CO 463/663: Convex Optimization and Analysis

Rui Gong

April 20, 2021

Acknowledgements

These notes are based on the CO463/663 lectures given by Professor *Walaa Moursi* in Winter 2021 at the University of Waterloo.

Contents

1 Convex Sets

1.1 Introduction

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable. Consider the problem

$$(P) : \begin{array}{c} min\ f(x) \\ s.t.\ x \in C \subseteq \mathbb{R}^n \end{array}$$

In the special case, when $C = \mathbb{R}^n$, the minimizers of f (if any) will occur at the critical points of f, namely, $x \in \mathbb{R}^n$ such that

$$\nabla f(x) = 0$$

This is known as "Fernat's Rule", which we will learn about more later.
In this course, we will discuss and learn ConVexity of sets and functions and how we can approach problem (P) in the more general settings of:

1. Absense of differentiability of the function f, f is convex (this is called the objective function) **and/or**

2. $\emptyset \neq C \subsetneq \mathbb{R}^n$, C convex (C is called the constraint set)

1.2 Affine sets and affine subspaces in $\mathbb{R}^n$

Definition 1.1

Let $S \subseteq \mathbb{R}^n$. Then:

1. *S is an affine set if*

$$\forall x, y \in S, \forall \lambda \in \mathbb{R}, \lambda x + (1 - \lambda)y \in S$$

   Observe that, trivially, $\emptyset, \mathbb{R}^n$ are affines sets.

2. *S is an affine subspace if*

$$S \neq \emptyset$$

   and

$$\forall x, y \in S, \forall \lambda \in \mathbb{R}, \; \lambda x + (1 - \lambda)y \in S$$

3. *Let $S \subseteq \mathbb{R}^n$. The affine hull of S, denoted by $aff(S)$ is the intersection of all affine sets containing S (i.e. the smallest affine set containing S)*

Example: Affine Sets of $\mathbb{R}^n$

1. *L, where $L \subseteq \mathbb{R}^n$ is a linear subsapce*
2. *$a + L$, where $a \in \mathbb{R}^n, L \subseteq \mathbb{R}^n$ is a linear subsapce*
3. *$\emptyset$, $\mathbb{R}^n$*

Geometrically Speaking:
A nonempty subset $S \subset \mathbb{R}^n$ is affine if the line connecting any two points in the set lies entirely in the set.

1.3 Convex Sets in $\mathbb{R}^n$

Definition 1.2

A subset C of $\mathbb{R}^n$ is convex if

$$\forall x, y \in C, \forall \lambda \in (0,1),\ \lambda x + (1-\lambda)y \in C$$

Example

Convex subsets of $\mathbb{R}^n$

1. $\emptyset$, $\mathbb{R}^n$
2. C, *where* C *is a ball*
3. C, *where* C *is an affine set*
4. C, *where* C *is a half-space. i.e.*

$$C := \{x \in \mathbb{R}^n | \langle x, u \rangle \leqslant \eta\}$$

where $u \in \mathbb{R}^n, \eta \in \mathbb{R}$ *are fixed*

Geometrically Speaking:
A subset $C \subseteq \mathbb{R}^n$ is convex if given any two points $x \in C$, $y \in C$, the line segment joining x and y, denoted by $[x, y]$, lies entirely in C

Theorem 1.3: Txtbook THM2.1

The intersection of an arbitrary collection of convex sets is convex.

Proof. Let I be an indexed set (not necessarily finite). Let $(C_i)_{i\in I}$ be a collecotion of convex subsets of $\mathbb{R}^n$. Set

$$C := \cap_{i\in I} C_i$$

Let $\lambda \in (0,1)$ and let $(x,y) \in C \times C$.
Since C_i is convex $(\forall i \in I)$, we learn that

$$\forall i \in I,\ \lambda x + (1-\lambda)y \in C_i$$

Hence,

$$\lambda x + (1-\lambda)y \in \cap_{i\in I} C_i = C$$

Hence, C is convex. □

Corollary: Txtbook Cor 2.1.1

Let $b_i \in \mathbb{R}^n$, $\beta_i \in \mathbb{R}$ for $i \in I$, where I is an arbitrary index set.
Then the set:

$$C = \{x \in \mathbb{R}^n | \langle x, b_i\rangle \leqslant \beta_i, \forall i \in I\}$$

is convex.

Proof. Set $\forall i \in I$,

$$C_i = \{x \in \mathbb{R}^n | \langle x, b_i\rangle \leqslant \beta_i\}$$

We claim that $\forall i \in I$, C_i is convex.
Indeed, let $i \in I$, let $(x,y) \in C_i \times C_i$, and let $\lambda \in (0,1)$. Set

$$z := \lambda x + (1-\lambda)y$$

Then

$$\begin{aligned}\langle z, b_i\rangle &= \langle \lambda x + (1-\lambda)y, b_i\rangle \\ &= \lambda \underbrace{\langle x, b_i\rangle}_{\leqslant \beta_i} + (1-\lambda)\underbrace{\langle y, b_i\rangle}_{\leqslant \beta_i} \\ &\leqslant \lambda\beta_i + (1-\lambda)\beta_i (\text{ Using } 1 > \lambda > 0,\ x, y \in C_i) \\ &= \beta_i\end{aligned}$$

Hence, $z \in C_i$
Consequently, C_i is convex, as claimed.
Now, combine with theorem 2.1 □

1.4 Convex Combinations of Vectors:

Definition 1.4

A vector sum

$$\lambda_1 x_1 + \ldots + \lambda_m x_m$$

is called a convex combination of vectors $x_1, \ldots, x_m$ *if* $\forall i \in \{1, \ldots, m\}, \lambda_i \geqslant 0$, *and* $\sum_{i=1}^{m} \lambda_i = 1$

Theorem 1.5: Txtbook THM2.2

A subset C *of* $\mathbb{R}^n$ *is convex iff it contains all the convex combination of its elements*

Proof. ($\Leftarrow$) Suppose C contains all the convex combinations of its elemtns.
Let $\lambda \in (0,1)$ and let $x \in C$, $y \in C$.
By assumption, the convex combination

$$\lambda x + (1-\lambda)y$$

lies in C.
Therefore, C is convex.
($\Rightarrow$) Suppose C is convex.
We proceed by induction on m, where m is the number of elements in the convex combination.
Base case: when $m = 2$, the conclusion is clear by the convexity of C.
Now, suppose that for some $m > 2$ it holds that any convex combination of m vectors lies in C.
Let $\{x_1, \ldots, x_m\} \subseteq C$, let $\lambda_1, \ldots, \lambda_m, \lambda_{m+1} \geqslant 0$, such that

$$\sum_{i=1}^{m+1} = 1$$

our goal is to show that

$$z := \sum_{i=1}^{m+1} \lambda_i x_i \in C$$

Observe that, there must exist at least one $\lambda_i \in [0,1)$ or else if all

$$\lambda_i = 1 \Rightarrow 1 = \sum_{i=1}^{m+1} \lambda_i = m + 1 > 3$$

which is a contradiction.

Without loss of generality, we can and do assume that $\lambda_{m+1} \in [0, 1)$. Now:

$$
\begin{aligned}
z &= \sum_{i=1}^{m+1} \lambda_i x_i \\
&= \sum_{i=1}^{m} \lambda_i x_i + \lambda_{m+1} x_{m+1} \\
&= (1 - \lambda_{m+1}) \sum_{i=1}^{m} \frac{\lambda_i}{1 - \lambda_{m+1}} x_i + \lambda_{m+1} x_{m+1} \\
&= (1 - \lambda_{m+1}) \sum_{i=1}^{m} \lambda'_i x_i + \lambda_{m+1} x_{m+1}
\end{aligned}
$$

Observe that, $\lambda'_i := \frac{\lambda_i}{1-\lambda_{m+1}} \geqslant 0$, and that

$$
\begin{aligned}
\sum_{i=1}^{m} \lambda'_i &= \frac{\lambda_1 + \ldots + \lambda_m}{1 - \lambda_{m+1}} \\
&= \frac{1 - \lambda_{m+1}}{1 - \lambda_{m+1}} \\
&= 1
\end{aligned}
$$

Using the inductive hypothesis, we learn that

$$
\sum_{i=1}^{m} \frac{\lambda_i}{1 - \lambda_{m+1}} x_i \in C
$$

Hence,

$$
z = \left[(1 - \lambda_{m+1}) \underbrace{\sum_{i=1}^{m} \frac{\lambda_i}{1 - \lambda_{m+1}} x_i}_{\in C} + \lambda_{m+1} \underbrace{x_{m+1}}_{\in C} \right] \in C
$$

so C is convex. □

Definition 1.6: Convex Hull

Let $S \subseteq \mathbb{R}^n$. The intersection of all convex sets containing S is called the convex hull of S and is denoted by $conv(S)$.
By theorem 2.1, $conv(S)$ is convex. In fact, it is the smallest convex set containing S.

Theorem 1.7: Txtbook THM2.3

Let $S \subseteq \mathbb{R}^n$. Then $conv(S)$ consists of all the convex combinations of the elements of S, i.e.,

$$
conv(S) = \left\{ \sum_{i \in I} \lambda_i x_i | I \text{ is a finite index set, } \forall i \in I, x_i \in S, \lambda_i \geqslant 0, \sum_{i \in I} \lambda_i = 1 \right\}
$$

Proof. Set

$$D := \left\{ \sum_{i \in I} \lambda_i x_i | I \text{ is a finite index set}, \forall i \in I, x_i \in S, \lambda_i \geqslant 0, \sum_{i \in I} \lambda_i = 1 \right\}$$

First, clearly, $S \subseteq D$. Moreover, we want to show that D is convex. Indeed, let $d_1, d_2 \in D$, and let $\lambda \in (0, 1)$
Then, there exist

$$\begin{aligned}
&\lambda_1, \ldots, \lambda_k \geqslant 0, \sum_{i=1}^{k} \lambda_i = 1 \\
&\mu_1, \ldots, \mu_r \geqslant 0, \sum_{j=1}^{r} \mu_j = 1 \\
&d_1 = \sum_{i=1}^{k} \lambda_i x_i, \{x_1, \ldots, x_k\} \subseteq S \\
&d_2 = \sum_{j=1}^{r} \mu_j y_j, \{y_1, \ldots, y_r\} \subseteq S
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\lambda d_1 + (1 - \lambda) d_2 \\
=&\lambda\lambda_1 x_1 + \ldots + \lambda\lambda_k x_k \\
&+ (1 - \lambda)\mu_1 y_1 + \ldots + (1 - \lambda)\mu_r y_r
\end{aligned}$$

Observe that

$$\lambda\lambda_i, (1 - \lambda_i)\mu_i \geqslant 0, \ i \in \{1, \ldots, k\}, j \in \{1, \ldots, r\}$$

and that

$$\begin{aligned}
&\lambda\lambda_1 + \ldots + \lambda\lambda_k + (1 - \lambda)\mu_1 + \ldots + (1 - \lambda)\mu_r \\
=&\lambda \sum_{i=1}^{k} \lambda_i + (1 - \lambda) \sum_{j=1}^{r} \mu_j \\
=&\lambda(1) + (1 - \lambda)(1) = \lambda + 1 - \lambda = 1
\end{aligned}$$

Although, we conclude that D is convex set $\subseteq S$. Hence, $conv(S) \subseteq D$
Secondly, observe that $S \subseteq conv(S)$.
Now, combine with theorem 2.2 to learn that the convex combinations of elements of S lie in $conv(S)$ □

Convex Hull: Examples

1.5 Convex Sets: Best Approximation

Definition: Distance Function

Let $S \subseteq \mathbb{R}^n$. The distance to S is the function

$$\begin{aligned} d_S :& \mathbb{R}^n \to [0, \infty] \\ & x \to \inf_{s \in S} \|x - s\| \end{aligned}$$

Definition: Projection onto a set

Let $\emptyset \neq C \subseteq \mathbb{R}^n$, let $x \in \mathbb{R}^n$, and let $p \in C$. Then p is a projection of x onto C, if

$$d_C(x) = \|x - p\|$$

If every point in $\mathbb{R}^n$ has exactly one projection onto C, the projection operator onto C, denoted by P_C, is the opearator that maps every point in $\mathbb{R}^n$ to its unique projection in C.

Recall:
Let $(x_n)_{n \in \mathbb{N}}$ be a sequence in $\mathbb{R}^n$. Then $(x_n)_{n \in \mathbb{N}}$ is a Cauchy sequence if $\|x_m - x_n\| \to 0$ as $\min\{m, n\} \to \infty$

Fact:
In $\mathbb{R}^n$, every Cauchy sequence converges.

Recall:
Let $f : \mathbb{R}^n \to \mathbb{R}$ and let $\overline{x} \in \mathbb{R}^n$. Then f is continuous at $\overline{x}$ if and only if for every sequence $(x_n)_{n \in \mathbb{N}}$ such that $x_n \to \overline{x}$ we have

$$f(x_n) \to f(\overline{x})$$

Fact:
Let $y \in \mathbb{R}^n$, and let $\|\cdot\|$ be the Euclidean norm on $\mathbb{R}^n$. Then the function

$$f : \mathbb{R}^n \to \mathbb{R} : \; x \to \|x - y\|$$

is continuous.

Proof. Only for illustration, you don't need to know the proof.
Let $(x_n)_{n \in \mathbb{N}}$ be a sequence in $\mathbb{R}^n$ such that $x_n \to \overline{x}$. Now:

$$\begin{aligned} f(x_n) - f(\overline{x}) &= \|x_n - y\| - \|\overline{x} - y\| \\ &= \|x_n - \overline{x} + \overline{x} - y\| - \|\overline{x} - y\| \\ &\leqslant \|x_n - \overline{x}\| + \|\overline{x} - y\| - \|\overline{x} - y\| \\ &= \|x_n - \overline{x}\| \end{aligned}$$

Similarly,

$$\begin{aligned} f(\overline{x}) - f(x_n) &= \|\overline{x} - y\| - \|x_n - y\| \\ &\leqslant \|\overline{x} - x_n\| + \|x_n - y\| - \|x_n - y\| \\ &= \|\overline{x} - x_n\| \end{aligned}$$

Altogether, we have

$$0 \leqslant |f(x_n) - f(\overline{x})| \leqslant \|x_n - \overline{x}\|$$

Now, take the limit as $n \to \infty$ to learn that

$$|f(x_n) - f(\overline{x})| \to 0$$

equivalently,

$$f(x_n) \to f(\overline{x})$$

Explicitly, this means $(\forall y \in \mathbb{R}^n)$ if $x_m \to \overline{x}$, then

$$\|x_m - y\| \to \|\overline{x} - y\|$$

□

Lemma 1.8

Let x, y, z be vectors in $\mathbb{R}^n$. Then

$$\|x-y\|^2 = 2\|z-x\|^2 + 2\|z-y\|^2 - 4\|z - \frac{x+y}{2}\|^2$$

Proof.

$$2\|z-x\|^2 = 2\|z\|^2 - 4\langle z, x\rangle + 2\|x\|^2 \tag{1.1}$$

$$2\|z-y\|^2 = 2\|z\|^2 - 4\langle z, y\rangle + 2\|y\|^2 \tag{1.2}$$

$$4\|z - \frac{x+y}{2}\|^2 = 4\|z\|^2 + \|x+y\|^2 - 4\langle z, x\rangle - 4\langle z, y\rangle \tag{1.3}$$

$$\begin{aligned} R.H.S &= (1.1) + (1.2) - (1.3) \\ &= 2\|x\|^2 + 2\|y\|^2 - \|x+y\|^2 \\ &= 2\|x\|^2 + 2\|y\|^2 - \|x\|^2 - \|y\|^2 - 2\langle x, y\rangle \\ &= \|x\|^2 + \|y\|^2 - 2\langle x, y\rangle \\ &= \|x-y\|^2 = L.H.S \end{aligned}$$

□

Lemma 1.9

Let $x \in \mathbb{R}^n$, $y \in \mathbb{R}^n$. Then

$$\langle x, y\rangle \leqslant 0 \iff (\forall \lambda \in [0,1]), \|x\| \leqslant \|x - \lambda y\|$$

Proof. Observe that

$$\begin{aligned} \|x-\lambda y\|^2 - \|x\|^2 &= \|x\|^2 - 2\lambda\langle x, y\rangle + \lambda^2\|y\|^2 - \|x\|^2 \\ &= \lambda(\lambda\|y\|^2 - 2\langle x, y\rangle) \ldots (*) \end{aligned}$$

($\Rightarrow$) Suppose $\langle x, y\rangle \leqslant 0$. Then

$$\|x-\lambda y\|^2 - \|x\|^2 = \lambda(\lambda\|y\|^2 - 2\langle x, y\rangle) \geqslant 0$$

($\Leftarrow$) Suppose that for every $\lambda \in (0,1]$, $\|x-\lambda y\| \geqslant \|x\|$. Then $(*)$ implies $\langle x, y\rangle \leqslant \frac{\lambda}{2}\|y\|^2$. Taking the limit as $\lambda \downarrow 0$ yields the desired result. □

Theorem 1.10: The projection theorem

Let C be a nonempty, closed, convex subset of $\mathbb{R}^n$. Then the following hold:

1. $(\forall x \in \mathbb{R}^n)$ *the projection of x onto C exists and is unique.*
2. *For every $x \in \mathbb{R}^n$ and every $p \in \mathbb{R}^n$:*

$$p = P_C x \Leftrightarrow [p \in C \text{ and } (\forall y \in C)\ \langle y-p, x-p\rangle \leqslant 0]$$

Proof. Let $x \in \mathbb{R}^n$.

1. Our goal is to show that x has a unique projection onto C.
   Existence:
   Recall that $(\forall x \in \mathbb{R}^n)$
   $$d_C(x) = \inf_{c \in C} \|x - c\|$$
   Therefore, there exists a sequence $(c_n)_{n\in\mathbb{N}}$ in C such that
   $$d_C(x) = \lim_{n\to\infty} \|c_n - x\| \ldots (1)$$
   Now, let m and n be in $\mathbb{N}$. By convexity of C ,we know that
   $$\frac{1}{2}(c_m + c_n) \in C$$
   Hence,
   $$d_C(x) = \inf_{c\in C} \|x - c\| \leqslant \|x - \frac{1}{2}(c_m + c_n)\|$$
   Applying the auxuliary Lemma 1 with (x, y, z) replaced by (c_m, c_n, x) we learn that :
   $$\begin{aligned}\|c_n - c_m\|^2 &= 2\|c_n - x\|^2 + 2\|c_m - x\|^2 - 4\|x - \frac{c_n + c_m}{2}\|^2 \\ &\leqslant 2\|c_n - x\|^2 + 2\|c_m - x\|^2 - 4d_C^2(x)\end{aligned}$$
   Letting $m \to \infty$, $n \to \infty$, we learn that
   $$0 \leqslant \|c_n - c_m\|^2 \leqslant 2d_C^2(x) + 2d_C^2(x) - 4d_C^2(x) = 0$$
   That is $\|c_n - c_m\|^2 \to 0$, hence $(c_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in C, hence $(c_n)_{n\in\mathbb{N}}$ converges to some point say $p \in C$ (by the closedness of C).
   We will now show that
   $$d_C(x) = \|x - p\|$$
   Observe that, $\|x - \cdot\|$ is continuous. Combining with $c_n \to p$ and (1), we learn that $d_C(x) \leftarrow \|x - c_n\| \to \|x - p\|$, hence
   $$d_C(x) = \|x - p\|$$
   This proves the existence.
   Uniqueness:
   Suppose that $q \in C$ satisfies that $d_C(x) = \|q - x\|$. By convexity of C,
   $$\frac{1}{2}(p + q) \in C$$
   Now, using the auxiliary lemma 1 with (x, y, z) replaced by $(p, q, \frac{1}{2}(p + q))$ we learn that:
   $$\begin{aligned}0 &\leqslant \|p - q\|^2 \\ &= 2\|p - x\|^2 + 2\|q - x\|^2 - 4\|x - \frac{p+q}{2}\|^2 \\ &\leqslant 2d_C^2(x) + 2d_C^2(x) - 4d_C^2(x) \\ &= 0\end{aligned}$$
   Hence, $\|p - q\| = 0$; equivalently $p = q$. This proves uniqueness.

2. We want to show that, for every $x \in \mathbb{R}^n$ and every $p \in \mathbb{R}^n$,

$$p = P_C(x) \Leftrightarrow [p \in C \text{ and } (\forall y \in C)\, \langle y - p, x - p \rangle \leqslant 0]$$

Indeed, $p = P_C(x) \Leftrightarrow [p \in C$ and $\|x - p\|^2 = d_C^2(x)]$.
Observe that, for every $y \in C$, $\alpha \in [0, 1]$,

$$y_\alpha := \alpha y + (1 - \alpha)p \in C$$

Therefore,

$$\begin{aligned}
&\|x - p\|^2 = d_C^2(x) \\
&\Leftrightarrow \forall y \in C, \forall \alpha \in [0,1] \|x - p\|^2 \leqslant \|x - y_\alpha\|^2 \\
&\Leftrightarrow \forall y \in C, \forall \alpha \in [0,1] \|x - p\|^2 \leqslant \|x - p - \alpha(y - p)\|^2 \\
&\Leftrightarrow \forall y \in C, \langle x - p, y - p \rangle \leqslant 0 \text{ (by the lemma 2)}
\end{aligned}$$

□

Example

Let $\epsilon > 0$, and let $C = ball(0, \epsilon) = \{c \in \mathbb{R}^n | \|x\|^2 \leqslant \epsilon^2\}$, i.e., the closed ball in $\mathbb{R}^n$ centered at 0 with radius ϵ. Show that

$$\forall x \in \mathbb{R}^n,\ P_C(x) = \frac{\epsilon}{\max\{\|x\|, \epsilon\}} x$$

Proof. Let $x \in \mathbb{R}^n$ and set $p = \frac{\epsilon}{\max\{\|x\|, \epsilon\}} x$. Using the projection theorem, it suffices to show that:

1. $p \in C$
2. $\forall y \in C$, $\langle x - p, y - p \rangle \leqslant 0$

We examine two cases, show $p \in C$

1. $\|x\| \leqslant \epsilon$. Then $x \in C$ and $p = \frac{\epsilon}{\epsilon} x = x \in C$
2. $\|x\| > \epsilon$, and $\|p\| = \epsilon \frac{\|x\|}{\|x\|} = \epsilon$, hence $p \in C$

Then, we show $\forall y \in C$,

$$\langle x - p, y - p \rangle \leqslant 0$$

Indeed, let $y \in C$.

1. $\|x\| \leqslant \epsilon \Rightarrow p = x$ and
$$0 = \langle x - p, y - p \rangle \leqslant 0$$

2. $\|x\|> \epsilon \Rightarrow = \frac{\epsilon}{\|x\|}x.$
   Moreover,

$$\begin{aligned}
\langle x-p, y-p\rangle &= \left\langle x-\frac{\epsilon}{\|x\|}x, y-\frac{\epsilon}{\|x\|}x\right\rangle \\
&= \left(1-\frac{\epsilon}{\|x\|}\right)\left\langle x, y-\frac{\epsilon}{\|x\|}x\right\rangle \\
&= \left(1-\frac{\epsilon}{\|x\|}\right)\left(\langle x, y\rangle-\frac{\epsilon}{\|x\|}\|x\|^2\right) \\
&= \left(1-\frac{\epsilon}{\|x\|}\right)(\langle x, y\rangle-\epsilon\|x\|) \\
&\leqslant \left(1-\frac{\epsilon}{\|x\|}\right)(\|x\|\|y\|-\epsilon\|x\|) \\
&\leqslant \underbrace{\left(1-\frac{\epsilon}{\|x\|}\right)}_{\geqslant 0}\left(\|x\|\underbrace{\epsilon}_{\|y\|\leqslant\epsilon, y\in C}-\epsilon\|x\|\right) \\
&= 0
\end{aligned}$$

$\square$

Definition: Minkowski sum of two sets

Let C and D be two subsets of $\mathbb{R}^n$. The Minkowski sum of C and D, denoted by $C + D$ is

$$C + D := \{c + d | c \in C, d \in D\}$$

Theorem 1.11: Minkowski sum of Convex sets, Txtbook THM3.1

Let C_1, C_2 be convex subsets of $\mathbb{R}^n$. Then $C_1 + C_2$ is convex.

Proof. If $C_1 = \emptyset$ or $C_2 = \emptyset$, then $C_1 + C_2 = \emptyset$ and the conclusion follows.
Now suppose that $C_1 \neq \emptyset, C_2 \neq \emptyset \Rightarrow C_1 + C_2 \neq \emptyset$
Let x, y be in $C_1 + C_2$ and let $\lambda \in (0, 1)$.
Since $x \in C_1 + C_2$, there exist $x_1 \in C_1, x_2 \in C_2$ such that

$$x = x_1 + x_2$$

.
Similarly, there exists $y_1 \in C_1, y_2 \in C_2$ such that $y = y_1 + y_2$.
Now,

$$\begin{aligned}\lambda x + (1-\lambda)y &= \lambda(x_1 + x_2) + (1-\lambda)(y_1 + y_2) \\ &= \lambda x_1 + (1-\lambda)y_2 + \lambda x_1 + (1-\lambda)y_2 \\ &\in C_1 + C_2\end{aligned}$$

The proof is complete. □

Proposition 1.12

Let C and D be nonempty, closed convex subsets of $\mathbb{R}^n$ such that D ***is bounded****. Then*

$C + D$ is nonempty, closed, convex

Proof.

$$C \neq \emptyset, D \neq \emptyset \implies C + D \neq \emptyset$$

C convex, D convex $\implies C + D$ is convex by theorem 3.1
It remains to show taht $C + D$ is closed.
Take a convergent sequence $(x_n + y_n)_{n\in\mathbb{N}}$ in $C + D$ such that $(x_n)_{n\in\mathbb{N}}$ lies in C, $(y_n)_{n\in\mathbb{N}}$ lies in D and $x_n + y_n \to z$ (say). Our goal is to show that $z \in C + D$.
By assumption, D is bounded, hence $(y_n)_{n\in\mathbb{N}}$ is bounded.
Using Bolzano-Weierstrass, we know that there exists a subsequence

$$(y_{k_n})_{n\in\mathbb{N}},\ y_{k_n} \to y \in D$$

Therefore, $z - y \leftarrow x_{k_n} \to \overline{x} \in C$
That is, $z \in C + y \subseteq C + D$ □

Quesion: What happens if we drop the assumption that D is bounded?

Example 1.13

Let

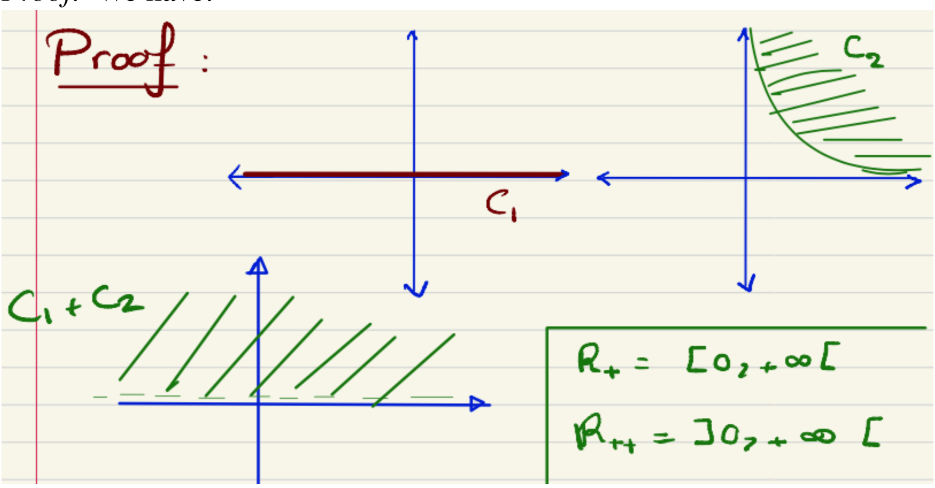
$$C_1 = \mathbb{R} \times \{0\}$$
$$C_2 = \{(x, y) \in \mathbb{R}^2_{++} | xy \geqslant 1\}$$

Then C_1, C_2 are closed and convex. However,

$$C_1 + C_2 = \mathbb{R} \times \mathbb{R}_{++}$$

which is convex but open.

Proof. We have:



- $(\subseteq)$, $C_1 + C_2 \subseteq \mathbb{R} \times \mathbb{R}_{++}$
  Indeed, let $(z_1, z_2) \in C_1 + C_2$. Then, there exists $(x_1, x_2) \in C_1, (y_1, 0) \in C_2$, such that
  $$z_1 = x_1 + y_1, z_2 = x_2$$
  Clearly, $z_1 = x_1 + y_1 \in \mathbb{R}$. And $z_2 = x_2 > 0$. Hence,
  $$C_1 + C_2 \subseteq \mathbb{R} \times \mathbb{R}_{++}$$
- $(\supseteq)$, $C_1 + C_2 \supseteq \mathbb{R} \times \mathbb{R}_{++}$
  Let $(x, y) \in \mathbb{R} \times \mathbb{R}_{++}$, set
  $$c_1 := \left(x - \frac{1}{y}, 0\right), c_2 := \left(\frac{1}{y}, y\right)$$
  Then we have $c_1 \in C_1$, $c_2 \in C_2$ and
  $$(x, y) = c_1 + c_2 \in C_1 + C_2$$

□

Theorem 1.14: Txtbook THM3.2

Let C to be a convex set, let $\lambda_1 \geqslant 0$ and let $\lambda_2 \geqslant 0$. Then

$$(\lambda_1 + \lambda_2)C = \lambda_1 C + \lambda_2 C$$

Proof. We prove two directions:

- $(\subseteq)$: Obvious. Indeed, let $x \in (\lambda_1 + \lambda_2)C$. Then $\exists c \in C$, such that

  $$x = (\lambda_1 + \lambda_2)c = \lambda_1 c + \lambda_2 \in \lambda_1 C + \lambda_2 C$$

  This direction is always true even in the absence of convexity.

