,

$$\begin{cases} \sum_i \exp^*(x_i) & \text{if} \quad \sum_i x_i = 1 \\ +\infty & \text{otherwise.} \end{cases}$$

Proof:

- Let $\epsilon > 0$ be arbitrary and fix $\lambda \in [0, 1]$. We show for all $z, z' \in \mathbb{R}^m$

$$g(\lambda z + (1-\lambda)z') \leq \lambda g(z) + (1-\lambda)g(z') + 2\epsilon.$$

Let $\tilde{x_i}$ be such that

$$\sum \exp^*(\tilde{x_i}) \leq g(z) + \epsilon.$$

Similarly, let $\tilde{y}_i$ be such that

$$\sum \exp^*(\tilde{y}_i) \leq g(z') + \epsilon.$$

Let $\tilde{z}_i = \lambda\tilde{x}_i + (1-\lambda)\tilde{y}_i$, so

$$g(\tilde{z}_i) \leq \sum \exp^*(\tilde{z}_i) \leq \lambda g(z) + (1-\lambda)g(z') + 2\epsilon.$$

3.3.20. Pointed cones and bases

Consider a closed convex cone K in $\mathbb{E}$. A base for K is a convex set C with $0 \notin cl(C)$ and $K = \mathbb{R}_+ C$. Prove the following properties are equivalent

- (a) K is pointed, i.e. $K \cap -K = \{0\}$.
- (b) $cl(K^\circ - K^\circ) = \mathbb{E}$.
- (c) $K^\circ - K^\circ = \mathbb{E}$.
- (d) K° has non-empty interior.
- (e) There exists a vector y in $\mathbb{E}$ and real $\epsilon > 0$ with $\langle y, x\rangle \geq \epsilon ||x||$ for all points x in K.
- (f) K has a bounded base.

Proof:

- $(a) \Rightarrow (b)$. Suppose $cl(K^\circ - K^\circ) \neq \mathbb{E}$ and let $x \in \mathbb{E}\backslash cl(K^\circ - K^\circ)$. Then due to Hyperplane separation theorem, there exists $0 \neq \phi \in \mathbb{E}$ and $\alpha \in \mathbb{R}$ such that

  $$\langle x, \phi\rangle > \alpha \geq \langle z_1 - z_2, \phi\rangle \quad \forall z_1, z_2 \in K^\circ.$$

  Thus for all $z \in K^\circ$, $\langle z, \phi\rangle$ is bounded above and thus $\langle z, \phi\rangle \leq 0$ as K° is a cone. On the other hand, due to the above equation, $\langle -z, \phi\rangle$ is bounded as well for all $z \in K^\circ$ and so $\langle z, -\phi\rangle \leq 0$ for all $z \in K^\circ$. So, ϕ and $-\phi$ both belongs to $K^{\circ\circ}$ which equals to K. Thus, since K is pointed we conclude that $\phi = 0$. This contradiction completes the proof.

- $(b) \Rightarrow (c)$. Note that $K^\circ - K^\circ$ is a subspace. In fact, it is clearly a convex cone and since $-(K^\circ - K^\circ) = K^\circ - K^\circ$ and it contains 0, it is also a subspace. Thus, since every subspace is closed, $K^\circ - K^\circ = cl(K^\circ - K^\circ)$.

- $(c) \Rightarrow (d)$. We already now that every nonempty convex set in $\mathbb{E}$ has a nontrivial relative interior. Now note that

  $$\text{aff}(K^\circ) = \text{aff}(-K^\circ) = \text{aff}(K^\circ - K^\circ).$$

  Thus, since K° has a nonempty interior and its affine hull is $\mathbb{E}$, we conclude that K° has a nonempty interior.

- $(d) \Rightarrow (e)$. Let $y \in (K^\circ)$, and $\epsilon > 0$ such that $y + td \in K^\circ$ for all $t \in [-\epsilon, \epsilon]$ and any $d \in \mathbb{E}$ with $||d|| = 1$. So, we have

$$\langle y + td, x\rangle \leq 0 \; \forall t \in [-\epsilon, \epsilon], \forall ||d|| = 1, \forall x \in K,$$

or equivalently for $0 \neq x \in K$,

$$|t\langle d, x\rangle| \leq \langle -y, x\rangle \Rightarrow |t\langle d, \frac{x}{||x||}\rangle| \leq \langle -y, \frac{x}{||x||}\rangle.$$

Now let $d = \frac{x}{||x||}$ in above, and let $t = \epsilon$ we realize that

$$\epsilon \leq \langle -y, \frac{x}{||x||}\rangle.$$

Thus, $\epsilon||x|| \leq \langle -y, x\rangle$ holds true for all $x \in K$.

- $(e) \Rightarrow (f)$. Now suppose $y \in \mathbb{E}$ and $\epsilon > 0$ are such that

$$\langle y, x\rangle \geq \epsilon||x|| \; \forall x \in K.$$

Define

$$C = \{x \in K : \langle x, y\rangle = 1\}.$$

First note that C is bounded, as if $x \in C$, then $\epsilon||x|| \leq 1$ and so $||x|| \leq \frac{1}{\epsilon}$. Also, note that clearly C is closed and also $0 \notin C$. Last, note that $\mathbb{R}_+ C = K$ as in fact for $0 \neq x \in K$, $\langle y, x\rangle \geq \epsilon||x|| > 0$ and thus $\langle y, x\rangle > 0$ and so there exists $\lambda > 0$ such that $\langle y, \lambda x\rangle = 1$. Thus, $\lambda x \in C$.

- $(f) \Rightarrow (a)$. Let C be a bounded base for K and suppose $a \in K \cap -K$. Then there exists $\lambda, \mu \in \mathbb{R}_+$ such that $a = \lambda c_1 = -\mu c_2$ for some $c_1, c_2 \in C$. Now, if $a \neq 0$ then λ and μ are both nonzero and thus $0 = \frac{\lambda}{\lambda+\mu} c_1 + \frac{\mu}{\mu+\lambda} c_2 \in C$ as C is convex. But this contradicts the fact that $0 \notin cl(C)$. So, K is pointed.

Chapter IV
Convex Analysis

March 11, 2023

1 Continuity of Convex Functions

4.1.6 Polar sets and strict separation

Fix a nonempty set C in $\mathbb{E}$.

- For points x in $\text{int}(C)$ and ϕ in C°, prove $\langle \phi, x\rangle < 1$.
- Assume further that C is a convex set. Prove γ_C is sublinear.
- Assume in addition that $0 \in \text{core}(C)$. Deduce
$$\text{cl}(C) = \{x : \gamma_C(x) \leq 1\}$$
- Finally, suppose in addition that $D \subseteq \mathbb{E}$ is a convex set disjoint from the interior of C. By considering the Fenchel problem $\inf\{\delta_D + \gamma_C\}$, prove there is a closed halfspace containing D but disjoint from the interior of C.

Proof:

- Note that $\langle \phi, z\rangle \leq 1$ for all $z \in C$. Now since $x \in C^\circ$, we have $x + \epsilon d \in C$ for all $||d|| = 1$, for some $\epsilon > 0$. Thus, $\langle \phi, x\rangle + \langle \phi, \epsilon\frac{\phi}{||\phi||}\rangle \leq 1$. Hence, $\langle \phi, x\rangle < 1$.
- Note that $\gamma(\mu c) = \inf\{\lambda : \mu x \in \lambda C\} = \inf\{\lambda\mu : \mu x \in \lambda\mu C\} = \mu \inf\{\lambda : x \in \lambda C\}$, for $\mu > 0$. Thus, γ_C is homogeneous. Now notice
$$\{\lambda_1 : x \in \lambda_1 C\} + \{\lambda_2 : y \in \lambda_2 C\} \subseteq \{\lambda : x + y \in \lambda C\},$$
as C is convex we have $\lambda_1 C + \lambda_2 C = (\lambda_1 + \lambda_2)C$, and thus
$$\inf\{\lambda : x+y \in \lambda C\} \leq \inf\{\lambda_1 : x \in \lambda_1 C\}+\inf\{\lambda_2 : x \in \lambda_2 C\} \Rightarrow \gamma_C(x+y) \leq \gamma_C(x)+\gamma_C(y).$$
- Note that for $x \in C$, we have $\gamma_C(x) \leq 1$ and thus $C \subseteq \{x : \gamma_C(x) \leq 1\}$. Since the latter is closed, as γ is everywhere finite continuous, we have $\text{cl}(C) \subseteq \{x : \gamma_C(x) \leq 1\}$. Now let $x \in \{x : \gamma_C(x) \leq 1\}$. Then since $0 \in C$, $\lambda_1 C \subseteq \lambda_2 C$ for all $\lambda_1 \leq \lambda_2$. Thus, $x \in (1 + \epsilon)C$ for all $\epsilon > 0$. Hence, $x \in \text{cl}(C)$.
- Note that
$$\inf_{x\in\mathbb{E}}\{\delta_D(x) + \gamma_C(x)\} \geq \sup_{\phi\in Y}\{-\delta_D^*(\phi) - \gamma_C^*(-\phi)\}.$$
Note that $\text{dom}\,\gamma_C - \text{dom}\,\delta_D = \mathbb{E} - D = \mathbb{E}$ hence CQ holds. However, note that if $\inf_{x\in\mathbb{E}}\{\delta_D(x) + \gamma_C(x)\} < 1$ then there exists $x \in D$ such that
$$x \in \lambda C \subseteq C$$
for some $\lambda < 1$. But, $\lambda C \subseteq \text{int}(C)$ and hence $x \in D \cap \text{int}(C)$ which is a contradiction. Thus, $\inf_{x\in\mathbb{E}}\{\delta_D(x) + \gamma_C(x)\} \geq 1$.

  But, this contradicts the fact that $D \cap \text{int}(C) = \emptyset$. Thus, $\inf_{x\in\mathbb{E}}\{\delta_D(x) + \gamma_C(x)\} > 0$. Hence, there exists $\phi^* \in Y$ such that $-\delta_D^*(\phi^*) - \gamma_C^*(-\phi^*) > 0$ or $\delta_D^*(\phi^*) + \gamma_C^*(-\phi^*) < 0$. Note that $\gamma_C^* = \delta_{C^\circ}$. Thus, $-\phi^* \in C^\circ$ and also $\langle \phi^*, y\rangle \leq -1$ for all $y \in D$. Thus,
$$\langle \phi^*, y\rangle \leq -1 < \langle \phi^*, x\rangle\ \forall x \in C^\circ, y \in D.$$

4.1.7. Polar calculus
Suppose C and D are subsets of $\mathbb{E}$.

- Prove $(C \cup D)^\circ = C^\circ \cap D^\circ$.
- If C and D are convex, prove

$$\operatorname{conv}(C \cup D) = \cup_{\lambda \in [0,1]} (\lambda C + (1-\lambda)D).$$

- If C is a convex cone and the convex set D contains 0, prove

$$C + D \subseteq cl\, \operatorname{conv}(C \cup D).$$

  Now suppose the closed convex sets K and H of $\mathbb{E}$ both contain 0.

- Prove $(K \cap H)^\circ = cl\, \operatorname{conv}(K^\circ \cup H^\circ)$.

Proof:

- Note that

  $\phi \in (C \cup D)^\circ \iff \langle \phi, x\rangle \leq 1\ \forall x \in C \cup D \iff \langle \phi, x\rangle \leq 1\ \forall x \in C\ \&\ \langle \phi, y\rangle \leq 1\ \forall y \in D.$

  Thus, $x \in (C \cup D)^\circ$ if and only if $x \in C^\circ \cap D^\circ$.

- For $\lambda \in [0,1]$, let $X_\lambda = \lambda C + (1-\lambda)D$. Now since $\operatorname{conv}(C \cup D)$ is convex and $C, D \subseteq \operatorname{conv}(C \cup D)$, $X_\lambda \subseteq \operatorname{conv}(C \cup D)$ for all $\lambda \in [0,1]$. So, $\cup_{\lambda \in [0,1]} X_\lambda \subseteq \operatorname{conv}(C \cup D)$.