- $(\supseteq)$: Without loss of generality, we can and do assume that $\lambda_1 + \lambda_2 > 0$ (o/w, the condition is trivial)
  Now, by convexity we have

  $$\frac{\lambda_1}{\lambda_1 + \lambda_2} C + \frac{\lambda_2}{\lambda_1 + \lambda_2} C \subseteq C$$

  Equivalently, $\lambda_1 C + \lambda_2 C \subseteq (\lambda_1 + \lambda_2)C$

$\square$

1.6 Convex Sets: Topological properties

Throughout this course we use:

$$B(x, \varepsilon) = \{y \in \mathbb{R}^n |\ \|y - x\|^2 \leqslant \varepsilon\}$$

and

$$B := B(0, 1) = \{y \in \mathbb{R}^n |\ \|y\| \leqslant 1\}$$

i.e., the closed unit ball.
Let $C \subseteq \mathbb{R}^n$,
the **interior** of C is

$$int(C) = \{x | \exists \varepsilon > 0,\ s.t.\ x + \varepsilon B \subseteq C\}$$

the **closure** of C is $\overline{C}$ (textbook uses $cl(C)$,

$$\overline{C} = cl(C) = \cap\{C + \varepsilon B | \varepsilon > 0\}$$

The **relative interior** of a convex set C is

$$ri(C) = \{x \in aff(C) | \exists \varepsilon > 0,\ s.t.\ (x + \varepsilon B) \cap aff(C) \subseteq C\}$$

Example 1.15

- *On the real line:*

  1.
  $$\begin{aligned} &C_1 = \{0\} \subseteq \mathbb{R} \\ &int(C_1) = \emptyset, \overline{C}_1 = \{0\} \\ &ri(C_1) = \{0\} \end{aligned}$$

  2.
  $$\begin{aligned} &C_2 = [a, b) \\ &int(C_2) = (a, b), \overline{C}_2 = [a, b] \\ &ri(C_2) = (a, b) \end{aligned}$$

- *in* $\mathbb{R}^2$*:*

  1. $C_1 = \{(0, 0)\}$, $int(C_1) = \emptyset$, $\overline{C}_1 = \{(0, 0)\}$, *and* $ri(C_1) = \{(0, 0)\}$
  2. *even for* $x \in \mathbb{R}^n$*, say* $C = \{x\}$, $int(C) = \emptyset$, $\overline{C} = ri(C) = \{x\}$
  3. $C_2 = [a, b] \times \{0\}$
  $$\begin{aligned} int(C_2) &= \emptyset \\ \overline{C}_2 &= C_2 \\ &= [a, b] \times \{0\} \\ ri(C_2) &= (a, b) \times \{0\} \end{aligned}$$
  4. $C_3 = [-1.1] \times [-1.1]$*, then*
  $$\begin{aligned} int(C_3) &= (-1, 1) \times (-1, 1) \\ \overline{C}_3 &= C_3 \\ ri(C_3) &= int(C_3) \\ &= (-1, 1) \times (-1, 1) \end{aligned}$$

Remark. 1. Let $C \subseteq \mathbb{R}^n$. Suppose that $int(C) \neq \emptyset$. Then $int(C) = ri(C)$

Proof. Let $x \in int(C)$. Then $\exists \varepsilon > 0$ such that

$$B(x; \varepsilon) \subseteq C$$

Hence,

$$\mathbb{R}^n = aff(B; \varepsilon)) \subseteq \; aff(C) \subseteq \mathbb{R}^n$$

Therefore, $aff(C) = \mathbb{R}^n$, and the conclusion follows by recalling that

$$\begin{aligned} ri(C) &= \{x \in \ aff(C) | \exists \varepsilon > 0, \ s.t. \ (x + \varepsilon B) \cap \ aff(C) \subseteq C\} \\ &= \{x \in \mathbb{R}^n | \exists \varepsilon > 0, \ s.t. \ (x + \varepsilon B) \cap \ \mathbb{R}^n \subseteq C\} \\ &= \{x | \exists \varepsilon > 0, \ s.t. \ x + \varepsilon B \subseteq C\} \\ &= int(C) \end{aligned}$$

□

2. Let $C \neq \emptyset$ be convex. The dimension of C, denoted $\dim(c)$, is the dimension of the affine hull of C "$aff(C)$". Observe that

$$L := aff(C) - aff(C)$$

is a linear subspace

$$\dim(aff(C)) = \dim L$$

Proposition 1.16: *

Let C be a convex set in $\mathbb{R}^n$. Then $\forall x \in int(C), \forall y \in \overline{C}$

$$[x, y) \subseteq \ int(C)$$

Proof. The above statement is equivalent to $\forall x \in int(C), \forall y \in \overline{C}, \forall \lambda \in [0, 1)$,

$$(1 - \lambda)x + \lambda y \in int(C)$$

Let $x \in int(C), y \in \overline{C}, \lambda \in [0, 1)$. We need to show that

$$(1 - \lambda)x + \lambda y + \varepsilon B \subseteq C$$

for some $\varepsilon > 0$.
Observe that, because $y \in \overline{C}$,

$$\forall \varepsilon > 0, y \in C + \varepsilon B$$

Hence, for every $\varepsilon > 0$, we have

$$\begin{aligned} &(1 - \lambda)x + \lambda y + \varepsilon B \\ &\subseteq (1 - \lambda)x + \lambda(C + \varepsilon B) + \varepsilon B \\ &= (1 - \lambda)x + \lambda C + \lambda \varepsilon B + \varepsilon B \\ &= (1 - \lambda)x + \lambda C + (1 + \lambda)\varepsilon B \\ &= (1 - \lambda)\left[\underbrace{x}_{\in int(C)} + \frac{1 + \lambda}{1 - \lambda}\varepsilon B\right] + \lambda C \\ &\subseteq (1 - \lambda)C + \lambda C \text{ (for suff. small } \varepsilon) \\ &= C \end{aligned}$$

□

Theorem 1.17: Txtbook THM6.1

Let C be a convex set in $\mathbb{R}^n$. Then $\forall x \in ri(C), \forall y \in \overline{C}$

$$[x, y) \subseteq ri(C)$$

Proof. We have just shown that if $int(C) \neq \emptyset$, then $\forall x \in int(C), \forall y \in \overline{C}$

$$[x, y) \subseteq \ int(C)$$

1. $int(C) \neq \emptyset$.
   Combine the previous proposition and remark 1, $int(C) = ri(C)$
2. $int(C) = \emptyset$
   In this case we must have $\dim C = m < n$.
   Let $L = aff(C) - aff(C)$, then L is a linear subspace whose dimension $= m$. Hence, L can be regarded as a copy of $\mathbb{R}^m$
   After translating C with $-c \in C$ (if necessary), we can and do assume that $C \subseteq \mathbb{R}^m$, and the interiors of $C - c$ with respect to $\mathbb{R}^m$ is $ri(C)$ (in $\mathbb{R}^n$). Now, apply case 1).

□

Theorem 1.18

Let C be a convex subset of $\mathbb{R}^n$, then the following hold:

1. *$\overline{C}$ is convex*
2. *$int(C)$ is convex*
3. *Suppose that $int(C) \neq \emptyset$. Then $int(C) = int(\overline{C})$ and $\overline{C} = \overline{int(C)}$*

Proof. We prove each of the above:

1. Let $x, y \in \overline{C}$, and let $\lambda \in (0, 1)$. Then there exist sequences $(x_n)_{n \in \mathbb{N}}$ and $(y_n)_{n \in \mathbb{N}}$ in C such that
   $$x_n \to x, \ y_n \to y$$
   Consequently,
   $$C \ni \lambda x_n + (1 - \lambda) y_n \longrightarrow \lambda x + (1 - \lambda) y$$
   which implies
   $$\lambda x + (1 - \lambda) y \in \overline{C}$$
   Hence, $\overline{c}$ is convex.
2. If $int(C) = \emptyset$, the conclusion is clear. Otherwise, use the previous proposition with $x, y \in int(C) \subseteq \overline{C}$. Observe that:
   $$\begin{aligned} [x, y] &= [x, y) \cup \{y\} \\ &\subseteq int(C) \cup int(C) \\ &= int(C) \end{aligned}$$

3. Clearly, $C \subseteq \overline{C}$. Hence,

$$int(C) \subseteq int(\overline{C})$$

Conversely, let $y \in int(\overline{C})$.
Then $\exists \varepsilon > 0$, such that $B(y, \varepsilon) \subseteq \overline{C}$. Now, let $x \in int(C), \lambda > 0$ such that $x \neq y$, and $y + \lambda(y - x) \in B(y; \varepsilon) \subseteq \overline{C}$. By the proposition $*$ applied with y replace by $y + \lambda(y - x)$, we learn that

$$y \in [x, y + \lambda(y - x)) \subseteq int(C)$$

To see $y \in [x, y + \lambda(y - x))$: set $\alpha := \frac{1}{1+\lambda} \in (0, 1)$
Observe that

$$\begin{aligned} y &= (1 - \alpha)x + \alpha(y + \lambda(y - x)) \\ &\neq y + \lambda(y - x) \end{aligned}$$

Indeed,

$$\begin{aligned} &(1 - \alpha)x + \alpha(y + \lambda(y - x)) \\ &= (1 - \alpha(1 + \lambda))x + \alpha(1 + \lambda)y \\ &= y \end{aligned}$$

Therefore, $int(\overline{C}) \subseteq int(C)$
Altogether, $int(C) = int(\overline{C}$ We now turn to the second identity. Clearly $\overline{int(C)} \subseteq \overline{C}$.
Conversely, let $y \in \overline{C}$ and let $x \in int(C)$.
Define, $\forall \lambda \in [0, 1)$

$$y_\lambda = (1 - \lambda)x + \lambda y$$

Again, proposition $*$ tells us that the $(y_\lambda)_{\lambda \in [0,1)}$ lies in $[x, y) \subseteq int(C)$ Hence, $y = \lim_{\lambda \downarrow 0} y_\lambda \in \overline{int(C)}$. That is,

$$\overline{C} \subseteq \overline{int(C)}$$

Altogether, we learn that

$$\overline{C} = \overline{int(C)}$$

$\square$

Fact(textbook THM6.2):
Let C be a convex subset of $\mathbb{R}^n$. Then $ri(C)$ and $\overline{C}$ are convex subsets of $\mathbb{R}^n$. Moreover,

$$C \neq \emptyset \Leftrightarrow ri(C) \neq \emptyset$$

2 Separation Theorems

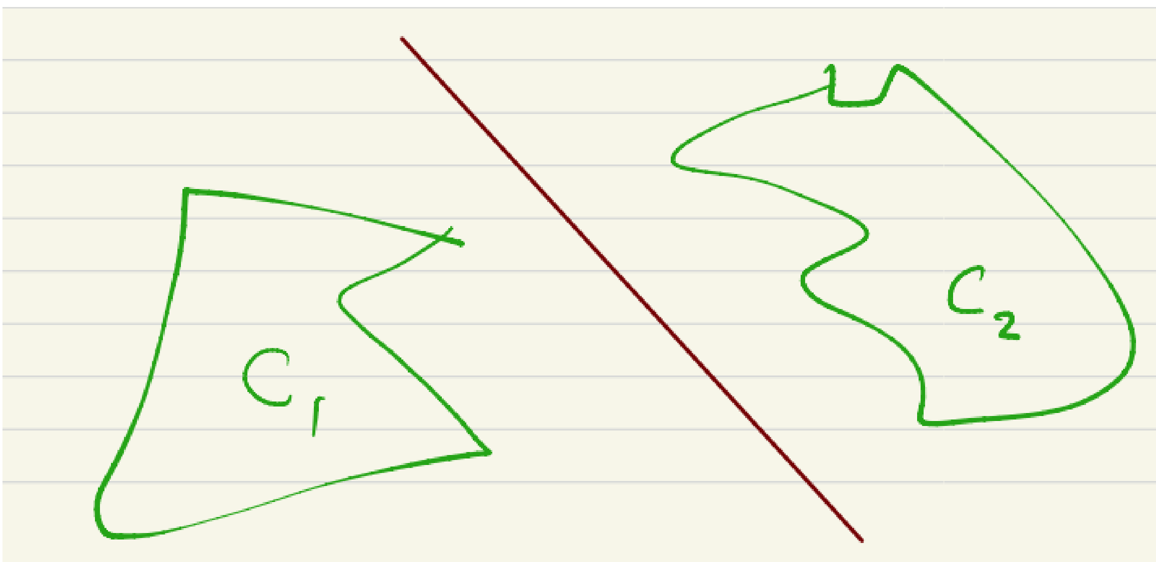
Definition 2.1

Let C_1, C_2 be subsets of $\mathbb{R}^n$. Then C_1 and C_2 are separated if $\exists b \in \mathbb{R}^n \setminus \{0\}$ such that

$$\sup_{c_1 \in C_1} \langle c_1, b\rangle \leqslant \inf_{c_2 \in C_2} \langle c_2, b\rangle$$

C_1 and C_2 are strongly separated if $\exists b \in \mathbb{R}^n \setminus \{0\}$ such that

$$\sup_{c_1 \in C_1} \langle c_1, b\rangle < \inf_{c_2 \in C_2} \langle c_2, b\rangle$$



We say that $x \in \mathbb{R}^n$ is (strongly) separated from $C \subseteq \mathbb{R}^n$ if the set $\{x\}$ is (strongly) separated from C.

Theorem 2.2

Let C be a nonempty, closed, convex subset of $\mathbb{R}^n$ and suppose that $x \notin C$. Then x is strongly separated from C.

Proof. We need to guarantee the existence of $\mathbb{R}^n \ni b \neq 0$ such that

$$\sup \langle c, b\rangle < \inf \langle x, b\rangle = \langle x, b\rangle$$

Set

$$b := x - P_C x \neq 0 \Leftrightarrow P_C x = x - b \neq x \ (x \notin C)$$

Let $y \in C$. By the projection theorem we have

$$p = P_C x \Leftrightarrow [p \in C \text{ and } \forall y \in C, \langle y - p, x - p\rangle \leqslant 0]$$

$$
\begin{aligned}
&\langle y-(x-b), x-(x-b)\rangle \leqslant 0 \\
\Leftrightarrow &\langle y-x+b, b\rangle \leqslant 0 \\
\Leftrightarrow &\langle y-x, b\rangle \leqslant -\langle b, b\rangle = -\|b\|^2
\end{aligned}
$$

Consequently,

$$
\sup_{y\in C}\langle y, b\rangle - \langle x, b\rangle \leqslant -\|b\|^2 < 0
$$

Hence,

$$
\sup_{y\in C}\langle y, b\rangle < \langle x, b\rangle
$$

□

Corollary 2.3

Let C_1, C_2 be nonempty subsets of $\mathbb{R}^n$ such that $C_1 \cap C_2 = \emptyset$ and $C_1 - C_2$ is closed and convex. Then C_1 and C_2 are strongly separated.

Proof. Observe that by definition C_1, C_2 are strongly separated if and only if $C_1 - C_2$ and 0 are strongly separated.
Indeed, $C_1 - C_2$ and 0 are strongly separated $\Leftrightarrow \exists b \neq 0$ such that

$$
\begin{aligned}
&\sup_{\substack{c_1\in C_1\\ c_2\in C_2}} \langle c_1 - c_2, b\rangle < \inf \langle 0, b\rangle = 0 \\
\Leftrightarrow &\sup_{\substack{c_1\in C_1\\ c_2\in C_2}} \{\langle c_1, b\rangle + \langle -c_2, b\rangle\} < 0 \\
\Leftrightarrow &\sup_{c_1\in C_1} \langle c_1, b\rangle + \sup_{c_2\in C_2} \langle -c_2, b\rangle < 0 \\
\Leftrightarrow &\sup_{c_1\in C_1} \langle c_1, b\rangle < -\sup_{c_2\in C_2} \langle -c_2, b\rangle = \inf_{c_2\in C_2} \langle c_2, b\rangle
\end{aligned}
$$

The conclusion follows by noting that $C_1 \cap C_2 = \emptyset \Rightarrow 0 \notin C_1 - C_2$, and combining with the previous theorem(12). □

Corollary 2.4

Let C_1, C_2 be nonempty closed convex subsets of $\mathbb{R}^n$ such that $C_1 \cap C_2 = \emptyset$ and C_2 is bounded. Then C_1 and C_2 are strongly separated.

Proof. Observe that $-C_2$ is nonempty closed and convex. Therefore, by proposition $*$, $C_1 - C_2$ is nonempty, closed and convex. Now we combine with the last corollary and get what's required. □

Theorem 2.5

Suppose that C_1 and C_2 are nonempty closed convex subsets of $\mathbb{R}^n$ such that $C_1 \cap C_2 = \emptyset$. Then C_1 and C_2 are separated.

Proof. Set $(\forall n \in \mathbb{N})$

$$D_n = C_2 \cap B(0; n)$$

Observe that $(\forall n \in \mathbb{N})$,

$$C_1 \cap D_n = \emptyset$$

Indeed, $D_n \subseteq C_2$. Hence, $C_1 \cap D_n \subseteq C_1 \cap C_2 = \emptyset$
D_n is bounded, because $D_n \subseteq B(0; n)$
Apply the corollary(15) from the previous lecture with C_2 replaced by D_n we learn that ($\forall n \in \mathbb{N}$, there exists a hyperplane that strongly separates C_1 and D_n. Equivalently.

$$\forall n \in \mathbb{N}, \exists u_n \in \mathbb{R}^n \setminus \{0\}, \|u_n\|= 1$$

and

$$\sup \langle C_1, u_n \rangle < \inf \langle D_n, u_n \rangle$$

Because $(u_n)_{n\in\mathbb{N}}$ is bounded, there exists a convergent subsequence $(u_{K_n})_{n\in\mathbb{N}}$ of $(u_n)_{n\in\mathbb{N}}$ such that $u_{K_n} \to u$ (say), and $\|u\|= 1$.
Now let $x \in C_1, y \in C_2$. Then, eventually $y \in B(0; K_n)$, hence eventually $y \in D_{K_n}$ and by $\sup \langle C_1, u_n \rangle < \inf \langle D_n, u_n \rangle$, we have

$$\left\langle \underbrace{x}_{\in C_1}, u_{K_n} \right\rangle < \left\langle \underbrace{y}_{\in D_{K_n}}, u_{K_n} \right\rangle$$

Taking the limit as $n \to \infty$, we learn that $\langle x, u \rangle \leqslant \langle y, u \rangle$. The proof is complete. $\square$

2.1 More Convex Sets: Cones

Definition 2.6

Let C be a subset of $\mathbb{R}^n$, then

1. *C is a cone if*
$$C = \mathbb{R}_{++}C$$

2. *The conical hull of C, denoted by $cone(C)$, is the intersection of all the cones of $\mathbb{R}^n$ containing C. It is the smallest cone in $\mathbb{R}^n$ containing C.*

3. *The closed conical hull of C, denoted by $\overline{cone}(C)$ is the smallest closed cone in $\mathbb{R}^n$ containing C.*

The definition of cone above means that

$$\forall x \in C, \forall \alpha \in \mathbb{R}_{++}, \alpha x \in C \implies C \text{ is a cone}$$

Example 2.7

1.

$$K_1 = \{(x_1, \ldots, x_n) \in \mathbb{R}^n | x_i \geqslant 0, 1 \leqslant i \leqslant n\}$$

It's a closed convex cone.

2.

$$K_2 = \{(x_1, \ldots, x_n) \in \mathbb{R}^n | x_i > 0, 1 \leqslant i \leqslant n\}$$

It's a convex cone.

3.

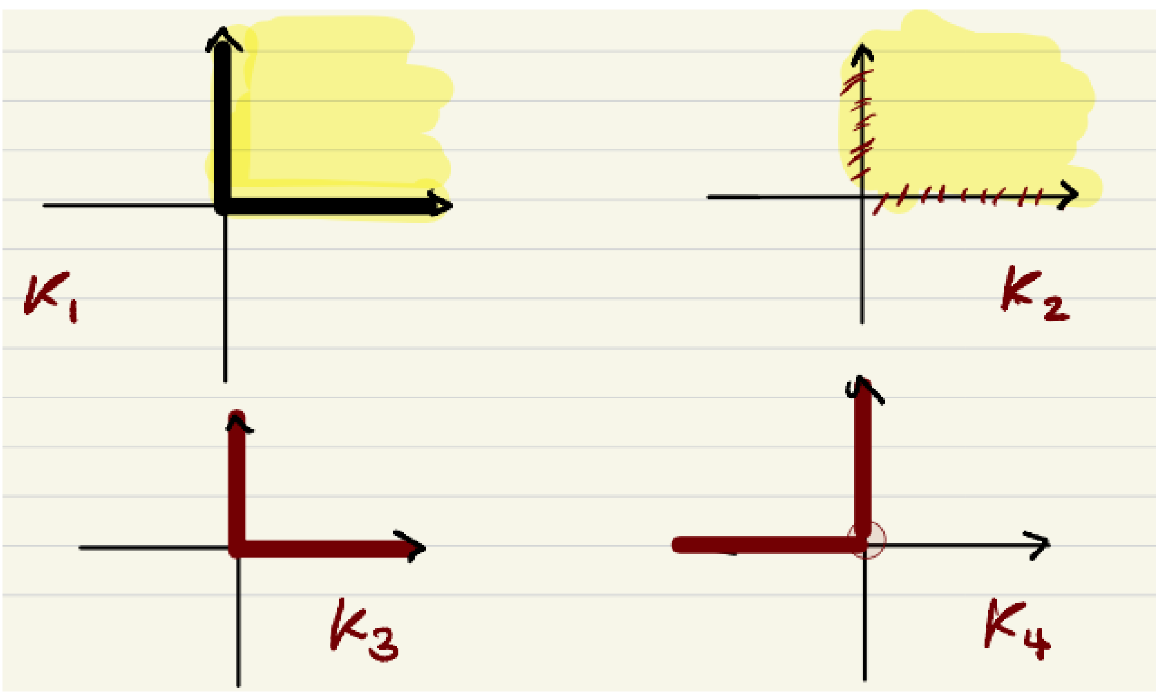
$$K_3 = (\{0\} \times \mathbb{R}_+) \cup (\mathbb{R}_+ \times \{0\}) \subseteq \mathbb{R}^2$$

It's a closed cone but not convex

4.

$$K_4 = (\{0\} \times \mathbb{R}_{++}) \cup (\mathbb{R}_{--} \times \{0\})$$

It's a closed cone but it's not closed nor convex.



Proposition 2.8

Let C be a subset of $\mathbb{R}^n$. Then the following hold:

1. $cone(C) = \mathbb{R}_{++}C$
2. $\overline{cone(C)} = \overline{cone}(C)$
3. $cone(conv(C)) = conv(cone(C))$
4. $\overline{cone}(conv(C)) = \overline{conv}(cone(C))$

Proof. If $C = \emptyset$, then the conclusion is obvious. Now, suppose that $C \neq \emptyset$.

1. Set $D = \mathbb{R}_{++}C$, and observe that $C \subseteq D$, and D is a cone.

$$\implies cone(C) \subseteq cone(D) = D = \mathbb{R}_{++}C$$

Conversely, let $y \in D$. Then $\exists \lambda > 0, c \in C$ such that

$$y = \lambda c$$

Then $y \in cone(C)$. Hence,

$$\mathbb{R}_{++}C = D \subseteq cone(C)$$

Altogether,

$$cone(C) = \mathbb{R}_{++}C$$

2. Observe that $\overline{cone}(C)$ is closed cone. Clearly, $C \subseteq \overline{cone}(C)$. Hence,

$$\overline{cone(C)} \subseteq \overline{cone(\overline{cone}(C))} = \overline{cone}(C)$$

Conversely, since $\overline{cone(c)}$ is a cone ,

$$\overline{cone}(C) \subseteq \overline{cone(C)}$$

Altogether,

$$\overline{cone}(C) = \overline{cone(C)}$$

3. We want to show that

$$cone(conv(C)) = conv(cone(C))$$

- ($\subseteq$) let $x \in cone(conv(C))$. Then by 1), $\exists \lambda > 0$ and $y \in conv(C)$ such that

$$x = \lambda y$$

Since $y \in conv(C)$, there exist $\lambda_1, \ldots, \lambda_m \in \mathbb{R}_{++}$, $\sum_{i=1}^m \lambda_i = 1$, $x_1, \ldots, x_m \in C$, such that

$$y = \sum_{i=1}^m \lambda_i x_i$$

Hence

$$\begin{aligned} x &= \lambda \sum_{i=1}^{m} \lambda_i x_i \\ &= \sum_{i=1}^{m} \lambda_i (\underbrace{\lambda x_i}_{\in cone(C)}) \\ &\in conv(cone(C)) \end{aligned}$$

- ($\supseteq$), conversely, let $x \in conv(cone(C))$. In view of 1) $cone(C) = \mathbb{R}_{++}C$, we learn that there exist $\lambda_1, \ldots, \lambda_m > 0$, there exist $\mu_1, \ldots, \mu_m > 0$ with $\sum_{i=1}^{m} \mu_i = 1, \{x_1, \ldots, x_m\} \subseteq C$ such that

$$\begin{aligned} x &= \sum_{i=1}^{m} \mu_i \lambda_i x_i \\ &= \underbrace{\left(\sum_{i=1}^{m} \lambda_i \mu_i \right)}_{:=\alpha} \left[\sum_{i=1}^{m} \underbrace{\frac{\lambda_i \mu_i}{\sum \lambda_i \mu_i}}_{:=\beta_i} x_i \right] \\ &= \alpha \sum_{i=1}^{m} \beta_i x_i \end{aligned}$$

Then $\alpha > 0$, $\beta_i > 0, \forall i \in \{1, \ldots, m\}$ and $\sum_{i=1}^{m} \beta_i = 1$. Hence

$$x = \alpha \underbrace{\sum_{i=1}^{m} \beta_i x_i}_{\in conv(C)} \in cone(conv(C))$$

4. This is a <u>direct</u> consequence of 3) and 2),

$$\overline{cone}(conv(C)) = \overline{conv}(cone(C))$$

$\square$

Lemma 2.9

Let C be a convex subset of $\mathbb{R}^n$ such that $int(C) \neq \emptyset$ and $0 \in C$. Then the following are equivalent.

1. $0 \in int(C)$
2. $cone(C) = \mathbb{R}^n$
3. $\overline{cone}(C) = \mathbb{R}^n$

Proof.

- $1) \implies 2)$: Indeed, $0 \in int(c) \Leftrightarrow \exists \varepsilon > 0$ such that $B(0;\varepsilon) \subseteq C$. Hence,

$$\begin{aligned}\mathbb{R}^n &= cone(B(0;\varepsilon))\\ &\subseteq cone(C) \subseteq \mathbb{R}^n\\ \implies cone(c) &= \mathbb{R}^n\end{aligned}$$

- $2) \implies 3)$ By an earlier Proposition

$$\overline{cone(C)} = \overline{cone}(C)$$

Nowe,

$$\mathbb{R}^n \overset{2)}{=} cone(c) \subseteq \overline{cone(C)} = \overline{cone}C$$

- $3) \implies 1)$: $\overline{cone}(C) = \mathbb{R}^n \overset{??}{\Rightarrow} 0 \in int(C)$

By an earlier result, we proved that for any set C we have

$$cone(conv(C)) = conv(cone(C))$$

Since C is convex, we have

$$C = conv(C)$$

Hence,

$$cone(C) = conv(cone(C))$$

implies that $cone(C)$ is convex. By assumption

$$\emptyset \neq int(C) \subseteq int(cone(C))$$

Hence, $cone(C)$ is a convex set,

$$int(cone(C)) \neq \emptyset$$

By an earlier result

$$int(cone(C)) = int(\overline{cone(C)}) = int(\overline{cone}(C))$$

Hence,

$$\begin{aligned}\mathbb{R}^n &= int(\mathbb{R}^n)\\ &= int(\overline{cone}(C))\\ &= int(cone(C))\\ &= cone(int(C))\\ \implies & 0 \in cone(int(C))\\ \implies & 0 \in \lambda int(C), \text{ for some } \lambda > 0\\ \implies & 0 \in int(C)\end{aligned}$$

Fact: Let C be a convex subset of $\mathbb{R}^n$ such that $int(C) \neq \emptyset$ and $0 \in C$, then

$$int(cone(C)) = cone(int(C))$$

□

Definition 2.10: Tangent and Normal Cones

Let C be a nonempty convex subset of $\mathbb{R}^n$ and let $x \in \mathbb{R}^n$.
The tangent cone to C at x is

$$T_c(x) = \begin{cases} \overline{cone}(C - x) = \overline{\cup_{\lambda \in \mathbb{R}_{++}} \lambda(C - x)}, & x \in C; \\ \emptyset, & x \notin C \end{cases}$$

and the normal cone of C at x is

$$N_c(x) = \begin{cases} \{u \in \mathbb{R}^n | \sup_{c \in C} \langle c - x, u \rangle \leqslant 0\}, & x \in C; \\ \emptyset, & x \notin C \end{cases}$$

Example 2.11

Let $C = B = B(0; 1) \subseteq \mathbb{R}^n$.

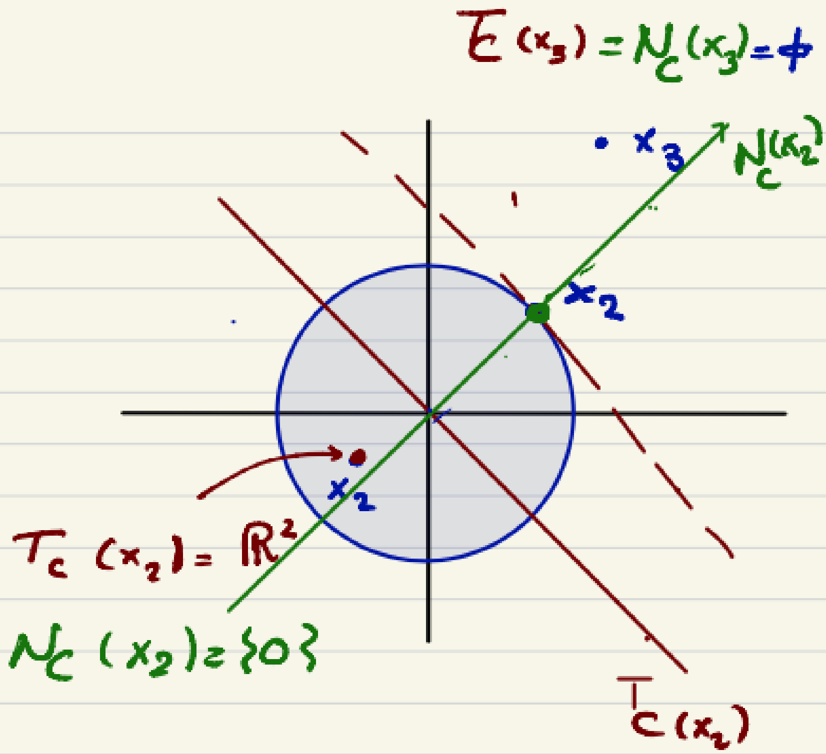
$$T_C(x) = \begin{cases} \{y \in \mathbb{R}^n | \langle x, y \rangle \leqslant 0\}, & \|x\| = 1; \\ \mathbb{R}^n, & \|x\| < 1; \\ \emptyset, \ \textit{otherwise} \end{cases}$$

Theorem 2.12

Let C be a nonempty closed convex subset of $\mathbb{R}^n$ and let $x \in \mathbb{R}^n$. Prove that $N_C(x), T_C(x)$ are closed convex cones.

Proof. See A2. □

15



$$N_C(x) = \begin{cases} \mathbb{R}_+ x, & \|x\| = 1; \\ \{0\}, & \|x\| < 1; \\ \emptyset & \text{otherwise.} \end{cases}$$

Lemma 2.13

Let C be a nonempty closed convex subset of $\mathbb{R}^n$ and let $x \in C$. Then

$$n \in N_C(x) \Leftrightarrow \forall t \in T_C(x), \langle n, t\rangle \leqslant 0$$

Proof.

- $(\rightarrow)$ Let $n \in N_C(x)$, and let $t \in T_C(x)$. Recall that

$$T_C(x) = \overline{cone}(C - x)$$

Therefore, there exists $\lambda_K > 0$, $(t_K)_{k\in\mathbb{N}}$ in $\mathbb{R}^n$ such that

$$\forall k \in \mathbb{N}, x + \lambda_k t_k \in C,\ t_k \to t$$

Since $n \in N_C(x)$, and $x + \lambda_k t_k \in C$, we learn that

$$\begin{aligned} &\forall k \in \mathbb{N}, \langle n, \lambda t_k\rangle \\ &= \langle n, x + \lambda_k t_k - x\rangle \leqslant 0 \\ &\overset{\lambda_k > 0}{\Rightarrow} \forall k \in \mathbb{N}, \langle n, t_k\rangle \leqslant 0 \end{aligned}$$

Letting $k \to \infty$,

$$\implies \langle n, t\rangle \leqslant 0$$

- $(\Leftarrow)$ Suppose that $\forall t \in T_C(x)$, we have $\langle n, t\rangle \leqslant 0$.
Let $y \in C$ and observe that

$$y - x \in T_C(x)$$

$$(y - x \in C - x \subseteq cone(C - x) \subseteq \overline{cone}(C - x))$$

Therefore,

$$\langle n, y - x\rangle \leqslant 0 \implies n \in N_C(x)$$

□

Theorem 2.14

Let C be a convex subset of $\mathbb{R}^n$ such that $int(C) \neq \emptyset$, and let $x \in C$. Then,

$$x \in int(C) \overset{(1)}{\Leftrightarrow} T_C(x) = \mathbb{R}^n \overset{(2)}{\Leftrightarrow} N_C(x) = \{0\}$$

Proof.

- (1) Observe that

$$x \in int(C) \Leftrightarrow 0 \in int(C - x)$$

Applying the earlier result (lemma 18) with C replaced by $C - x$.

$$0 \in int(C - x) \Leftrightarrow \overline{cone}(C - x) = \mathbb{R}^n \Leftrightarrow T_C(x) = \mathbb{R}^n$$

- (2) Recalling the earlier Lemma 20
  Let $T_C(x) = \mathbb{R}^n$.

$$\begin{aligned} n \in N_C(x) &\Leftrightarrow \forall t \in T_C(x) = \mathbb{R}^n, \langle n, t\rangle \leqslant 0 \\ &\implies \langle n, n\rangle \leqslant 0 \\ &\Leftrightarrow \|n\|^2 = 0 \Leftrightarrow n = 0 \end{aligned}$$

  Hence, $N_C(x) \subseteq \{0\}$. Clearly, $\{0\} \subseteq N_C(x)$.
  Hence, $N_C(x) = \{0\}$ as claimed.
  Conversely, if $N_C(x) = \{0\}$, for simplicity, set $K = T_C(x)$. Recall that K is a closed convex cone , $0 \in K$.
  Let $x \in \mathbb{R}^n$ and set $p = P_K(x)$.
  By the projection theorem

$$\forall y \in K, \langle x - p, y - p\rangle \leqslant 0$$

  In particular,

$$\begin{aligned} &\langle x - p, -p\rangle \leqslant 0 \text{ By setting } y = 0 \\ &\langle x - p, p\rangle \leqslant 0 \text{ By setting } y = 2p \in K \text{ as } K \text{ is a cone} \\ \implies &\langle x - p, p\rangle = 0 \end{aligned}$$

  Hence the projection theorem gives

$$\forall y \in K, \langle x - p, y\rangle \leqslant 0$$

  It follows from the lemma 20 that $x - p \in N_C(x) = \{0\}$.
  Hence, $x - p = 0$; equivalently

$$x = p = P_K(x) \in K$$

  so $\mathbb{R}^n \subseteq K \implies \mathbb{R}^n = K = T_C(x)$

$\square$

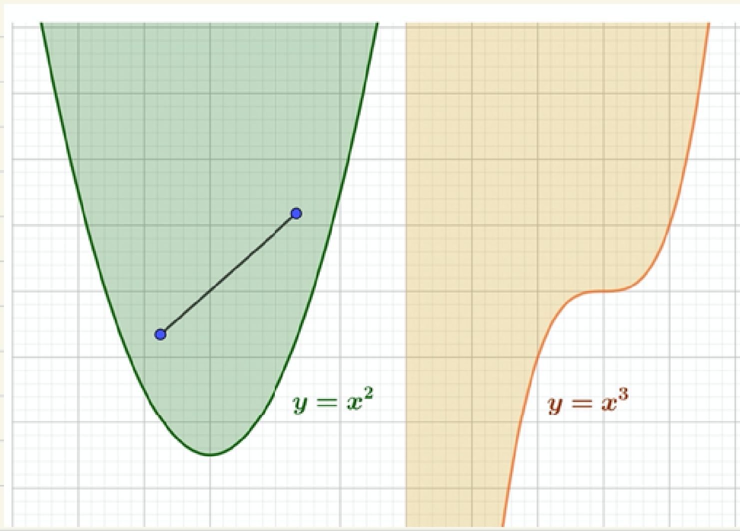
3 Convex Function

Definition 3.1: Epigraph

Let $f : \mathbb{R}^n \to [-\infty, \infty]$. *The epigraph of* f *is*

$$epi(f) = \{(x, \alpha) | f(x) \leqslant \alpha\} \subseteq \mathbb{R}^n \times \mathbb{R}$$

Example 3.2



$*$ $[-\infty, +\infty] = \mathbb{R} \cup \{\pm\infty\}$

Definition 3.3

Let $f : \mathbb{R}^n \to [-\infty, \infty]$*. Then*

$$dom(f) = \{x \in \mathbb{R}^n | f(x) < \infty\}$$

f *is* ***proper*** *if* $dom(f) \neq \emptyset$ *and*

$$\forall x \in \mathbb{R}^n,\ f(x) > -\infty$$

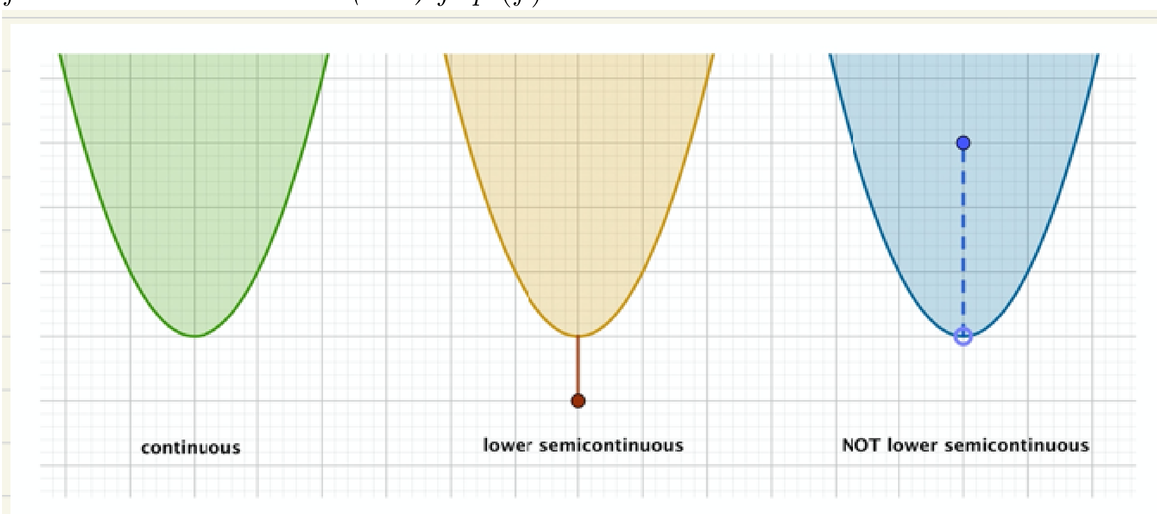
Example 3.4

- *Let* $f : \mathbb{R}^m \to (-\infty, \infty)$ *be continuous. Then* f *is proper.*
- *Let* C *be a subset of* $\mathbb{R}^m$*. The indicator function of* C *at* $x \in \mathbb{R}^m$ *(see txtbook p 28) is*

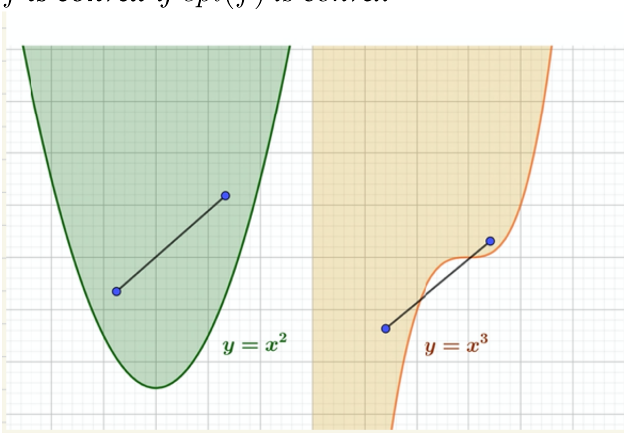
$$\delta_C(x) = \begin{cases} 0, & x \in C \\ \infty, & o/w \end{cases}$$

Clearly, δ_C *is proper whenever* $C \neq \emptyset$*.*

f *is* ***lower semicontinuous*** *(l.s.c) if* $epi(f)$ *is closed.*



f ***is convex*** *if* $epi(f)$ *is convex*



Proposition 3.5: L5-1

Let $f : \mathbb{R}^m \to [-\infty, \infty]$ *be convex. Then* $dom(f) = \{x \in \mathbb{R}^n | f(x) < \infty\}$ *is convex.*

Proof. Fact: Let C be subset of $\mathbb{R}^n$ and let $A : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. If C is a convex subset of $\mathbb{R}^n$ then $A(C)$ is a convex subset of $\mathbb{R}^m$
Recall that

$$epi(f) = \{(x, \alpha) | f(x) \leqslant \alpha\} \subseteq \mathbb{R}^{n+1}$$

Consider the linear map (transformation)

$$L : \mathbb{R}^{n+1} \to \mathbb{R}^n : (x, \alpha) \to x$$

Then $dom(f) = L(epi(f))$, and the conclusion follows in view of the above Fact. □

Theorem 3.6: L5-2

Let $f : \mathbb{R}^m \to [-\infty, \infty]$. *Then* f *is convex if and only if*

$$\forall x, y \in dom(f), \forall \lambda \in (0,1), f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda) f(y)$$

Proof. Observe that $f = \infty \Leftrightarrow epi(f) = \emptyset \Leftrightarrow dom(f) = \emptyset$ and the conclusion follows.
Now, suppose $dom(f) \neq \emptyset$,

- ($\implies$) Let $(x, y) \in dom(f) \times dom(f)$ and let $\lambda \in (0,1)$. Observe that $(x, f(x)) \in epi(f)$, $(y, f(y)) \in epi(f)$. By convexity of $epi(f)$ we have

$$\begin{aligned} \lambda(x, f(x)) + (1-\lambda)(y, f(y)) &= (\lambda x + (1-\lambda)y, \lambda f(x) + (1-\lambda) f(y)) \in epi(f) \\ \implies f(\lambda x + (1-\lambda)y) &\leqslant \lambda f(x) + (1-\lambda) f(y) \end{aligned}$$

- ($\impliedby$). Let $(x, \alpha) \in epi(f), (y, \beta) \in epi(f), \lambda \in (0,1)$
Observe that this implies that

$$f(x) \leqslant \alpha, f(y) \leqslant \beta$$

Now,

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &\leqslant \lambda f(x) + (1-\lambda) f(y) \\ &\leqslant \lambda \alpha + (1-\lambda)\beta \end{aligned}$$

Hence,

$$(\lambda x + (1-\lambda)y, \lambda\alpha + (1-\lambda)\beta) \in epi(f)$$

which implies

$$\lambda(x, \alpha) + (1-\lambda)(y, \beta) \in epi(f)$$

That is, $epi(f)$ is convex. Equivalent, f is convex.

□

3.1 Lower Semicontinuity

Definition 3.7: Lower semicontinuity(Alter. Defn)

Let $f : \mathbb{R}^m \to [-\infty, \infty]$, and let $x \in \mathbb{R}^m$. Then f is lower semicontinuous (l.s.c.) at x if , for every sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathbb{R}^m$,

$$x_n \to x \implies f(x) \leqslant \liminf f(x_n)$$

Moreover, f is l.s.c. if f is l.s.c. at every point in $\mathbb{R}^m$.

Remark.
1. If f is continuous then f is l.s.c.
2. One can show the equivalence of the definition(s) of l.s.c. However, we will omit the proof.

Example 3.8: The indicator function

Let $C \subseteq \mathbb{R}^m$. Then indicator function $\delta_C : \mathbb{R}^m \mapsto (-\infty, \infty]$ of C is defined by

$$\delta_C(x) = \begin{cases} 0, & x \in C \\ \infty, & x \notin C \end{cases}$$

Theorem 3.9: L5-3

Let $C \subseteq \mathbb{R}^m$. Then the following hold

1. *$C \neq \emptyset \iff \delta_C$ is proper*
2. *C is convex $\iff \delta_C$ is convex*
3. *C is closed $\iff \delta_C$ is l.s.c*

Proof.
1. See A2
2. See A2
3. Observe that $C = \emptyset \iff epi(\delta_C) = \emptyset$ which is closed. Now suppose $C \neq \emptyset$
   - ($\implies$) Suppose C is closed.
     We want to show that $epi(\delta_C)$ is closed. Let $((x_n, \alpha_n))_{n\in\mathbb{N}}$ be a sequence in $epi(\delta_C)$, such that $(x_n, \alpha_n) \to (x, \alpha)$.
     Observe that:
     $$(x_n)_{n\in\mathbb{N}} \text{ is a sequence in } C,\ x_n \to x$$
     Hence, $x \in C$ (C closed). And $(\alpha_n)_{n\in\mathbb{N}}$ is a sequence in $[0, \infty)$, $\alpha_n \to \alpha$. Hence $\alpha \geqslant 0$. Indeed,
     $$\forall n \in \mathbb{N}, 0 = \delta_C(x_n) \leqslant \alpha_n$$

Consequently,

$$\begin{aligned} 0 = \delta_C(x) \leqslant \alpha \\ \implies (x, \alpha) \in epi(\delta_C) \end{aligned}$$

- ($\Longleftarrow$) Conversely, suppose that δ_C is l.s.c. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in C, $x_n \to x$. We want to show that $x \in C$. By definition of δ_C, it is sufficient to show that $\delta_C(x) = 0$. Observe that
$$0 \leqslant \delta_C(x) \leqslant \liminf \delta_C(x_n) = 0$$
Hence, $\delta_C(x) = 0 \implies x \in C$

$\square$

Why optimizers like indicator functions? Consider the problem

$$(P)\min f(x),\ s.t.\ x \in C \subseteq \mathbb{R}^m$$

f convex, l.s.c proper, C convex closed $\neq \emptyset$

Then (P) is equivalent to

$$\min_{x\in\mathbb{R}^m}\ h(x) := f(x) + \delta_C(x)$$

where $h(x) = \begin{cases} f(x), & x \in C \\ \infty, & x \notin C \end{cases}$.

The problem is now "unconstrained" minimization of "a sum of two" functions.

- f is not necessarily smooth
- δ_C is Not smooth (whenever $C \neq \mathbb{R}^m$)

Proposition 3.10: L5-4

let I be an indexed set and let $(f_i)_{i\in I}$ be a family of l.s.c convex functions on $\mathbb{R}^n$. Then $\sup_{i\in I} f_i$ is convex and l.s.c

Proof. Set $F = \sup_{i\in I} f_i$
We claim that

$$epi(F) = \cap_{i\in I} epi(f_i) \ldots (*)$$

Indeed, let $(x, \alpha) \in \mathbb{R}^m \times \mathbb{R}$. Then

$$\begin{aligned}(x, \alpha) \in epi(F) &\iff \sup_{i\in I} f_i(x) \leqslant \alpha \\ &\iff \forall i \in I, f_i(x) \leqslant \alpha \\ &\iff \forall i \in I, (x, \alpha) \in epi(f_i) \\ &\iff (x, \alpha) \in \cap_{i\in I} epi(f_i)\end{aligned}$$

This proves $(*)$

- F is l.s.c.
  Since $\forall i \in I$, f_i is l.s.c., we conclude that $\forall i \in I$, $epi(f_i)$ is closed. Now combine with $(*)$ to learn that
  $$epi(F) = \cap_{i\in I} epi(f_i) \text{ is closed} \implies F \text{ is l.s.c}$$
- F is convex
  Since $\forall i \in I$, f_i is convex, we conclude that $\forall i \in I$, $epi(f_i)$ is convex. Now combine with $(*)$ and an earlier result to learn that
  $$epi(F) = \cap_{i\in I} epi(f_i) \text{ is convex}$$

□

3.2 The Support Function (txtbook p-28)

Definition 3.11

Let C be a subset of $\mathbb{R}^m$. The support function of C is

$$\begin{aligned} \sigma &: \mathbb{R}^m \mapsto [-\infty, \infty] \\ &: u \mapsto \sup_{c \in C} \langle c, u \rangle \end{aligned}$$

Proposition 3.12: L5-5

Let C be a nonempty subset of $\mathbb{R}^n$ Then σ_C is convex, l.s.c and proper

Proof. Let $c \in C$ and set

$$f_C : \mathbb{R}^m \mapsto \mathbb{R} \ : x \mapsto \langle x, c \rangle$$

Then f_C is proper, l.s.c and convex (In fact, f_C is linear). Moreover,

$$\sigma_C = \sup_{c \in C} f_c$$

Now combine with the earlier result (L5-4) to learn that σ_C is convex and l.s.c.
Finally, observe that, since $C \neq \emptyset$,

$$\sigma_C(0) = \sup_{c \in C} \langle 0, c \rangle = 0 < \infty$$

Hence, $0 \in \ dom(\sigma_C) \neq \emptyset$. Moreover, let $\bar{c} \in C$. Then $\forall u \in \mathbb{R}^m$,

$$\begin{aligned} \sigma_C(u) &= \sup_{c \in C} \langle u, c \rangle \\ &\geqslant \langle u, \bar{c} \rangle \\ &> -\infty \end{aligned}$$

Hence, σ_C is proper. $\square$

Example 3.13: L5-6

Let $C = [a, b] \subseteq \mathbb{R}_+$. *Then* $\forall x \in \mathbb{R}$

$$\sigma_C(x) = \sup_{c \in [a,b]} cx = \begin{cases} bx, & x \geqslant 0 \\ ax, & x < 0 \end{cases}$$

Example

Let $C = [0, \infty) \subseteq \mathbb{R}$. *We examine two cases:*

1. $x \leqslant 0$, *then*
$$\sigma_C(x) = \sup_{c \in [0,\infty)} cx = 0$$

2. $x > 0$, *then*
$$sup_{c \in [0,\infty)} cx = \infty$$

Hence $dom(\sigma_C) = (-\infty, 0]$. *Moreover,*

$$\forall x \in (-\infty, 0], \sigma_C(x) = 0$$

Definition 3.14

Let $f : \mathbb{R}^m \mapsto (-\infty, \infty]$ *be proper. Then* f *is*

1. *Strictly convex if*
$$\forall x, y \in dom(f), x \neq y, \lambda \in (0, 1) \implies f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y)$$

2. *strongly convex with constant* β*, if for some* $\beta > 0$ *we have:*
$$\forall x, y \in dom(f), x \neq y, \lambda \in (0, 1)$$
$$\implies f(\lambda x + (1 - \lambda)y) \leqslant \lambda f(x) + (1 - \lambda)f(y) - \frac{\beta}{2}\lambda(1 - \lambda)\|x - y\|^2$$

Clearly,

$$\textit{Strong Convexity} \implies \textit{Strict Convexity} \implies \textit{Convexity}$$

and example for f *being strictly convex but not strongly convex is* $f(x) = e^x$.

3.3 Operations That Preserves Convexity

Proposition 3.15: L6-1

Let I be a finite indexed set, let $(f_i)_{i\in I}$ be a family of Convex functions from $\mathbb{R}^m$ to $[-\infty, \infty]$, then

$$\sum_{i\in I} f_i \text{ is convex}$$

Proof. See A2 □

Proposition 3.16: L6-2

Let f be convex and l.s.c and let $\lambda > 0$. Then λf is convex and l.s.c

Proof. See A2 □

Definition 3.17: Minimizers of Functions

Let $f : \mathbb{R}^n \mapsto (-\infty, \infty]$ be proper and let $x \in \mathbb{R}^m$. Then x is a (global) minimizer of f if

$$f(x) = \min f(\mathbb{R}^m) \in \mathbb{R}$$

Throughout this course we will use $\arg\min f$ to denote the set of minimizers of f.

Definition 3.18: Local and Global Minimizers/Maximizers

Let $f : \mathbb{R}^m \mapsto (-\infty, \infty]$ be proper and let $\overline{x} \in \mathbb{R}^m$. Then:

- *$\overline{x}$ is a local minimum of f if $\exists\delta > 0$ such that*

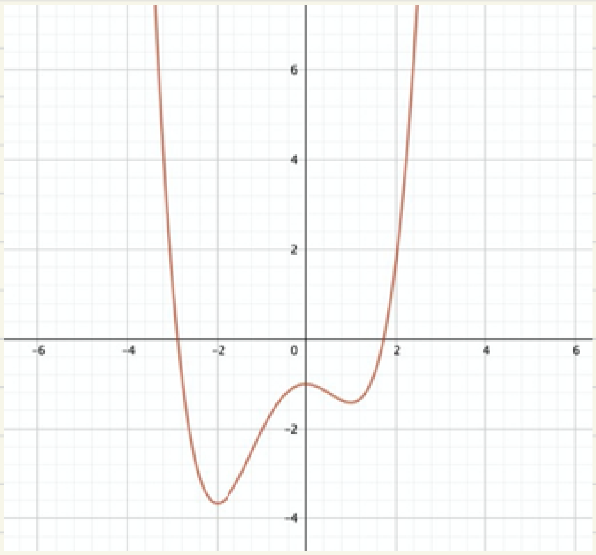
$$\|x - \overline{x}\| < \delta \implies f(\overline{x}) \leqslant f(x)$$

- *$\overline{x}$ is a global minimum of f if*

$$\forall x \in dom(f),\ f(\overline{x}) \leqslant f(x)$$

Analogously, we define local/global max.

Example 3.19: L6-3



$f'(x)$

$= x^3 + x^2 - 2x$

$= x(x-1)(x+2)$

$f(x) = \frac{1}{4}x^4 + \frac{1}{3}x^3 - x^2 - 1$

Local minimums at $x=1$, $x=-2$

Global minimum at $x=-2$

local max at $x=0$

No global maximum.

Why do we "love" convex functions?

Proposition 3.20: L6-4

Let $f : \mathbb{R}^m \mapsto (-\infty, \infty]$ be proper and convex. Then every local minimizer of f is a global minimizer.

Proof. Let x be a local minimizer of f. Then $\exists p > 0$ such that

$$f(x) = \min f(B(x;p))$$

Let $y \in dom(f)$ and observe that if $y \in B(x;p)$ (i.e. $\|x-y\| \leqslant p$) then $f(x) \leqslant f(y)$.
Now, suppose that $y \in dom(f) \setminus B(x;p)$. Observe that $\lambda := 1 - \frac{p}{\|x-y\|} \in (0,1)$, set

$$z = \lambda x + (1-\lambda)y \in dom(f)$$

note $dom(f)$ is convex by L5-1. Moreover:

$$\begin{aligned} z - x &= \lambda x + (1-\lambda)y - x \\ &= (1-\lambda)y - (1-\lambda)x \\ &= (1-\lambda)(y-x) \end{aligned}$$

Hence,

$$\begin{aligned} \|z - x\| &= \|(1-\lambda)(y-x)\| \\ &= (1-\lambda)\|y-x\| \\ &= \frac{p}{\|y-x\|}\|y-x\| = p \end{aligned}$$

Hence,

$$z \in B(x;p)$$

Moreover, because f is convex, it follows from Jensen's Inequality that

$$\begin{aligned} f(x) &\leqslant f(z) \\ &= f(\lambda x + (1-\lambda)y) \\ &\leqslant \lambda f(x) + (1-\lambda)f(y) \end{aligned}$$

Hence,

$$(1-\lambda)f(x) \leqslant (1-\lambda)f(y) \implies f(x) \leqslant f(y)$$

□

Proposition 3.21: L6-5

let $f : \mathbb{R}^m \mapsto (-\infty, \infty]$ be proper and convex and let C be a subset of $\mathbb{R}^m$. Suppose that x is a minimizer of f over C such that $x \in \ int(C)$. Then x is a minimizer of f

Proof. Since $x \in int(C)$, $\exists \varepsilon > 0$ such that $B(x;\varepsilon) \subseteq C$.
Since x is a minimizer of f over $C \supseteq B(xl\varepsilon)$ we learn that

$$f(x) = \inf f(B(x;\varepsilon))$$

That is, x is a local minimizer of f. Now we combine with (L6-4) to get the result. □

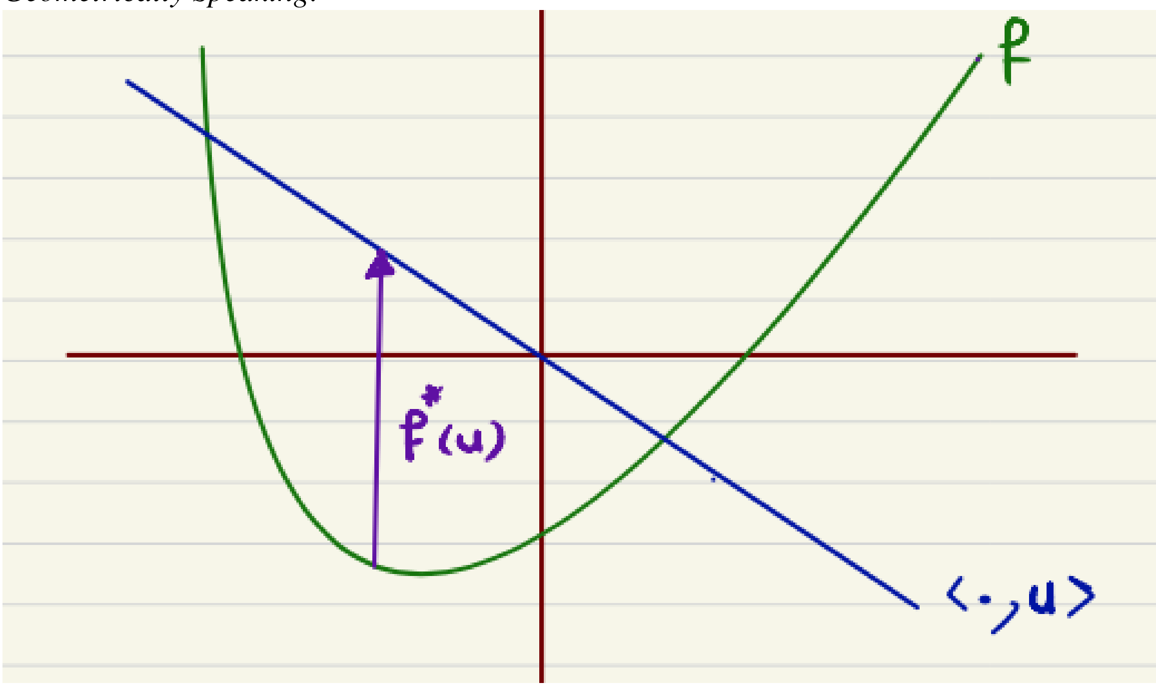
3.4 Conjugates of Convex Functions

Definition 3.22: Conjugates of Convex Functions

Let $f : \mathbb{R}^m \mapsto [-\infty, \infty]$. The Fenchel-Legendre/ConVex Conjugate of f is

$$\begin{aligned} f^* &: \mathbb{R}^m \mapsto [-\infty, \infty] \\ &: u \mapsto \sup_{x \in \mathbb{R}^m} (\langle x, u \rangle - f(x)) \end{aligned}$$

Geometrically Speaking:



Proposition 3.23: L6-6

Let $f : \mathbb{R}^m \mapsto [-\infty, \infty]$. Then f^* is convex and l.s.c.

Proof. Observe that if $f \equiv \infty \iff dom(f) = \emptyset$. Hence, $\forall u \in \mathbb{R}^m$,

$$\begin{aligned} f^*(u) &= \sup_{x \in \mathbb{R}^m} (\langle x, u \rangle - f(x)) \\ &= \sup_{x \in dom(f)} (\langle x, u \rangle - f(x)) \\ &= -\infty \end{aligned}$$

i.e. $f^* = -\infty$ which is l.s.c. and convex.
Now suppose that $f \not\equiv \infty$. we claim that $\forall u \in \mathbb{R}^m$,

$$f^* = \sup_{(x,\alpha) \in epi(f)} (\langle x, u \rangle - \alpha) \ldots (*)$$

$f(x,\alpha) := \langle x, \cdot \rangle - \alpha$ is an affine function.
Indeed, let $u \in \mathbb{R}^m$.
On the one hand, $\forall (x,\alpha) \in epi(f)$, we have

$$\langle x, u\rangle - f(x) \geqslant \langle x, u\rangle - \alpha$$

Hence,

$$\sup_{x\in\mathbb{R}^m} (\langle x, u\rangle - f(x)) \geqslant \sup_{(x,\alpha)\in epi(f)} (\langle x, u\rangle - \alpha)$$

On the other hand,

$$G = \{(x, f(x)) | x \in dom(f)\} \subseteq epi(f)$$

Hence,

$$\begin{aligned}\sup_{x\in\mathbb{R}^m} (\langle x, u\rangle - f(x)) &= \sup_{x\in dom(f)} (\langle x, u\rangle - f(x)) \\ &= \sup_{(x,f(x))\in G} (\langle x, u\rangle - f(x)) \\ &\leqslant \sup_{(x,\alpha)\in epi(f)} (\langle x, u\rangle - \alpha)\end{aligned}$$

Altogether, we learn that $(*)$ holds. This implies $\forall u \in \mathbb{R}^m$,

$$f^*(u) = \sup_{\substack{(x,\alpha)\in epi(f) \\ \subseteq \mathbb{R}^m\times\mathbb{R}}} (f_{(x,\alpha)}(u))$$

Now by L5-4, we get required result. □

Example 3.24: L6-7

let $p > 1$ *and set* $q = \frac{p}{p-1}$. *Let*

$$f : \mathbb{R} \mapsto \mathbb{R} : \ x \mapsto \frac{|x|^p}{p}$$

Then

$$f^* : \mathbb{R} \mapsto \mathbb{R} : \ u \mapsto \frac{|u|^q}{q}$$

Proof. Observe that $f(x)$ is differentiable on $\mathbb{R}$, $f(x) = \begin{cases} \frac{x^p}{p}, & x \geqslant 0 \\ \frac{(-x)^p}{p}, & x < 0 \end{cases}$.

Now, let $u \in \mathbb{R}$

$$\begin{aligned} f^*(u) &= \sup_{x\in\mathbb{R}} (xu - f(x)) \\ &= \sup_{x\in\mathbb{R}} \left(\underbrace{xu - \frac{|x|^p}{p}}_{:=g(x)} \right) \end{aligned}$$

so

$$g'(x) = u - \begin{cases} x^{p-1}, & x \geqslant 0 \\ -(-x)^{p-1}, & x < 0 \end{cases}$$

If $u \geqslant 0$, then setting $g'(x) = 0$ yields $x^{p-1} = u$, and $x \geqslant 0$; equivalently

$$x = u^{1/(p-1)}$$

If $u < 0$, then setting $g'(x) = 0$ yields $u = -(|x|)^{p-1}$, and $x < 0$; equivalently

$$|u| = -u = |x|^{1/(p-1)}$$

Altogether, $|x| = |u|^{1/(p-1)}$ and $sign(x) = sign(u)$. Hence,

$$\begin{aligned} f^*(u) &= |u|^{\frac{1}{p-1}}|u| - \frac{|u|^{\frac{p}{p-1}}}{p} \\ &= (1 - 1/p)|u|^{1/(1-p)+1} \\ &= \frac{p-1}{p}|u|^{\frac{p}{p-1}} \\ &= \frac{|u|^q}{q} \end{aligned}$$

□

Example 3.25: L6-8

Let $f : \mathbb{R} \mapsto \mathbb{R},\ f(x) = e^x$. *Then*

$$f^*(u) = \begin{cases} u\ln(u) - u, & u > 0 \\ 0, & u = 0 \\ \infty, & u < 0 \end{cases}$$

Proof. Let $u \in \mathbb{R}$, then

$$f^*(u) = \sup_{x \in \mathbb{R}} (\underbrace{xu - e^x}_{:= g(x)})$$

Hence,

$$\text{If } u = 0 \implies f^*(u) = \sup_{x \in \mathbb{R}} (-e^x) = 0$$
$$\text{If } u > 0 \implies f^*(u) = u\ln(u) - u$$

Also, $g'(x) = u - e^x$. Setting $g'(x) = 0$

$$\implies e^x = u \iff x = \ln(u)$$

If $u < 0 \implies g'(x) < 0, \forall x \in \mathbb{R}$. Therefore, $g(x)$ is decreasing on $\mathbb{R}$. Hence

$$\sup_{x \in \mathbb{R}} g(x) = \lim_{x \to -\infty} g(x) = \infty$$

□

Example 3.26: L6-9

Let C be a subset of $\mathbb{R}^m$. Then $\delta_c^* = \sigma_C$.

Proof. Indeed, recall that:

$$\delta_C(x) = \begin{cases} 0, & x \in C \\ \infty, & x \notin C \end{cases}$$
$$\sigma_C(x) = \sup_{y \in C} \langle x, y \rangle$$

Now,

$$\begin{aligned} \delta_C^*(u) &= \sup_{y \in C} \left(\langle x, y \rangle - \delta_C(y) \right) \\ &= \sup_{y \in C} \langle x, y \rangle \end{aligned}$$

□

3.5 The Subdifferential Operator

Definition 3.27

Let $f : \mathbb{R}^m \mapsto (-\infty, \infty]$ be proper. The subdifferential of f is the set-valued operator

$$
\begin{aligned}
\partial f &: \mathbb{R}^m \rightrightarrows \mathbb{R}^m \\
&: x \mapsto \{u \in \mathbb{R}^m | \forall y \in \mathbb{R}^m,\ f(y) \geqslant f(x) + \langle u, y - x\rangle\}
\end{aligned}
$$

Let $x \in \mathbb{R}^m$. Then f is subdifferentiable at x if $\partial f(x) \neq \emptyset$. The elements of $\partial f(x)$ are called the subgradient of f at x.