  Conversely, since, $X_1 = C$ and $X_0 = D$. So, $C \cup D \subseteq \cup_{\lambda \in [0,1]} X_\lambda$. Thus, in order to prove, $\operatorname{conv}(C \cup D) \subseteq \cup_{\lambda \in [0,1]} X_\lambda$, we just need to show that $\cup_{\lambda \in [0,1]} X_\lambda$ is convex. Let $\lambda_1, \lambda_2 \in [0,1]$ and

$$\lambda_1 c_1 + (1-\lambda_1)d_1 \in X_{\lambda_1}\ \&\ \lambda_2 c_2 + (1-\lambda_2)d_2 \in X_{\lambda_2},$$

  in which $c_1, c_2 \in C$ and $d_1, d_2 \in D$. Now we need to show for any $\mu \in [0,1]$ we have

$$\mu(\lambda_1 c_1 + (1-\lambda_1)d_1) + (1-\mu)(\lambda_2 c_2 + (1-\lambda_2)d_2) \in \cup_{\lambda \in [0,1]} X_\lambda.$$

  However, $\mu(\lambda_1 c_1 + (1-\lambda_1)d_1) + (1-\mu)(\lambda_2 c_2 + (1-\lambda_2)d_2)$ equals to

$$(\mu\lambda_1 c_1 + (1-\mu)\lambda_2 c_2) + (\mu(1-\lambda_1)d_1 + (1-\mu)(1-\lambda_2))d_2,$$

  and if $t = \mu\lambda_1 + (1-\mu)\lambda_2$, then the above equals to, notice $1-t = \mu(1-\lambda_1) + (1-\mu)(1-\lambda_2)$,

$$t(\frac{\mu\lambda_1}{t}c_1 + \frac{(1-\mu)\lambda_2}{t}c_2) + (1-t)(\frac{\mu(1-\lambda_1)}{1-t}d_1 + \frac{(1-\mu)(1-\lambda_2)}{1-t}d_2).$$

  Note that $0 \leq t, 1-t$ and thus $t \in [0,1]$.

- Note that C is a cone and thus $\lambda C = C$ for all $\lambda \in (0,1]$. So

$$\cup_{\lambda\in[0,1]}(\lambda C + (1-\lambda)D) = D \cup \left(\cup_{\lambda\in(0,1]}(C + (1-\lambda)D)\right).$$

  Now since, $0 \in D$ we should have $(1-\lambda)D \subseteq D$ for all $\lambda \in [0,1]$. Thus,

$$\cup_{\lambda\in(0,1]}(C + (1-\lambda)D) = C + D.$$

  So,

$$\operatorname{conv}(C \cup D) = \cup_{\lambda\in[0,1]}(\lambda C + (1-\lambda)D) = (C + D) \cup D.$$

  So, the proof is complete.

- Note that since K, H are closed convex set we obtain

$$K = K^{\circ\circ}, H = H^{\circ\circ}.$$

  Thus, due to part 1,

$$(K \cap H)^\circ = (K^\circ \cup H^\circ)^\circ$$

4.1.13. Existence of extreme points

Prove any nonempty compact convex set $C \subseteq \mathbb{E}$ has an extreme point, without using Minkowski's theorem, by considering the furthest point in C from the origin.

Proof: Since C is compact closed, there exists $\bar{x} \in C$ such that $||\bar{x}|| = \sup_{a\in C} ||a||$. We prove $\bar{x}$ is an extreme point. Let $\bar{x} = \lambda a + (1-\lambda)b$ for some $a, b \in C$. Then $||\bar{x}|| \leq \lambda||a|| + (1-\lambda)||b|| \leq \lambda||\bar{x}|| + (1-\lambda)||\bar{x}|| = ||\bar{x}||$. Thus $||a|| = ||b|| = ||\bar{x}||$ and hence $a, b, \bar{x}$ are collinear and since they have the same norm we conclude that $a = b = \bar{x}$.

Note that fixing any point $c \in \mathbb{E}$, there exists $\bar{x} \in C$ such that $||\bar{x} - c|| = \sup_{a\in C} ||a - c||$. Thus, $\bar{x}$ is also an extreme point.

Remark: All the extreme points can be obtained this way.

4.1.14.

Given a supporting hyperplane H of a convex set $C \subseteq \mathbb{E}$, any extreme points of $C \cap H$ is also an extreme point of C.

Proof: Let $\bar{x}$ be an extreme point of $C \cap H$ in which

$$H = \{x \in \mathbb{E} : \langle a, x - \bar{x}\rangle = 0\},$$

and for all $x \in C$, $\langle a, x\rangle \geq \langle a, \bar{x}\rangle$. Now let $\bar{x} = \lambda x_1 + (1-\lambda)x_2$ for some $x_1, x_2 \in C$. Then

$$\langle a, x_1\rangle, \langle a, x_2\rangle \geq \langle a, \bar{x}\rangle \;\&\; \langle a, \bar{x}\rangle = \lambda\langle a, x_1\rangle + (1-\lambda)\langle a, x_2\rangle.$$

Thus, $\langle a, x_1\rangle = \langle a, x_2\rangle = \langle a, \bar{x}\rangle$, and hence $x_1, x_2 \in H$. However, $\bar{x}$ is an extreme point of $C \cap H$. Thus, $\bar{x} = x_1 = x_2$. This completes the proof.

4.1.15.

For any compact convex set $C \subseteq \mathbb{E}$, prove $C = \operatorname{conv}(\operatorname{bd} C)$.

Proof: Clearly, $\mathrm{conv}(\partial C) \subseteq C$. Now let $\bar{x} \in C \setminus \mathrm{conv}(\partial C)$. $\partial C = C \setminus \mathrm{int}(C)$ and thus ∂C is closed. Also $\partial C \subseteq C$ and thus it is compact. We know that the convex hull of a closed set is closed, we realize that $\mathrm{conv}(\partial C)$ is closed ,convex and bounded.

Now consider the following general case:

Question: Let $C_1 \subsetneq C_2$ be two compact convex sets in $\mathbb{E}$. We know there exists $\bar{x} \in C_2$ such that $d(\bar{x}, C_1) = \sup_{a \in C_2} d(a, C_1)$ as C_2 is compact. Is it true that $\bar{x}$ is an extreme point of C_2?

Answer: Suppose $\bar{x} = \lambda x_1 + (1-\lambda)x_2$, with $x_1, x_2 \in C_2$. Now, since C_1 is compact, there exists $c \in C_1$ such that $||\bar{x} - c|| = \inf_{a \in C_1} ||x - a||$. Thus, due to the way we picked $\bar{x}$, we have

$$||\bar{x} - a|| \geq ||x_1 - a||, ||\bar{x} - a|| \geq ||x_2 - a||.$$

But,

$$||x-a|| = ||\lambda(x_1-a)+(1-\lambda)(x_2-a)|| \leq \lambda||x_1-a||+(1-\lambda)||x_2-a|| \leq \lambda||x-a||+(1-\lambda)||x-a|| = ||x-a||.$$

Thus, $||x - a|| = ||x_1 - a|| = ||x_2 - a||$. Since in the triangle inequality above, equality holds, $x - a, x_1 - a, x_2 - a$ are collinear, and since $||x - a|| = ||x_1 - a|| = ||x_2 - a||$, we have $x_1 = x_2 = x$. Hence, x is an extreme point of C_2.

Now, back to the main question, since $\mathrm{conv}(\partial C) \subseteq C$, if $\mathrm{conv}(\partial C) \neq C$, then the above discussion gives us an extreme point $\bar{x}$ of C lying outside of $\mathrm{conv}(\partial C)$. However, we know that no point in $\mathrm{int}(C)$ can be an extreme point. Thus $\bar{x} \in \partial(C)$ and this is a contradiction.

4.1.16. A converse of Minkowski's theorem

Suppose D is a subset of a compact convex set $C \subseteq \mathbb{E}$ satisfying $\mathrm{cl}(\mathrm{conv}(D)) = C$. Prove $\mathrm{ext}\, C \subseteq \mathrm{cl}\, D$.

Proof: Since $\mathrm{conv}(\mathrm{cl}\, D) = \mathrm{cl}(\mathrm{conv}(D))$, we need to prove for $D \subseteq \mathbb{E}$ closed with $C = \mathrm{conv}(D)$, we should have $\mathrm{ext}(C) \subseteq D$. Suppose that $\mathrm{aff}(C) = \mathbb{E}$. Then, let $\bar{x} \in \mathrm{ext}(C)$. Then $\bar{x} = \sum_{i=1}^m \lambda_i x_i$ for some $\lambda_i \geq 0$ with $\sum_{i=1}^m \lambda_i = 1$ and some $x_i \in D$. Since, $s \in \mathrm{bd}(C)$, there exists a supporting hyperplane for C at $\bar{x}$. Clearly, $x_i \in H$ and thus $x \in \mathrm{conv}(D \cap H)$. x must be an extreme point of $\mathrm{conv}(D \cap H)$ and thus due to induction, $x \in D \cap H$. Note that $\dim(D \cap H) < \dim(D)$. This completes the proof.

4.1.17 Extreme points

Consider a compact convex set $C \subseteq \mathbb{E}$.

- If $\dim \mathbb{E} \leq 2$, prove the set $\mathrm{ext}(C)$ is closed.
- If $\mathbb{E}$ is $\mathbb{R}^3$ and C is the convex hull of the set

  $$\{(x, y, 0) : x^2 + y^2 = 1\} \cup \{(1, 0, 1), (1, 0, -1)\},$$

  prove $\mathrm{ext}(C)$ is not closed.

Proof:

- Suppose that $\operatorname{aff}(C) = \mathbb{E}$. If $\dim \mathbb{E} = 0$ then $\operatorname{ext}(C) = C = \mathbb{E} = \{0\}$. If $\dim \mathbb{E} = 1$, then C is closed and compact, $C = [a, b]$ for some $a, b \in \mathbb{R}$. In this case, $\operatorname{ext}(C) = \{a, b\}$.

  Now suppose $\dim(E) = 2$. Let $x_i \in \operatorname{ext}(C) \to x$ for some $x \in \operatorname{bd}(C)$. Then there exists a hyperplane
  $$H := \{\phi : \langle a, \phi\rangle = \beta\},$$
  such that $x \in H$ and $\langle a, y\rangle \geq \beta$ for all $y \in C$. If $x \in \operatorname{ext}(C \cap H)$ then $x \in \operatorname{ext}(C)$. So suppose that $x \notin \operatorname{ext}(C \cap H)$. Then if $||b|| = 1$ such that $b \in H$ and $\langle b, a\rangle = 0$, then for some $\epsilon > 0$, $x + tb \in C$ for all $t \in [-\epsilon, +\epsilon]$. Now let $\bar{x} \in C \backslash H$. Then let $D = \operatorname{conv}\{x + \epsilon b, x - \epsilon b, \bar{x}\}$. Now there exists $\delta > 0$ such that $B_\delta(x) \cap C \subseteq D$. So since, $x_i \to x$, for some n, $x_n \in D$. So, x_n can't be an extreme point as it is a convex combination of $x + \epsilon, x - \epsilon, \bar{x}$.

- We show that $P = (1, 0, 0) \in \operatorname{cl}(\operatorname{ext}(C))$ but at the same time $(1, 0, 0)$ is not an extreme point. In fact, P is not an extreme point is clear: $P = \frac{1}{2}((1, 0, 1) + (1, 0, -1))$ and thus P is not extreme point. Now we show $Q = (x, y, 0) \in \operatorname{ext}(C)$ for all $Q \neq P$. Let
  $$H_Q = \{z \in \mathbb{R}^3 : \langle z, Q\rangle = 1\}.$$
  Then since $x^2 + y^2 = 1$ and $Q \neq P$, $x < 1$ and hence
  $$\langle (1, 0, -1), Q\rangle = \langle (1, 0, 1), Q\rangle = x < 1 \Rightarrow (1, 0, -1), (1, 0, 1) \in \{z \in \mathbb{R}^3 : \langle z, Q\rangle < 1\}.$$
  However, let $(x', y', 0) \in C$ with $(x', y') \neq (x, y)$. Then due to Cauchy-Schwartz inequality we have
  $$\langle (x', y'), (x, y)\rangle < \sqrt{||(x', y')||||(x, y)||} = 1 \Rightarrow \langle (x', y', 0), Q\rangle < 1.$$
  Thus, C which is the convex hull of $(x', y', 0)$ with $x'^2 + y'^2 = 1$ and $(1, 0, 1), (1, 0, -1)$ lie inside $\{z \in \mathbb{R}^3 : \langle z, Q\rangle \leq 1\}$ and P is the only point of $C \cap H$. Hence, P is a vertex.