Theorem 3.28: Fermat L6-10

Let $f : \mathbb{R}^m \mapsto (-\infty, \infty]$ be proper. Then

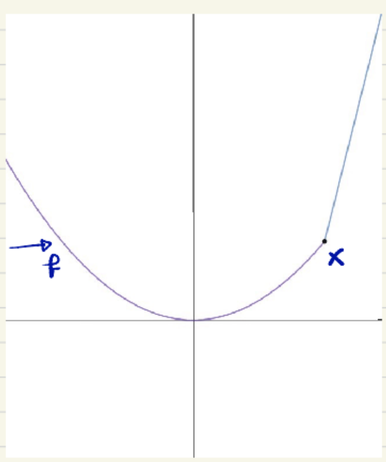
$$\arg\min f = \{x \in \mathbb{R}^m | 0 \in \partial f(x)\} := zero(\partial f)$$

Proof. Indeed, let $x \in \mathbb{R}^m$. Then

$$
\begin{aligned}
x \in \arg\min f &\iff \forall y \in \mathbb{R}^m, f(x) \leqslant f(y) \\
&\iff \forall y \in \mathbb{R}^m\ \langle 0, y - x\rangle + f(x) \leqslant f(y) \\
&\iff 0 \in \partial f(x)
\end{aligned}
$$

□

Consider the function



u is a subgradient of f at x if for all $y \in \mathbb{R}^m$ we have

$f(y) \geqslant f(x) + \langle u, y-x \rangle =: h(y)$

h(y) is an affine function, with slope = u

$h(x) = f(x)$

Example 3.29: L6-11

Let $f : \mathbb{R} \mapsto \mathbb{R} \ : x \mapsto |x|$, *then*

$$\partial f(x) = \begin{cases} \{-1\}, & x < 0 \\ [-1, 1], & x = 0 \\ \{1\}, & x > 0 \end{cases}$$

Proof. See A2 □

Lemma 3.30: L6-12

$f : \mathbb{R}^m \mapsto (-\infty, \infty]$ *proper* $\implies$ $dom(\partial f) \subseteq dom(f)$

Proof. Indeed, if $f(x) = \infty \implies \partial f(x) = \emptyset$.
"Contrapositive: $x \notin dom(f) \implies x \notin dom(\partial f)$" □

Example 3.31: L6-13

Let C *be a convex closed nonempty subset of* $\mathbb{R}^m$. *Let* $x \in \mathbb{R}^m$, *then*

$$\partial \delta_C(x) = N_C(x)$$

Proof. Indeed, let $u \in \mathbb{R}^m$ and let $x \in C$ ($dom(\partial f) \subseteq dom(f)$), then

$$\begin{aligned}
& u \in \partial \delta_C(x) \\
\iff & \forall y \in \mathbb{R}^m, \delta_C(y) \geqslant \delta_C(x) + \langle u, y - x \rangle \\
\iff & \forall y \in C, \delta_C(y) \geqslant \delta_C(x) + \langle u, y - x \rangle \\
\iff & \forall y \in C, 0 \geqslant \langle u, y - x \rangle \\
\iff & u \in N_C(x)
\end{aligned}$$

$\square$

Casually Speaking
Recall the problem

$$(P) \min f(x), \ s.t. \ x \in C \subseteq \mathbb{R}^m$$

f convex, l.s.c proper, C convex closed $\neq \emptyset$

Then (P) is equivalent to

$$\min_{x\in\mathbb{R}^m} \ h(x) := f(x) + \delta_C(x)$$

where $h(x) = \begin{cases} f(x), & x \in C \\ \infty, & x \notin C \end{cases}$.
In view of Fermat's Theorem:

$$x \text{ is a minimizer of } h(x) \iff 0 \in \partial h(x)$$

Goal: Find x such that $0 \in \partial h(x)$

$$\begin{aligned} \partial h(x) &= \partial(f + \delta_C)(x) \\ &= (\partial f + \partial\delta_C)(x) \text{ requires more assumptions} \\ &= \partial f(x) + \partial\delta_C(x) \\ &= \partial f(x) + N_C(x) \end{aligned}$$

3.6 Calculus of Subdifferentials

Let $f, g : \mathbb{R}^m \mapsto (-\infty, \infty]$ be proper and let $x \in \mathbb{R}^m$. Suppose that f, g are differentiable at x, then

$$\nabla(f+g)(x) = \nabla f(x) + \nabla g(x)$$

Question:
Let $f.g : \mathbb{R}^m \mapsto (-\infty, \infty]$ be proper, convex l.s.c let $x \in \mathbb{R}^m$. Suppose that f, g are subdifferentiable at x, then

$$\partial(f+g)(x) \stackrel{?}{=} \partial f(x) + \partial g(x)$$

Fact (L7-1):
Let $f : \mathbb{R}^m \mapsto (-\infty, \infty]$ be convex l.s.c and proper, then

$$\emptyset \neq ri(dom(f)) \subseteq dom(\partial f)$$

In particular,

$$ri(dom(f)) = ri(dom(\partial f))$$
$$\overline{dom(f)} = \overline{dom(\partial f)}$$

Separation Theorem revisited:
Let C_1, C_2 be nonempty subsets in $\mathbb{R}^m$. Then,

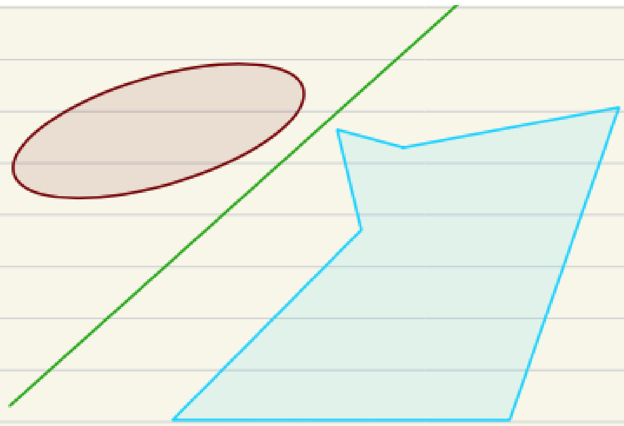
- C_1, C_2 are separated if $\exists b \neq 0$ such that

$$\sup_{c_1 \in C_1} \langle b, c_1 \rangle \leqslant \inf_{c_2 \in C_2} \langle b, c_2 \rangle$$

- C_1, C_2 are properly separated if $\exists b \neq 0$ such that C_1 and C_2 are separated and

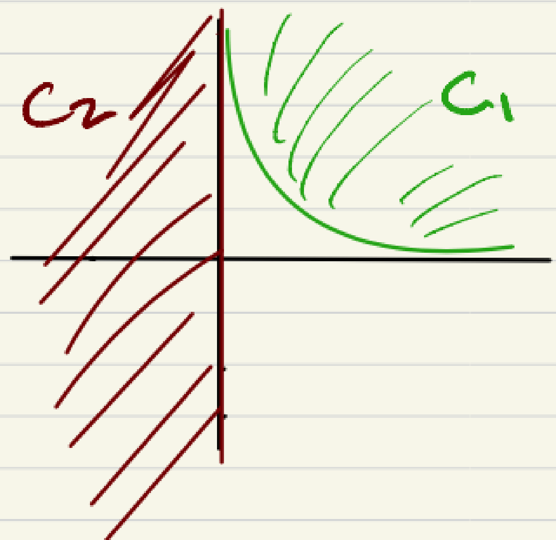
$$\inf_{c_1 \in C_1} \langle b, c_1 \rangle < \sup_{c_2 \in C_2} \langle b, c_2 \rangle$$

Strongly Separated



$\Downarrow$

Properly separated But Not strongly separated

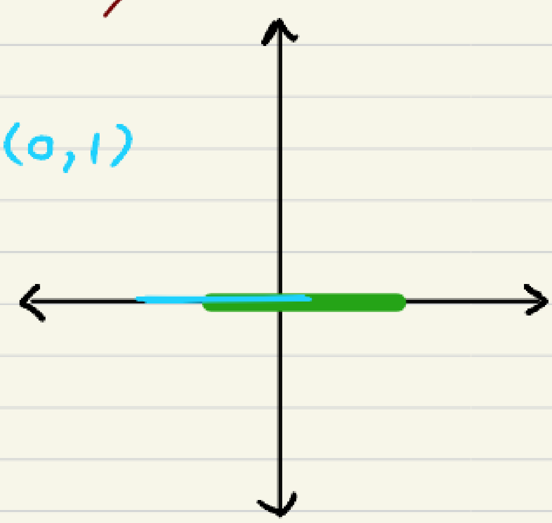


$\Downarrow$

Separated by $(0,1)$

Not properly separated

Not strongly separated.



Fact (L7-2): [txtbook Thm 11.3]
Let C_1, C_2 be nonempty convex subsets of $\mathbb{R}^m$, then C_1 and C_2 are "properly" separated if and only if

$$ri(C_1) \cap ri(C_2) = \emptyset$$

Fact (L7-3): [txtbook Cor 6.6.2]
Let C_1, C_2 be convex subsets of $\mathbb{R}^m$, then

$$ri(C_1 + C_2) = ri(C_1) + ri(C_2)$$

Let $\lambda \in \mathbb{R}$, then $ri(\lambda C) = \lambda ri(C)$
Fact (L7-4): [txtbook top of page 49]
Let $C_1 \subseteq \mathbb{R}^m, C_2 \subseteq \mathbb{R}^p$ be convex, then

$$ri(C_1 \oplus C_2) = ri(C_1) \oplus ri(C_2)$$

and

$$C_1 \oplus C_2 \simeq C_1 \times C_2 = \{(c_1, c_2) | c_1 \in C_1, c_2 \in C_2\}$$

Theorem 3.32: L7-2

et C_1, C_2 be convex subsets of $\mathbb{R}^m$ such that $ri(C_1) \cap ri(C_2) \neq \emptyset$. Let $x \in C_1 \cap C_2$, then

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

Proof.

- " $\supseteq$ ", see A2
- " $\subseteq$ ": Let $x \in C_1 \cap C_2$ and let $n \in N_{C_1 \cap C_2}(x)$, then $\forall y \in C_1 \cap C_2$, we have

$$\langle n, y - x \rangle \leqslant 0$$

$$E_1 = epi(\delta_{C_1}) = C_1 \times [0, \infty) \subseteq \mathbb{R}^m \times \mathbb{R}$$
$$E_2 = \{(y, \alpha) | y \in C_2, \alpha \leqslant \langle n, y - x \rangle\} \subseteq \mathbb{R}^m \times \mathbb{R}$$

Using Fact(L7-4), applied with C_2 replaced by $[0, \infty) \subseteq \mathbb{R}$, we learn that

$$ri(E_1) = ri(C_1) \times (0, \infty)$$

One can also show that

$$ri(E_2) = \{(y, \alpha) | y \in ri(C_2), \alpha < \langle n, y - x \rangle\}$$

We claim that

$$ri(E_1) \cap ri(E_2) = \emptyset \ldots (*)$$

Indeed, suppose for eventual contradiction that

$$\exists (z, \alpha) \in ri(E_1) \cap ri(E_2)$$

The $0 < \alpha < \langle n, z - x\rangle \leqslant 0$, which is absend. Hence $(*)$ holds. Applying Fact L7-2 with C_i's replaced by E_i's yield.
$\exists (b, \gamma) \in \mathbb{R}^m \times \mathbb{R} \setminus \{0\}$ such that

$$\begin{aligned} \forall (x, \alpha) \in E_1 & \qquad \forall (y, \beta) \in E_2 \\ \langle (x, \alpha), (b, \gamma)\rangle \leqslant & \qquad \langle (y, \beta), (b, \gamma)\rangle \\ \langle x, b\rangle + \alpha\gamma \leqslant & \qquad \langle y, b\rangle + \beta\gamma \ldots (1) \end{aligned}$$

Moreover, $\exists (\overline{x}, \overline{\alpha}) \in E_1, \exists (\overline{y}, \overline{\beta}) \in E_2$ such that

$$\langle \overline{x}, b\rangle + \overline{\alpha}\gamma < \langle \overline{y}, b\rangle + \overline{\beta}\gamma \ldots (2)$$

We claim that $\gamma < 0$. Indeed, observe that:

$$(x, 1) \in E_1,\ (x, 0) \in E_2 \ldots (3)$$

Combining with we obtain

$$\begin{gathered} \langle x, b\rangle + \gamma \leqslant \langle x, b\rangle \\ \implies \gamma \leqslant 0 \end{gathered}$$

Next, we show that $\gamma \neq 0$
Suppose on the contrary that $\gamma = 0$. Observe that this implies that (1) and (2), $\exists b \neq 0$

$$\begin{aligned} \forall (x, \alpha) \in C_1 & \qquad \forall (y, \beta) \in C_2 \\ \langle x, b\rangle \leqslant & \qquad \langle y, b\rangle \\ \exists \overline{x} \in C_1 & \qquad \exists \overline{y} \in C_2 \\ \langle \overline{x}, b\rangle < & \qquad \langle \overline{y}, b\rangle \end{aligned}$$

That is, C_1, C_2 are properly separated. By the earlier Fact(L7-2), we learn that $ri(C_1) \cap ri(C_2) = \emptyset$, which is a contradiction.
Altogether,

$$\gamma < 0$$

We will show that,

$$N_{C_1 \cap C_2} \ni n = \underbrace{-\frac{b}{\gamma}}_{\in N_{C_1}(x)} + \underbrace{n + \frac{b}{\gamma}}_{\in N_{C_2}(x)}$$

Recall

$$\langle x, b\rangle + \alpha\gamma \leqslant \langle y, b\rangle + \beta\gamma \ldots (1)$$

Next, we claim that $\forall y \in C_1$,

$$\langle b, y\rangle \leqslant \langle b, x\rangle \ldots (4)$$

Indeed, observe that $\forall y \in C_1, (y, 0) \in E_1$, and by (3) $(x, 0) \in E_2$. Therefore, (1) yields (4). This implies that $b \in N_{C_1}(x)$. Hence,

$$-\frac{b}{\gamma} = -\frac{1}{\gamma} b \in N_{C_1}(x)$$

Finally, using (3) $(x,0) \in E_1$, and

$$\forall y \in C_2, (y, \langle n, y-x \rangle) \in E_2$$

Therefore, (1) yields

$$\forall y \in C_2, \langle b, x \rangle \leqslant \langle b, y \rangle + \gamma \langle n, y-x \rangle$$

Eauivalently,

$$\forall y \in C_2, \left\langle \frac{b}{\gamma} + n, y-x \right\rangle \leqslant 0$$

Therefore,

$$\frac{b}{\gamma} + n \in N_{C_2}(x)$$

Altogether, we conclude that

$$n = -\frac{b}{\gamma} + \frac{b}{\gamma} + n \in N_{C_1}(x) + N_{C_2}(x)$$

□

Proposition 3.33: L7-3

Let $f : \mathbb{R}^m \mapsto (-\infty, \infty]$ be convex l.s.c and proper. Let $x \in \mathbb{R}^m$ and let $u \in \mathbb{R}^m$ then the following are equivalent:

$$u \in \partial f(x) \iff (u, -1) \in N_{epi(f)}(x, f(x))$$

Proof. Observe tha $epi(f) \neq \emptyset$ and convex (because f is proper+convex). Now let $u \in \mathbb{R}^m$, then

$$\begin{aligned} & (u, -1) \in N_{epi(f)}(x, f(x)) \\ \iff & [x \in dom(f), \text{ and } \forall (y, \beta) \in epi(f), \langle (y, \beta) - (x, f(x)), (u, -1) \rangle \leqslant 0] \\ \iff & [x \in dom(f), \text{ and } \forall (y, \beta) \in epi(f), \langle (y - x, \beta - f(x)), (u, -1) \rangle \leqslant 0] \\ \iff & \forall (y, \beta) \in epi(f), \langle y - x, u \rangle + f(x) \leqslant \beta \\ \underset{(?)}{\iff} & \forall y \in dom(f), \langle y - x, u \rangle + f(x) \leqslant f(y) \end{aligned}$$

For $(?)$, clearly $\implies$ holds, so $(y, f(y)) \in epi(f)$, and $\impliedby$ hold because $(y, \beta) \in epi(f) \iff f(y) \leqslant \beta$, so

$$u \in \partial f(x)$$

□

Theorem 3.34: L7-4(txtbook THM 23.9)

Let $f : \mathbb{R}^m \mapsto (-\infty, \infty]$, $g : \mathbb{R}^m \mapsto (-\infty, \infty]$ be convex l.s.c. and proper. Suppose that $ri(dom(f)) \cap ri(dom(g)) \neq \emptyset$. Then $\forall x \in \mathbb{R}^m$, we have

$$\partial f(x) + \partial g(x) = \partial (f + g)(x)$$

Proof. Let $x \in \mathbb{R}^m$. If $x \notin dom(f) \cap dom(g) \supseteq dom(\partial f) \cap dom(\partial g)$,

$$\implies \partial f(x) + \partial g(x) = \emptyset$$

Also, $\partial (f + g)(x) = \emptyset$.
Now, let $x \in dom(f) \cap dom(g) = dom(f + g)$, one can easily verify that

$$\partial f(x) + \partial g(x) \subseteq \partial (f + g)(x) \ldots (A2)$$

We now verify the opposite inclusion.
Suppose that $u \in \partial (f + g)(x)$,

$$\forall y \in \mathbb{R}^m, \ (f + g)(y) \geqslant (f + g)(x) + \langle u, y - x \rangle \ldots (1)$$

Consider the closed convex sets:

$$\begin{aligned} \emptyset \neq E_1 &= \{(x, \alpha, \beta) \in \mathbb{R}^m \times \mathbb{R} \times \mathbb{R} | f(x) \leqslant \alpha\} = epi(f) \times \mathbb{R} \\ \emptyset \neq E_2 &= \{(x, \alpha, \beta) \in \mathbb{R}^m \times \mathbb{R} \times \mathbb{R} | g(x) \leqslant \beta\} \end{aligned}$$

We claim that

$$(u, -1, -1) \in N_{E_1 \cap E_2}(x, f(x), g(x)) \ldots (2)$$

Indeed, let $(y, \alpha, \beta) \in E_1 \cap E_2$, then $f(y) \leqslant \alpha, g(y) \leqslant \beta$

$$\implies f(y) - \alpha \leqslant 0,\ g(y) - \beta \leqslant 0$$

Now,

$$\begin{aligned}
&\langle (u, -1, -1), (y, \alpha, \beta) - (x, f(x), g(x)) \rangle \\
&= \langle u, y - x \rangle - (\alpha - f(x)) - (\beta - g(x)) \\
&= \langle u, y - x \rangle + f(x) + g(x) - \alpha - \beta \\
&= \langle u, y - x \rangle + (f + g)(x) - (\alpha + \beta) \\
&\leqslant (f + g)(y) - \alpha - \beta \\
&= f(y) - \alpha + g(y) - \beta \\
&\leqslant 0
\end{aligned}$$

This proves (2).
Next we claim that:

$$ri(E_1) \cap ri(E_2) \neq \emptyset$$

Using that Fact (L7-4), we know that

$$\begin{aligned}
ri(E_1) &= ri(epi(f) \times \mathbb{R}) \\
&= ri(epi(f)) \times ri(\mathbb{R}) \\
&= ri(epi(f)) \times \mathbb{R}
\end{aligned}$$

Moreover, we can show that

$$ri(E_2) = \{(x, \alpha, \beta) \in \mathbb{R}^m \times \mathbb{R} \times \mathbb{R} | g(x) < \beta\}$$

Now, let $z \in ri(dom(f)) \cap ri(dom(g))$, then

$$(z, f(z) + 1, g(z) + 1) \in ri(E_1) \cap ri(E_2) \neq \emptyset$$

Therefore, E_1, E_2 are nonempty closed convex,$ri(E_1) \cap ri(E_2) \neq \emptyset$. Hence by Theorem L7-2, we have

$$N_{E_1 \cap E_2}(x, f(x), g(x)) = N_{E_1}(x, f(x), g(x)) + N_{E_2}(x, f(x), g(x))$$

Therefore,

$$(u, -1, -1) = \underbrace{(u_1, -\alpha, 0)}_{\in N_{E_1}(x, f(x), g(x))} + \underbrace{(u_2, 0, -\beta)}_{\in N_{E_2}(x, f(x), g(x))}$$

Observe that $E_1 = epi(f) \times \mathbb{R}$. Hence

$$N_{E_1}(x, f(x), g(x)) = N_{epi(f)}(x, f(x)) \times N_{\mathbb{R}}(g(x)) = N_{epi(f)}(x, f(x)) \times \{0\}$$

This yield: $u = u_1 + u_2,\ \alpha = \beta = 1$, hence,

$$(u_1, -1) \in N_{epi(f)}(x, f(x))$$
$$(u_2, -1) \in N_{epi(f)}(x, g(x))$$

Recalling Proposition L7-3, we conclude that

$$u_1 \in \partial f(x),\ u_2 \in \partial g(x)$$

Hence,

$$u = u_1 + u_2 \in \partial f(x) + \partial g(x)$$

The proof is complete. □

Example 3.35: L7-5

Let $f : \mathbb{R}^m \mapsto (-\infty, \infty]$ be convex l.s.c. and proper and let $\emptyset \neq C \subseteq \mathbb{R}^m$ be convex and closed. Suppose that

$$ri(C) \cap ri(dom(f)) \neq \emptyset$$

Consider the problem:

$$(P)\ \min f(x),\ s.t.\ x \in C$$

Let $\overline{x} \in \mathbb{R}^m$, then $\overline{x}$ solved (P) if and only if $(\partial f(\overline{x})) \cap (-N_C(\overline{x})) \neq \emptyset$

Proof. Write (P) as

$$\min_{x \in \mathbb{R}^m} f(x) + \delta_C(x)$$

Observe that $f + \delta_C$ is convex l.s.c. and proper.
By Fermat's Theorem

$$\overline{x} \text{ solves } p \iff 0 \in \partial(f + \delta_C)(\overline{x})$$

Now,

$$\begin{aligned} & ri(dom(f)) \cap ri(dom(\delta_C)) \\ = & ri(dom(f)) \cap ri(C) \\ \neq & \emptyset \end{aligned}$$

Therefore, by Theorem L7-4, we conclude that

$$\begin{aligned} \overline{x} \text{ solves } p & \iff 0 \in \partial(f + \delta_C)(\overline{x}) = \partial f(\overline{x}) + \partial \delta_C(\overline{x}) = \partial f(\overline{x}) + N_C(\overline{x}) \\ & \iff \exists u \in \partial f(\overline{x}),\ -u \in N_C(\overline{x}) \\ & \iff \partial f(\overline{x}) \cap (-N_C(\overline{x})) \neq \emptyset \end{aligned}$$

□

Example 3.36: L7-6

Let $d \in \mathbb{R}^m$, and let $\emptyset \neq C \subseteq \mathbb{R}^m$ be convex and closed. Consider the problem

$$(P)\ \min \langle d, x \rangle\,,\ s.t.\ x \in C$$

Let $\overline{x} \in \mathbb{R}^m$. Then $\overline{x}$ solved

$$p \iff -d \in N_C(\overline{x})$$

3.7 Differentiability of Convex Functions

Definition 3.37: L8-1

et $f : \mathbb{R}^m \mapsto (-\infty, \infty]$ *be proper, and let* $x \in dom(f)$. *The directional derivative of* f *at* x *in the direction of* d *is*

$$f'(x; d) := \lim_{y \downarrow 0} \frac{f(x + td) + f(x)}{t}$$

f *is differentiable at* x *if there exists an operator* $\nabla f(x) : \mathbb{R}^m \mapsto \mathbb{R}^m$, *called the derivative (or gradient) of* f *at* x *that satisfies*

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

Remark. If f is differentiable at x, then the directional derivative of f at x in the direction of d is

$$f'(x; d) = \langle \nabla f(x), d \rangle$$

Theorem 3.38: txtbook Thm 23.2

Let $f : \mathbb{R}^m \mapsto (-\infty, \infty]$ *be convex and proper and let* $x \in dom(f)$. *Let* $u \in \mathbb{R}^m$. *Then* u *is a subgradient of* f *at* x *of and only if*

$$\forall y \in \mathbb{R}^m, f'(x; y) \geqslant \langle u, y \rangle$$

Proof. Using the subgradient inequality we have

$$\begin{aligned} u \in \partial f(x) &\iff \forall y \in \mathbb{R}^m, \lambda > 0,\ f(x + \lambda y) \geqslant f(x) + \langle u, x + \lambda y - x \rangle \\ &\iff \forall y \in \mathbb{R}^m, \lambda > 0,\ \frac{f(x + \lambda y) - f(x)}{\lambda} \geqslant \langle u, y \rangle \end{aligned}$$

Taking the limit as $\lambda \downarrow 0$ in view of Theorem 23.1 in the textbook yields the desired result. □

Theorem 3.39: Txtbook 25.2

Let $f : \mathbb{R}^m \mapsto (-\infty, \infty]$ *be convex and proper and let* $x \in dom(f)$. *If* f *is differentiable at* x, *then* $\nabla f(x)$ *is the unique subgradient of* f *at* x.

Proof. Recall that $\forall y \in \mathbb{R}^m$,

$$f'(x; y) = \langle \nabla f(x), y \rangle$$

Let $u \in \mathbb{R}^m$, using the previous theorem we have

$$u \in \partial f(x) \iff \forall y \in \mathbb{R}^m,\ f'(x; y) \geqslant \langle u, y \rangle$$

Altogether,

$$u \in \partial f(x) \iff \forall y \in \mathbb{R}^m \langle \nabla f(x), y \rangle \geqslant \langle u, y \rangle$$

Clearly, we have $\{\nabla f(x)\} \subseteq \partial f(x)$. Moreover, letting $y = u - \nabla f(x)$ yields

$$\begin{aligned} &\|u - \nabla f(x)\|^2 = 0 \\ \implies &u = \nabla f(x) \\ \implies &\partial f(x) \subseteq \{\nabla f(x)\} \end{aligned}$$

Hence,

$$\partial f(x) = \{\nabla f(x)\}$$

□

Lemma 3.40: L8-4

Let $\varphi : \mathbb{R} \mapsto (-\infty, \infty]$ be a proper function that is differentiable on a nonempty open interval $I \subseteq dom(\varphi)$, then:

$$\varphi' \text{ is increasing on } I \implies \varphi \text{ is convex on } I$$

Proof. Fix $x, y \in I$, and $\lambda \in (0, 1)$. Set

$$\begin{aligned} \psi &: \mathbb{R} \mapsto (-\infty, \infty] \\ &: z \mapsto \lambda\varphi(x) + (1-\lambda)\varphi(z) - \varphi(\lambda x + (1-\lambda)z) \end{aligned}$$

Then

$$\psi'(z) = (1-\lambda)\varphi'(z) - (1-\lambda)\varphi'(\lambda x + (1-\lambda)x)...(*)$$

and $\psi'(x) = 0\psi(x)$.
And $(*)$ implies that

$$\begin{aligned} \psi'(z) &\leqslant \text{ whenever } z < x \\ \psi"(z) &> 0 \text{ whenever } z \geqslant x \end{aligned}$$

Therefore, ψ achieves its infimum on I at x.
That is $\forall y \in I, \psi(y) \geqslant \psi(x) = 0$.
That is $\forall y \in I$,

$$\lambda\varphi(x) + (1-\lambda)\varphi(y) \geqslant \varphi(\lambda x + (1-\lambda)y)$$

□

Proposition 3.41: L8-5

et $f : \mathbb{R}^m \mapsto (-\infty, \infty]$ be proper. Suppose that $dom(f)$ is open and convex, and that f is differentiable on $dom(f)$. Then the following are equivalent:

1. *f is convex*
2. *$\forall x, y \in dom(f),\ \langle x - y, \nabla f(y)\rangle + f(y) \leqslant f(x)$*
3. *$\forall x, y \in dom(f),\ \langle x - y, \nabla f(x) - \nabla f(y)\rangle \geqslant 0$*

Proof.

- 1) $\implies$ 2): Combine the subgradient inequality with the previous result
- 2) $\implies$ 3): See A2 for a proof in a more general setting
- 3) $\implies$ 1): Fix $x \in dom(f),\ y \in domf(f),\ z \in \mathbb{R}^m$. By assumption, $dom(f)$ is open. Therefore, $\exists \varepsilon > 0$ such that

$$\begin{aligned} y + (1+\varepsilon)(x-y) = x + \varepsilon(x-y) &\in dom(f) \\ y - \varepsilon(x-y) = y + \varepsilon(y-x) &\in dom(f) \end{aligned}$$

Hence, by convexity of $dom(f)$ we have

$$\forall \alpha \in (-\varepsilon, 1+\varepsilon),\ x + \alpha(x-y) \in dom(f)$$

Set $C = (-\varepsilon, 1+\varepsilon) \subseteq \mathbb{R}$ and set $\varphi : \mathbb{R} \mapsto (-\infty, \infty]$, where

$$\varphi(\alpha) = f(y + \alpha(x-y)) + \delta_C(x)$$

Then φ is differentiable on C, and $\forall \alpha \in C$,

$$\varphi'(\alpha) = \langle \nabla f(y + \alpha(x-y)), x-y \rangle$$

Now, take $\alpha \in C, \beta \in C,\ \alpha < \beta$. Set

$$\left.\begin{aligned} y_\alpha &= y + \alpha(x-y) \\ y_\beta &= y + \beta(x-y) \end{aligned}\right\} \implies y_\beta - y_\alpha = (\beta - \alpha)(x-y)$$

Then,

$$\begin{aligned} \varphi'(\beta) - \varphi'(\alpha) &= \langle \nabla f(y + \beta(x-y)), x-y \rangle - \langle \nabla f(y + \alpha(x-y)), x-y \rangle \\ &= \langle \nabla f(y_\beta) - \nabla f(y_\alpha), x-y \rangle \\ &= \left\langle \nabla f(y_\beta) - \nabla f(y_\alpha), \frac{y_\beta - y_\alpha}{\beta - \alpha} \right\rangle \\ &= \frac{1}{\beta - \alpha} \langle \nabla f(y_\beta) - \nabla f(y_\alpha), y_\beta - y_\alpha \rangle \\ &\geqslant 0 \end{aligned}$$

That is φ' is increasing on C. By lemma L8-4, we know φ is convex on C. Recalling

$$\varphi(\alpha) = f(y + \alpha(x-y)) + \delta_C(\alpha)$$

We learn that

$$\begin{aligned} f(\alpha x + (1-\alpha)y) &= \varphi(\alpha) \\ &\leqslant \alpha\varphi(1) + (1-\alpha)\varphi(0) \\ &= \alpha f(x) + (1-\alpha) f(y) \end{aligned}$$

$\square$

Example 3.42: L8-5

Let A be $m \times m$ matrix, and set $f : \mathbb{R}^m \mapsto \mathbb{R}$, $f(x) = \langle x, Ax \rangle$. Then the followings hold:

1. $\nabla f(x) = (A + A^T)(x), \forall x \in \mathbb{R}^m$
2. *f is convex if and only if $A + A^T$ is positive semidefinite.*

Proof. 1. See A3

2. Recall Prop L8-5. Therefore f is convex if and only if

$$\begin{aligned} & \forall x, y \in \mathbb{R}^m \ \langle \nabla f(x) - \nabla f(y), x - y \rangle \geqslant 0 \\ \iff & \forall x, y \in \mathbb{R}^m \left\langle (A + A^T)x - (A + A^T)y, x - y \right\rangle \geqslant 0 \\ \iff & \forall z \in |R^m \left\langle (A + A^T)z, z \right\rangle \geqslant 0 \end{aligned}$$

$\square$

3.8 Subdifferentiability and Conjugacy

Recall that, for a function $f : \mathbb{R}^m \mapsto [-\infty, \infty]$, the Fenchel Conjugate of f is

$$f^* : \mathbb{R}^m \mapsto [-\infty, \infty]$$
$$f^*(u) = \sup_{x \in \mathbb{R}^m} (\langle x, u \rangle - f(x))$$

Proposition 3.43: L8-6

Let f, g be functions from $\mathbb{R}^m$ to $[-\infty, \infty]$. Then

1. $f^{**} := (f^*)^* \leqslant f$
2. $f \leqslant g \implies [f^* \geqslant g^*$ *and* $f^{**} \leqslant g^{**}]$

Proof. See A3 □

Proposition 3.44: L8-7

et $f : \mathbb{R}^m \mapsto (-\infty, \infty]$ be proper. Then $\forall x, y \in \mathbb{R}^m$

$$f(x) + f^*(u) \geqslant \langle x, u \rangle$$

Fenchel-Young inequality

Proof. Observe that the definition of f^* yields:

$$f \equiv \infty \iff f^* \equiv -\infty$$

Therefore, by assumption we know that

$$\forall u \in \mathbb{R}^m, f^*(u) \neq -\infty$$

Now let $(x, u) \in \mathbb{R}^m \times \mathbb{R}^m$. If $f(x) = \infty$, the desired inequality clearly holds, else, if $f(x) < \infty$, we have

$$f^*(u) = \sup_{y \in \mathbb{R}^m} (\langle y, u \rangle, f(u)) \geqslant \langle y, x \rangle - f(x)$$

□

Proposition 3.45: L8-8

Let $f : \mathbb{R}^m \mapsto (-\infty, \infty]$ be convex l.s.c and proper. Let $x \in \mathbb{R}^m$ and let $u \in \mathbb{R}^m$. Then the following are equivalent:

$$u \in \partial f(x) \iff f(x) + f^*(u) = \langle x, u \rangle$$

Proof.

$$
\begin{aligned}
u \in \partial f(x) &\iff \forall y \in dom(f),\ \langle y - x, u\rangle + f(x) \leqslant f(y)\\
&\iff \forall y \in dom(f),\ \langle y, u\rangle - f(y) \leqslant \langle x, u\rangle - f(x) \leqslant f^*(u)\\
&\iff f^*(u) = \sup_{y\in\mathbb{R}^m} (\langle y, u\rangle - f(y)) \leqslant \langle x, u\langle - f(x) \leqslant f^*(u)\\
&\iff f(x) + f^*(u) = \langle x, u\rangle
\end{aligned}
$$

□

Proposition 3.46: L8-9

Let $f : \mathbb{R}^m \mapsto (-\infty, \infty]$ be convex and proper, let $x \in \mathbb{R}^m$ and suppose that $\partial f(x) \neq \emptyset$. Then

$$(f^*)^* := f^{**}(x) = f(x)$$

where

$$f^{**}(x) = \sup_{y\in\mathbb{R}^m} \{\langle y, x\rangle - f^*(y)\}$$

Proof. Let $u \in \partial f(x)$. By Prop L8-8

$$
\begin{aligned}
&\langle u, x\rangle = f(x) + f^*(u)\\
\implies &f(x) = \langle u, x\rangle - f^*(u)
\end{aligned}
$$

Consequently,

$$
\begin{aligned}
f^{**}(x) &= \sup_{y\in\mathbb{R}^m} \{\langle x, y\rangle - f^*(y)\}\\
&\geqslant \langle x, u\rangle - f^*(u)\\
&= f(x)
\end{aligned}
$$

Conversely,

$$
\begin{aligned}
f^{**}(x) &= \sup_{y\in\mathbb{R}^m} \{\langle y, x\rangle - f^*(y)\}\\
&= \sup_{y\in\mathbb{R}^m} \{\langle y, x\rangle - \sup_{z\in\mathbb{R}^m} \{\langle z, y\rangle - f(z)\}\}\\
&= \sup_{y\in\mathbb{R}^m} \{\langle y, x\rangle + \inf_{z\in\mathbb{R}^m} \{f(z) - \langle z, y\rangle\}\}\\
&= \sup_{y\in\mathbb{R}^m} \{\inf_{z\in\mathbb{R}^m} \{f(z) + \langle y, x - z\rangle\}\}\\
&\leqslant \sup_{y\in\mathbb{R}^m} \{f(x) = \langle y, x - x\rangle\}\\
&= \sup_{y\in\mathbb{R}^m} f(x)\\
&= f(x)
\end{aligned}
$$

Altogether,

$$f(x) = f^{**}(x)$$

□

Fact(L8-10) Let $f : \mathbb{R}^m \mapsto (-\infty, \infty]$ be proper. Then

$$[f \text{ is convex and l.s.c}] \iff f = f^{**}$$

In this case, f^* is also proper.

Corollary 3.47

Let $f : \mathbb{R}^m \mapsto (-\infty, \infty]$ be convex l.s.c and proper. Then

1. *f^* is convex l.s.c and proper*
2. *$f^{**} = f$*

Proof.

- Combine Fact L8-10 and Prop L6-6
- Follows from Fact L8-10

□

Proposition 3.48: L8-12

Let $f : \mathbb{R}^m \mapsto (-\infty, \infty]$ be convex l.s.c and proper. Then

$$u \in \partial f(x) \iff x \in \partial f^*(u)$$

Proof. Recall that

$$u \in \partial f(x) \iff f(x) + f^*(u) = \langle x, u \rangle$$

by Proposition L8-8.
Set $g := f^*$. Then Corollary L8-11 imply that g is convex l.s.c and proper. Moreover, $g^* = f$.
Hence,

$$\begin{aligned} u \in \partial f(x) &\iff f(x) + f^*(u) = \langle x, u \rangle \\ &\iff g^*(x) + g(u) = \langle x, u \rangle \\ &\iff x \in \partial g(u) = \partial f^*(u) \end{aligned}$$

□

Theorem 3.49: L9-1

Let $f : \mathbb{R}^m \to \mathbb{R}$ be proper, l.s.c and let C be a compact subset of $\mathbb{R}^m$ such that $C \cap dom(f) \neq \emptyset$. Then the following holds:

1. *f is bounded below over C*
2. *f attains its minimal value over C*

Proof.

1. Suppose for eventual contradiction that f is not bounded below over C. Then there exists a sequence $(x_n)_{n\in\mathbb{N}}$ in C such that $\lim_{n\to\infty} f(x_n) = -\infty$. Recall that C is compact, equivalently, C is closed and bounded (finite-dim). Since $(x_n)_{n\in\mathbb{N}}$ is a sequence in C, $(x_n)_{n\in\mathbb{N}}$ must be bounded. By Bolzano-Weierstrass theorem, there exists a convergent subsequence say $x_{k_n} \to \overline{x} \in C$ because C is closed.
   Since f is l.s.c, we learn that,
   $$f(\overline{x}) \leqslant \liminf_{n\to\infty} f(x_{k_n})$$
   but, $f(\overline{x}) \in \mathbb{R}$ by definition, contradiction.

2. Let $f_{\min}$ be the minimal value of f over C. Then there exists a sequence $(x_n)_{n\in\mathbb{N}}$ in C such that
   $$f(x_n) \to f_{\min}$$
   AND C is bounded $\implies$ $(x_n)_{n\in\mathbb{N}}$ is bounded.
   Let $\overline{x}$ be a cluster point of $(x_n)_{n\in\mathbb{N}}$, say $x_{k_n} \to \overline{x} \in C$. Then by l.s.c.
   $$f(\overline{x}) \leqslant \liminf_{n\to\infty} f(x_{k_n}) = f_{\min}$$
   Hence, $\overline{x}$ is a minimizer of f over C.

$\square$

Definition 3.50: L9-2

Let $f : \mathbb{R}^m \to (-\infty, \infty]$. Then f is coercive if

$$\lim_{\|x\|\to\infty} f(x) = \infty$$

and f is super coercive if

$$\lim_{\|x\|\to\infty} \frac{f(x)}{\|x\|} = \infty$$

Theorem 3.51: L9-3

Let $f : \mathbb{R}^m \to (-\infty, \infty]$ be proper, l.s.c. and coercive and let C be a closed subset of $\mathbb{R}^m$ satisfying that $C \cap dom(f) \neq \emptyset$. Then f attains its minimal value over C.

Proof. Let $x \in C \cap dom(f)$. Since f is coercive, $\exists M > 0$ such that

$$f(y) > f(x) \text{ whenever} \|y\| > M \ldots (1)$$

observe that if $\overline{x}$ is a minimizer of f over C, we have $f(\overline{x}) \leqslant f(x)$. In view of (1) above, we learn that the set of minimizers of f over C is the same as the set of minimizers of f over $C \cap B(0; M)$. The latter is closed and bounded. Hence, it is compact, then apply the previous result with the set C replaced by $C \cap B(0; M)$ we conclude that f attains its minimal value over $C \cap B(0; M)$ say at $\tilde{x}$. Altogether, $\tilde{x}$ is a minimizer of f over C. $\square$

3.9 Differentiability and Strong Convexity:

Definition 3.52: L9-4

Let $T : \mathbb{R}^m \to \mathbb{R}^m$, and let $L \geqslant 0$. Then T is L-Lipschitz if $\forall x \in \mathbb{R}^m, \forall y \in \mathbb{R}^m$,

$$\|Tx - Ty\| \leqslant L\|x - y\|$$

Example 3.53: L9-5

Let $f : \mathbb{R}^m \to \mathbb{R} : x \to \frac{1}{2}\langle x, Ax\rangle + \langle b, x\rangle + c$, where $A \succcurlyeq 0$ (A is positive semi-definite), $b \in \mathbb{R}^m$, $C \in \mathbb{R}$. Then the following hold:

1. $\forall x \in \mathbb{R}^m,\ \nabla f(x) = Ax + b$
2. ∇f *is Lipschitz with a constant* $L = \|A\|$, *where* $\|A\| = \max_{\|x\|\neq 0} \frac{\|Ax\|}{\|x\|}$

Proof.

1. It follows from lecture 8 that $\forall x \in \mathbb{R}^m$,

$$\nabla f(x) = \frac{1}{2}(A + A^T)x + b = \frac{1}{2}(A + A)x + b = Ax + b$$

2. Indeed,

$$\begin{aligned}\|\nabla f(x) - \nabla f(y)\| &= \|Ax - Ay\| \\ &= \|A(x - y)\| \\ &\leqslant \|A\|\|x - y\|\end{aligned}$$

and the conclusion follows.

□

Example 3.54: L9-6

Let C be a nonempty closed convex subset of $\mathbb{R}^m$. Then P_C is Lipschitz Continuous with a constant 1.

Proof. If C is a singleton, the conclusion is trivial. Now, suppose that C is not a singleton. Let $\{x, y\} \subseteq \mathbb{R}^m,\ x \neq y$. If $P_C(x) = P_C(y)$,

$$0 = \|P_C(x) - P_C(y)\| \leqslant \|x - y\|$$

Else, if $P_C(x) \neq P_C(y)$, then,

$$\begin{aligned}
\|P_C(x) - P_C(y)\|^2 &= \langle P_C(x) - P_C(y), P_C(x) - P_C(y)\rangle \\
&= \langle P_C(x) - P_C(y), P_C(x) - x\rangle + \langle P_C(x) - P_C(y), y - P_C(y)\rangle \\
&\quad + \langle P_C(x) - P_C(y), x - y\rangle \\
&= \underbrace{\langle P_C(x) - P_C(y), P_C(x) - x\rangle}_{\leqslant 0} + \underbrace{\langle P_C(y) - P_C(x), P_C(y) - y\rangle}_{\leqslant 0} \\
&\quad + \langle P_C(x) - P_C(y), x - y\rangle \text{ by projection theorem} \\
&\leqslant \langle P_C(x) - P_C(y), x - y\rangle \\
&\leqslant \|P_C(x) - P_C(y)\| \|x - y\|
\end{aligned}$$

so by $\|P_C(x) - P_C(y)\| \neq 0$, we have

$$\|P_C(x) - P_C(y)\| \leqslant \|x - y\|$$

□

Lemma 3.55: (descent lemma) L9-7

Let $f : \mathbb{R}^m \to (-\infty, \infty]$ be differentiable on $\emptyset \neq D \subseteq int(dom(f))$ such that ∇f is L-Lipschitz over D, D is convex. Then $\forall x, y \in D$,

$$f(y) \leqslant f(x) + \langle \nabla f(x), y - x\rangle + \frac{L}{2}\|x - y\|^2$$

Proof. Recall that the fundamental theorem of calculus implies that

$$\begin{aligned}
f(y) - f(x) &= \int_0^1 \langle \nabla f(x + t(y - x)), y - x\rangle \, dt \\
&= \langle \nabla f(x), y - x\rangle + \int_0^1 \langle \nabla f(x + t(y - x)) - \nabla f(x), y - x\rangle \, dt
\end{aligned}$$

Hence,

$$
\begin{aligned}
& |f(y) - f(x) - \langle \nabla f(x), y - x\rangle| \\
=& \left| \int_0^1 \langle \nabla f(x + t(y - x)) - \nabla f(x), y - x\rangle \, dt \right| \\
\leqslant& \int_0^1 |\langle \nabla f(x + t(y - x)) - \nabla f(x), y - x\rangle| \, dt \\
\leqslant& \int_0^1 \|\nabla f(x + t(y - x)) - \nabla f(x)\| \|y - x\| dt \\
\leqslant& \int_0^1 L\|x + t(y - x) - x\| \|y - x\| dt \text{ by } \nabla f \text{ is L-Lipschitz} \\
=& \int_0^1 tL\|y - x\|^2 dt \\
=& L\|x - y\|^2 \int_0^1 t dt \\
=& \frac{L}{2}\|x - y\|^2
\end{aligned}
$$

Hence,

$$f(y) \leqslant f(x) + \langle \nabla f(x), y - x\rangle + \frac{L}{2}\|x - y\|^2$$

□

Theorem 3.56: L9-8

Let $f : \mathbb{R}^m \to \mathbb{R}$ be convex and differentiable, and let $L > 0$. Then the followings are equivalent:

1. *∇f is L-Lipschitz*
2. *$\forall x, y \in \mathbb{R}^m$,*
$$f(y) \leqslant f(x) + \langle \nabla f(x), y - x\rangle + \frac{L}{2}\|x - y\|^2$$
3. *$\forall x, y \in \mathbb{R}^m$,*
$$f(y) \geqslant f(x) + \langle \nabla f(x), y - x\rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2$$
4. *$\forall x, y \in \mathbb{R}^m$,*
$$\langle \nabla f(x) - \nabla f(y), x - y\rangle \geqslant \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2$$

Proof.

- $1) \implies 2)$ This is the descent lemma applied with $D = \mathbb{R}^m$

- 2) $\implies$ 3) Without loss of generality, we can and do assume that $\nabla f(x) \neq \nabla f(y)$. Otherwise, the conclusion follows immediately using the subgradient inequality and the fact that $\partial f(X) = \{\nabla f(x)\}$
  Fix $x \in \mathbb{R}^m$ and set,

  $$h_x : \mathbb{R}^m \to \mathbb{R},\ h_x(y) = f(y) - f(x) - \langle \nabla f(x), y - x\rangle$$

  Observe that h_x is convex, differentiable and

  $$\nabla h_x(y) = \nabla f(y) - \nabla f(x)$$

  We claim that $\forall y, z \in \mathbb{R}^m$,

  $$h_x(z) \leqslant h_x(y) + \langle \nabla h_x(y), z - y\rangle + \frac{L}{2}\|z - y\|^2$$

  Indeed,

  $$\begin{aligned} h_x(z) =& f(z) - f(x) - \langle \nabla f(x), z - x\rangle \\ \leqslant& f(y) + \langle \nabla f(y), z - y\rangle + \frac{L}{2}\|z - y\|^2 - f(x) - \langle \nabla f(x), z - x\rangle \\ =& f(y) - f(x) - \langle \nabla f(x), y - x\rangle - \langle \nabla f(x), z - y\rangle \\ &+ \langle \nabla f(y), z - y\rangle + \frac{L}{2}\|z - y\|^2 \\ =& f(y) - f(x) - \langle \nabla f(x), y - x\rangle + \langle \nabla f(y) - \nabla f(x), z - y\rangle + \frac{L}{2}\|z - y\|^2 \\ =& h_x(y) + \langle \nabla h_x(y), z - y\rangle + \frac{L}{2}\|z - y\|^2 \ldots (1) \end{aligned}$$

  Observe that $\nabla h_x(x) = \nabla f(x) - \nabla f(x) = 0$. Hence, because h_x is convex, x is a global minimizer of h_x.
  That is, $\forall z \in \mathbb{R}^m$,

  $$h_x(x) \leqslant h_x(z) \ldots (2)$$

  Let $y \in \mathbb{R}^m$ and let $v \in \mathbb{R}^m$ be such that $\|v\| = 1$ and $\langle \nabla h_x(y), v\rangle = \|\nabla h_x(y)\|$. Set $z = y - \frac{\|\nabla h_x(y)\|}{L} v \ldots (3)$.
  On the one hand applying (2) with z as defined in (3) yields

  $$0 = h_x(x) \leqslant h_x\left(y - \frac{\|\nabla h_x(y)\|}{L} v\right)$$

On the other hand, (1) implies that

$$\begin{aligned}
0 &= h_x(x) \\
&\leqslant h_x(y) - \frac{\|\nabla h_x(y)\|}{L} \langle \nabla h_x(y), v \rangle + \frac{1}{2L} \|\nabla h_x(y)\|^2 \|v\|^2 \\
&= h_x(y) - \frac{\|\nabla h_x(y)\|^2}{L} + \frac{1}{2L} \|\nabla h_x(y)\|^2 \\
&= h_x(y) - \frac{1}{2L} \|\nabla h_x(y)\|^2 \\
&= f(y) - f(x) - \langle \nabla f(x), y - x \rangle - \frac{1}{2L} \|\nabla f(x) - \nabla f(y)\|^2
\end{aligned}$$

- $3) \implies 4)$: Using $3)$ we have

$$\begin{aligned}
f(y) &\geqslant f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L} \|\nabla f(x) - \nabla f(y)\|^2 \\
f(x) &\geqslant f(y) + \langle \nabla f(y), x - y \rangle + \frac{1}{2L} \|\nabla f(y) - \nabla f(x)\|^2
\end{aligned}$$

  Adding the above two inequalities yield $4)$.

- $4) \implies 1)$, Without loss of generality we can and do assume that $\nabla f(x) \neq \nabla f(y)$ (otherwise the conclusion is trivial). Now $4)$ implies

$$\|\nabla f(x) - \nabla f(y)\|^2 \leqslant L \langle \nabla f(x) - \nabla f(y), x - y \rangle \leqslant L \|\nabla f(x) - \nabla f(y)\| \|x - y\|$$

  Since $\nabla f(x) \neq \nabla f(y)$,

$$\|\nabla f(x) - \nabla f(y)\| \leqslant L \|x - y\|$$

$\square$

Example 3.57: L10-1

Let C be nonempty closed convex subset of $\mathbb{R}^m$. Then $\forall x, y \in \mathbb{R}^m$,

$$\|P_C(x) - P_C(y)\|^2 \leqslant \langle P_C(x) - P_C(y), x - y\rangle \dots (*)$$

Proof. Observe that $(*)$ can be rewritten as:

$$\langle P_C(x) - P_C(y), P_C(x) - P_C(y) - (x - y)\rangle \leqslant 0$$

Now:

$$\begin{aligned}
&\langle P_C(x) - P_C(y), P_C(x) - P_C(y) - (x - y)\rangle \\
=&\langle P_C(x) - P_C(y), P_C(x) - x\rangle - \langle P_C(x) - P_C(y), P_C(y) - y\rangle \\
=&\langle P_C(x) - P_C(y), P_C(x) - x\rangle + \langle P_C(y) - P_C(x), P_C(y) - y\rangle \\
\leqslant& 0 \text{ by projection theorem}
\end{aligned}$$

□

The above property is know as ”Firm nonexpansivenes” of the projection onto convex sets.

Example 3.58: L10-2

Let C be nonempty closed and convex subset of $\mathbb{R}^m$. Consider the function $f : \mathbb{R}^m \to \mathbb{R}$, where $f(x) = \frac{1}{2}d_C^2(x)$. Then the following holds:

1. *f is differentiable over $\mathbb{R}^m$ and $\forall x \in \mathbb{R}^m$,*

$$\nabla f(x) = x - P_C(x)$$

2. *∇f is 1-Lipschitz*

Proof.

1. Let $x \in \mathbb{R}^m$. Define $\forall y \in \mathbb{R}^m$,

$$h_x(y) = f(x + y) - f(x) - \langle y, x - P_C(x)\rangle$$

Clearly, h_x is convex.
By the definition of $\nabla f(x)$, it is sufficient to show that

$$\frac{|h_x(y)|}{\|y\|} \to 0 \text{ as } y \to 0$$

Observe that, $\forall x \in \mathbb{R}^m$,

$$f(x) = \frac{1}{2}d_C^2(x) = \frac{1}{2}\|x - P_C(x)\|^2$$

Now, on the one hand:

$$\begin{aligned} h_x(y) &= \frac{1}{2}\|(x+y) - P_C(x+y)\|^2 - \frac{1}{2}\|x - P_C(x)\|^2 - \langle y, x - P_C(x)\rangle \\ &\leqslant \frac{1}{2}\|(x+y) - P_C(x)\|^2 - \frac{1}{2}\|x - P_C(x)\|^2 - \langle y, x - P_C(x)\rangle \\ &= \frac{1}{2}\|x - P_C(x)\|^2 + \langle y, x - P_C(x)\rangle + \frac{1}{2}\|y\|^2 - \frac{1}{2}\|x - P_C(x)\|^2 - \langle y, x - P_C(x)\rangle \\ &= \frac{1}{2}\|y\|^2 \ldots (1) \end{aligned}$$

On the other hand, by the above argument $h_x(-y) \leqslant \frac{1}{2}\|y\|^2$. Therefore,

$$\begin{aligned} 0 = h_x(0) = h_x\left(\frac{1}{2}(y + (-y))\right) \leqslant \frac{1}{2}h_x(y) + \frac{1}{2}h_x(-y) \\ \implies h_x(y) \geqslant -h_x(-y) \geqslant -\frac{1}{2}\|y\|^2 \ldots (2) \end{aligned}$$

(1) and (2) imply $|h_x(y)| \leqslant \frac{1}{2}\|y\|^2$, and

$$\frac{|h_x(y)|}{\|y\|} = \frac{1}{2}\|y\| \to 0 \text{ as } y \to 0$$

2. To show that ∇f is 1-Lipschitz, let $x, y \in \mathbb{R}^m$. Now:

$$\begin{aligned} \|\nabla f(x) - \nabla f(y)\|^2 &= \|x - P_C(x) - (y - P_C(y))\|^2 \\ &= \|(x-y) - (P_C(x) - P_C(y))\|^2 \\ &= \|x-y\|^2 - 2\langle x - y, P_C(x) - P_C(y)\rangle + \|P_C(x) - P_C(y)\|^2 \\ &\leqslant \|x-y\|^2 - 2\|P_C(x) - P_C(y)\|^2 + \|P_C(x) - P_C(y)\|^2 \\ &= \|x-y\|^2 - \|P_C(x) - P_C(y)\|^2 \\ &\leqslant \|x-y\|^2 \end{aligned}$$

□

Theorem 3.59: Second Order Characterization, L10-3

Let $f : \mathbb{R}^m \to \mathbb{R}$ be twice continuously differentiable over $\mathbb{R}^m$, and let $L \geqslant 0$. Then the following are equivalent:

1. *∇f is L-Lipschitz*
2. *$\forall x \in \mathbb{R}^m$, $\|\nabla^2 f(x)\| \leqslant L$*

Proof.

- 1) $\implies$ 2). Suppose that ∇f is L-Lipschitz continuous. Observe that for any $y \in \mathbb{R}^m$, $\alpha > 0$, we have

$$\|\nabla f(x + \alpha y) - \nabla f(x)\| \leqslant L\|x + \alpha y - x\| = aL\|y\|$$

That is,

$$
\begin{aligned}
\|\nabla^2 f(x)(y)\| &= \lim_{\alpha \downarrow 0} \frac{\|\nabla f(x+\alpha y) - \nabla f(x)\|}{\alpha} \\
&\leqslant \lim_{\alpha \downarrow 0} \frac{L\|x+\alpha y - x\|}{\alpha} \\
&= \lim_{\alpha \downarrow 0} \frac{\alpha L\|y\|}{\alpha} \\
&= L\|y\|
\end{aligned}
$$

Equivalently, $\|\nabla^2 f(x)\| \leqslant L$ as desired.

- 2) $\implies$ 1): Suppose that for any $x \in \mathbb{R}^m$, $\|\nabla^2 f(x)\| \leqslant L$. Using the fundamental theorem of calculus we have $\forall x, y \in \mathbb{R}^m$,

$$
\begin{aligned}
\nabla f(x) &= \nabla f(y) + \int_0^1 \nabla^2 f(y+\alpha(x-y))(x-y)d\alpha \\
&= \nabla f(y) + \left[\int_0^1 \nabla^2 f(y+\alpha(x-y))d\alpha\right](x-y)
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\|\nabla f(x) - \nabla f(y)\| &= \left\|\left[\int_0^1 \nabla^2 f(y+\alpha(x-y))d\alpha\right](x-y)\right\| \\
&\leqslant \left\|\int_0^1 \nabla^2 f(y+\alpha(x-y))d\alpha\right\| \|(x-y)\| \\
&\leqslant \int_0^1 \left\|\nabla^2 f(y+\alpha(x-y))\right\| d\alpha \, \|(x-y)\| \\
&\leqslant L\|x-y\|
\end{aligned}
$$

$\square$

Fact(L10-4):
Ket A be an $m \times m$ symmetric matrix. Then $\|A\| = \sup_{\|x\|=1} \|Ax\| = \max_{1 \leqslant i \leqslant m} |\lambda_i|$, where $\lambda_1, \ldots, \lambda_m$ are the eigenvalues of A.

Proposition 3.60: L10-5

Let $f : \mathbb{R}^m \to \mathbb{R}$ be twice continuously differentiable. Then f is convex if and only if $\forall x \in \mathbb{R}^m, \nabla^2 f(x)$ is positive semi-definite.

Proof. See A3. ☐

Corollary 3.61: L10-6

Let $f : \mathbb{R}^m \to \mathbb{R}$ be convex and twice continuously differentiable and let $L \geqslant 0$. Then ∇f is L-Lipschitz $\iff \forall x \in \mathbb{R}^m,\ \lambda_{\max}(\nabla^2 f(x)) \leqslant L$.

Proof. Since f is convex and twice continuouly differentiable, we have $\forall x \in \mathbb{R}^m, \nabla^2 f(x)$ is positive semi-definite. Now combine with earlier result 1 to learn that

$$L \geqslant \|\nabla^2 f(x)\| = |\lambda_{\max}(\nabla^2 f(x))| = \lambda_{\max}(\nabla^2 f(x))$$

☐

Example 3.62: L10-7

Let $f : \mathbb{R}^m \to \mathbb{R}$ be given by $\forall x \in \mathbb{R}^m$,

$$f(x) = \sqrt{1 + \|x\|^2}$$

Prove that:

1. *f is convex*
2. *∇f is L-Lipschitz*

Proof. See A3 ☐

Strong Convexity:
Recall that a function $f : \mathbb{R}^m \to \mathbb{R}$ is strongly convex(2) with constant β, if for some $\beta > 0$ we have: $\forall x, y \in dom(f), \forall \lambda \in (0, 1)$,

$$f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda) f(y) - \frac{\beta}{2}\lambda(1-\lambda)\|x - y\|^2$$

Proposition 3.63: L10-8

Let $\beta > 0$, $f : \mathbb{R}^m \to (-\infty, \infty]$ is β-strongly convex $\iff f - \frac{\beta}{2}\|\cdot\|^2$ is convex.

Proof. See A3 ☐

Proposition 3.64: L10-9

Let $f : \mathbb{R}^m \to (-\infty, \infty],\ g : \mathbb{R}^m \to (-\infty, \infty]$ and let $\beta > 0$. Suppose that f is β-strongly convex and that g is convex. Then $f + g$ is β-strongly convex.

Proof. Set

$$h = f + g - \frac{\beta}{2}\|\cdot\|^2 = \left(\underbrace{f - \frac{\beta}{2}\|\cdot\|^2}_{\text{convex by prev. prop.}} \right) + g$$

Then h is convex being the sum of two convex functions (see A2). Therefore, applying the previous proposition again with f replaced by $f + g$ yields the desired result. □

Fact (L10-10):
Let $f : \mathbb{R}^m \to (-\infty, \infty]$ be strongly convex l.s.c. and proper. Then has a unique minimizer.

3.10 The Proximal Operator

Definition 3.65: L10-11

Let $f : \mathbb{R}^m \to (-\infty, \infty]$. The proximal point mapping of f is the operator

$$Prox_f : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$$

given by

$$Prox_f(x) = \arg\min_{u\in\mathbb{R}^m} \left\{ f(u) + \frac{1}{2}\|u - x\|^2 \right\}$$

Theorem 3.66: L10-12

Let $f : \mathbb{R}^m \to (-\infty, \infty]$ be convex l.s.c. and proper. Then $\forall x \in \mathbb{R}^m$, $Prox_f(x)$ is a singleton.

Proof. Observe that for a fixed $x \in \mathbb{R}^m$. $h_x := \frac{1}{2}\|\cdot - x\|^2$ is β-strongly convex for every $\beta < 1$. Set $g_x := f + h_x$, we learn that g_x is strongly convex for every $x \in \mathbb{R}^m$. Using A2, we know that $\forall x \in \mathbb{R}^m$, g_x is l.s.c (because f is l.s.c. and h_x is l.s.c). And, $\forall x \in \mathbb{R}^m$, g_x is proper (because f, h are proper and $dom(f) \cap dom(h_x) = dom(f) \cap \mathbb{R}^m \neq \emptyset$). Therefore, applying earlier Fact, we learn that $\forall x \in \mathbb{R}^m$, $\arg\min_{u\in\mathbb{R}^m} g_x = Prox_f(x)$ exists and is unique. □

Example 3.67: L10-13

Let C be a nonempty closed convex subset of $\mathbb{R}^m$. Then $Prox_{\delta_C} = P_C$

Proof. Let $x \in \mathbb{R}^m$. By definition,

$$\begin{aligned}
& p = Prox_{\delta_C}(x) \\
\iff & p = \arg\min_{u\in\mathbb{R}^m} \left\{ \delta_C(x) + \frac{1}{2}\|x - u\|^2 \right\} \\
\iff & \forall u \in \mathbb{R}^m,\ \delta_C(p) + \frac{1}{2}\|x - p\|^2 \leqslant \delta_C(u) + \frac{1}{2}\|x - u\|^2 \\
\iff & p \in C, \forall u \in C,\ \|x - p\|^2 \leqslant \|x - u\|^2 \\
\iff & p \in C, \forall v \in C, \|x - p\| \leqslant \|x - u\| \\
\iff & p = P_C(x)
\end{aligned}$$

□

Proposition 3.68: L10-14

Let $f : \mathbb{R}^m \to (-\infty, \infty]$ be convex l.s.c. and proper. Let $x \in \mathbb{R}^m$, let $p \in \mathbb{R}^m$. Then

$$p = Prox_f(x) \iff \forall y \in \mathbb{R}^m,\ \langle y - p, x - p \rangle + f(p) \leqslant f(y)$$

Proof. Let $y \in \mathbb{R}^m$.

- ($\implies$) Suppose that $p = Prox_f(x)$ and set $\forall\lambda \in (0,1),\ P_\lambda = \lambda y + (1-\lambda)p$. Then

  $$f(p) + \frac{1}{2}\|x-p\|^2 \leqslant f(p_\lambda) + \frac{1}{2}\|x-p_\lambda\|^2$$

  which implies that

  $$\begin{aligned}
  f(p) &\leqslant f(p_\lambda) + \frac{1}{2}\|x-p_\lambda\|^2 - \frac{1}{2}\|x-p\|^2 \\
  &= f(p_\lambda) + \frac{1}{2}\|x-\lambda y-(1-\lambda)p\|^2 - \frac{1}{2}\|x-p\|^2 \\
  &= f(p_\lambda) + \frac{1}{2}\langle x-p-\lambda(y-p)-(x-p), x-p-\lambda(y-p)+(x-p)\rangle \\
  &= f(p_\lambda) + \frac{1}{2}\langle -\lambda(y-p), 2(x-p)-\lambda(y-p)\rangle \\
  &= f(p_\lambda) + \frac{\lambda^2}{2}\|y-p\|^2 - \lambda\langle x-p, y-p\rangle \\
  &= f(\lambda y+(1-\lambda)p) + \frac{\lambda^2}{2}\|y-p\|^2 - \lambda\langle x-p, y-p\rangle
  \end{aligned}$$

  By convexity of f we have for every $\lambda \in (0,1)$,

  $$f(p) \leqslant \lambda f(y) + (1-\lambda)f(p) + \frac{\lambda^2}{2}\|y-p\|^2 - \lambda\langle x-p, y-p\rangle$$

  Rearranging yields

  $$\lambda\langle x-p, y-p\rangle + \lambda f(p) \leqslant \lambda f(y) + \frac{\lambda^2}{2}\|y-p\|^2$$

  Dividing by λ and taking the limit as $\lambda \to 0$ yields the desired inequality.