  Now let $P_n = (\sqrt{1 - \frac{1}{n}}, \sqrt{\frac{1}{n}}, 0)$ for $n \in \mathbb{N}$. Then $\lim_n P_n = P$ and thus $P \in \operatorname{cl}(\operatorname{ext}(C))$.

4.1.21 Essential smoothness

For any convex function f and any point $x \in \operatorname{bd} \operatorname{dom} f$, prove $\partial f(x)$ is either empty or unbounded. Deduce that a function is essentially smooth if and only if its subdifferential is always singleton or empty.

Proof: Let $C = \operatorname{dom} f$. Note that according to problem 4.1.20, $N_C(\bar{x}) = \{0\}$ implies $\bar{x} \in \operatorname{ri}C$. Thus, there exists $s \in N_C(\bar{x})$ and if $\phi \in \partial f(\bar{x})$ then
$$\langle \phi + s, y - \bar{x}\rangle + f(\bar{x}) \leq f(y) \ \forall y \in \mathbb{E}.$$
Note that the above holds for $y \notin C$ obviously, and if $y \in C$, then $\langle s, y - \bar{x}\rangle \leq 0$ and the result follows from the fact that $\phi \in \partial f(\bar{x})$. Thus, $\phi + ts \in \partial f(\bar{x})$ for all $t \in \mathbb{R}_+$.

---

2 Fenchel Biconjugation

4.2.12 Compact bases for cones

Consider a closed convex cone K. Using Moreau-Rockafellar theorem, show that a point x lies in $\mathrm{int}(K)$ if and only if the set $\{\phi \in K^- : \langle \phi, x\rangle \geq -1\}$ is bounded. If the set $\{\phi \in K^- : \langle \phi, x\rangle = -1\}$ is nonempty and bounded, prove $x \in \mathrm{int}(K)$.

Proof: First, suppose $x \in \mathrm{int}(K)$. Then there exists $\epsilon > 0$ such that $x + \epsilon d \in K$ for all $d \in \mathbb{E}$ with $||d|| = 1$. So, if $\langle \phi, x\rangle \geq -1$ for some $0 \neq \phi \in K^-$, then since $x + \epsilon\phi/||\phi|| \in K$, then $\langle \phi, x + \epsilon\phi/||\phi||\rangle \leq 0$. Thus, $\epsilon||\phi|| - 1 \leq 0$ and hence $||\phi|| \leq 1/\epsilon$.

Conversely, suppose $\{\phi \in K^- : \langle \phi, x\rangle \geq -1\}$ is bounded. Thus first note that if $\langle \phi, x\rangle = 0$ for some $\phi \in K^-$ then $\phi = 0$. Now, let $f(.) = \langle ., x\rangle + \delta_{K^-}(.)$. Then clearly,

$$f^*(\psi) = \sup_{\phi \in K^-} \langle \psi, \phi\rangle - \langle x, \phi\rangle = \sup_{\phi \in K^-} \langle \psi - x, \phi\rangle.$$

Note that $f^*(0) = 0$ and thus f^* is bounded about 0. Thus, for all $\psi_i \to x$, $\langle \psi_i - x, \phi\rangle \leq 0$ for all $\phi \in K^-$. So, $\langle \psi_i, \phi\rangle \leq \langle x, \phi\rangle \leq 0$ for all $\phi \in K^-$. Thus, $\psi_i \in K$, and hence x lies inside the interior of K.

Now suppose the set $\{\phi \in K^- : \langle \phi, x\rangle = -1\}$ is nonempty and bounded. Then let $0 \neq \phi \in K^-$ such that $\langle \phi, x\rangle \geq -1$. Then since $\langle \phi, x\rangle \neq 0$, then $\langle \phi/|\langle \phi, x\rangle|, x\rangle = -1$. So, $||\phi||/|\langle \phi, x\rangle| \leq M$. Thus, $||\phi|| \leq M|\langle \phi, x\rangle| \leq M$.

4.2.13

For any function $h : \mathbb{E} \to [-\infty, +\infty]$, prove the set $\mathrm{cl}(\mathrm{epi}\, h)$ is the epigraph of some function.

Proof: Let

$$f(x) := \inf\{y : (x, y) \in \mathrm{cl}(\mathrm{epi}\, h)\}.$$

Then, note that if $(x, y_i) \in \mathrm{cl}(\mathrm{epi}\, h) \to (x, f(x))$, then $(x, f(x)) \in \mathrm{cl}(\mathrm{epi}\, h)$ as the latter is closed. Now, note that for $r > 0$ and some $(x, y) \in \mathrm{cl}(\mathrm{epi}\, h)$, since there exists $(x_i, y_i) \in \mathrm{epi}\, h$ such that $(x_i, y_i) \to (x, y)$. Then since $(x_i, y_i + r) \in \mathrm{epi}\, h$ and $(x_i, y_i + r) \to (x, y + r)$ we realize that $(x, y + r) \in \mathrm{cl}(\mathrm{epi}\, h))$. Thus, $\mathrm{cl}(\mathrm{epi}\, h) = \mathrm{epi}\, f$.

4.2.14 Lower semicontinuity and closure

For any convex function $h : \mathbb{E} \to [-\infty, +\infty]$ and any point x^0 in $\mathbb{E}$, prove

$$(\mathrm{cl}\, h)(x^0) = \lim_{\delta \downarrow 0} \inf_{||x - x^0|| \leq \delta} h(x).$$

Deduce

Proposition 4.2.7 If a function $f : \mathbb{E} \to [-\infty, +\infty]$ is convex then it is lower semicontinuous at a point x where it is finite if and only if $f(x) = \mathrm{cl}\, f(x)$. In this case, f is proper.

Proof:

4.2.15

For any point x in $\mathbb{E}$ and any function $h : \mathbb{E} \to (-\infty, +\infty]$ with a sub-gradient at x, prove h is lower semicontinuous at x.

Proof: Note that
$$(\mathrm{cl}\, f)(x) = \lim_{\delta \downarrow 0} \inf_{||y-x|| \le \delta} f(y) \le f(x).$$
However, suppose $s \in \partial f(x)$. Then
$$\langle s, y-x\rangle + f(x) \le f(y)\ \forall y \in \mathbb{E},$$
and thus
$$\inf_{||y-x||\le\delta} f(y) \ge f(x) - \delta ||s|| \Rightarrow (\mathrm{cl}\, f)(x) \ge f(x).$$
Thus, $\mathrm{cl}\, f(x) = f(x)$ and due to Proposition 4.2.7, since f is finite at x as otherwise $f \equiv +\infty$ and in that case obviously f is lower semicontinuous everywhere, we have f is lowersemicontinuous at x.

4.2.16. Von Neumann's minmax theorem

Suppose Y is a Euclidean space. Suppose that the sets $C \subseteq \mathbb{E}$ and $D \subseteq Y$ are nonempty and convex with D closed and that the map $A : \mathbb{E} \to Y$ is linear.

Proof:

- By considering the Fenchel problem
$$\inf_{x\in\mathbb{E}} \{\delta_C(x) + \delta_D^*(Ax)\}$$
prove
$$\inf_{x\in\mathbb{E}} \sup_{y\in D} \langle y, Ax\rangle = \max_{y\in D} \inf_{x\in C} \langle y, Ax\rangle$$
(where the max is attained if finite), under the assumption
$$0 \in \mathrm{core}(\mathrm{dom}\, \delta_D^* - AC).$$

- Prove property above holds in either of the two cases
  1. D is bounded, or
  2. A is surjective and 0 lies in $\mathrm{int}\, C$.

- Suppose both C and D are compact. Prove
$$\min_{x\in C} \max_{y\in D} \langle y, Ax\rangle = \max_{y\in D} \min_{x\in C} \langle A^* y, x\rangle.$$

Proof:

- Note that under the assumption $0 \in \mathrm{core}(\mathrm{dom}\, \delta_D^* - AC)$.
$$\inf_{x\in\mathbb{E}} \{\delta_C(x) + \delta_D^*(Ax)\} = \sup_{y\in Y} \{-\delta_C^*(A^* y) - \delta_D(-y)\},$$
since D is closed, δ_D is a closed convex function. Also,
$$\inf_{x\in\mathbb{E}} \{\delta_C(x) + \delta_D^*(Ax)\} = \inf_{x\in C} \{\delta_D^*(Ax)\} = \inf_{x\in C} \sup_{y\in D} \langle y, Ax\rangle = \inf_{x\in C} \max_{y\in D} \langle y, Ax\rangle.$$

Also,

$$\sup_{y\in Y}\{-\delta^*_C(A^*y)-\delta_D(-y)\} = \sup_{y\in -D}\{-\delta^*_C(A^*y)\} = \sup_{y\in -D}\{-\sup_{x\in C}\langle x, A^*y\rangle\} = \sup_{y\in -D}\inf_{x\in C}\langle x, -A^*y\rangle,$$

which equals to $\max_{y\in D}\inf_{x\in C}\langle x, A^*y\rangle$.

- Note that
$$\delta^*_D(x) = \sup_{y\in D}\langle x, y\rangle \Rightarrow \operatorname{dom}\delta^*_D = \mathbb{E}.$$
Also, if A is surjective and 0 lies in interior of C, then since $0 \in \operatorname{dom}\delta^*_D$, then due problem 4.1.9 we are done.

- Clear!

4.2.8. Closed subdifferential
If a function $h : \mathbb{E} \to (\infty, +\infty]$ is closed, prove the multifunction ∂h is closed: that is,

$$\phi_r \in \partial h(x_r), x_r \to x, \phi_r \to \phi \Rightarrow \phi \in \partial h(x).$$

Deduce that if h is essentially smooth and a sequence of points x_r in $\operatorname{int}(\operatorname{dom} h)$ approaches a point in $\partial(\operatorname{dom} h)$ then $||\nabla h(x_r)|| \to \infty$.

Proof: Let $y \in \mathbb{E}$ then

$$\langle \phi_r, y - x_r\rangle + h(x_r) \le h(y).$$

But, since h is lowersemicontinuous at x, we have

$$h(x) \le \liminf_{r\to+\infty} h(x_r) \le h(y) - \langle \phi, y - x\rangle.$$

Hence, $\phi \in \partial h(x)$.

Note that if $\partial h(x) \neq \emptyset$ then h has Gâteaux differential at x and hence $x \in \operatorname{int}\operatorname{dom}(h)$. So $\partial h(x) = \emptyset$ and thus if $||\nabla h(x_r)|| \le C$ for some C, w can assume that $\nabla h(x_r) \to \phi$ for some $\phi \in \mathbb{E}$ and thus $\phi \in \partial h(x)$ which is a contradiction.

4.2.9. Support functions
Prove that if the set $C \subseteq \mathbb{E}$ is nonempty then δ^*_C is a closed sublinear function and $\delta^{**}_C = \delta_{\operatorname{cl}\operatorname{conv} C}$. Prove that if C is also bounded then δ^*_C is everywhere finite.

- Prove that any sets $C, D \subseteq \mathbb{E}$ satisfy
$$\delta^*_{C+D} = \delta^*_C + \delta^*_D \text{ and}$$
$$\delta^*_{\operatorname{conv}(C\cup D)} = \max(\delta^*_C, \delta^*_D).$$

- Suppose the function $h : \mathbb{E} \to (-\infty, +\infty]$ is positively homogeneous, and define a closed convex set
$$C = \{\phi \in \mathbb{E} : \langle \phi, d\rangle \le h(d)\ \forall d\}.$$
Prove $h^* = \delta_C$. Prove that if h is in fact sublinear and everywhere finite then C is nonempty and compact.