- ($\impliedby$) Suppose that

  $$\langle y-p, x-p\rangle + f(p) \leqslant f(y)$$

  Then

  $$f(p) \leqslant f(y) - \langle y-p, x-p\rangle = f(y) + \langle x-p, p-y\rangle$$

  Therefore,

  $$\begin{aligned}
  f(p) + \frac{1}{2}\|x-p\|^2 &\leqslant f(y) + \langle x-p, p-y\rangle + \frac{1}{2}\|x-p\|^2 \\
  &\leqslant f(y) + \langle x-p, p-y\rangle + \frac{1}{2}\|x-p\|^2 + \frac{1}{2}\|p-y\|^2 \\
  &= f(y) + \frac{1}{2}\|(x-p)+(p-y)\|^2 \\
  &= f(y) + \frac{1}{2}\|x-y\|^2
  \end{aligned}$$

$\square$

Example 3.69: L10-15

Let $f : \mathbb{R} \to \mathbb{R} : x \to |x|$, *then*

$$Prox_f(x) = \begin{cases} x-1, & x > 1 \\ 0, & -1 \leqslant x \leqslant 1 \\ x+1, & x < -1 \end{cases}$$

Proof. Let $p \in \mathbb{R}$. Recall that $p = Prox_{|\cdot|}(x)$

$$\iff \forall y \in \mathbb{R},\ (y-p)(x-p) + |p| \leqslant |y| \ldots (1)$$

Setting $y = 0$, $y = 2p$ respectively yield

$$\begin{aligned} -p(x-p) + |p| \leqslant 0,\ p(x-p) + |p| \leqslant 2|p| \\ \implies p(x-p) \geqslant |p|,\ p(x-p) \leqslant |p| \\ \implies p(x-p) = |p| \ldots (2) \end{aligned}$$

Therefore, (1) becomes

$$\begin{aligned} \forall y \in \mathbb{R}, (y-p)(x-p) + p(x-p) \leqslant |y| \\ \implies \forall y \in \mathbb{R},\ y(x-p) \leqslant |y| \\ \implies x-p \leqslant 1,\ x-p \geqslant -1 \\ \implies p \geqslant x-1,\ p \leqslant x+1 \ldots (3) \end{aligned}$$

- If $x > 1$: Then (3) implies $p \geqslant x - 1 > 0$. Hence, (2) implies that $x - p = 1$. Equivalently, $p = x - 1$.
- If $x < -1$: Proceed similar to the above case
- If $-1 \leqslant x \leqslant 1$: It follows from (3) that

  $$x - p \leqslant 1, x - p \geqslant -1 \implies (x-p)^2 = |x-p|^2 \geqslant 1$$

  Now, using (1) with $y = x$ yields

  $$|x| \geqslant |p| + (x-p)^2 \geqslant |p| + 1$$

  That is

  $$[0 \leqslant |p| \leqslant |x| - 1 \leqslant 1 - 1 \leqslant 0] \iff p = 0$$

$\square$

Proposition 3.70: L10-16

Let $f : \mathbb{R}^m \to (-\infty, \infty]$ *be convex l.s.c. and proper. Then*

$$x \text{ minimizes } f \text{ over } \mathbb{R}^m \iff x = Prox_f(x)$$

Proof. Recall the prop L10-14. Let $x \in \mathbb{R}^m$, then

$$\begin{aligned} & x = Prox_f(x) \\ \iff & \forall y \in \mathbb{R}^m, \langle y - x, x - x \rangle + f(x) \leqslant f(y) \\ \iff & \forall y \in \mathbb{R}^m, f(x) \leqslant f(y) \end{aligned}$$

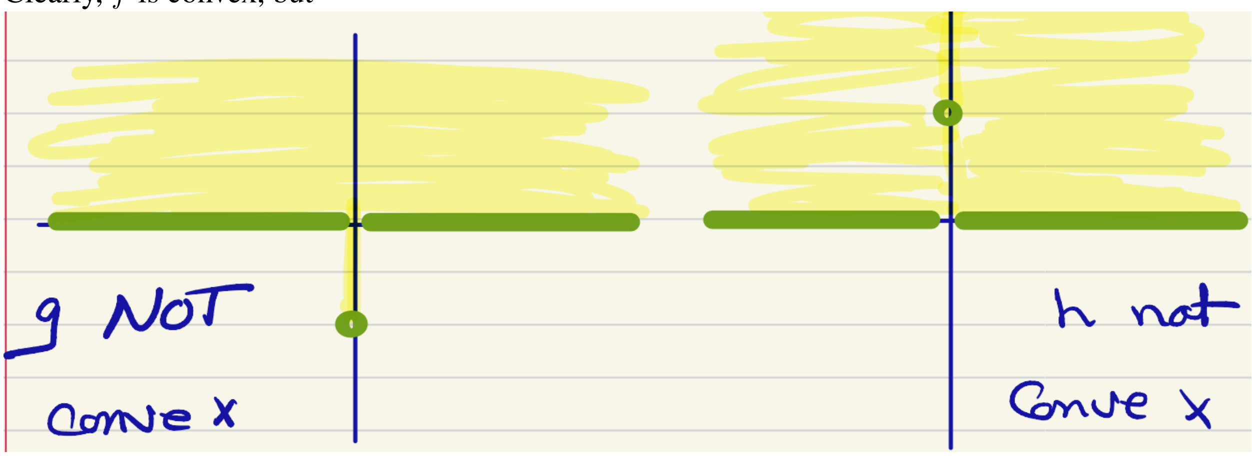
$\square$

3.11 More on Proximal Operators

Why Proximal Operators of convex functions are really "nice"? Consider the functions f, g, h defined on the real line:

$$\begin{aligned}
&\forall x \in \mathbb{R}, \lambda > 0 \\
&f(x) = 0 \\
&g(x) = \begin{cases} 0, & x \neq 0 \\ -\lambda, & x = 0 \end{cases} \\
&h(x) = \begin{cases} 0, & x \neq 0 \\ \lambda, & x = 0 \end{cases}
\end{aligned}$$

Clearly, f is convex, but



- $Prox_f$: Let $x \in \mathbb{R}$. $Prox_f(x)$ is the "unique" minimizer of the function $\frac{1}{2}(y-x)^2 \geqslant 0$. Clearly, $\forall x \in \mathbb{R}$, $Prox_f(x) = x$
- $Prox_g$: $g(x) = \begin{cases} 0, & x \neq 0 \\ -\lambda, & x = 0 \end{cases}$. Let $x \in \mathbb{R}$. $Prox_g(x)$ is the minimizer of function

$$\begin{aligned}
k(y) &= g(y) + \frac{1}{2}(y-x)^2 \\
&= \begin{cases} \frac{1}{2}(y-x)^2, & y \neq 0 \\ \frac{1}{2}x^2 - \lambda, & y = 0 \end{cases}
\end{aligned}$$

Let k_{opt} be the minimum value of $k(y)$. Observe that if $x^2 \geqslant 2\lambda$, then $k_{opt} \geqslant 0$.
If $x^2 > 2\lambda$ (equivalently $|x| > \sqrt{2\lambda}$), then $k_{opt} = 0$ and is attained $\iff y = x$.
If $x^2 = 2\lambda$ (equivalently $|x| = \sqrt{2\lambda}$), then $k_{opt} = 0$ and is attained $\iff y \in \{0, x\}$.
If $x^2 < 2\lambda$ (equivalently $|x| < \sqrt{2\lambda}$), then $k_{opt} = \frac{1}{2}x^2 - 2\lambda$ and is attained $\iff y = 0$
Therefore,

$$Prox_g(x) = \begin{cases} \{x\}, & |x| > \sqrt{2\lambda} \\ \{0, x\}, & |x| = \sqrt{2\lambda} \\ \{0\}, & |x| < \sqrt{2\lambda} \end{cases}$$

which shows that $Prox_g$ is NOT necessarily single valued.

- $$Prox_h(x) = \begin{cases} \{x\}, & x \neq 0 \\ \emptyset, & x = 0 \end{cases}$$

  i.e., $Prox_h(x)$ is not defined at $x = 0$.

So convexity is critical for the Proximal Operator to be well defined.

Proof. See A3 $\square$

Example 3.71: L11-1

Let $f : \mathbb{R} \to \mathbb{R} : x \to \lambda|x|$, $\lambda \geqslant 0$. *Then* f *is convex. We claim that* $\forall x \in \mathbb{R}$

$$Prox_f(x) = \begin{cases} x - \lambda, & x > \lambda \\ 0, & -\lambda \leqslant x \leqslant \lambda \\ x + \lambda, & x < -\lambda \end{cases}$$

This is know as the soft threshold. The above formula is often written as

$$Prox_f(x) = sgn(x)(|x| - \lambda)_+$$

where $\forall y \in \mathbb{R}$,

$$sgn(y) = \begin{cases} 1, & y \geqslant 0 \\ -1, & y < 0 \end{cases}$$

$$\begin{aligned}(y)_+ &= \begin{cases} y, & y \geqslant 0 \\ 0, & y < 0 \end{cases} \\ &= \max\{y, 0\}\end{aligned}$$

Theorem 3.72: L11-2

Let $f : \mathbb{R}^m \to (-\infty, \infty]$ *be given by* $\forall x = (x_1, \dots, x_m) \in \mathbb{R}^m$,

$$f(x_1, x_2, \dots, x_m) = \sum_{i=1}^{m} f_i(x_i)$$

where $\forall i \in \{1, \dots, m\}$,

$$f_i : \mathbb{R} \to (-\infty, \infty] \text{ is convex l.s.c and proper}$$

Then $\forall x = (x_1, \dots, x_m) \in \mathbb{R}^m$,

$$Prox_f(x) = (Prox_{f_i}(x_i))_{i=1}^m = (Prox_{f_1}(x_1), \dots, Prox_{f_m}(x_m))$$

Proof. It follows from A2 that f is convex l.s.c and proper.
Let $p = (p_1, p_2, \dots, p_m) \in \mathbb{R}^m$. Then

$$\begin{aligned}
&p = Prox_f(x)\\
\iff &\forall y = (y_1, \dots, y_m) \in \mathbb{R}^m,\\
&f(y) \geqslant f(p) + \langle y - p, x - p\rangle \text{ by L10-14}\\
\iff &\forall \{y_1, \dots, y_m\} \subseteq \mathbb{R},\\
&f_1(y_1) + \dots f_m(y_m) \geqslant f_1(p_1) + \dots f_m(p_m) + (y_1 - p_1)(x_1 - p_1) + \dots + (y_m - p_m)(x_m - p_m)
\end{aligned}$$

Setting $\forall i \in \{2, \dots, m\}$, $y_i = p_i$, we learn that $\forall y_1 \in \mathbb{R}$,

$$f_1(y_1) \geqslant f_1(p_1) + (y_1 - p_1)(x_1 - p_1) \iff p_1 = Prox_{f_1}(x_1)$$

Similar arguments yield

$$\forall i \in \{1, \dots, m\},\ p_i = Prox_{f_i}(x)$$

The proof is complete. □

Example 3.73: L11-3

Let $g : \mathbb{R}^m \to (-\infty, \infty]$ be given by $\alpha > 0$,

$$g(x) = \begin{cases} -\alpha \sum_{i=1}^m \log(x_i), & x > 0 \\ \infty, & \textit{otherwise} \end{cases}$$

Then,

$$Prox_g(x) = \left(\frac{x_i + \sqrt{x_i^2 + 4\alpha}}{2}\right)_{i=1}^m$$

Proof. Consider the function $f : \mathbb{R} \to (-\infty, \infty]$ where $\forall x \in \mathbb{R}$,

$$f(x) = \begin{cases} -\alpha \log(x), & x > 0 \\ \infty, & \text{otherwise} \end{cases}$$

Then f is convex, l.s.c and proper.
Indeed,

$$\begin{aligned}
&\forall x > 0,\ f \text{ is differentiable} \implies \text{l.s.c}\\
&\forall x > 0,\ f''(x) = \frac{\alpha}{x^2} > 0 \implies \text{convex}\\
&\forall x > 0,\ f(x) > -\infty,\ dom(f) \neq \emptyset \implies \text{proper}
\end{aligned}$$

We claim that $\forall x \in \mathbb{R}$,

$$Prox_f(x) = \frac{x + \sqrt{x^2 + 4\lambda}}{2}$$

Indeed, recall that $p = Prox_f(x)$ is the unique minimizer of the function.

$$\begin{aligned} h(y) &= f(y) + \frac{1}{2}(y-x)^2 \\ &= \begin{cases} -\alpha \log(y) + \frac{1}{2}(y-x)^2, & y > 0 \\ \infty, & \text{otherwise} \end{cases} \end{aligned}$$

Clearly, h is differentiable on its $domain = (0, \infty)$. Therefore,

$$\begin{aligned} p = Prox_f(x) &\iff h'(p) = 0 \\ &\iff (-\alpha \log(p) + \frac{1}{2}(p-x)^2)' = 0 \\ &\iff -\frac{\alpha}{p} + p - x = 0 \\ &\iff p^2 - xp - \alpha = 0,\ p > 0 \\ &\iff p > 0, p = \frac{x \pm \sqrt{x^2 + 4\alpha}}{2} \\ &\implies p = \frac{x + \sqrt{x^2 + 4\alpha}}{2} \end{aligned}$$

Now combine with $L11 - 2$,

$$f_1 = f_2 = \ldots = f_m = f$$

$\square$

Theorem 3.74: L11-4

Let $g : \mathbb{R}^m \to (-\infty, \infty]$ be proper, let $c > 0$, let $a \in \mathbb{R}^m$, let $\gamma \in \mathbb{R}$, and set $\forall x \in \mathbb{R}^m$,

$$f(x) = g(x) + \frac{c}{2}\|x\|^2 + \langle a, x \rangle + \gamma$$

Then $\forall x \in \mathbb{R}^m$,

$$Prox_f(x) = Prox_{\frac{1}{c+1}g}\left(\frac{x-a}{c+1}\right)$$

Proof. Indeed, recall that

$$\begin{aligned} Prox_f(x) &= \arg\min_{u \in \mathbb{R}^m} \left\{ f(u) + \frac{1}{2}\|u - x\|^2 \right\} \\ &= \arg\min_{u \in \mathbb{R}^m} \left\{ g(u) + \underbrace{\frac{c}{2}\|u\|^2 + \langle a, u \rangle + \gamma + \frac{1}{2}\|u - x\|^2}_{(1)} \right\} \end{aligned}$$

Now,

$$
\begin{aligned}
&\frac{c}{2}\|u\|^2+\langle a,u\rangle+\frac{1}{2}\|u-x\|^2\\
=&\frac{c}{2}\|u\|^2+\langle a,u\rangle+\frac{1}{2}\|u\|^2-\langle u,x\rangle+\frac{1}{2}\|x\|^2\\
=&\frac{c+1}{2}\|u\|^2-\langle u,x-a\rangle+\frac{1}{2}\|x\|^2\\
=&\frac{c+1}{2}\left[\|u\|^2-2\left\langle u,\frac{x-a}{c+1}\right\rangle+\frac{1}{c+1}\|x\|^2\right]\\
=&\frac{c+1}{2}\left[\left\|u-\frac{x-a}{c+1}\right\|^2-\frac{\|x-a\|^2}{(c+1)^2}+\frac{1}{c+1}\|x\|^2\right]\dots(2)
\end{aligned}
$$

Observe that for any function $h, c\in\mathbb{R}, \alpha>0$,

$$\arg\min_{u\in\mathbb{R}^m}\{\alpha h(u)+c\}=\arg\min_{u\in\mathbb{R}^m}\{h(u)\}$$

Combining (1), (2),

$$
\begin{aligned}
Prox_f(x)&=\arg\min_{u\in\mathbb{R}^m}\left\{g(u)+\frac{c+1}{2}\left\|u-\frac{x-a}{c+1}\right\|^2+\gamma-\frac{\|x-a\|^2}{(c+1)^2}+\frac{1}{c+1}\|x\|^2\right\}\\
&=\arg\min_{u\in\mathbb{R}^m}\left\{g(u)+\frac{c+1}{2}\left\|u-\frac{x-a}{c+1}\right\|^2\right\}\\
&=\arg\min_{u\in\mathbb{R}^m}\left\{(c+1)\left[\frac{1}{c+1}g(u)+\frac{1}{2}\left\|u-\frac{x-a}{c+1}\right\|^2\right]\right\}\\
&=\arg\min_{u\in\mathbb{R}^m}\left\{\frac{1}{c+1}g(u)+\frac{1}{2}\left\|u-\frac{x-a}{c+1}\right\|^2\right\}\\
&=Prox_{\frac{1}{c+1}g}\left(\frac{x-a}{c+1}\right)
\end{aligned}
$$

□

Example 3.75: L11-5

Let $\alpha\in[0,\infty)$*, let* $C=[0,\alpha]$*, set* $f=\delta_C$*. Then* $\forall x\in\mathbb{R}$*,*

$$
\begin{aligned}
Prox_f(x)&=P_C(x)\\
&=\begin{cases}0, & x\leqslant 0\\ x, & 0<x<\alpha\\ \alpha, & x\geqslant\alpha\end{cases}\\
&=\min\{\max\{x,0\},\alpha\}
\end{aligned}
$$

Proof. Recall L10-13: If C is a nonempty, closed convex subset of $\mathbb{R}^m$, then $Prox_{\delta_C}=P_C$. □

Example 3.76: L11-6

Let $f:\mathbb{R}\to(-\infty,\infty]$ *be given by* $\forall x\in\mathbb{R}$,

$$f(x)=\begin{cases}\mu x, & 0\leqslant x\leqslant\alpha\\ \infty, & \textit{otherwise}\end{cases}$$

where $\mu\in\mathbb{R},\alpha\geqslant 0$.
Then $\forall x\in\mathbb{R}$,

$$f(x)=\mu x+\delta_{[0,\alpha]}(x)\ldots(1)$$

Moreover,

$$Prox_f(x)=\min\{\max\{x-\mu,0\},\alpha\}$$

Proof. (1) follows from the definition of

$$\delta_{[0,\alpha]}(x)=\begin{cases}0, & x\in[0,\alpha]\\ \infty, & \text{otherwise}\end{cases}$$

f is proper, convex and l.s.c.
Then apply Theorem L11-4 with $c=\gamma=0$, $g=\delta_{[0,\alpha]}$, $a=\mu$, $C=[0,\alpha]$. In the view of L11-5, we yield

$$Prox_f(x)=Prox_g(x-\mu)=P_C(x-\mu)=\min\{\max\{x-\mu,0\},\alpha\}$$

□

Theorem 3.77: L12-1

Let $g : \mathbb{R} \to (-\infty, \infty]$ be convex l.s.c and proper such that $dom(g) \subseteq [0, \infty)$ and let $f : \mathbb{R}^m \to \mathbb{R}$ be given by

$$f(x) = g(\|x\|)$$

Then

$$Prox_f(x) = \begin{cases} Prox_g(\|x\|)\frac{x}{\|x\|}, & x \neq 0 \\ \{u \in \mathbb{R}^m | \|u\| = Prox_g(0)\}, & x = 0 \end{cases}$$

Proof.

- $x = 0$: By definition we have $Prox_f(0)$ is the set:

$$\arg\min_{u \in \mathbb{R}^m} \left\{ f(u) + \frac{1}{2}\|u\|^2 \right\}$$

  Using the change of variable, $w = \|u\|$, the above set of minimizers is the same as

$$\arg\min_{w \in \mathbb{R}} \left\{ g(w) + \frac{1}{2}w^2 \right\} = Prox_g(0)$$

  That is,

$$Prox_f(0) = \{u \in \mathbb{R}^m | \|u\| = Prox_g(0)\}$$

- $x \neq 0$: In this case $Prox_f(x)$ is the set of solutions of the problem

$$\begin{aligned} &\min_{u \in \mathbb{R}^m} \left\{ g(\|u\|) + \frac{1}{2}\|u - x\|^2 \right\} \\ &= \min_{u \in \mathbb{R}^m} \left\{ g(\|u\|) + \frac{1}{2}\|u\|^2 - \langle u, x \rangle + \frac{1}{2}\|x\|^2 \right\} \\ &= \min_{\alpha \geqslant 0} \min_{\substack{u \in \mathbb{R}^m \\ \|u\| = \alpha}} \left\{ g(\alpha) + \frac{1}{2}\alpha^2 - \langle u, x \rangle + \frac{1}{2}\|x\|^2 \right\} \end{aligned}$$

Observe that

$$-\langle u, x \rangle = -\|u\|\|x\|\cos(\theta_{u,x}) \geqslant -\|u\|\|x\|$$

Therefore,

$$\min_{\substack{u \in R^m \\ \|u\| = \alpha}} -\langle u, x \rangle = -\|u\|\|x\| = -\alpha\|x\|$$

and it is attained at $u = \alpha\frac{x}{\|x\|}$.
The corresponding optimal value of the inner minimization problem is therefore

$$g(\alpha) + \frac{1}{2}\alpha^2 - \alpha\|x\| + \frac{1}{2}\|x\|^2 = g(\alpha) + \frac{1}{2}(\alpha - \|x\|)^2$$

Therefore, $Prox_f(x) = \overline{\alpha}\frac{x}{\|x\|}$, where

$$\begin{aligned}\overline{\alpha} &= \min_{\alpha \geqslant 0}\left\{g(\alpha) + \frac{1}{2}(\alpha - \|x\|)^2\right\} \\ &= \min_{\alpha \in \mathbb{R}}\left\{g(\alpha) + \frac{1}{2}(\alpha - \|x\|)^2\right\} \\ &= Prox_g(\|x\|)\end{aligned}$$

The proof is complete. □

Example 3.78: L12-2

Let $\alpha > 0$*,* $f : \mathbb{R} \to (-\infty, \infty]$ *be given by* $\forall x \in \mathbb{R}$*,*

$$f(x) = \begin{cases} \lambda|x|, & |x| \leqslant \alpha \\ \infty, & \textit{otherwise} \end{cases}$$

where $\lambda \geqslant 0$*. Then* f *is convex l.s.c and proper. Moreover,* $\forall x \in \mathbb{R}$*,*

$$Prox_f(x) = \min\{\max\{|x| - \lambda, 0\}, \alpha\}sgn(x)$$

where $\forall x \in \mathbb{R}$*,*

$$sgn(x) = \begin{cases} 1, & x \geqslant 0 \\ -1, & x < 0 \end{cases}$$

Proof. Define $\forall x \in \mathbb{R}$,

$$g(x) = \begin{cases} \lambda x, & 0 \leqslant x \leqslant \alpha \\ \infty, & \text{otherwise} \end{cases}$$

$$dom(g) = [0, \alpha] \subseteq [0, \infty)$$

Moreover, $\forall x \in \mathbb{R},\ f(x) < g(|x|)$, using theorem L12-1, we learn that

$$Prox_f(x) = \begin{cases} Prox_g(|x|)\frac{x}{|x|}, & x \neq 0 \\ \{u \in \mathbb{R} |\ |u| = Prox_g(0)\}, & x = 0 \end{cases}$$

Recalling

$$g(x) = \begin{cases} \lambda x, & 0 \leqslant x \leqslant \alpha \\ \infty, & \text{otherwise} \end{cases}$$

and example L11-6, we obtain

$$|u| = Prox_g(0) \iff |u| = \min\{\max\{-\lambda, 0\}, \alpha\} = 0 \iff u = 0$$

Hence,

$$\begin{aligned}Prox_f(x) &= \begin{cases} Prox_g(|x|)sgn(x), & x \neq 0 \\ 0, & x = 0 \end{cases} \\ &= \min\{\max\{|x| - \lambda, 0\}, \alpha\}sgn(x)\end{aligned}$$

□

Example 3.79: L12-3

Let $w = (w_1, \ldots, w_m) \in \mathbb{R}^m_+$*, let* $\alpha = (\alpha_1, \ldots, \alpha_m) \in \mathbb{R}^m_+$*. Let* $f : \mathbb{R} \to (-\infty, \infty]$ *be given by*

$$f(x) = \begin{cases} \sum_{i=1}^m w_i|x_i|, & -\alpha \leqslant x \leqslant \alpha \\ \infty, & \textit{otherwise} \end{cases}$$

Then,

1. $Prox_f(x) = (\min\{\max\{|x_i|-w_i, 0\}, \alpha_i\} sgn(x_i))_{i=1}^m$
2. *Let* $x_0 \in \mathbb{R}^m$*.* $\forall n \in \mathbb{N}$*, update via*

$$x_{n+1} = Prox_f(x_n)$$

Then $x_n \to \overline{x}$ *where* $\overline{x}$ *solves the problem*

$$\min \sum_{i=1}^m w_i|x_i|$$
$$\textit{subject to } |x_i| \leqslant \alpha_i,\ i \in \{1, \ldots, m\}$$

Proof.

1. See A3
2. See A3 for numerical illustration. Proof later.

□

4 Nonexpansive, Firmly Nonexpansive and Averaged Operators

From now on, we shall use I_d to denote the $m \times m$ identity matrix on $\mathbb{R}^m$, i.e.,

$$\begin{aligned} I_d &: \mathbb{R}^m \to \mathbb{R}^m \\ &: x \to x \end{aligned}$$

Definition 4.1: L12-4

1. *Let $T : \mathbb{R}^m \to \mathbb{R}^m$. Then T is nonexpansive if $\forall x, y \in \mathbb{R}^m$,*

$$\|Tx - Ty\| \leqslant \|x - y\|$$

2. *T is firmly nonexpansive if $\forall x, y \in \mathbb{R}^m$,*

$$\|Tx - Ty\|^2 + \|(I_d - T)x - (I_d - T)y\|^2 \leqslant \|x - y\|^2$$

3. *Let $\alpha \in (0, 1)$, then T is α-averaged if*

$$\begin{gathered} \exists N : \mathbb{R}^m \to \mathbb{R}^m, N \text{ is nonexpansive} \\ T = (1 - \alpha)I_d + \alpha N \end{gathered}$$

We can show that Firmly nonexpansive (f.n.e) $\implies$ Averaged $\implies$ (Triangle Inequality) nonexpansive.

Proposition 4.2: L12-5

Let $T : \mathbb{R}^m \to \mathbb{R}^m$. Then the following are equivalent:

1. *T is f.n.e.*
2. *$I_d - T$ is f.n.e.*
3. *$2T - I_d$ is nonexpansive.*
4. *$\forall x, y \in \mathbb{R}^m$, $\|Tx - Ty\|^2 \leqslant \langle x - y, Tx - Ty \rangle$*
5. *$\forall x, y \in \mathbb{R}^m$, $\langle Tx - Ty, (I_d - T)x - (I_d - T)y \rangle \geqslant 0$*

Proof.

- $(1) \iff (2)$: clear from the definition
- $(1) \iff (3) \iff (4) \iff (5)$ See A3

$\square$

For linear operators the previous Proposition can be defined as follows:

Proposition 4.3: L12-6

Let $T : \mathbb{R}^m \to \mathbb{R}^m$ be linear. Then the following are equivalent

1. *T is f.n.e*
2. $\|2T - I_d\| \leqslant 1$
3. $\forall x \in \mathbb{R}^m$, $\|Tx\|^2 \leqslant \langle x, Tx\rangle$
4. $\forall x \in \mathbb{R}^m$, $\langle Tx, x - Tx\rangle \geqslant 0$

Proof.

- Using Prop L12-5, we have T is f.n.e $\iff$ $2T - I_d$ is nonexpansive. Since T ie linear, so is $2T - I_d$. Therefore, $2T - I_d$ is nonexpansive $\iff$ $\forall x, y \in \mathbb{R}^m$,

$$\begin{aligned}
&\|(2T - I_d)x - (2T - I_d)y\| \leqslant \|x - y\| \\
\iff &\forall z \in \mathbb{R}^m \ \|(2T - I_d)z\| \leqslant \|z\| \\
\implies &\forall z \in \mathbb{R}^m \setminus \{0\}, \ \frac{\|(2T - I_d)z\|}{\|z\|} \leqslant 1 \\
\implies &\sup \frac{\|(2T - I_d)z\|}{\|z\|} \leqslant 1 \\
\implies &\|2T - I_d\| \leqslant 1
\end{aligned}$$

- Conversely, suppose that $\|2T - I_d\| \leqslant 1$, then $\forall z \in \mathbb{R}^m \setminus \{0\}$,

$$\frac{\|(2T - I_d)z\|}{\|z\|} \leqslant \sup_{z \neq 0} \frac{\|(2T - I_d)z\|}{\|z\|} = \|2T - I_d\| \leqslant 1$$

  which implies

$$\forall z \in \mathbb{R}^m, \ \|(2T - I_d)z\| \leqslant \|z\|$$

  let $x, y \in \mathbb{R}^m$, setting $z = x - y$ shows that $2T - I_d$ is nonexpansive, so we yield the desired results.

$\square$

Remark. L12-7 It follows from the equivalence,

$$T \text{ is f.n.e} \iff 2T - I_d \text{ is nonexpansive}$$

that T is f.n.e $\iff$ T is $\frac{1}{2}$ − averaged.
Indeed,

$$\begin{aligned}
T \text{ is f.n.e} &\iff 2T - I_d =: N \text{ is nonexpansive} \\
&\iff 2T = I_d + N, \ N \text{ nonexpansive} \\
&\iff T = \frac{1}{2} I_d + \frac{1}{2} N, \ N \text{ nonexpansive}
\end{aligned}$$

Example 4.4: L12-8

Let C be convex closed nonempty subset of $\mathbb{R}^m$. Then P_C is f.n.e. Simply recall L10-1 and L12-5.

Example 4.5: L12-9

Suppose that $T = -\frac{1}{2}I_d$. Then T is averaged but NOT f.n.e.
Indeed,

$$T = \frac{1}{4}I_d + \frac{3}{4}(-I_d) \implies T \text{ is } \frac{3}{4}\text{-averaged}$$

T is NOT f.n.e as $\forall x \in \mathbb{R}^m$,

$$\|Tx\|^2+\|x-Tx\|^2= \frac{1}{4}\|x\|^2+\frac{9}{4}\|x\|^2= \frac{10}{4}\|x\|^2= \frac{5}{2}\|x\|^2> \|x\|^2$$

whenever $x \neq 0$.

Example 4.6: L12-10

Suppose that $T = -I_d$. Then T is nonexpansive, but T is NOT average. Indeed,

$$\begin{aligned}
& T \text{ is averaged} \\
\iff & \exists \alpha \in (0,1),\ N : \mathbb{R}^m \to \mathbb{R}^m \text{ nonexpansive}, T = (1-\alpha)I_d + \alpha N \\
\iff & \exists \alpha \in (0,1),\ -I_d = (1-\alpha)I_d + \alpha N \\
\iff & \exists \alpha \in (0,1),\ (-2+\alpha)I_d = \alpha N \\
\iff & \exists \alpha \in (0,1),\ N = \frac{\alpha-2}{\alpha}I_d
\end{aligned}$$

and so

$$\begin{aligned}
& N \text{ is nonexpansive} \\
\iff & \left|\frac{\alpha-2}{\alpha}\right| \leqslant 1 \\
\iff & \frac{2-\alpha}{\alpha} \leqslant 1 \\
\iff & 2-\alpha \leqslant \alpha \\
\iff & 2\alpha \geqslant 2 \iff \alpha \geqslant 1
\end{aligned}$$

which is absurd (contradiction).

Proposition 4.7: L12-11

Let $T : \mathbb{R}^m \to \mathbb{R}^m$ be nonexpansive. Then T is continuous.

Proof. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}^m$ such that $x_n \to \overline{x}$. Goal: $T(x_n) \to T(\overline{x})$.
Indeed, $\forall n \in \mathbb{N}$,

$$0 \leqslant \|T(x_n) - T(\overline{x})\| \leqslant \|x_n - \overline{x}\|$$

Letting $n \to \infty$,

$$0 \leqslant \lim_{n\to\infty} \|T(x_n) - T(\overline{x})\| \leqslant 0$$

which shows

$$T(x_n) - T(\overline{x})$$

, as claimed. $\square$

4.1 Fixed Points

Definition 4.8: L12-12

Let $T : \mathbb{R}^m \to \mathbb{R}^m$. Then

$$Fix(T) = \{x \in \mathbb{R}^m | x = Tx\}$$

Rui Gong

Definition 4.9: L13-1

Let C be a nonempty subset of $\mathbb{R}^m$ and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}^m$. Then $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to C if $\forall c \in C, n \in \mathbb{N}$,

$$\|x_{n+1} - c\| \leqslant \|x_n - c\|$$

Example 4.10: L13-2

Recall $Fix(T) = \{x|Tx = x\}$. Say $T : \mathbb{R}^m \to \mathbb{R}^m$ nonexpansive, $Fix(T) \neq \emptyset$. Let $x_0 \in \mathbb{R}^m$, $\forall n \in \mathbb{N}$ update via

$$x_{n+1} = T(x_n)$$

Then $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to $Fix(T)$.
Indeed, observe that $\forall f \in Fix(T)$,

$$f = T(f) = T^2(f) = T^3(f) = \dots$$

Observe also that $\forall n \in \mathbb{N}$,

$$x_{n+1} = T(x_n) = T(T(x_{n-1})) = T^2(x_{n-1}) = \dots T^n(x_0)$$

Now, let $n \in \mathbb{N}$, let $f \in Fix(T)$.
Then

$$\begin{aligned}\|x_{n+1} - f\| &= \|T^n(x_0) - T^n(f)\| \\ &= \|T(T^{n-1}(x_0)) - T(T^{n-1}(f))\| \\ &= \|T(x_n) - T(f)\| \\ &\leqslant \|x_n - f\|\end{aligned}$$

Proposition 4.11: L13-3

Let $\emptyset \neq C \subseteq \mathbb{R}^m$, let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}^m$. Suppose $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to C. Then the following hold:

1. *$(x_n)_{n\in\mathbb{N}}$ is bounded.*
2. *For every $c \in C$, $(\|x_n - c\|)_{n\in\mathbb{N}}$ converges.*
3. *$(d_C(x_n))_{n\in\mathbb{N}}$ is decreasing and converges.*

Proof. 1. Let $c \in C$. By the triangle inequality, $\forall n \in \mathbb{N}$, have

$$\begin{aligned}\|x_n\| &\leqslant \|c\|+\|x_n - c\| \\ &\leqslant \|c\|+\|x_{n-1} - c\| \\ &\vdots \\ &\leqslant \|c\|+\|x_0 - c\|\end{aligned}$$

Hence, $(x_n)_{n\in\mathbb{N}}$ is bounded as claimed.