Proof:

- Let $c \in \mathbb{R}$, then
$$\delta_C^*(x) \le c \iff \langle x, y\rangle - \delta_C(y) \le c \; \forall y \in C \iff x \in \cap_{y\in C} H_y,$$
where $H_y = \{\phi \in \mathbb{E} : \langle \phi, y\rangle \le c\}$. Thus δ_C^* is closed.
Also, let $\lambda \in \mathbb{R}_+$, then
$$\delta_C^*(\lambda x) = \sup_{y\in\mathbb{E}} \langle \lambda x, y\rangle - \delta_C(y) = \lambda \sup_{y\in\mathbb{E}} \langle x, y\rangle - \delta_C(y) = \lambda\delta_C^*(x).$$
Now let $x, y \in \mathbb{E}$, then
$$\langle x + y, z\rangle - \delta_C(z) = \langle x + y, z\rangle - 2\delta_C(z).$$
On the other hand,
$$\delta_C^*(x) + \delta_C(z) \ge \langle x, z\rangle \;,\; \delta_C^*(y) + \delta_C(z) \ge \langle y, z\rangle.$$
Thus, it can be derived from the above two statements
$$\langle x + y, z\rangle - \delta_C(z) \le \delta_C^*(x) + \delta_C^*(y).$$
Now, let $\bar{x} \in \operatorname{cl}\operatorname{conv}(C)$, then we want to show that
$$\delta_C^{**}(\bar{x}) = 0.$$
Note that $\delta_C^*(0) = 0$ as C is nonempty, thus $\delta_C^{**}(\bar{x}) \ge 0$. Now we want to show for all $y \in \mathbb{E}$, $\langle \bar{x}, y\rangle - \delta_C^*(y) \le 0$, or $\langle \bar{x}, y\rangle \le \delta_C^*(y)$. So, we can suppose that $x \in \operatorname{conv}(C)$ and hence $x = \sum_i \lambda_i x_i$ for some $x_i \in C$. Now note that $\langle x_i, y\rangle \le \delta_C^*(y)$ and hence
$$\lambda_i\langle x_i, y\rangle \le \lambda_i\delta_C^*(y) \Rightarrow \langle \bar{x}, y\rangle \le \delta_C^*(y),$$
for all $y \in C$. Thus $\delta_C^{**}(\bar{x}) \le 0$.
Now let $\bar{x} \in C\setminus \operatorname{cl}\operatorname{conv}(C)$. So, there exists $\phi \in \mathbb{E}$ such that $\langle \bar{x}, \phi\rangle > b \ge \langle y, \phi\rangle$ for all $y \in C$. Thus $\delta_C^*(y) \le b$
$$\delta_C^{**}(\bar{x}) \ge \langle \bar{x}, \phi\rangle - \delta_C^*(\phi) \ge \langle \bar{x}, \phi\rangle - b > 0.$$
Thus $\delta_C^{**}(\bar{x}) \ge \langle \bar{x}, \lambda\phi\rangle - \delta_C^*(\lambda\phi) = \lambda(\langle \bar{x}, \phi\rangle - \delta_C^*(\phi)) \to +\infty$. Thus, $\delta_C^{**}(\bar{x}) = +\infty$. Now, suppose $C \subseteq rB$ for some $r > 0$. Then $\delta_C^*(\bar{x}) = \sup_{y\in C}\langle x, y\rangle \le \sup_{y\in rB}\langle x, y\rangle \le r||x|| < +\infty$.

- Very easy! Omitted.

- C is trivially closed and convex, and if $\phi \notin C$, then $h^*(\phi) > 0$ by definition of C. However, if $\langle \phi, d\rangle - h(d) > 0$ for some $d \in \mathbb{E}$, then $\langle \phi, td\rangle - h(td) \to +\infty$ as $t \to +\infty$. Thus, $h^*(\phi) = \infty$. Now, let $\phi \in C$, then $\langle \phi, d\rangle - h(d) \le 0$ by definition and hence $h^*(\phi) \le 0$. However, $\langle \phi, 0\rangle - h(0) = 0$. Thus, $h^*(\phi) = 0$. Therefore, we have proved that $h^* = \delta_C$. Now, since h is sublinear and everywhere finite $h(d) \le M$, for some $M > 0$, for all $||d|| = 1$. Thus, for $\phi \in C$, we have $\langle \phi, d\rangle \le M$ for all $||d|| = 1$ and hence $||\phi|| \le M$. Now, if $C = \emptyset$, then $h^* \equiv +\infty$ and thus since h is closed and convex we have $h = h^{**}$, we obtain $h \equiv -\infty$ which is a contradiction.

4.2.21 cofiniteness

Consider a function $h : \mathbb{E} \to (\infty, +\infty]$ and the following properties:

1. $h(.) - \langle \phi, . \rangle$ has bounded level sets for all ϕ in $\mathbb{E}$.
2. $\lim_{||x|| \to \infty} ||x||^{-1} h(x) = +\infty$.
3. h^* is everywhere finite.

Complete the following steps.

- Prove properties 1 and 2 are equivalent.
- If h is closed, convex and proper, use and Moreau-Rockafellar theorem to prove properties 1 and 3 are equivalent.

Proof:

- Suppose 1 holds. Then if $\lim_{||x|| \to \infty} ||x||^{-1} h(x) = +\infty$ does not hold true then there exists x_i with $||x_i|| \to +\infty$ such that $||x_i||^{-1} h(x_i) \leq C$ for some constant $C > 0$. Thus, since $||x_i|| \to +\infty$ and hence

$$h(x_i) - \langle \phi, x_i \rangle \to +\infty,$$

for all $\phi \in \mathbb{E}$. However, without loss of generality, suppose that $||x_i||^{-1} x_i \to v$ for some $v \in \mathbb{E}$. Then, let $\phi = Cv$. We have

$$||x_i||^{-1} \langle \phi, x_i \rangle \uparrow C.$$

However, $||x_i||^{-1}(h(x_i) - \langle \phi, x_i \rangle) \leq C - C \leq 0$ and hence $h(x_i) - \langle \phi, x_i \rangle \to +\infty$ cannot hold true.

Conversely, suppose that 2 holds, and $\phi \in \mathbb{E}$. Then $\lim_{||x|| \to \infty} ||x||^{-1}(h(x) - \langle \phi, x \rangle) = +\infty$. and thus without loss of generality we can assume that $\phi = 0$. Now suppose $h(x) \leq M$ is not bounded for some $M \in \mathbb{R}$ and hence there exists $||x_i|| \to +\infty$ such that $h(x_i) \leq M$. So $\lim_{i \to \infty} ||x_i||^{-1} h(x_i) = 0$, which is a contradiction.

- Let $h_\phi(.) := h(.) - \langle \phi, . \rangle$. Then $h_\phi^*(\psi) = h^*(\phi + \psi)$. Thus h^* is continuous at ϕ if and only if h_ϕ^* is continuous at 0. But, we know h_ϕ^* has bounded level set if and only if h_ϕ^* is continuous at zero. Hence, h^* is finite everywhere.

4.2.22 Computing closures

- Prove that any closed convex function $g : \mathbb{R} \to (\infty, +\infty]$ is continuous on its domain.
- Consider a convex function $f : \mathbb{E} \to (\infty, +\infty]$. For any point $x \in \mathbb{E}$ and any $y \in \operatorname{int}(\operatorname{dom} f)$, prove

$$f^{**}(x) = \lim_{t \uparrow 1} f(y + t(x - y)).$$

Proof:

- Without loss of generality assume that $\operatorname{aff}(\operatorname{dom} f) = \mathbb{E}$. Note that $f = f^{**}$ is continuous at 0 if and only if f^* has bounded level sets.

  Suppose this does not hold. Then there exists $x_i \in \mathbb{E}$ such that $||x_i|| \to +\infty$ and

  $$f^*(x_i) \leq M \text{ for some } M > 0.$$

  So, for every $y \in \mathbb{E}$ and every $i = 1, 2, \cdots$ we have

  $$\langle x_i, y\rangle \leq f(y) + M \Rightarrow \langle x, y\rangle \leq f(y) + M \ \forall x \in \operatorname{conv}\{x_1, x_2, \cdots\}.$$

  But since $||x_i|| \to +\infty$, $C = \operatorname{conv}\{x_1, x_2, \cdots\}$ is unbounded and thus $d \in 0^+(C)$ for some $d \neq 0$. Thus, for all $t \in \mathbb{R}_+$,

  $$\langle x_1 + td, y\rangle \leq f(y) + M \Rightarrow \langle d, y\rangle \leq 0 \ \forall y \in \operatorname{ri}(\operatorname{dom} f).$$

  But, then $d = 0$ as $\operatorname{aff}(\operatorname{dom} f) = \mathbb{E}$. This contradiction completes the proof.

4.2.24 Fisher information function

Let $f : \mathbb{R} \to (\infty, +\infty]$ be a given function, and define a function $g : \mathbb{R}^2 \to (\infty, +\infty]$ by

$$g(x, y) = \begin{cases} yf(\frac{x}{y}) \text{ if } y > 0 \\ +\infty \qquad \text{otherwise.} \end{cases}$$

- Prove g is convex if and only if f is convex.

Proof:

- Suppose f is convex with $\operatorname{epi} f = C \subseteq \mathbb{R}^2$. Then C is convex. Now $((x, y), r) \in \operatorname{epi} g$ for some $y > 0$ if and only if

  $$r \geq yf(\frac{x}{y}) \iff (\frac{x}{y}, \frac{r}{y}) \in C \iff (x, r) \in yC.$$

  Now for some $\lambda \in (0, 1)$ and $((x_1, y_1), r_1), ((x_2, y_2), r_2) \in \operatorname{epi} g$ we have $\lambda((x_1, y_1), r_1) + (1 - \lambda)((x_2, y_2), r_2) \in \operatorname{epi} g$ if and only if

  $$(\lambda x_1 + (1 - \lambda)x_2, \lambda r_1 + (1 - \lambda)r_2) \in (\lambda y_1 + (1 - \lambda)y_2)C,$$

  which holds true as C is convex and $(x_1, r_1) \in y_1C$ and $(x_2, r_2) \in y_2C$.

  Now conversely suppose that $\operatorname{epi} g$ is convex, then $(x, r) \in C$ if and only if $((x, 1), r) \in \operatorname{epi} g$. Thus, C is the image of the projection of $\operatorname{epi} g \cap \{y = 1\}$ onto $\mathbb{R}^2$. Hence, C is convex.

---

3 Lagrangian Duality

4.3.1 Weak duality

Prove that the primal and dual values p and d defined by equations

$$p = \inf_{x \in \mathbb{E}} \sup_{\lambda \in \mathbb{R}^n_+} L(x; \lambda), d = \sup_{\lambda \in \mathbb{R}^n_+} \inf_{x \in \mathbb{E}} L(x; \lambda),$$

satisfies $d \leq p$.

Proof: We only need to show $\inf_{x \in \mathbb{E}} L(x; \tilde{\lambda}) \leq \sup_{\lambda \in \mathbb{R}^n_+} L(\tilde{x}; \lambda)$ for fixed $\tilde{\lambda} \in \mathbb{R}^n_+$ and $\tilde{x} \in \mathbb{E}$. But, $\inf_{x \in \mathbb{E}} L(x; \tilde{\lambda}) \leq L(\tilde{x}; \tilde{\lambda}) \leq \sup_{\lambda \in \mathbb{R}^n_+} L(\tilde{x}; \lambda)$.

4.3.2

Calculate the Lagrangian dual of the problem:

$$\inf_{x \in \mathbb{R}^n_{++}} \{\sum_{i=1}^n \frac{c_i}{x_i} : \sum_{i=1}^n a_i x_i \leq b\},$$

where $a_1, c_1, \cdots, a_n, c_n, b \in \mathbb{R}_{++}$.

Proof: Define $\Phi(\lambda) = \inf_{x \in \mathbb{R}^n_{++}} L(x; \lambda)$. Fix $\tilde{\lambda} \in \mathbb{R}_{++}$. Then

$$L(x; \tilde{\lambda}) = \sum_{i=1}^n \frac{c_i}{x_i} + \tilde{\lambda}(\sum_{i=1}^n a_i x_i - b) = -\tilde{\lambda} b + \sum_{i=1}^n \frac{c_i}{x_i} + \tilde{\lambda} a_i x_i \geq -\tilde{\lambda} b + 2 \sum_{i=1}^n \sqrt{\tilde{\lambda} c_i a_i}.$$

Note that equality happens if $x_i = \frac{c_i}{\tilde{\lambda} a_i}$ and $x_i = +\infty$ if $\tilde{\lambda} = 0$. Thus, $\Phi(\tilde{\lambda}) = -\tilde{\lambda} b + 2 \sum_{i=1}^n \sqrt{\tilde{\lambda} c_i a_i}$. Note that $\Phi(\lambda^2)$ is a concave function in λ. Thus,

$$\Phi(\lambda^2) = -\lambda^2 b + 2\lambda \sum_{i=1}^n \sqrt{c_i a_i} \Rightarrow \sup_{\lambda \in \mathbb{R}^n_+} \Phi(\lambda^2) = \sup_{\lambda \in \mathbb{R}^n_+} -\lambda^2 b + 2\lambda \sum_{i=1}^n \sqrt{c_i a_i}.$$

However, the supremum happens at $\lambda^* = \frac{\sum \sqrt{c_i a_i}}{b}$. Thus $d = \frac{(\sum \sqrt{c_i a_i})^2}{b}$.

4.3.3 (Slater and compactness)

Prove the Slater condition holds for problem

$$\inf\{f(x) : g(x) \leq 0, x \in \mathbb{E}\},$$

if and only if there exists $\hat{x} \in \mathbb{E}$ for which the level sets

$$\{\lambda \in \mathbb{R}^n_+ : -L(\hat{x}; \lambda) \leq \alpha\},$$

is compact for all $\alpha \in \mathbb{R}$.

Proof: Suppose there exists a Slater point, then $-\lambda^T g(\hat{x}) \leq \alpha$ has compact level sets for all $\alpha \in \mathbb{R}$. In fact, for each $i = 1, \cdots, m$, we have $-\lambda_i g_i(\hat{x}) \leq \alpha$. Thus, $\lambda_i \leq \frac{\alpha}{-g_i(\hat{x})}$. Thus, λ is bounded above.

Conversely, suppose $\{\lambda \in \mathbb{R}^n_+ : -L(\hat{x};\lambda) \leq \alpha\}$, is compact for all $\alpha \in \mathbb{R}$. Then, if some i, $g_i(\hat{x}) \geq 0$, then if $\mu \in \{\lambda \in \mathbb{R}^n_+ : -L(\hat{x};\lambda) \leq \alpha\}$, then so is $\mu + te_i$ for all $t \geq 0$. Thus, $\{\lambda \in \mathbb{R}^n_+ : -L(\hat{x};\lambda) \leq \alpha\}$ is empty for all real α, which is a contradiction.

4.3.4 (Examples of duals)
Calculate the Lagrangian duals for the following problem:

- The *linear program*

$$\inf_{x\in\mathbb{R}^n}\{\langle c, x\rangle : \langle a^i, x\rangle \leq b_i \text{ for } i = 1, \cdots, m\}.$$

Proof: Let $A = [a^1|\cdots|a^m]$, then $\langle a^i, x\rangle \leq b_i$ translates into $A^T x \leq b$. We have

$$\langle c, x\rangle + \lambda^T(A^T x - b) = \langle c + A\lambda, x\rangle - \lambda^T b.$$

Thus the dual problem is as follows:

$$\sup_{A\lambda+c=0,\lambda\in\mathbb{R}^m_+} -\lambda^T b.$$

- Another linear program

$$\inf_{x\in\mathbb{R}^n}\{\langle c, x\rangle + \delta_{\mathbb{R}^n_+(x)} : \langle a^i, x\rangle \leq b_i \text{ for } i = 1, \cdots, m\}.$$

Proof: Again

$$\langle c, x\rangle + \lambda^T(A^T x - b) = \langle c + A\lambda, x\rangle - \lambda^T b.$$

Thus, the dual problem is a follows:

$$\sup_{A\lambda+c\geq 0,\lambda\in\mathbb{R}^m_+} -\lambda^T b.$$

- The *quadratic program* for some $C \in \mathbb{S}^n_{++}$

$$\inf_{x\in\mathbb{R}^n}\{\frac{1}{2}(x^T Cx) : \langle a^i, x\rangle \leq b_i \text{ for } i = 1, \cdots, m\}.$$

Proof: We have

$$\frac{1}{2}(x^T Cx) + \lambda^T(A^T x - b) \text{ is strictly convex w.r.t } x,$$

thus the dual function equals to (for $Cx^* + A\lambda = 0$)

$$\inf_{x\in\mathbb{E}}\Phi(\lambda) = -\frac{1}{2}x^{*T}A\lambda + \lambda^T A^T x^* - \lambda^T b = \frac{1}{2}x^{*T}A\lambda - \lambda^T b = -[\frac{1}{2}(C^{-1}A\lambda)^T A\lambda + \lambda^T b],$$

Hence,

$$\sup_{\lambda\geq 0} -[\frac{1}{2}(C^{-1}A\lambda)^T A\lambda + \lambda^T b] = -\inf_{\lambda\geq 0}\frac{1}{2}(\lambda^T A^T C^{-T} A\lambda) + \lambda^T b = \frac{1}{2}b^T A^{-1} C A^{-T} b.$$

- The *separable* problem

$$\inf_{x\in\mathbb{R}^n}\{\sum_{j=1}^n p(x_j) : \langle a^i, x\rangle \le b_i \text{ for } i = 1, \cdots, m\}$$

for a given function $p : \mathbb{R} \to (\infty, +\infty]$.

Proof: $\Phi(\lambda) = \inf_{x\in\mathbb{E}} \sum_{j=1}^n (p(x_j) + \lambda_j(\langle a^j, x\rangle - b_j))$

4.3.7
Given a matrix C in $\mathbb{S}^n_{++}$, calculate

$$\inf_{X\in\mathbb{S}^n_{++}}\{\mathrm{Tr}(CX) : -\log\det(X) \le 0\}$$

by the Lagrangian duality.
Proof: Suppose that $b < n$ and consider the problem

$$\inf_{X\in\mathbb{S}^n_{++}}\{\mathrm{Tr}(CX) : -\log\det(X) \le b\}$$

Let $\lambda \in \mathbb{R}_+$. Then the Lagrangian equals to

$$\mathrm{Tr}(CX) - \lambda\log\det(X) - \lambda b \Rightarrow \Phi(\lambda) = -\lambda b + \inf_{X\in\mathbb{S}^n_{++}} \mathrm{Tr}(CX) - \lambda\log\det(X).$$

Note that $\nabla^2(\mathrm{Tr}(CX)-\lambda\log\det(X)) = \lambda X^{-2} \succeq 0$. Thus, since $\nabla(\mathrm{Tr}(CX)-\lambda\log\det(X)) = C - \lambda X^{-1}$, we have $\Phi(\lambda) = -\lambda b + n\lambda + \lambda\log\det(C) - n\lambda\log\lambda$ for $\lambda > 0$ and $\Phi(0) = 0$.
Now note that $(n\lambda + \lambda\log\det(C) - n\lambda\log\lambda)'' = -n/\lambda < 0$ for $\lambda \in \mathbb{R}_{++}$. Now notice $(-\lambda b + n\lambda + \lambda\log\det(C) - n\lambda\log\lambda)' = -b + \log\det(C) - n\log\lambda$, thus, $\lambda^* = \sqrt[1/n]{e^{-b}\det(C)}$. Notice that $\sup_{\lambda\ge 0}\Phi(\lambda) = \Phi(\lambda^*) = n\lambda^* - \lambda^* b > 0$.

4.3.8. Mixed constraints
Explain why an appropriate dual for the problem

$$\inf\{f(x) : g(x) \le 0, h(x) = 0\}$$

for a function $h : \mathrm{dom}\, f \to \mathbb{R}^k$ is

$$\sup_{\lambda\in\mathbb{R}^m_+,\mu\in\mathbb{R}^k}\ \inf_{x\in\mathrm{dom}\, f}\{f(x) + \lambda^T g(x) + \mu^T h(x)\}.$$

Proof: Come on!

4.3.9. Fenchel and Lagrangian duality
Let Y be a Euclidean space. By suitably rewriting the problem Fenchel problem

$$\inf_{x\in\mathbb{E}}\{f(x) + g(Ax)\}$$

for given function $f : \mathbb{E} \to (\infty, +\infty]$, $g : Y \to (\infty, +\infty]$ and linear map $A : \mathbb{E} \to Y$, interpret the dual Fenchel problem

$$\sup_{\phi \in Y}\{-f^*(A^*\phi) - g^*(-\phi)\}$$

as a Lagrangian dual problem.

Proof: Note that

$$\inf_{x \in \mathbb{E}}\{f(x) + g(Ax)\} = \inf_{(x,y) \in \mathbb{E}^2}\{f(x) + g(y) : Ax = y\}.$$

Thus, $L(x, y, \phi) = f(x) + g(y) + \langle \phi, Ax - y \rangle$ with $L : \mathbb{E}^2 \times \mathbb{R} \to [-\infty, +\infty]$. Then

$$\Phi(\phi) = \inf_{(x,y)} f(x) + g(y) + \langle \phi, Ax - y \rangle = -\sup_{x}[\langle -A^*\phi, x \rangle - f(x)] - \sup_{y \in \mathbb{E}}[\langle \phi, y \rangle - g(y)]$$

which equals to $-f^*(-A^*\phi) - g^*(\phi)$. This completes the proof.

Chapter V
Special cases

March 11, 2023

1 Functions of Eigenvalues

5.2.11 Semidefinite complementarity Suppose matrices X and Y lie in $\mathbb{S}^n_+$.

- If $\mathrm{Tr}(XY) = 0$, prove $-Y \in \partial\delta_{\mathbb{S}^n_+}(X)$.
- Hence prove the following properties are equivalent:
  1. $\mathrm{Tr}(XY) = 0$.
  2. $XY = 0$.
  3. $XY + YX = 0$.
- Prove for any matrices U and V in $\mathbb{S}^n$

$$(U^2 + V^2)^{\frac{1}{2}} = U + V \iff U, V \succeq 0, \mathrm{Tr}(UV) = 0.$$

Proof:

- Note that $\langle -Y, Z - X\rangle = \langle -Y, Z\rangle = -\langle Y, Z\rangle \leq 0$ for all $Z \in \mathbb{S}^n_+$. Thus, $-Y \in \partial\delta_{\mathbb{S}^n_+}(X)$.
- Suppose $\mathrm{Tr}(XY) = 0$ and hence $-Y \in \partial\delta_{\mathbb{S}^n_+}(X)$. Thus, we have $\lambda(-Y) \in \delta_{\mathbb{R}^+_n}(\lambda(X))$ and so $\langle\lambda(-Y), \lambda(X)\rangle = 0$. Thus, $\mathrm{Tr}(X(-Y)) = \langle\lambda(-Y), \lambda(X)\rangle = 0$. Thus, X and $-Y$ have common spectral decomposition. But, $\lambda(X) \geq 0$ and $\lambda(-Y) \leq 0$ and hence $\langle\lambda(-Y), \lambda(X)\rangle$ along with the fact that X and $-Y$ have common spectral decomposition implies $-XY = 0$. The rest is clear.
- Suppose $(U^2 + V^2)^{\frac{1}{2}} = U + V$ then since $U^2 + V^2 \succeq (U^2)^{\frac{1}{2}}$, we should have $U + V = (U^2 + V^2)^{\frac{1}{2}} \succeq (U^2)^{\frac{1}{2}} \succeq U$. Note that if $U = Q\,\mathrm{Diag}(\lambda)Q^T$ for some $\lambda \in \mathbb{R}^n$ and some $Q \in O(n)$, then $(U^2)^{\frac{1}{2}} = Q\,\mathrm{Diag}(|\lambda|)Q^T$. Hence, $V \succeq 0$ and similarly $U \succeq 0$. Now since $(U^2 + V^2)^{\frac{1}{2}} = U + V$ we have $U^2 + V^2 = U^2 + V^2 + UV + VU$ and hence $\mathrm{Tr}(UV) = 0$. The other way is clear.

Chapter VI
Nonsmooth Optimization

March 11, 2023

1 Generalized Derivatives

6.1.2 Continuity of Dini derivative

For a point in $\mathbb{E}$, prove the function $f^-(x;.)$ is Lipschitz if f is locally Lipschitz around x.