2. Observe that $\forall n \in \mathbb{N},\ c \in C$,

$$0 \leqslant \|x_{n+1} - c\| \leqslant \|x_n - c\|$$

That is the sequence $(\|x_n - c\|)_{n\in\mathbb{N}}$ is a non increasing sequence of real number, bounded below implies that $(\|x_n - c\|)_{n\in\mathbb{N}}$ converges.

3. Recall that $\forall n \in \mathbb{N},\ c \in C$,

$$\|x_{n+1} - c\| \leqslant \|x_n - c\|$$

Now take the infimum over $c \in C$ to learn that

$$0 \leqslant d_C(x_{n+1}) \leqslant d_C(x_n)$$

so it converges.

□

Lemma 4.12: L13-4

Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}^m$ and let $C \neq \emptyset$ subset of $\mathbb{R}^m$. Suppose that for every $c \in C$, $(\|x_n - c\|)_{n\in\mathbb{N}}$ converges and that every cluster point of $(x_n)_{n\in\mathbb{N}}$ lies in C. Then $(x_n)_{n\in\mathbb{N}}$ converges to a point in C.

Proof. Observe that $(x_n)_{n\in\mathbb{N}}$ is bounded, because $\|x_n\| \leqslant \|x_n - c\|+\|c\|$ where $\|x_n - c\|$ converges and $\|c\|$ is a constant.
Let x, y be two cluster points of $(x_n)_{n\in\mathbb{N}}$. That is

$$x_{k_n} \to x,\ x_{l_n} \to y$$

By assumption $x \in C$, $y \in C$, observe that

$$\begin{aligned}&\|x_n - y\|^2 - \|x_n - x\|^2 + \|x\|^2 - \|y\|^2 \\ =&\|x_n\|^2+\|y\|^2-2\langle x_n, y\rangle - \|x_n\|^2-\|x\|^2+2\langle x_n, x\rangle + \|x\|^2-\|y\|^2 \\ =&2\langle x_n, x-y\rangle\end{aligned}$$

Since $(x_n - y)$ and $(x_n - x)$ converges, we have $\langle x_n, x - y\rangle$ converges say to l.
Taking the limit along x_{k_n} and x_{l_n} respectively yield

$$\begin{aligned}\langle x, x-y\rangle &= \langle y, x-y\rangle = l \\ \implies \|x-y\|^2 &= \langle x, x-y\rangle - \langle y, x-y\rangle \\ &= 0 \\ \implies x &= y\end{aligned}$$

□

Theorem 4.13: L13-5

Let $\emptyset \neq C \subseteq \mathbb{R}^m$ and let (x_n) be a sequence in $\mathbb{R}^m$. Suppose that $(x_n)_{n\in\mathbb{N}}$ if Fejér with respect to C, and that every cluster of $(x_n)_{n\in\mathbb{N}}$ lies in C. Then $(x_n)_{n\in\mathbb{N}}$ converges to a point in C.

Proof. By Fejér monotonicity of (x_n) we have

$$\text{For every } c \in C,\ (\|x_n - c\|)_{n\in\mathbb{N}} \text{ converges}$$

Now combine with Lemma 13-4 □

Let $x \in \mathbb{R}^m$, let $y \in \mathbb{R}^m$ and let $\alpha \in \mathbb{R}$. One could directly verify that

$$\|\alpha x + (1-\alpha)y\|^2 + \alpha(1-\alpha)\|x-y\|^2 = \alpha\|x\|^2 + (1-\alpha)\|y\|^2$$

Indeed:

$$\begin{aligned}\|\alpha x + (1-\alpha)y\|^2 &= \alpha^2\|x\|^2 + 2\alpha(1-\alpha)\langle x, y\rangle + (1-\alpha)^2\|y\|^2\\ \alpha(1-\alpha)\|x-y\|^2 &= \alpha(1-\alpha)\|x\|^2 + \alpha(1-\alpha)\|y\|^2 - 2\alpha(1-\alpha)\langle x, y\rangle\end{aligned}$$

Adding yields:

$$\begin{aligned}&\|\alpha x + (1-\alpha)y\|^2 + \alpha(1-\alpha)\|x-y\|^2\\ =&(\alpha^2 + (\alpha-\alpha^2))\|x\|^2 + (1-\alpha)(1-\alpha+\alpha)\|y\|^2\\ =&\alpha\|x\|^2 + (1-\alpha)\|y\|^2\end{aligned}$$

Theorem 4.14: L13-6

Let $\alpha \in (0,1)$ and let $T : \mathbb{R}^m \to \mathbb{R}^m$ be α-averaged, such that $Fix(T) \neq \emptyset$. Let $x_0 \in \mathbb{R}^m$. Update via $\forall n \in \mathbb{N}$,

$$x_{n+1} = T(x_n)$$

Then the following hold:

1. *$(x_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to $Fix(T)$*
2. *$\frac{1}{\alpha}(T - (1-\alpha)I_d)(x_n) - x_n \to 0$*
3. *$(x_n)_{n\in\mathbb{N}}$ converges to a point in $Fix(T)$.*

Proof.

1. T is averaged implies that T is nonexpansive. Now we use the example L13-2
2. By assumption, $\exists N : \mathbb{R}^m \to \mathbb{R}^m$, N is nonexpansive, such that

$$T = (1-\alpha)I_d + \alpha N \implies N = \frac{1}{\alpha}(T - (1-\alpha)I_d)$$

Rui Gong

Hence $\forall n \in \mathbb{N}$,

$$x_{n+1} = T(x_n) = (1-\alpha)x_n + \alpha N(x_n)$$

Now let $f \in Fix(T)$,

$$\begin{aligned}
\|x_{n+1} - f\|^2 &= \|(1-\alpha)x_n + \alpha N(x_n) - f\|^2 \\
&= \|(1-\alpha)(x_n - f) + \alpha(N(x_n - f)\|^2 \\
&= (1-\alpha)\|x_n - f\|^2 + \alpha\|N(x_n) - N(f)\|^2 - \alpha(1-\alpha)\|N(x_n) - x_n\|^2 \\
&\leqslant (1-\alpha)\|x_n - f\|^2 + \alpha\|x_n - f\|^2 - \alpha(1-\alpha)\|N(x_n) - x_n\|^2 \\
&= \|x_n - f\|^2 - \alpha(1-\alpha)\|N(x_n) - x_n\|^2
\end{aligned}$$

Telescoping, yields

$$\sum_{n=0}^{\infty} \alpha(1-\alpha)\|N(x_n) - x_n\|^2 \leqslant \|x_0 - f\|^2 < \infty$$

That is,

$$\begin{aligned}
&\alpha(1-\alpha)\|N(x_n) - x_n\|^2 \to 0 \\
\iff &\|N(x_n) - x_n\| \to 0
\end{aligned}$$

Recall that $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to $Fix(T)$, Observe also that

$$Fix(T) = Fix(N)$$

Indeed, let $x \in \mathbb{R}^m$, then

$$\begin{aligned}
x \in Fix(T) &\iff x = T(x) \\
&\iff x = (1-\alpha)x + \alpha N(x) \\
&\iff x = x - \alpha x + \alpha N(x) \\
&\iff \alpha x = \alpha N(x) \\
&\iff x = N(x) \\
&\iff x \in Fix(N)
\end{aligned}$$

Altogether, we learn that $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to $Fix(N)$.

3. Let $\overline{x}$ be a cluster point of $(x_n)_{n\in\mathbb{N}}$ say $x_{k_n} \to \overline{x}$. Observe that N is nonexpansive implies N is continuous. Now, recall

   $$Nx_n - x_n \to 0$$

   Taking the limit along the subsequence x_{k_n}, we learn that

   $$N\overline{x} - \overline{x} = 0$$

   equivalently, $N\overline{x} = \overline{x}$.
   That is, every cluster point of $(x_n)_{n\in\mathbb{N}}$ lies in $Fix(N) = Fix(T)$. Now combine with theorem L13-5

Rui Gong

□

Corollary 4.15: L14-1

Let $T : \mathbb{R}^m \to \mathbb{R}^m$ be f.n.e and suppose that $Fix(T) \neq \emptyset$. Let $x_0 \in \mathbb{R}^m$. $\forall n \in \mathbb{N}$, update via

$$x_{n+1} = T(x_n)$$

Then $\exists \overline{x} \in Fix(T)$ such that

$$x_n \to \overline{x}$$

Proof. Since T is f.n.e T is averaged. Now combine with Theorem L13-6 □

Proposition 4.16: L14-2

Let $f : \mathbb{R}^m \to (-\infty, \infty]$ be convex lsc and proper. Then $Prox_f$ is f.n.e.

Proof. Let $x, y \in \mathbb{R}^m$. Set

$$p = Prox_f(x),\ q = Prox_f(y)$$

Using we have Prop L10-14, $\forall z \in \mathbb{R}^m$,

$$\langle z - p, x - p\rangle + f(p) \leqslant f(z) \tag{4.1}$$
$$\langle z - q, y - q\rangle + f(q) \leqslant f(z) \tag{4.2}$$

Choosing $z = q$ in (4.1), $z = p$ in (4.2), we obtain

$$\langle q - p, x - p\rangle + f(p) \leqslant f(q)$$
$$\langle p - q, y - q\rangle + f(q) \leqslant f(p)$$

Adding the last two inequalities yields

$$\langle q - p, (x - p) - (y - p)\rangle \leqslant 0$$

Equivalently,

$$\langle p - q, (x - p) - (y - p)\rangle \geqslant 0$$

Now recall that

$$p = Prox_f(x),\ q = Prox_f(y)$$

Combining L12-5(5) with the conclusion yields the desired results. □

Corollary 4.17: L14-3

Let $f : \mathbb{R}^m \to (-\infty, \infty]$ be convex lsc and proper, such that $\arg\min f \neq \emptyset$. Let $x_0 \in \mathbb{R}^m$, $\forall n \in \mathbb{N}$, update via

$$x_{n+1} = Prox_f(x_n)$$

Then $\exists \overline{x} \in \arg\min f$ such that

$$x_n \to \overline{x}$$

Proof. Observe that by L10-16,

$$\arg\min f = Fix(Prox_f) \neq \emptyset$$

Recall the $Prox_f$ is f.n.e by L14-2
Now combine with L14-1 applied with T replaced by $Prox_f$ $\square$

The following simple identity will be used in the next result.
Let $x, y \in \mathbb{R}^m$, $\alpha \in \mathbb{R} \setminus \{0\}$, then

$$\alpha^2 \left(\|x\|^2 - \left\| (1-\frac{1}{\alpha})x + \frac{1}{\alpha}y \right\|^2 \right) = \alpha \left(\|x\|^2 - \frac{1-\alpha}{\alpha} \|x-y\|^2 - \|y\|^2 \right)$$

Indeed,

$$\begin{aligned}
LHS &= \alpha^2 \left(\|x\|^2 - \left\| (1-\frac{1}{\alpha})x + \frac{1}{\alpha}y \right\|^2 \right) \\
&= \alpha^2 \left(\|x\|^2 - \left(1-\frac{1}{\alpha}\right)^2 \|x\|^2 - \frac{1}{\alpha^2}\|y\|^2 + 2\frac{\alpha-1}{\alpha-2} \langle x, y \rangle \right) \\
&= \alpha^2 \left(\left(\frac{2}{\alpha} - \frac{1}{\alpha^2}\right) \|x\|^2 - \frac{1}{\alpha^2}\|y\|^2 + 2\frac{\alpha-1}{\alpha-2} \langle x, y \rangle \right) \\
&= (2\alpha-1)\|x\|^2 - \|y\|^2 + 2(\alpha-1) \langle x, y \rangle
\end{aligned}$$

$$\begin{aligned}
RHS &= \alpha \left(\|x\|^2 - \frac{1-\alpha}{\alpha}\|x-y\|^2 - \|y\|^2 \right) \\
&= \alpha \left(\|x\|^2 - \frac{1-\alpha}{\alpha}\|x\|^2 - \frac{1-\alpha}{\alpha}\|y\|^2 + \frac{2(1-\alpha)}{\alpha} \langle x, y \rangle - \|y\|^2 \right) \\
&= \alpha\|x\|^2 - (1-\alpha)\|x\|^2 - (1-\alpha)\|y\|^2 + 2(1-\alpha) \langle x, y \rangle - \alpha\|y\|^2 \\
&= (2\alpha-1)\|x\|^2 - \|y\|^2 + 2(\alpha-1) \langle x, y \rangle \\
&= LHS
\end{aligned}$$

4.1.1 Composition of Averaged Operators

Proposition 4.18: L14-4

Let $T : \mathbb{R}^m \to \mathbb{R}^m$ be nonexpansive and let $\alpha \in (0,1)$. Then the following are equivalent:

1. T *is* α*-average*
2. $\left(1 - \frac{1}{\alpha}\right) I_d + \frac{1}{\alpha} T$ *is nonexpansive*
3. $\forall x, y \in \mathbb{R}^m$

$$\|T(x) - T(y)\|^2 \leqslant \|x-y\|^2 - \frac{1-\alpha}{\alpha}\|(I_d - T)(x) - (I_d - T)(y)\|^2$$

Proof.

1. 1) $\iff$ 2):

$$\begin{aligned}
& T \text{ is } \alpha\text{-averaged} \\
\iff & \exists N : \mathbb{R}^m \to \mathbb{R}^m,\ N \text{nonexpansive} \\
& T = (1-\alpha) I_d + \alpha N \iff \qquad N = \frac{1}{\alpha}(T - (1-\alpha)I_d) \text{ is nonexpansive} \\
\iff & \left(1 - \frac{1}{\alpha}\right) I_d + \frac{1}{\alpha} T \text{ is nonexpansive}
\end{aligned}$$

2. Recalling the previous identity

$$\alpha^2 \left(\|x\|^2 - \left\| (1 - \frac{1}{\alpha})x + \frac{1}{\alpha} y \right\|^2 \right) = \alpha \left(\|x\|^2 - \frac{1-\alpha}{\alpha}\|x-y\|^2 - \|y\|^2 \right)$$

$$\left(\|x\|^2 - \left\| (1 - \frac{1}{\alpha})x + \frac{1}{\alpha} y \right\|^2 \right) = \frac{\alpha}{\alpha^2} \left(\|x\|^2 - \frac{1-\alpha}{\alpha}\|x-y\|^2 - \|y\|^2 \right)$$

Now, 2) $\iff \forall x, y \in \mathbb{R}^m$,

$$\left\| \left(1 - \frac{1}{\alpha}\right) x + \frac{1}{\alpha} T(x) - \left(1 - \frac{1}{\alpha}\right) y - \frac{1}{\alpha} T(y) \right\|^2 \leqslant \|x-y\|^2$$

We then rewrite the left hand side as

$$\begin{aligned}
& \left\| \left(1 - \frac{1}{\alpha}\right)(x-y) + \frac{1}{\alpha}(T(x) - T(y)) \right\|^2 \\
= & \|x-y\|^2 - \frac{1}{\alpha}\left(\|x-y\|^2 - \frac{1-\alpha}{\alpha}\|(x - T(x)) - (y - T(y))\|^2 - \|T(x) - T(y)\|^2 \right) \\
\leqslant & \|x-y\|^2
\end{aligned}$$

Now, we have

$$-\frac{1}{\alpha}\left(\|x-y\|^2-\frac{1-\alpha}{\alpha}\|(x-T(x))-(y-T(y))\|^2-\|T(x)-T(y)\|^2\right)\leqslant 0$$
$$\underset{\alpha>0}{\iff}\|x-y\|^2-\frac{1-\alpha}{\alpha}\|(x-T(x))-(y-T(y))\|^2-\|T(x)-T(y)\|^2\geqslant 0$$

□

Theorem 4.19: L14-5

Let $\alpha_1,\alpha_2\in(0,1)$, let $T_i:\mathbb{R}^m\to\mathbb{R}^m$ be α_i-averaged. Set

$$T:=T_1T_2,\ \alpha:=\frac{\alpha_1+\alpha_2-2\alpha_1\alpha_2}{1-\alpha_1\alpha_2}$$

Then T is α-averaged.

Proof. First observe that $\alpha\in(0,1)$. Indeed, clearly $\alpha_1,\alpha_2\in(0,1)$.
Now,

$$\begin{aligned}\alpha\in(0,1)&\iff\alpha_1+\alpha_2-2\alpha_1\alpha_2<1-\alpha_1\alpha_2\\&\iff\alpha_1+\alpha_2<1+\alpha_1\alpha_2\\&\iff\alpha_1-\alpha_1\alpha_2<1-\alpha_2\\&\iff\alpha_1(1-\alpha_2)<1-\alpha_2\end{aligned}$$

Hence, $\alpha\in(0,1)$ as claimed. Recalling L14-4
Now, call the inequality below (4.3),

$$\begin{aligned}\|T(x)-T(y)\|^2&=\|T_1(T_2(x))-T_1(T_2(y))\|^2\\&\leqslant\|T_2(x)-T_2(y)\|^2-\frac{1-\alpha_1}{\alpha_1}\|(I_d-T_1)(T_2(x))-(I_d-T_1)(T_2(y))\|^2\\&\leqslant\|x-y\|^2-\underbrace{\frac{1-\alpha_2}{\alpha_2}\|(I_d-T_2)(x)-(I_d-T_2)(y)\|^2}_{(1)}\\&\quad\underbrace{-\frac{1-\alpha_1}{\alpha_1}\|(I_d-T_1)(T_2(x))-(I_d-T_1)(T_2(y))\|^2}_{(2)}\end{aligned}$$

Set $\beta=\frac{1-\alpha_1}{\alpha_1}+\frac{1-\alpha_2}{\alpha_2}>0$, we claim that

$$(1)+(2)\geqslant\frac{(1-\alpha_1)(1-\alpha_2)}{\beta\alpha_1\alpha_2}\|(I_d-T)(x)-(I_d-T)(y)\|^2\ldots(3)$$

Indeed, we have

$$\begin{aligned}
\frac{1}{\beta}((1)+(2)) =& \frac{1-\alpha_2}{\beta\alpha_2}\|(I_d-T_2)(x)-(I_d-T_2)(y)\|^2 \\
&+\frac{1-\alpha_1}{\beta\alpha_1}\|(I_d-T_1)(T_2(x))-(I_d-T_1)(T_2(y))\|^2 \\
=&\left\|\frac{1-\alpha_1}{\beta\alpha_1}\left((I_d-T_1)(T_2(x))-(I_d-T_1)(T_2(y))\right)-\frac{1-\alpha_2}{\beta\alpha_2}\left((I_d-T_2)(x)-(I_d-T_2)(y)\right)\right\|^2 \\
&+\frac{(1-\alpha_1)(1-\alpha_2)}{\beta^2\alpha_1\alpha_2}\|(I_d-\underbrace{T_1T_2}_{T})(x)-(I_d-T_1T_2)(y)\|^2 \\
\geqslant&\frac{(1-\alpha_1)(1-\alpha_2)}{\beta^2\alpha_1\alpha_2}\|(I_d-T)(x)-(I_d-T)(y)\|^2
\end{aligned}$$

Note we go from the first equation to second one by, $\overline{\alpha}\in\mathbb{R}$,

$$\|\overline{\alpha}x-(1-\overline{\alpha})y\|^2+\overline{\alpha}(1-\overline{\alpha})\|x-y\|^2=\overline{\alpha}\|x\|^2+(1-\overline{\alpha})\|y\|^2$$

So we have proved (3).
Consequently, (4.3) becomes

$$\|T(x)-T(y)\|^2\leqslant\|x-y\|^2-\frac{(1-\alpha_1)(1-\alpha_2)}{\beta\alpha_1\alpha_2}\|(I_d-T)(x)-(I_d-T)(y)\|^2$$

Finally, recalling that

$$\alpha=\frac{\alpha_1+\alpha_2-2\alpha_1\alpha_2}{1-\alpha_1\alpha_2}$$

we can verify that

$$\frac{(1-\alpha_1)(1-\alpha_2)}{\beta\alpha_1\alpha_2}=\frac{1-\alpha}{\alpha}$$

Indeed,

$$\begin{aligned}
\frac{(1-\alpha_1)(1-\alpha_2)}{\alpha_1\alpha_2\left(\frac{1-\alpha_1}{\alpha_1}+\frac{1-\alpha_2}{\alpha_2}\right)} =& \frac{(1-\alpha_1)(1-\alpha_2)}{\alpha_2(1-\alpha_2)+\alpha_1(1-\alpha_2)} \\
=&\frac{1-\alpha_1-\alpha_2+\alpha_1\alpha_2}{\alpha_1+\alpha_2-2\alpha_1\alpha_2}
\end{aligned}$$

$$\begin{aligned}
\frac{1-\alpha}{\alpha} =& \frac{1-\frac{\alpha_1+\alpha_2-2\alpha_1\alpha_2}{1-\alpha_1\alpha_2}}{\frac{\alpha_1+\alpha_2-2\alpha_1\alpha_2}{1-\alpha_1\alpha_2}} \\
=&\frac{1-\alpha_1\alpha_2-\alpha_1-\alpha_2+2\alpha_1\alpha_2}{\alpha_1+\alpha_2-2\alpha_1\alpha_2} \\
=&\frac{1-\alpha_1-\alpha_2+\alpha_1\alpha_2}{\alpha_1+\alpha_2-2\alpha_1\alpha_2}
\end{aligned}$$

Now we use L14-4 to get our result. $\square$

5 Constrained Convex Optimization

We now consider the problem

$$(P)\quad \begin{array}{l}\min f(x)\\ \text{subject to } x \in C\end{array}$$

- $f : \mathbb{R}^m \to (-\infty, \infty]$ convex, l.s.c., proper
- $C \neq \emptyset$, convex and closed.

Recall L7-5, we shall see now some weaker results, in the absence of convexity.

Theorem 5.1: L15-2

$f : \mathbb{R}^m \to (-\infty, \infty]$ proper, $g : \mathbb{R}^m \to (-\infty, \infty]$ convex l.s.c. proper. $dom(g) \subseteq int(dom(f))$
Consider the problem:

$$\min_{x \in \mathbb{R}^m} f(x) + g(x)$$

1. *If $x^* \in dom(g)$ is a local optimal of (P) and f is differentiable at x^*, then*

$$-\nabla f(x^*) \in \partial g(x^*)$$

2. *Suppose that f is convex. If f is differentiable at $x^* \in dom(g)$ then*

$$x^* \text{ is a global minimizer of } (P) \iff -\nabla f(x^*) \in \partial g(x^*)$$

Proof.

1. Let $y \in dom(g)$. Since g is convex, we know that $dom(g)$ is convex. Hence $\forall \lambda \in (0, 1)$:

$$x^* + \lambda(y - x^*) = \underbrace{(1 - \lambda)x^* + \lambda y}_{:= x_\lambda} \in dom(g)$$

Therefore, for sufficiently small λ

$$\begin{aligned} &f(x_\lambda) + g(x_\lambda) \geqslant f(x^*) + g(x^*) \\ \implies &f((1 - \lambda)x^* + \lambda y) + g((1 - \lambda)x^* + \lambda y) \geqslant f(x^*) + g(x^*) \end{aligned}$$

By the convexity of g we learn that

$$f((1 - \lambda)x^* + \lambda y) + (1 - \lambda)g(x^*) + \lambda g(y) \geqslant f(x^*) + g(x^*)$$

Rearranging yield

$$\lambda g(x^*) - \lambda g(y) \leqslant f((1 - \lambda)x^* + \lambda y) - f(x^*)$$

Equivalently,

$$g(x^*) - g(y) \leqslant \frac{f((1 - \lambda)x^* + \lambda y) - f(x^*)}{\lambda}$$

Taking the limit as $\lambda \to 0^+$, we obtain

$$g(x^*) - g(y) \leqslant f'(x^*; y - x^*) = \langle \nabla f(x^*), y - x^* \rangle$$

That is: for any $y \in dom(g)$,

$$g(y) \geqslant g(x^*) + \langle -\nabla f(x^*), y - x^* \rangle \implies -\nabla f(x^*) \in \partial g(x^*)$$

2. Suppose that f is convex. Observe that $1)$ prove ($\impliedby$). Now suppose that $-\nabla f(x^*) \in \partial g(x^*)$. On the one hand, for any $y \in dom(g)$,

$$g(y) \geqslant g(x^*) + \langle -\nabla f(x^*), y - x^* \rangle \ldots (1)$$

On the other hand, since f is convex, differentiable at x^*, then, $\forall y \in dom(g) \subseteq dom(f)$,

$$f(y) \geqslant f(x^*) + \langle \nabla f(x^*), y - x^* \rangle \ldots (2)$$

Adding (1) and (2) yields for any $y \in dom(g)$,

$$f(y) + g(y) \geqslant f(x^*) + g(x^*)$$

That is, x^* is optimal solution of (P)

$\square$

5.1 KKT Conditions

In the following we assume, $f, g_1, \ldots, g_n$ are functions from $\mathbb{R}^m \to \mathbb{R}$ (full domain). $I = \{1, \ldots, n\}$
Consider the problem,

$$\begin{aligned} &\min f(x) \\ &s.t.\ g_i(x) \leqslant 0,\ (\forall i \in I) \end{aligned}$$

We assume that (P) has at least one solution and that

$$\mu := \min\{f(x) | \forall i \in I,\ g_i(x) \leqslant 0\} \in \mathbb{R}$$

is the **optimal value**.
Define

$$F(x) := \max\{\underbrace{f(x) - \mu}_{=:g_0(x)}, g_1(x), \ldots, g_n(x)\}$$

Lemma 5.2: L15-3

We have $\forall x \in \mathbb{R}^m,\ F(x) \geqslant 0$. Moreover, solutions of (P) is

$$\textit{minimizers of } F = \{x | F(x) = 0\}$$

Proof. Let $x \in \mathbb{R}^m$.

1. x does not solve (P)
   (a) x is infeasible for (P), i.e., x doesn't satisfy the constraints. Then
   $$\begin{aligned}&\implies \exists j \in I \text{ such that } g_j(x) > 0\\&\implies F(x) \geqslant g_j(x) > 0\end{aligned}$$
   (b) x is feasible ($g_i(x) \leqslant 0, \forall i$), but not optimal $\implies f(x) > \mu$,
   $$\implies F(x) \geqslant g_0(x) = f(x) - \mu > 0$$
2. x solves (P), implies that
   $$x \text{ is feasible and } f(x) = \mu$$
   also,
   $$\begin{aligned}x \text{ is feaisble} &\iff \forall i \in I,\ g_i(X) \leqslant 0\\ f(x) = \mu &\iff g_0(x) = f(x) - \mu = 0\end{aligned}$$
   Hence, $F(x) = 0$

□

Fact L15-4 (max rule for subdifferential calculus):
Let $g_1, \ldots, g_n : \mathbb{R}^m \to (-\infty, \infty]$ be convex l.s.c. and proper. Define

$$\begin{aligned}g(x) &= \max\{g_1(x), \ldots, g_n(x)\}\\ A(x) &= \{i \in \{1, \ldots, n\} | g_i(x) = g(x)\}\end{aligned}$$

Let $x \in \cap_{i=1}^n (int(dom(g_i)))$, then

$$\partial g(x) = Conv\left(\cup_{i \in A(x)} \partial g_i(x)\right)$$

Theorem 5.3: L15-5(Fritz-John necessary optimality conditions

Suppose that $f, g_1, \ldots, g_n$ are convex and x^* solves (P). Then $\exists \alpha_0 \geqslant 0, \ldots, \alpha_n \geqslant 0$ not all 0, for which

$$0 \in \alpha_0 \partial f(x^*) + \sum_{i \in I} \alpha_i \partial g_i(x^*)$$

amd $\forall i \in I$,

$$\alpha_i g_i(x^*) = 0 \leftarrow \textit{complementary slackness}$$

Proof. Recall that
$$F(x) = \max\{f(x) - \mu, g_1(x), \dots, g_n(x)\}$$
By the previous lemma,
$$F(x^*) = 0 = \min F(\mathbb{R}^m)$$
Hence,
$$0 \in \partial F(x^*) = Conv_{i \in A(x^*)} \partial g_i(x^*)$$
where
$$A(x^*) := \left\{ i \in \{0, 1, \dots, n\} | g_i(x^*) = \underbrace{0}_{F(X^*)=0} \right\}$$
Observe that $0 \in \partial F(x^*)$ because $g_0(x^*) = f(x^*) - \mu = 0 = \min F(\mathbb{R}^m)$. Moreover, $\partial g_0 = \partial f$ $(g_0 = f - \mu)$
Hence, $\forall i \in A(x^*), \exists \alpha_i \geqslant 0$,
$$\sum_{i \in A(x^*)} \alpha_i = 1$$
and
$$\begin{aligned} 0 &\in \sum_{i \in A(x^*)} \alpha_i \partial g_i(x^*) \\ &= \alpha_0 \partial g_0(x^*) + \sum_{i \in A(x^*) \setminus \{0\}} \alpha_i \partial g_i(x^*) \\ &= \alpha_0 \partial f(x^*) + \sum_{i \in A(x^*) \setminus \{0\}} \alpha_i \partial g_i(x^*) \end{aligned}$$
Now, for $i \in I \setminus A(x^*)$, set $\alpha_i = 0$.
If $i \in A(x^*) \cap I$, then
$$g_i(x^*) = 0$$
Hence,
$$\begin{aligned} i \in A(x^*) \cap I &\implies \alpha_i \underbrace{g_i(x^*)}_{=0} = 0 \\ i \notin A(x^*) \cap I = I \setminus A(x^*) &\implies \underbrace{\alpha_i}_{=0} g_i(x^*) = 0 \end{aligned}$$
Altogether, $\forall i \in I$,
$$\alpha_i g_i(x^*) = 0 \leftarrow \text{complementary slackness}$$
$\square$

5.1.1 KKT conditions

KKT: Karush-Kuhn-Tucker conditions.

In the following we assume, $f, g_1, \dots, g_n$ are functions from $\mathbb{R}^m \to \mathbb{R}$ (full domain). $I = \{1, \dots, n\}$
Consider the problem,

$$(P) \begin{array}{l} \min f(x) \\ s.t.\ g_i(x) \leqslant 0,\ (\forall i \in I) \end{array}$$

Theorem 5.4: L16-1:KKT Conditions Necessary Part

Suppose $f, g_1, \dots, g_n$ are convex, x^* solved (P). Suppose that Slater's conditions holds, i.e.,

$$\exists s \in \mathbb{R}^m, \forall i \in I = \{1, 2, \dots, n\},\ g_i(s) < 0$$

Then $\exists \lambda_1, \dots, \lambda_n \geqslant 0$ such that the KKT conditions:

1. $0 \in \partial f(x^*) + \sum_{i \in I} \lambda_i \partial g_i(x^*)$, *stationarity condition*
2. $\forall i \in I, \lambda_i g_i(x^*) = 0$, *complementary slackness condition*

hold.

Proof. Recalling Fritz-John.
$\exists \alpha_0, \alpha_1, \dots, \alpha_n \geqslant 0$, Not all 0, such that

$$0 \in \alpha_0 \partial f(x^*) + \sum_{i \in I} \alpha_i \partial g_i(x^*) \dots (*)$$

and

$$\forall i \in I, \alpha_i g_i(x^*) = 0$$

Done if we can show that $\alpha_0 > 0$!
Suppose for eventual contradiction that $\alpha_0 = 0$.
By $(*)$, $\forall i \in I, \exists y_i \in \partial g_i(x^*)$

$$\sum_{i \in I} \alpha_i y_i = 0$$

Hence, $i \in I, \forall y \in \mathbb{R}^m$,

$$g_i(x^*) + \langle y_i, y - x^* \rangle \leqslant g_i(y)$$

In particular:

$$g_i(x^*) + \langle y_i, s - x^* \rangle \leqslant g_i(s)$$

Multiplying the inequality above by $\alpha_i \geqslant 0$, then $\forall i \in I$,

$$\alpha_i g_i(x^*) + \langle \alpha_i y_i, s - x^* \rangle \leqslant \alpha_i g_i(s)$$

Adding the above inequalities,

$$\sum_{i\in I}\underbrace{\alpha_i g_i(x^*)}_{=0}+\left\langle\underbrace{\sum_{i\in I}\alpha_i y_i}_{=0}, s-x^*\right\rangle\leqslant\underbrace{\sum_{i\in I}\underbrace{\alpha_i}_{\geqslant 0}\underbrace{g_i(s)}_{<0}}_{<0}$$

which implies

$$0<0$$

which is a contradiction. Hence, $\alpha_0>0$.
Now divide $(*)$ and $\alpha_i g_i(x^*)=0$ by α_0 and set $\forall i\in I$,

$$\lambda_i=\frac{\alpha_i}{\alpha_0}\geqslant 0$$

□

Theorem 5.5: L16-2 KKT Conditions, Sufficient Parts

Suppose $f, g_1, \ldots, g_n$ are convex and $x^*\in\mathbb{R}^m$ satisfies:

1. $\forall i\in I,\ g_i(x^*)\leqslant 0$, *Primal feasibility.*
2. $\forall i\in I, \lambda_i\geqslant 0$, *Dual feasibility.*
3. $0\in\partial f(x^*)+\sum_{i\in I}\lambda_i\partial g_i(x^*)$, *Stationarity.*
4. $\forall i\in I,\ \lambda_i g_i(x^*)=0$, *Complementary Slackness.*

Then x^* solves (P).