Proof: Note that

$$|f^-(x;h_1) - f^-(x;h_2)| = |\liminf_{t\downarrow 0} \frac{f(x+th_1)-f(x)}{t} - \liminf_{t\downarrow 0} \frac{f(x+th_2)-f(x)}{t}|.$$

But, note that lim sup is sublinear and thus

$$\liminf x^r \geq \liminf x^r + y^r + \liminf -y^r \Rightarrow \liminf x^r + y^r - \liminf x^r \leq \limsup y^r.$$

Thus,

$$|\liminf_{t\downarrow 0} \frac{f(x+th_1)-f(x)}{t} - \liminf_{t\downarrow 0} \frac{f(x+th_2)-f(x)}{t}| \leq ||h_1 - h_2||.$$

Similarly,

$$|\liminf_{t\downarrow 0} \frac{f(x+th_2)-f(x)}{t} - \liminf_{t\downarrow 0} \frac{f(x+th_1)-f(x)}{t}| \leq ||h_1 - h_2||.$$

Hence, $|f^-(x;h_1) - f^-(x;h_2)| \leq ||h_1 - h_2||$.

6.1.4 Surjective Dini subdifferentials

Suppose the continuous function $f : \mathbb{E} \to \mathbb{R}$ satisfies the growth condition

$$\lim_{||x||\to+\infty} \frac{f(x)}{||x||} = +\infty. \tag{1}$$

For any element $\phi \in \mathbb{E}$, prove there is a point x in $\mathbb{E}$ with $\phi \in \partial_- f(x)$.

Proof: Note that $\langle \phi, . \rangle - f(.)$ also satisfies 1. Thus without loss of generality suppose $\phi = 0$. Thus, we need to show f has a local minimum if f satisfies 1. Suppose not! Then let $B_i := \{x \in \mathbb{E} : ||x|| \leq i\}$. Then B_i is compact and thus $f|_{B_i}$ obtains its minimum on B_i and assume it happens at $x_i \in B_i$. If $||x_i|| < i$, then x_i is a local minimum for f and we are done. Thus, suppose $||x_i|| = i$. Hence, we have $f(x_{i+1}) < f(x_i)$ by definition of x_{i+1} and the fact that $x_i \in B_{i+1}$ and also x_i is not a local minimum for $f|_{B_{i+1}}$ (and thus the strict inequality). Now note that $||x_i|| \to +\infty$ and thus $\lim_{i\to+\infty} \frac{f(x_i)}{||x_i||} = +\infty$ which is a contradiction as $f(x_i)$ is decreasing.

6.1.6. Failure of Dini calculus

Show that the inclusion

$$\partial_-(f+g)(x) \subseteq \partial_- f(x) + \partial_- g(x)$$

can fail for locally Lipschitz functions f and g.

Proof: Let $f(x) = ||x|| - ||x||^2$ and also $g(x) = ||x||^2$. Then we claim

$$\partial_- g(0) = \{0\}.$$

If $\phi \in \partial_- g(0)$ then for any $||h|| = 1$, and small enough $t > 0$ we have

$$\langle s, th \rangle \leq t^2 ||h||^2 \Rightarrow \langle s, h \rangle \leq t||h||^2 \Rightarrow \langle s, h \rangle = 0.$$

Thus, $s = 0$. It is clear that $0 \in \partial_- g(0)$. However, $(f+g)(x) = ||x||$ and hence $\partial_-(f+g)(0) = B$. Now if $\partial_-(f+g)(0) \subseteq \partial_- f(0) + \partial_- g(0)$ then $B \subseteq \partial_- f(0)$. Now, let $s \in B$ with $||s|| = 1$ and hence $\langle s, ts \rangle \leq t - t^2$ for small enough t. Thus, $t \leq t - t^2$ for small enough t, which is a contradiction.

Side: Note that the function

$$f : B_{\frac{1}{2}} \subseteq \mathbb{E} \to \mathbb{R}, f(x) = \sqrt{1 - ||x||^2}$$

is Lipschitz and also has no Dini subgradient at 0. In fact, suppose $\phi \in \partial_- f(0)$ then for each h with $||h|| = 1$,

$$\langle \phi, th \rangle \leq f(th) - 1 \text{ for small enough } t > 0.$$

But, $f(th) - 1 \leq 0$ and thus $\langle \phi, h \rangle \leq 0$ for all h and hence $\phi = 0$. Thus, $f(th) = 1$ if and only if $th = 0$ or $h = 0$.

6.1.9. Mean value theorem

- Suppose the function $f : \mathbb{E} \to \mathbb{R}$ is locally Lipschitz. For any points x and y, prove there is a real t in $(0, 1)$ satisfying

$$f(x) - f(y) \in \langle x - y, \partial_\diamond f(tx + (1-t)y) \rangle$$

- **Monocity and convexity** If the set C in $\mathbb{E}$ is open and convex and the function $f : C \to \mathbb{R}$ is locally Lipschitz, prove f is convex if and only if it satisfies

$$\langle x - y, \phi - \psi \rangle \geq 0 \text{ for all } x, y \in C, \phi \in \partial_\diamond f(x) \ \& \ \psi \in \partial_\diamond f(y).$$

- If $\partial_\diamond f(y) \subseteq kB$ for all points y near x, prove f has local Lipschitz constant k about x.

Proof:

- To be done!
- Suppose that f is convex on C, then $\partial_\diamond f(x) = \partial f(x)$ and thus

$$f(y) \geq f(x) + \langle \phi, y - x \rangle, f(x) \geq f(y) + \langle \psi, y - x \rangle \Rightarrow \langle \phi - \psi, x - y \rangle \geq 0.$$

Conversely, suppose that the above statement holds. Then, let $\phi \in \partial_\diamond f(x)$, note that this set is nonempty since f is locally Lipschitz on $\mathbb{E}$. Then,

$$f(y) - f(x) = \langle y - x, \psi \rangle \text{ for some } \psi \in f(ty + (1-t)x) \text{ for some } t \in (0, 1).$$

Now it suffices to prove that $\langle y - x, \psi \rangle \geq \langle y - x, \phi \rangle$. But we have

$$\langle tx + (1-t)y - x, \psi - \phi \rangle \geq 0 \Rightarrow \langle y - x, \psi - \phi \rangle \geq 0.$$

- Let y, z be in a small neighborhood about x, then

  $f(y) - f(z) = \langle y - z, \phi \rangle$ for some $\phi \in \partial_{\diamond} f(w)$ wherein w lies on the line segment $[y, z]$.

  Thus, $|f(y) - f(x)| \leq k||y - z||$.

6.1.11 Order statistics

Calculate the Dini, the Michel-Penot, and the Clarke directional derivatives and differentials of the function

$$x \in \mathbb{R}^n \to [x]_k.$$

Proof:

Dini directional derivative

Suppose that

$$[x]_1 = \cdots = [x]_{l_1} > [x]_{l_1+1} = \cdots = [x]_{l_1+l_2} > \cdots > [x]_{l_1+l_2+\cdots+l_{t-1}+1} = \cdots = [x]_{l_1+\cdot+l_t},$$

and assume

$$h_{i_{1,1}} \geq h_{i_{1,2}} \geq \cdots \geq h_{i_{1,l_1}}, h_{i_{2,1}} \geq \cdots \geq h_{i_{2,l_2}}, \cdots, h_{i_{t,1}} \geq \cdots \geq h_{i_{t,l_t}},$$

wherein, for all $1 \leq j \leq t$

$$S_j := \{i_{j,l_l+\cdots+l_{j-1}+1}, \cdots, i_{j,l_l+\cdots+l_{j-1}+l_j}\} = \{l_1+\cdots+l_{j-1}+1, l_1+\cdots+l_{j-1}+2, \cdots, l_1+\cdots+l_{j-1}+l_j\}.$$

Then, for $t > 0$ small enough,

$$[x + th]_k = [x]_k + th_{i_{j,k}} \text{ where } k \in S_j.$$

Thus, $[.]_k^-(x; h) = h_{i_{j,k}}$.

Michel-Penot directional derivative

From the above discussion we have $[.]_k^\diamond(x; h) = h_{i_{j,1}}$ where $k \in S_j$.

6.1.12 Closed subdifferentials

- Suppose the function $f : \mathbb{E} \to (\infty, +\infty]$ is convex, and the point x lies in $\text{int}(\text{dom}\, f)$. Prove the convex subdifferential $\partial f(.)$ is closed at x; in other words, $x^r \to x$ and $\phi^r \to \phi$ in $\mathbb{E}$ with ϕ^r in $\partial f(x^r)$ implies $\phi \in \partial f(x)$.

  Suppose the real function f is locally Lipschitz around the point x in $\mathbb{E}$.

- For any direction h in $\mathbb{E}$, prove the Clarke directional derivative has the property that $-f^\circ(.; h)$ is lower semicontinuous at x.

- Deduce the Clarke subdifferential is closed at x.

- Deduce further the inclusion $\subseteq$ in the Intrinsic Clarke subdifferential theorem:

  $$\partial_\circ f(x) = \text{conv}\{\lim_r \nabla f(x^r) : x^r \to x, x^r \notin S\},$$

  wherein outside of the measure zero set S, f is Gateaux differentiable.

- Show that Dini and Michel-Penot subdifferentials are not necessary closed.

Proof:

- Note that f on int(dom f) is continuous and thus closed. Hence, the proof is complete due to 4.2.8.

- We need to show for any $x^r \to x$ we have
$$\liminf_r -f^\circ(x^r;h) \geq -f^\circ(x;h) \iff \limsup_r f^\circ(x^r;h) \leq f^\circ(x;h).$$
This holds if and only if
$$\limsup_r \limsup_{t\downarrow 0, y\to x^r} \frac{f(y+th)-f(y)}{t}.$$
Let $\epsilon > 0$, then let y^r, t_r be such that $||y^r - x^r|| \leq \frac{1}{r}$ and $t_r < \frac{1}{r}$.
$$|\frac{f(y^r+t_rh)-f(y^r)}{t_r} - \limsup_{t\downarrow 0, y\to x^r} \frac{f(y+th)-f(y)}{t}| \leq \frac{\epsilon}{2^r}.$$
Then,
$$\limsup_r \limsup_{t\downarrow 0, y\to x^r} \frac{f(y+th)-f(y)}{t} = \limsup_r \frac{f(y^r+t_rh)-f(y^r)}{t_r} \leq \limsup_{t\downarrow 0, y\to x} \frac{f(y+th)-f(y)}{t} = f^\circ(x;h)$$

- Let $x^r \to x$ and also $\phi^r \to \phi$ where $\phi^r \in \partial_\circ f(x^r)$. Then we wish to show that $\phi \in \partial_\circ f(x)$. This holds true if and only if
$$\langle \phi, h\rangle \leq f^\circ(x;h).$$
But,
$$\langle \phi, h\rangle = \lim_r \langle \phi^r, h\rangle \leq \limsup_r f^\circ(x^r,h) \leq f^\circ(x;h).$$
This completes the proof.

- Note that $\text{conv}\{\lim_r \nabla f(x^r) : x^r \to x, x^r \notin S\} \subseteq \partial_\circ f(x)$ as Clarke subdifferntials are convex and closed. Then we claim that $\text{conv}\{\lim_r \nabla f(x^r) : x^r \to x, x^r \notin S\}$ is compact. In fact, $\text{conv}\{\lim_r \nabla f(x^r) : x^r \to x, x^r \notin S\} \subseteq \partial_\circ f(x)$ and thus $\text{conv}\{\lim_r \nabla f(x^r) : x^r \to x, x^r \notin S\}$ is bounded as $\partial_\circ f(x)$ is compact. Let $s_i = \lim_r \nabla f(x_i^r) \in \partial_\circ f(x)$ with $x_i^r \to x$. Let $||x_i^{r_i} - x|| < \frac{1}{i}$ and $||s_i - \nabla f(x_i^{r_i})|| < \frac{1}{i}$. Now
$$\lim_j \nabla f(x_j^{r_j}) = \lim_j s_j.$$
Thus, $\lim_j s_j \in \text{conv}\{\lim_r \nabla f(x^r) : x^r \to x, x^r \notin S\}$.

Now let $s \in \partial_\circ f(x) \backslash \text{conv}\{\lim_r \nabla f(x^r) : x^r \to x, x^r \notin S\}$, then there exists $\phi \in \mathbb{E}$ such that
$$\langle s, \phi\rangle < a < b \leq \langle \lim_r \nabla f(x^r), \phi\rangle \text{ wherein } x^r \to x, x^r \notin S.$$
Let $\phi = y - x$. Then choose $x^r \to x$ with $x^r \notin S$ and $x^r \in [x,y]$. Thus,
$$0 < b - a \leq \langle \nabla f(x^r) - s, y - x\rangle.$$
We obtain a contradiction as we tend r to infinity.

2 Regularity and Strict Differentiability

6.2.6.

Prove that a unique Clarke subgradient implies regularity. Note that the function is Lipschitz about the point x.

Proof: Recall that Clarke subgradient is a unique vector ϕ if and only if

$$\limsup_{y\to x, t\downarrow 0} \frac{f(y+th)-f(y)}{t} = \langle \phi, h\rangle.$$

Note that

$$f^{-}(x;h) = -\limsup_{t\downarrow 0} \frac{f((x+th)-th)-f(x+th)}{t} = -\langle \phi, -h\rangle = \langle \phi, h\rangle.$$

This completes the proof.

6.2.7 Strict differentiability

A real function f has strict derivative ϕ at a point x in $\mathbb{E}$ if and only if it is locally Lipschitz around x with

$$\lim_{y\to x, t\downarrow 0} \frac{f(y+th)-f(y)}{t} = \langle \phi, h\rangle$$

for all direction h in $\mathbb{E}$. In particular, this holds if f is continuously differentiable around x with $\nabla f(x) = \phi$.

Proof: First, suppose that f has strict derivative ϕ at x. Then if f is not locally Lipschitz around x, then for any fixed $C \in \mathbb{R}_{++}$ and for every i, there exists $y_i, z_i \in B_{\frac{1}{i}}(x)$ such that $|f(y_i)-f(z_i)| > C||y_i - z_i||$. However,

$$0 = \lim_{i\to+\infty, t\downarrow 0} |\frac{f(y_i)-f(z_i)-\langle \phi, y_i - z_i\rangle}{||y_i - z_i||}| \geq C - \limsup_{i\to+\infty} \frac{\langle \phi, y_i - z_i\rangle}{||y_i - z_i||},$$

so,

$$||\phi|| \geq \limsup_{i\to+\infty} \frac{\langle \phi, y_i - z_i\rangle}{||y_i - z_i||} \geq C,$$

which is a contradiction. Thus, f is locally Lipschitz around x. Now, fix h and let $y \leftarrow y + th$ and $z \leftarrow y$. Thus,

$$0 = \lim_{y\to x, t\downarrow 0} \frac{f(y+th)-f(y)-t\langle \phi, h\rangle}{t} \Rightarrow \lim_{y\to x, t\downarrow 0} \frac{f(y+th)-f(y)}{t} = \langle \phi, h\rangle.$$

Conversely, suppose that

$$\lim_{y\to x, t\downarrow 0} \frac{f(y+th)-f(y)}{t} = \langle \phi, h\rangle$$

for all $\phi \in \mathbb{E}$ and also f is locally Lipschitz around x, then we wish to prove

$$\lim_{y,z\to x, y\neq z} \frac{f(y)-f(z)-\langle \phi, y-z\rangle}{||y-z||} = 0.$$

However, the above equals to,

$$\lim_{z\to x, w\in S^1, t\downarrow 0} \frac{f(z+tw)-f(z)-\langle \phi, tw\rangle}{t} = \lim_{z\to x, w\in S^1, t\downarrow 0} \frac{f(z+tw)-f(z)}{t} - \langle \phi, w\rangle = \lim_{z\to x, w\in S^1, t\downarrow 0} g(z,w,t).$$

Now suppose the above does not hold, then there exists (z_i, w_i, t_i) with $t_i \downarrow 0$ and $w_i \in S^1$ and also $||w_i|| \to ||w||$ such that $|g(z_i, w_i, t_i)| \geq \epsilon$ for some $\epsilon > 0$. However,

$$|g(z_i, w_i, t_i) - g(z_i, w, t_i)| \leq ||w - w_i|| + |\langle \phi, w - w_i\rangle|.$$

However, $|g(z_i, w, t_i)| \to 0$ and thus $g(z_i, w_i, t_i) \to 0$ as desired.

Now if f is continuously differentiable then $||\nabla f||$ is bounded above in a neighborhood of x and thus

$$||f(x+h) - f(x)|| \leq ||\nabla f(x+th)|| ||h|| \leq C||h||,$$

for some constant C; note that $t \in (0,1)$ comes from the Taylor expansion. Now for each $t \in (0, \epsilon)$ for some small enough ϵ, there exists $t^* \in (0,t)$ such that

$$\frac{f(y+th)-f(y)}{t} = \nabla f(y + t^* h)^T h.$$

Now if $y \to x$ and $t \downarrow 0$, then the above tends to $\nabla f(x)^T h = \langle \phi, h\rangle$.

6.1.8

Prove the following results:

- $f^\circ(x; -h) = (-f)^\circ(x; h)$
- $(\lambda f)^\circ(x; h) = \lambda f^\circ(x; h)$ for $0 \leq \lambda \in \mathbb{R}$.
- $\partial_\circ(\lambda f)(x) = \lambda \partial_\circ f(x)$ for all λ in $\mathbb{R}$.

Proof:

- Note that
$$f^\circ(x; -h) = \limsup_{y\to x, t\downarrow 0} \frac{f(y-th)-f(y)}{t},$$
and
$$(-f)^\circ(x;h) = \limsup_{y\to x, t\downarrow 0} \frac{-f(y+th)+f(y)}{t} = \limsup_{y-th\to x, t\downarrow 0} \frac{-f((y-th)+th)+f(y-th)}{t}.$$
Now note that $y \to x$ is the same as $y - th \to x$.
- $$(\lambda f)^\circ(x;h) = \limsup_{y\to x, t\downarrow 0} \frac{\lambda f(y+th)-\lambda f(y)}{t} = \lambda \limsup_{y\to x, t\downarrow 0} \frac{f(y+th)-f(y)}{t} = \lambda f^\circ(x;h)$$
- $$\partial_\circ(\lambda f)(x) = \{\phi : \langle \phi, h\rangle \leq \lambda f^\circ(x;h)\ \forall h \in \mathbb{E}\}$$

6.2.9. Mixed sum rules

Suppose that the real function f is locally Lipschitz around the point x in $\mathbb{E}$ and that the function $g : \mathbb{E} \to (infty, +\infty]$ is convex with $x in \operatorname{int}(\operatorname{dom} g)$. Prove:

- $\partial_\diamondsuit(f+g)(x) = \nabla f(x) + \partial g(x)$ if f is Gateaux differentiable at x.
- $\partial_\circ(f+g)(x) = \nabla f(x) + \partial g(x)$ if f is strictly differentiable at x.

Proof:

- We have

$$(f+g)^\diamondsuit(x;h) = \sup_{u\in\mathbb{E}} \limsup_{t\downarrow 0} \frac{(f+g)(x+th+tu)-(f+g)(x+th)}{t}$$

$$= \langle \nabla f(x), h\rangle + \sup_{u\in\mathbb{E}} \limsup_{t\downarrow 0} \frac{g(x+th+tu)-g(x+th)}{t} = \langle \nabla f(x), h\rangle + g'(x;h).$$

  Thus $\nabla f(x) + \phi \in \partial_\diamondsuit(f+g)(x)$ if and only if $\langle \phi, h\rangle \le g'(x;h)$.

- We have

$$(f+g)^\circ(x;h) = \limsup_{y\to x, x\downarrow 0} \frac{(f+g)(y+th)-(f+g)(y)}{t}$$

$$= \limsup_{y\to x, t\downarrow 0} \frac{f(y+th)-f(y)}{t} + \limsup_{y\to x, t\downarrow 0} \frac{g(y+th)-g(y)}{t}.$$

$$= \langle \nabla f(x), h\rangle + g'(x;h).$$

  Thus $\nabla f(x) + \phi \in \partial_\circ(f+g)(x)$ if and only if $\langle \phi, h\rangle \le g'(x;h)$.

6.2.13 Dense Dini subgradients

Suppose the real function f is locally Liptschitz around the point x in $\mathbb{E}$. By considering the closet point in $\operatorname{epi} f$ to the point $(x, f(x)-\delta)$ (for a small real $\delta > 0$), prove there are Dini Subgradients at points arbitrary close to x.

Proof:

Lemma: Let $B_{\frac{\sqrt{2}}{2}}$ be the ball of radius $\frac{\sqrt{2}}{2}$ around the origin. Then the function $f : \mathbb{E} \to \mathbb{R}$ with $f(x) = \sqrt{1-||x||^2}$ is Lipschitz on $B_{\frac{\sqrt{2}}{2}}$.

Proof of Lemma: Note that

$$|\sqrt{1-||x||^2} - \sqrt{1-||y||^2}| \le |\sqrt{1-||x||^2} - \sqrt{1-||y||^2}||\sqrt{1-||x||^2} + \sqrt{1-||y||^2}| = |||x||^2 - ||y||^2| \le 2|||x|| -$$

Note that f has no Dini subgradients at 0.

Now let (y, r) to be the closest point on the epigraph from $(x, f(x)-\delta)$. We claim $r = f(y)$. In fact, suppose that $r > f(y)$ and therefore

$$d^2 = (r - f(x) + \delta)^2 + ||x-y||^2 \le (f(y) - f(x) + \delta)^2 + ||x-y||^2 \Rightarrow (r - f(y))(r + f(y) + 2\delta - 2f(x)) \le 0.$$

Thus, $2f(y) + 2\delta - 2f(x) \leq r + f(y) + 2\delta - 2f(x) \leq 0$. Thus, $f(x) - \delta \geq f(y)$. Note that if $f(y) < f(x) - \delta$, then there exists y' closed enough to y such that $f(y') \leq f(x) - \delta$ and also $||x - y'|| < ||x - y||$. So, $(y', f(x) - \delta)$ is closer to $(x, f(x) - \delta)$ than $(y, f(x) - \delta)$. Thus, $f(y) = f(x) - \delta$. After all, $r = f(x) - \delta$ which is a contradiction. Thus, $r = f(y)$ and also $f(y) \geq f(x) - \delta$. Now if $y \neq x$, then choose $||x - y|| > \epsilon > 0$ small enough such that $|f(z) - f(x)| < \delta$ for all $||z - x|| < \epsilon$. Then we have

$$(f(z) - f(x) + \delta)^2 + ||x - z||^2 \geq (f(y) - f(x) + \delta)^2 + ||x - y||^2 \geq (f(y) - f(x) + \delta)^2 + ||x - z||^2.$$

Thus, $f(z) - f(x) + \delta \geq f(y) - f(x) + \delta$, or $f(z) \geq f(x)$. Thus, f is a local minimum of hence $0 \in \partial_- f(x)$. Thus suppose that $x = y$ and $d = \delta$. Hence,

$$(f(y) - f(x) + \delta)^2 + ||x - y||^2 \geq \delta^2 \Rightarrow f(y) - f(x) \geq \sqrt{\delta^2 - ||x - y||^2} - \delta \text{ for } y \text{ close enough to } x.$$

So for y closed enough to x we have

$$f(y) - f(x) \geq \delta[\sqrt{1 - (\frac{||y - x||}{\delta})^2} - 1].$$

But, the RHS has subgradients for points arbitrary close to x.

---

3 Tangent Cones

6.3.1 Exact penalization

For a set $U \subseteq \mathbb{E}$, suppose that the function $f : U \to \mathbb{R}$ has Lipschitz constant L', and that the set $S \subseteq U$ is closed. For any $L > L'$, if the point x minimizes $f + Ld_S$ on U, prove $x \in S$.

Proof: Suppose $x \in U$ is not in S and also $y \in S$ such that $||y - x|| = d_S(x)$. Then we have

$$(f + Ld_S)(x) \leq (f + Ld_S)(y) = f(y) \Rightarrow Ld_S(x) \leq f(y) - f(x) \leq L'||y - x|| < L||y - x||.$$

Thus, $d_S(x) < ||y - x||$. This contradiction completes the proof.

6.3.3 Examples of tangent cones

For the following sets $S \subseteq \mathbb{R}^2$, calculate $T_S(0)$ and $K_S(0)$:

- $\{(x, y) : y \geq x^3\}$.
- $\{(x, y) : x \geq 0, y \geq 0\}$.
- $\{(x, y) : x = 0 \text{ or } y = 0\}$.
- $\{r(\cos\theta, \sin\theta) : 0 \leq r \leq 1, \frac{\pi}{4} \leq \theta \leq \frac{7\pi}{4}\}$.

6.3.4 Topology of contingent cone

Prove that the contingent cone is closed, and derive the following topological description: Suppose $x \in S$. The contingent cone $K_S(x)$ consists of those vectors h in $\mathbb{E}$ such that there are sequences $t_r \downarrow 0$ in $\mathbb{R}$ and $h^r \to h$ in $\mathbb{E}$ such that $x + t_r h^r$ lies in S for all r.

Proof: Recall that

$$K_S(x) = \{h : d_S^-(x;h) = 0\}.$$

Since, $x \in S$, x is a local minimum for d_S and thus $0 \leq d_S^-(s;h)$ for all $h \in \mathbb{E}$, so $T_S(x) \subseteq K_S(x)$. However, suppose $d_S^-(x;h) = 0$ and hence $\liminf_{t\downarrow 0} \frac{d_S(x+th)}{t} = 0$. Thus, there exists $t_r \downarrow 0$ such that $d_S(x+t_rh)/t_r \to 0$. Thus, if $x + t_rh^r \in S$ such that $||x + t_rh - x - t_rh^r|| \leq d_S(x+t_rh) + t_r^2$. Thus, $||h - h^r|| \to 0$ and hence $h^r \to h$.

Conversely, suppose that $x + t_rh^r \in S$ and $h^r \to h$ and also $t_r \downarrow 0$. Then wish to show that $h \in K_S(x)$. Note that

$$d_S(x+t_rh) \leq ||x+t_rh-x-t_rh^r|| = t_r||h-h^r|| \Rightarrow 0 \leq \liminf_{t\downarrow 0} \frac{d_S(x+th)}{t} \leq \lim_{t_r \to +\infty} \frac{d_S(x+t_rh^r)}{t_r} = 0.$$

Thus, $d_S^-(x;h) = 0$ and $h \in K_S(x)$.

6.3.5 Topology of Clarke cone

Suppose that x lies in the set $S \subseteq \mathbb{E}$.

- Prove $d_S^\circ(x;.) \geq 0$.
- Prove
$$d_S^\circ(x;h) = \limsup_{y\to x \text{ in } S, t\downarrow 0} \frac{d_S(y+th)}{t}.$$
- Prove that the Clarke Tangent cone consists of those vectors h in $\mathbb{E}$ such that for any sequence $t_r \downarrow 0$ in $\mathbb{R}$ and $x^r \to x$ in S, there is a sequence $h^r \to h$ such that $x^r + t_rh^r$ lies in S for all r.

Proof:

- Let $h \in \mathbb{E}$, then
$$d_S^\circ(x;h) = \limsup_{y\to x, t\downarrow 0} \frac{d_S(y+th) - d_S(y)}{t} \geq \limsup_{t\downarrow 0} \frac{d_S(x+th)}{t} \geq 0.$$
- Note that
$$d_S^\circ(x;h) = \limsup_{y\to x, t\downarrow 0} \frac{d_S(y+th) - d_S(y)}{t} \geq \limsup_{y\to x \text{ in } S, t\downarrow 0} \frac{d_S(y+th)}{t}.$$
Now fix some $\epsilon > 0$. Note that $||y'+th-y''|| \leq d_S(y'+th)+\frac{1}{2}\epsilon$ and $||y-y'|| \leq d_S(y)+\frac{1}{2}\epsilon$ for some $y', y'' \in S$. Thus,
$$d_S(y+th) \leq ||y+th-y''|| \leq d_S(y'+th)+d_S(y)+\epsilon \Rightarrow d_S(y'+th)+\epsilon \geq d_S(y+th)-d_S(y).$$

Thus,

$$\limsup_{y\to x, t\downarrow 0} \frac{d_S(y+th)-d_S(y)}{t} \le \limsup_{y'\to x \text{ in } S, t\downarrow 0} \frac{d_S(y'+th)}{t} + \epsilon.$$

Thus,

$$\limsup_{y\to x, t\downarrow 0} \frac{d_S(y+th)-d_S(y)}{t} \le \limsup_{y\to x \text{ in } S, t\downarrow 0} \frac{d_S(y+th)}{t}.$$

This completes the proof.

- Suppose h has the aforementioned properties then

$$d_S^\circ(x;h) = \limsup_{y\to x \text{ in } S, t\downarrow 0} \frac{d_S(y+th)}{t} \le \limsup_{x^r\to x \text{ in } S, t_r\downarrow 0, h^r\to h} \frac{||x^r+t_rh-x^r-t_rh^r||}{t_r} = 0.$$

Thus, $h \in T_S(x)$. Conversely, suppose that $d_S^\circ(x;h)=0$. Then for every $x^r \to x$ and every $t_r \to 0$, we must have $\lim_{r\to+\infty} d_S(x^r+t_rh)/t_r = 0$. Let $y^r \in S$ such that $||x^r+t_rh-y^r|| \le d_S(x^r+t_rh)+t_r^2$. Suppose $h^r \in \mathbb{E}$ such that $y^r = x^r + t_rh^r$. Then $||h-h^r|| \le d_S(x^r+t_rh)/t_r + t_r$. Thus, $h^r \to h$. Since, $x^r+t_rh^r \in S$, we are done.

6.3.8 Isotonicity

Suppose $x \in U \subseteq V \subseteq \mathbb{E}$. Prove $K_U(x) \subseteq K_V(x)$, but give an example where $T_U(x) \not\subseteq T_V(x)$.

Proof: Recall that

$$K_S(x) = \{h : d_S^-(x;h) = 0\}.$$

Now we want to show that $d_U^-(x;h)=0$ implies $d_V^-(x;h)=0$. Note that

$$0 \le d_V^-(x;h) = \liminf_{t\downarrow 0} \frac{d_V(x+th)}{t} \le \liminf_{t\downarrow 0} \frac{d_U(x+th)}{t} = 0.$$

This completes the proof.

Now recall that

$$d_S^\circ(x;h) = \limsup_{y\to x \text{ in } S, t\downarrow 0} \frac{d_S(y+th)}{t}.$$

So we wish to find $U \subseteq V$ and some $h \in \mathbb{E}$ such that

$$\limsup_{y\to x \text{ in } U, t\downarrow 0} \frac{d_U(y+th)}{t} = 0 \text{ but } \limsup_{w\to x \text{ in } V, t\downarrow 0} \frac{d_V(w+th)}{t} \ne 0.$$

Let $U = S^1$ and $V = S^1 \cup \{(x,y) : x \ge 1, y = 0\}$ and also $h = (0,1)$ and $x = (1,0)$. Then we first show that

$$d_U(x;h) = 0.$$

In fact, let $p_i = (x_i, y_i) \to x$ inside U, then $x_i^2 + y_i^2 = 1$ and $x_i \to 1$ and $y_i \to 0$. Note that $d_U(p_i + th) = \sqrt{x_i^2 + (y_i+t)^2} - 1$. Hence,

$$d_U^\circ(x;h) = \limsup_{i\to+\infty, t\to 0} \frac{x_i^2+(y_i+t)^2-1}{t(\sqrt{x_i^2+(y_i+t)^2}+1)} = \limsup_{i\to+\infty, t\to 0} \frac{t^2+2ty_i}{t(\sqrt{1+t^2+2ty_i}+1)},$$

which equals to

$$\limsup_{i\to+\infty,t\to 0} \frac{t+2y_i}{\sqrt{1+t^2+2ty_i}+1} = 0.$$

Now let $p_t = (\sqrt{2t+1}, 0)$. Note that $p_t + th = (\sqrt{2t+1}, t)$ with distance $\sqrt{2t+1+t^2} - 1 = t$ to U. Hence, $d_V(p_t + th) = t$. Hence,

$$\lim_{t\downarrow 0} \frac{d_V(p_t+th)}{t} = 1 \le \limsup_{w\to x \text{ in } V, t\downarrow 0} \frac{d_V(w+th)}{t} \neq 0.$$

6.4.3 Local minimizers
Consider a function $f : \mathbb{E} \to [-\infty, +\infty]$ which is finite at the point $x \in \mathbb{E}$.

- If x is local minimizer, prove $0 \in \partial_- f(x)$.
- If $0 \in \partial_- f(x)$, prove for any $\delta > 0$ that x is a strict local minimizer of the function $f(.) - \delta||. - x||$.

Proof:

- We know that $0 \in \partial_- f(x)$ if and only if $f^-(x; h) \ge 0$ for all $h \in \mathbb{E}$. However,

$$f^-(x;h) = \liminf_{t\downarrow 0, h'\to h} \frac{f(x+th') - f(x)}{t} \ge 0 \text{ as } x \text{ is a local minimizer.}$$

- Now suppose that $0 \in \partial_- f(x)$. Then if x is not a strict local minimizer for $f() + \delta||.||$, then there exists $x_i \to x$ such that

$$f(x_i) + \delta||x_i|| \le f(x).$$

Let $x_i = x + t_i u_i$ where $u_i = \frac{x_i - x}{||x_i - x||}$ and also $t_i = ||x_i - x|| \to 0$. Also, assume $u_i \to u$. Then

$$0 \le f^-(x;u) = \liminf_{t\downarrow 0, v\to u} \frac{f(x+tv)-f(x)}{t} \le \liminf_i \frac{f(x+t_iu_i)-f(x)}{t_i} \le \liminf_i \frac{-\delta||x_i||}{t_i}.$$

If $x = 0$ then $\frac{-\delta||x_i||}{t_i} = -\delta < 0$ which is a contradiction. If $x \neq 0$, then $\liminf_i \frac{-\delta||x_i||}{t_i} = -\infty$, again a contradiction.

6.4.6. Prove a limiting sub differential sum rule for a finite number of lower semi continuous functions, with all but one being locally Lipschitz.
Proof:
Let $f_1, \cdots, f_k$ be lower semicontinuous at x and also g locally Lipschitz around x. Recall the Fuzzy sum rule:
Fuzzy sum rule: Fix $\delta > 0$. Then

$$\partial_-(\sum_{i=1}^k f_i + g)(x) \subseteq \delta B + \sum_{i=1}^k \partial_-(f_i)(U(f_i, x, \delta)) + \partial_-(g)(U(f_i, x, \delta)).$$

Let $\phi^r \in \partial_-(\sum_{i=1}^k f_i + g)$ and also $\phi_i^r \in \partial_-(f_i)(x^r)$ and $\psi^r \in \partial_- g(y^r)$ such that

$$||\phi^r - \sum_{i=1}^k \phi_i^r - \psi^r|| < \frac{1}{r} \ , \ ||x^r - x|| < \frac{1}{r}, ||f_i(x^r) - f_i(x)|| < \frac{1}{r}, ||g(y^r) - g(x)|| < \frac{1}{r}, ||y^r - y|| < \frac{1}{r}$$

Note that for all $\psi \in \partial_- g(x')$ we have $\langle \psi, v \rangle \leq C||v||$. Hence, $||\psi|| \leq C$. So, suppose that $\psi^r \to \psi$. 0

6.4.7 Limiting and Clarke sub differentials

Suppose the real function f is locally Lipschitz around the point x in $\mathbb{E}$.

- Use the fact that the Clarke sub differential is a closed multi-function to show $\partial_a f(x) \subseteq \partial_\circ f(x)$.
- Deduce from the Intrinsic Clarke sub differential theorem the property $\partial_\circ f(x) = \operatorname{conv} \partial_a f(x)$.
- Prove $\partial_a f(x) = \{\phi\}$ if and only if ϕ is the strict derivative of f at x.

Proof:

- Let $\phi \in \partial_a f(x)$, then there exists $\phi_i \in \partial_- f(x^i)$ for some $x^i \to x$ such that $\phi_i \to \phi$. Then $\phi_i \in \partial_\circ f(x^i)$ and thus $\phi \in \partial_a f(x)$ as Clarke sub differentials are closed under limit.
- Now since $\partial_\circ f(x)$ is convex we have $\operatorname{conv} \partial_a f(x) \subseteq \partial_\circ f(x)$. On the other hand,

$$\begin{aligned}\partial_\circ f(x) = \operatorname{conv}\{\lim \nabla f(x^i) : f \text{ is differentiable at } x^i \text{ and also } x^i \to x\} \subseteq \\ \operatorname{conv}\{\lim \phi_i : \phi_i \in \partial_- f(x^i) \text{ and also } x^i \to x\} = \operatorname{conv} \partial_a f(x).\end{aligned}$$

- Note that $\partial_\circ f(x)$ is a singleton if and only if $\partial_a f(x)$ is a singleton.