Proof. Define

$$h(x):=f(x)+\sum_{i\in I}\lambda_i g_i(x)$$

By $2)$, $h(x)$ is convex. Observe that the sum rule applies to the sum of convex functions $f,\ \lambda_i g_i,\ i\in I$.
Therefore, $\forall x\in\mathbb{R}^m$,

$$\partial h(x)=\partial\left(f+\sum_{i\in I}\lambda_i g_i\right)(x)$$
$$\underbrace{=}_{\text{sum rule}}\partial f(x)+\sum_{i\in I}\lambda_i\partial g_i(x)$$

Consequently,

$$0\in\partial h(x^*)\underbrace{=}_{3)}\partial f(x^*)+\sum_{i\in I}\lambda_i\partial g_i(x^*)$$

By Fermat: x^* is a global minimizer of h.
Now, let x be feasible for (P), i.e.,

$$\forall i \in I,\ g_i(x) \leqslant 0$$

Then,

$$\begin{aligned} f(x^*) &\underbrace{=}_{4)} f(x^*) + \sum_{i\in I} \lambda_i g_i(x^*) \\ &= h(x^*) \\ &\leqslant h(x) \\ &= f(X) + \sum_{i\in I} \underbrace{\lambda_i}_{\geqslant 0} \underbrace{g_i(x)}_{\leqslant 0} \\ &\underbrace{\leqslant}_{1),2)} f(x) \end{aligned}$$

□

5.2 Algorithms

Subgradient methods:
Gradient descent: classical theory
Consider the problem:

$$(P)\ \min_{x\in\mathbb{R}^m} f(x)$$

Definition 5.6: L16-3

Let $f : \mathbb{R}^m \to (-\infty, \infty]$ be proper and let $x \in int(dom(f))$, $d \in \mathbb{R}^m \setminus \{0\}$ is a descent direction of f at x if the directional derivative satisfies

$$f'(x; d) < 0 \ldots (*)$$

Remark. L16-4

1. $0 \neq \nabla f(x)$ exists at $x \implies -\nabla f(x)$ is a descent direction
   Indeed:

$$\begin{aligned} f'(x; -\nabla f(x)) &= \langle \nabla f(x), -\nabla f(x)\rangle \\ &= -\|\nabla f(x)\|^2 \\ &< 0 \end{aligned}$$

2. $(*) \implies \exists \varepsilon > 0, \forall 0 < t \leqslant \varepsilon,\ f(x + td) < f(x)$

Gradient/Steepest descent method: With f is differentiable, $x_0 \in \mathbb{R}^m$. $\forall n \in \mathbb{N}$, update via

$$\begin{aligned} x_{n+1} &:= x_n - t_n \nabla f(x_n) \\ t_n &\in \arg\min_{t\geqslant 0} f(x_n - t\nabla f(x_n)) \end{aligned}$$

If f is strictly convex and coercive,

$$x_n \to \text{unique minimizer of } f$$

"Peressini, Sullivan, Uhl"

In the lack of smoothness

Example 5.7: L16-5 (L.Vandenberghe)

Negative subgradients are NOT necessarily descent directions.
Consider

$$\begin{aligned} f &: \mathbb{R}^2 \to \mathbb{R}_+ \\ &: (x_1, x_2) \mapsto |x_1| + 2|x_2| \end{aligned}$$

f *convex (sum of convex functions), full domain* $\implies$ *continuous.*

$$\begin{aligned} \partial f(1,0) &= \{1\} \times [-2,2] \\ &\ni (1,2) \end{aligned}$$

Consider $d = -(1,2) = (-1,-2)$*, let* $t > 0$*, then*

$$\begin{aligned} f((1,0) + t * (-1,-2)) &= f(1-t, -2t) \\ &= |1-t| + 2|-2t| \\ &= |1-t| + 4|t| \\ &= \begin{cases} 1+3t, & 0 \leqslant t \leqslant 1; \\ -1-3t, & t < 0; \\ 5t-1, & t \geqslant 1; \end{cases} \end{aligned}$$

Therefore,

$$\begin{aligned} & f'((1,0);(-1,-2)) \\ =& \lim_{t \downarrow 0} \frac{f((1,0)+t(-1,2)) - f(1,0)}{t} \\ =& \lim_{t \downarrow 0} \frac{1+3t-1}{t} \\ =& 3 > 0 \end{aligned}$$

Hence $(-1,2)$ *is NOT a descent direction. Moreover,*

$$\forall t > 0,\ f(1,0) = 1 < f((1,0) + t(-1,-2))$$

Example 5.8: L16-6 (Wolfe)

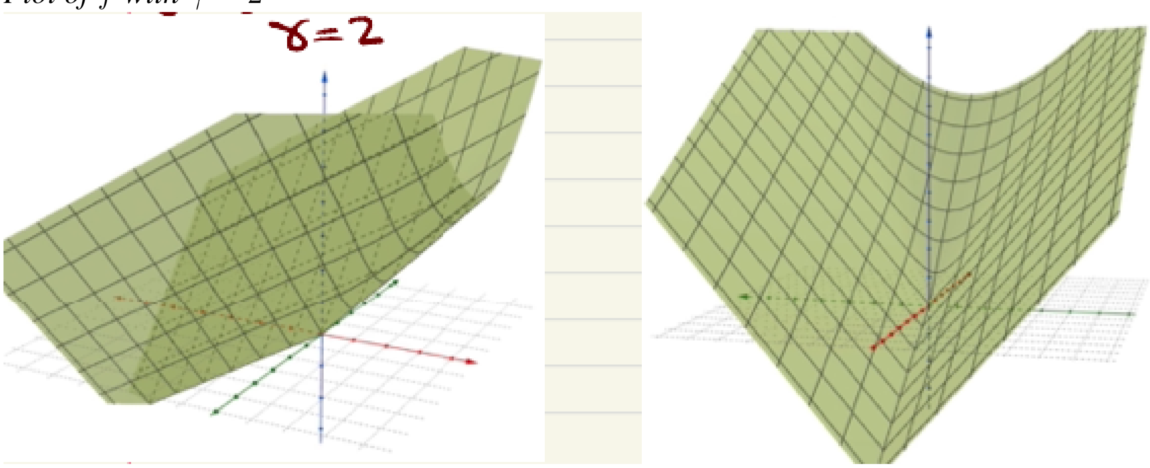
Let $\gamma > 1$. *Consider the function:*

$$f : \mathbb{R}^2 \to \mathbb{R}$$
$$: (x_1, x_2) \mapsto \begin{cases} \sqrt{x_1^2 + \gamma x_2^2}, & |x_2| \leqslant x_1; \\ \frac{x_1 + \gamma |x_2|}{\sqrt{1+\gamma}}, & \textit{otherwise} \end{cases}$$

Observe that $\arg\min_{x \in \mathbb{R}^m} f(x) = \varphi$
Indeed, $\inf_{x \in \mathbb{R}^m} f(x) = -\infty$*, as*

$$f(r, 0) = \frac{r}{\sqrt{1+\gamma}} \to -\infty, \ \textit{as}\ r \to -\infty$$

Plot of f *with* $\gamma = 2$



One can show that

$$f = \sigma_C$$

, where

$$C = \left\{ (x_1, x_2) \in \mathbb{R}^2 | x_1^2 + \frac{x_2^2}{\gamma} \leqslant 1, x_1 \geqslant \frac{1}{\sqrt{1+\gamma}} \right\}$$

Therefore, f *is convex.*
Also, f *is differentiable on*

$$\mathbb{R}^2 \setminus ((-\infty, 0] \times \{0\})$$

Now, let $x_0 = (\gamma, 1)$ *be in the set above.*
The steepest descent will generate a sequence (details omitted)

$$x_n = \left(\gamma \left(\frac{\gamma - 1}{\gamma + 1} \right)^n, (-1)^n \left(\frac{\gamma - 1}{\gamma + 1} \right)^n \right) \to (0, 0)$$

Observe that $(0, 0)$ *is NOT a minimizer of* f.
In the absence of smoothness a lot of pathologies happen.

5.3 Projected Subgradient Method

$$(P)\ \begin{array}{l}\min f(x)\\ s.t.\ x \in C\end{array}$$

where

- $f : \mathbb{R}^m \to (-\infty, \infty]$ is convex, l.s.c., proper.
- $C \neq \emptyset$ convex closed subset of $int(dom(f))$
- $S := \arg\min_{x \in C} f(x) \neq \emptyset$
- $\mu := \min_{x \in C} f(x)$
- $\exists L > 0,\ \sup\|\partial f(C)\| \leqslant L < \infty \iff \forall c \in C, \forall u \in \partial f(c),\ \|u\| \leqslant L$

Projected Subgradient Method

Get $x_0 \in C$. $\forall n \in \mathbb{N}$, note $int(dom(f)) \subseteq dom(\partial f)$, given x_n, pick a stepsize $t_n > 0$ and "$f'(x_n)$" $\in \partial f(x_n)$. Here $f'(x_n)$ means the subgradient of f at x_n.
Update via

$$x_{n+1} := P_C(x_n - t_n f'(x_n))$$

Recall that $C \subseteq int(dom(f))$, hence $\forall n \in \mathbb{N},\ x_n \in int(dom(f))$. Therefore $\partial f(x_n) \neq \emptyset$, and $(x_n)_{n\in\mathbb{N}}$ is well-defined.

Lemma 5.9: L17-1

Let $s \in S = \arg\min_{x\in C} f(x)$ and $f(s) = \mu$. Then

$$\|x_n - s\|^2 \leqslant \|x_n - s\|^2 - 2t_n(f(x_n) - \mu) + t_n^2\|f'(x_n)\|^2$$

Observe that $S \subseteq C$

Proof.

$$\begin{aligned}\|x_{n+1} - s\|^2 &= \|P_c(x_n - t_n f'(x_n)) - \underbrace{P_C(s)}_{s\in C, P_C(s)=s}\|^2\\ &\underbrace{\leqslant}_{P_C f.n.e,\ nonexp} \|x_n - t_n f'(x_n) - s\|^2\\ &= \|(x_n - s) - t_n f'(x_n)\|^2\\ &= \|x_n - s\|^2 + t_n^2\|f'(x_n)\|^2 - 2t_n\langle x_n - s, f'(x_n)\rangle\end{aligned}$$

Recall we want to show

$$\|x_n - s\|^2 + t_n^2\|f'(x_n)\|^2 - 2t_n(f(x_n) - \mu)$$

Done if

$$-2t_n\langle x_n - s, f'(x_n)\rangle \leqslant -2t_n(f(x_n) - \mu)$$

Equivalent lt,

$$\langle x_n - s, f'(x_n)\rangle \geqslant (f(x_n) - \mu)$$

which is true by the subgradient inequality which is

$$\underbrace{f(s)}_{\mu} \geqslant f(x_n) + \langle f'(x_n), s - x_n\rangle$$

$\square$

What is a good stepsize t_n
Let us minimizer the upper bound

$$\begin{aligned}
0 =& \frac{d}{dt_n} RHS \\
=& \frac{d}{dt_n}(-2t_n(f(x_n) - \mu) + t_n^2 \|f'(x_n)\|^2) \\
=& -2(f(x_n - \mu) + 2t_n \|f'(x_n)\|^2
\end{aligned}$$

Assuming $f'(x_n) \neq 0$ (else, $0 \in \partial f(x_n)$ and hence , by Fermat x_n is a global minimizer and we are DONE). Pick

$$t_n = \frac{f(x_n) - \mu}{\|f'(x_n)\|^2}$$

which is known as Polyak's rule.

$$(P)\quad \begin{array}{l}\min f(x)\\ s.t.\ x \in C\end{array}$$

where

- $f : \mathbb{R}^m \to (-\infty, \infty]$ is convex, l.s.c., proper.
- $C \neq \emptyset$ convex closed subset of $int(dom(f))$
- $S := \arg\min_{x\in C} f(x) \neq \emptyset$
- $\mu := \min_{x\in C} f(x)$
- $\exists L > 0,\ \sup\|\partial f(C)\| \leqslant L < \infty \iff \forall c \in C, \forall u \in \partial f(c),\ \|u\| \leqslant L$
- $x_0 \in C$,

  $$x_{n+1} := P_C(x_n - t_n f'(x_n))$$

  Polyak's stepsize

  $$t_n = \frac{f(x_n) - \mu}{\|f'(x_n)\|^2}$$

Theorem 5.10: L17-2

We have

1. $\forall s \in S, \forall n \in \mathbb{N}$, $\|x_{n+1} - s\| \leqslant \|x_n - s\|$. *"$(x_n)_{n\in\mathbb{N}}$ is Fejér monotone w.r.t. S"*
2. $f(x_n) \to \mu$
3. $\mu_n - \mu \leqslant \frac{L d_s(x_0)}{\sqrt{n+1}} = O(\frac{1}{\sqrt{n}})$, *where* $\forall n \in \mathbb{N},\ \mu_n := \min_{0\leqslant k\leqslant n} f(x_k)$
4. *Let* $\varepsilon > 0$. *If* $n \geqslant \frac{L^2 d_s^2(x_0)}{\varepsilon^2} - 1 \implies \mu_n \leqslant \mu + \varepsilon$

Proof. Let $s \in S,\ n \in \mathbb{N}$.

1.

$$\begin{aligned}\|x_{n+1} - s\|^2 &\leqslant \|x_n - s\|^2 - 2t_n(f(x_n) - \mu) + t_n^2\|f'(x_n)\|^2\\ &= \|x_n - s\|^2 - 2\frac{f(x_n) - \mu}{\|f'(x_n)\|^2}(f(x_n) - \mu) + \left(\frac{f(x_n) - \mu}{\|f'(x_n)\|^2}\right)^2 \|f'(x_n)\|^2\\ &= \|x_n - s\|^2 - 2\frac{(f(x_n) - \mu)^2}{\|f'(x_n)\|^2} + \frac{(f(x_n) - \mu)^2}{\|f'(x_n)\|^2}\\ &= \|x_n - s\|^2 - \frac{(f(x_n) - \mu)^2}{\|f'(x_n)\|^2}\\ &\leqslant \|x_n - s\|^2 - \frac{(f(x_n) - \mu)^2}{L^2}\\ &\leqslant \|x_n - s\|^2\end{aligned}$$

Note $\|f'(x_n)\|^2 \leqslant L^2 \implies \frac{1}{\|f'(x_n)\|^2} \leqslant \frac{1}{L^2} \implies -\frac{1}{\|f'(x_n)\|^2} \leqslant -\frac{1}{L^2}$

2. Observe that $\forall k \in \mathbb{N}$,

$$\frac{(f(x_k)-\mu)^2}{L^2} \leqslant \|x_k - s\|^2 - \|x_{k+1} - s\|^2$$

Summing the above inequalities over $k=0$ to $k=n$ yields

$$\frac{1}{L^2}\sum_{k=0}^{n}(f(x_k)-\mu)^2 \leqslant \|x_0 - s\|^2 - \|x_{n+1} - s\|^2 \leqslant \|x_0 - s\|^2 \ldots (*)$$

Letting $n \to \infty$, we learn that

$$0 \leqslant \sum_{k=0}^{\infty}(f(x_k)-\mu)^2 \leqslant L^2\|x_0 - s\|^2 < \infty$$

which implies

$$f(x_k) - \mu \to 0 \iff f(x_k) \to \mu$$

3. Recall $\forall n \in \mathbb{N}$, $\mu_n := \min_{0\leqslant k \leqslant n} f(x_k)$. Let $n \geqslant 0$.
Then $\forall k \in \{0, \ldots, n\}$,

$$\begin{aligned} (\mu_n - \mu)^2 &\leqslant (f(x_k)-\mu)^2 \\ \implies (n+1)\frac{(\mu_n-\mu)^2}{L^2} &\leqslant \frac{1}{L^2}\sum_{k=0}^{n}(f(x_k)-\mu)^2 \\ &\underbrace{\leqslant}_{(*)} \|x_0 - s\|^2 \end{aligned}$$

Minimizing over $s \in S$, we get

$$(n+1)\frac{(\mu_n-\mu)^2}{L^2} \leqslant d_S^2(x_0)$$

4.

$$\begin{aligned} n &\geqslant \frac{L^2 d_S^2(x_0)}{\varepsilon^2} - 1 \\ \iff \frac{d_S(x_0)^2 L^2}{(n+1)} &\leqslant \varepsilon^2 \end{aligned}$$

Then by 3), we have

$$(\mu_n - \mu)^2 \leqslant \frac{d_S^2(x_0)L^2}{n+1} \leqslant \varepsilon^2$$

which implies

$$\mu_n - \mu \leqslant \varepsilon \implies \mu_n \leqslant \mu + \varepsilon$$

$\square$

Recall that: Theorem L13-5

Theorem 5.11: L17-3 Convergence of Projected Subgradient

Suppose that $(x_n)_{n\in\mathbb{N}}$ *is generated as in* (P). *Then*

$$x_n \to \text{ a solution of } (P) \text{ in } S$$

Proof. By the previous theorem, $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone w.r.t. S.
Since $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone w.r.t to S, $(x_n)_{n\in\mathbb{N}}$ is bounded.
Also, by the previous theorem,

$$f(x_n) \to \mu = \min_{x\in C} f(x)$$

By Bolzano-Weirestrass, $\exists x_{k_n} \to \overline{x}$ and $\overline{x} \in C$ (because $(x_n)_{n\in\mathbb{N}}$ lies in C by construction, C is closed).
Now,

$$\mu = \min_{x\in C} f(x) \leqslant f(\overline{x}) \underbrace{\leqslant}_{f \text{ is lsc}} \liminf_{n\to\infty} f(x_{k_n}) = \mu$$

which implies $f(\overline{x}) = \mu$. Hence, $\overline{x} \in S$ That is, all cluster point of $(x_n)_{n\in\mathbb{N}}$ lies in S.
$x_n \to \overline{x} \in S$ by the Fejér monotone theorem. □

Example 5.12: L18-1

Let $C \subseteq \mathbb{R}^m$ be convex closed and nonempty and let $x \in \mathbb{R}^m$. Then

$$\partial d_C(x) = \begin{cases} \frac{x - P_C(x)}{d_C(x)}, & x \notin C \\ N_C(x) \cap B(0;1), & x \in C \end{cases}$$

Consequently, $\forall x \in \mathbb{R}^m$,

$$\sup \|\partial d_C(x)\| \leqslant 1$$

Proof. Omitted. The bound can be easily verified. □

Lemma 5.13: L18-2

Let f be convex, l.s.c, proper and let $\lambda > 0$. Then

$$\partial(\lambda f) = \lambda \partial f$$

Proof. Easy □

5.4 The Convex Feasibility Problem

- Given k closed convex subsets S_i of $\mathbb{R}^m$ such that

$$S = S_1 \cap S_2 \cap \dots S_k \neq \emptyset$$

- Problem: Find $x \in S$
- Can we use the Projected subgradient method for (P)? What is f? What is C? What is L?

Set $C = \mathbb{R}^m$, $P_C = I_d$. Set $f(x) = \max\{d_{S_1}(x), \dots, d_{S_k}(x)\}$, then $f(x) \geqslant 0, \forall x \in \mathbb{R}^m$. And

$$\begin{aligned}
&f(x) = 0 \\
\iff &\forall i \in \{1, \dots, k\},\ d_{S_i}(x) = 0 \\
\iff &\forall i \in \{1, \dots, k\},\ x \in S_i \\
\iff &x \in \bigcap_{i=1}^{k} S_i = S
\end{aligned}$$

$$s \neq \emptyset \implies \mu = \min_{x \in \mathbb{R}^m} f(x) = 0$$

$L = 1$ by the previous example.
Finally, observe that the max formula for subdifferentials implies that $x \notin S$.

$$\begin{aligned}
\partial f(x) &= Conv\{\partial d_{S_i}(x) | d_{S_i}(x) = f(x)\} \\
&= Conv\left\{\frac{x - P_{S_i}(x)}{d_{S_i}(x)} | d_{S_i}(x) = f(x)\right\}
\end{aligned}$$

What do we do with that?
Well, given x_n pick an index i_n such that

$$d_{S_{i_n}}(x_n) = f(x_n)$$

Set

$$f'(x_n) := \frac{x_n - P_{S_{i_n}}(x_n)}{d_{S_{i_n}}}$$

What about t_n?
Polyak's step size:

$$\begin{aligned}
t_n &= \frac{f(x_n) - \mu}{\|f'(x_n)\|^2} \\
&= \frac{d_{S_{i_n}}(x_n) - 0}{\left\|\frac{x_n - P_{S_{i_n}}}{d_{S_{i_n}}(x_n)}\right\|^2} \\
&= \frac{d_{S_{i_n}}(x_n)}{\frac{\|x_n - P_{S_{i_n}}\|^2}{d^2_{S_{i_n}}(x_n)}} \\
&= d_{S_{i_n}}(x_n)
\end{aligned}$$

The update leads to the Greedy Projection Algorithm.

$$\begin{aligned} x_{n+1} &= x_n - t_n f'(x_n) \\ &= x_n - d_{S_{i_n}}(x) \frac{x_n - P_{S_{i_n}}(x_n)}{d_{S_{i_n}}(x_n)} \\ &= x_n - (x_n - P_{S_{i_n}}(x_n)) \\ &= P_{S_{i_n}}(x_n) \end{aligned}$$

so

$$x_{n+1} := P_{S_{i_n}}(x_n)$$

where S_{i_n} is any set that is farthest away from x_n. And, by theorem L17-3,

$$x_n \to \text{some point in } S$$

5.4.1 The Case $k = 2$

We obtain that method of alternating projections "MAP".
$x_0 \in \mathbb{R}^m$. Update via

$$x_{n+1} = P_{S_2} P_{S_1} x_n$$

Example 5.14: L18-3

Find $x \in S$ where

$$S := \{x \in \mathbb{R}^m | Ax = b,\ x \geqslant 0\}$$

- *A is $k \times m$ matrix*
- $b \in \mathbb{R}^k$

We can use "MAP"! Set $S_1 = \mathbb{R}^m_+$,

$$\begin{aligned} P_{S_1}(x) &= x^+ = (\max\{\xi_i, 0\})_{i=1}^m,\ x = (\xi_1, \ldots, \xi_m) \\ S_2 &= \{x \in \mathbb{R}^m | Ax = b\} = A^{-1}(b)(\textit{ the inverse image of } b) \\ P_{S_2}(x) &= x - A^+(Ax - b) \end{aligned}$$

A^+ is the Moore-Penrose pseudo inverse (pinv). Let $x_0 \in \mathbb{R}^m$. Update via

$$\begin{aligned} x_{n+1} =& P_{S_2} P_{S_1}(x_n) \\ =& P_{S_2}(x_n^+) \\ =& x_n^+ - A^+(Ax_n^+ - b) \\ \implies & \overline{x} \in S \end{aligned}$$

Remark. L18-4 In practice, it is possible that $\mu = \min_{x \in C} f(x)$ is NOT known to us. In this case replace Polyak's stepsize by a sequence $(t_n)_{n \in \mathbb{N}}$ such that

$$\frac{\sum_{k=0}^n t_k^2}{\sum_{k=0}^n t_k} \to 0 \text{ as } n \to \infty$$

for example,

$$t_k = \frac{1}{k+1}$$

One can show that

$$\mu_n := \min\{f(x_0), \ldots, f(x_n)\} \to \mu$$

as $n \to \infty$

5.5 The Proximal Gradient Method(PGM)

Consider the problem

$$(P) \ \min_{x\in\mathbb{R}^m} F(x) := f(x) + g(x)$$

Assumptions:
(P) has solutions

$$S := \arg\min_{x\in\mathbb{R}^m} F(x) \neq \emptyset$$

and

$$\mu = \min_{x\in\mathbb{R}^m} F(x)$$

- f is "nice": convex, lsc, proper and differentiable on $int(dom(f)) \neq \emptyset$. ∇f is L-Lipschitz on $int(dom(f))$
- g is convex, lsc and proper.

  $$dom(g) \subseteq int(dom(f))$$

  implies that

  $$\emptyset \neq ri(dom(g)) \subseteq dom(g) \subseteq ri(dom(f))$$

  and implies

  $$ri(dom(g)) \cap ri(dom(f)) = ri(dom(g)) \neq \emptyset$$

Example 5.15: L18-5

$$\min_{x \in C} \; f(x)$$

where is $\emptyset \neq C \subseteq \mathbb{R}^m$ *convex, closed is equivalent to*

$$\min_{x \in \mathbb{R}^m} f(x) + \underbrace{\delta_C(x)}_{:=g}$$

PGM:

$$x \in int(dom(f)) \supseteq dom(g)$$

Update via

$$\begin{aligned} x_+ &= Prox_{\frac{1}{L}g}\left(x - \frac{1}{L}\nabla f(x)\right) \\ &= \arg\min_{y \in \mathbb{R}^m} \left\{ \frac{1}{L} g(y) + \frac{1}{2} \left\| y - \left(x - \frac{1}{L}\nabla f(x)\right) \right\|^2 \right\} \\ &\in dom(g) \subseteq int(dom(f)) = dom(f) \end{aligned}$$

Set

$$T = Prox_{\frac{1}{L}g}\left(I_d - \frac{1}{L}\nabla f\right)$$

i.e., $\forall x \in \mathbb{R}^m$

$$Tx = Prox_{\frac{1}{L}g}\left(x - \frac{1}{L}\nabla f(x)\right)$$

Theorem 5.16: L18-6

Let $x \in \mathbb{R}^m$*. Then*

$$\begin{gathered} x \in S = \arg\min\nolimits_{x \in \mathbb{R}^m} F = \arg\min\nolimits_{x \in \mathbb{R}^m} (f + g) \\ \iff \\ x = Tx \text{ (i.e., } x \in Fix(T)) \end{gathered}$$

Proof. Observe that by Fermat,

$$\begin{aligned} x \in S \iff 0 &\in \partial(f+g)(x) \\ &= \partial f(x) + \partial g(x) \\ &= \nabla f(x) + \partial g(x) \end{aligned}$$

Let $x \in \mathbb{R}^m$. Then $x \in S$

$$\begin{aligned}
&\iff 0 \in \partial(f+g)(x) \\
&\iff 0 \in \nabla f(x) + \partial g(x) \\
&\iff -\nabla f(x) \in \partial g(x) \\
&\iff -\frac{1}{L}\nabla f(x) \in \frac{1}{L}\partial g(x) \\
&\iff x - \frac{1}{L}\nabla f(x) \in x + \partial\left(\frac{1}{L}g\right)(x) = \left(I_d + \partial\left(\frac{1}{L}g\right)\right)(x) \\
&\iff x \in \left(I_d + \partial\left(\frac{1}{L}g\right)\right)^{-1}\left(x - \frac{1}{L}\nabla f(x)\right) \\
&\iff x = Prox_{\frac{1}{L}g}\left(I_d - \frac{1}{L}\nabla f\right)(x) = Tx
\end{aligned}$$

$\square$

Fact L18-7

Let $f : \mathbb{R}^m \to (-\infty, \infty]$ be convex lsc and proper and let $\beta > 0$. Then f is β-strongly convex $\iff$ $\forall x \in dom(\partial f)$, $\forall v \in \partial f(x)$,

$$f(y) \geqslant f(x) + \langle v, y - x\rangle + \frac{\beta}{2}\|y - x\|^2$$

5.6 The Prox-Grad Inequality

Proposition 5.17: L18-8

Let $x \in \mathbb{R}^m$, $y \in int(dom(f))$,

$$y_+ = Ty = Prox_{\frac{1}{L}g}(y - \nabla f(y))$$

Then

$$F(x) - F(y_+) \geqslant \frac{L}{2}\|x - y_+\|^2 - \frac{L}{2}\|x - y\|^2 + D_f(x, y)$$

where

$$D_f(x, y) := f(x) - f(y) - \langle \nabla f(y), x - y \rangle \geqslant 0$$

which is the "Bregman distance" by convexity of f.

Proof. Define

$$h(z) := f(y) + \langle \nabla f(y), z - y \rangle + g(z) + \frac{L}{2}\|z - y\|^2$$

Then h is L-strongly convex.
Let $z \in \mathbb{R}^m$. Then

$$z \text{ minimizes } h$$

$$\begin{aligned}
\iff \ 0 &\in \partial\left(f(y) + \langle \nabla f(y), z - y \rangle + g(z) + \frac{L}{2}\|z - y\|^2\right) \\
&= \partial\left(\langle \nabla f(y), z - y \rangle + g(z) + \frac{L}{2}\|z - y\|^2\right) \\
&= \nabla f(y) + \partial g(z) + L(z - y) \\
\iff \ 0 &\in \frac{1}{L}\nabla f(y) + \partial\left(\frac{1}{L}g\right)(z) + (z - y) \\
\iff \ y - \frac{1}{L}\nabla f(y) &\in z + \partial\left(\frac{1}{L}g\right)(z) \\
&= \left(I_d + \partial\left(\frac{1}{L}g\right)\right)(z) \\
\iff \ z &= \left(I_d + \partial\left(\frac{1}{L}g\right)\right)^{-1}\left(y - \frac{1}{L}\nabla f(y)\right) \\
&= Prox_{\frac{1}{L}g}\left(y - \frac{1}{L}\nabla f(y)\right) \\
&= Ty =: y_+
\end{aligned}$$

which implies

$$\arg\min h =: \{y_+\}$$

Recalling L18-7, $f \to h$, $\beta \to L$, $y \to x$, $x \to y_+$, then

$$h(x) - h(y_+) \geqslant \frac{L}{2}\|x - y_+\|^2 \ldots (1)$$

Moreover, by the descent lemma, we have

$$f(y_+) \leqslant f(y) + \langle \nabla f(y), y_+ - y \rangle + \frac{L}{2}\|y_+ - y\|^2$$

Therefore,

$$\begin{aligned} h(y_+) &= f(y) + \langle \nabla f(y), y_+ - y \rangle + g(y_+) + \frac{L}{2}\|y_+ - y\|^2 \\ &\geqslant f(y_+) + g(y_+) \\ &= F(y_+) \end{aligned}$$

Combining with (1),

$$h(x) - F(y_+) \geqslant h(x) - h(y_+) \geqslant \frac{L}{2}\|x - y_+\|^2$$

Using the definition of h, the inequality above becomes

$$f(y) + \langle \nabla f(y), x - y \rangle + g(x) + \frac{L}{2}\|x - y\|^2 - F(y_+) \geqslant \frac{L}{2}\|x - y_+\|^2$$

Adding $f(x)$ to both sides and rearranging yields:

$$f(x) + g(x) - F(y_+) \geqslant \frac{L}{2}\|x - y_+\|^2 - \frac{L}{2}\|x - y\|^2 + \underbrace{f(x) - f(y) + \langle \nabla f(y), x - y \rangle}_{D_f(x,y)}$$

□

The Proximal Gradient Method:
The problem is here:

$$(P) \min_{x\in\mathbb{R}^m} F(x) := f(x) + g(x)$$

Assumptions:
(P) has solutions

$$S := \arg\min_{x\in\mathbb{R}^m} F(x) \neq \emptyset$$

and

$$\mu = \min_{x\in\mathbb{R}^m} F(x)$$

- f is "nice": convex, lsc, proper and differentiable on $int(dom(f)) \neq \emptyset$. ∇f is L-Lipschitz on $int(dom(f))$
- g is convex, lsc and proper.

  $$dom(g) \subseteq int(dom(f))$$

  implies that

  $$\emptyset \neq ri(dom(g)) \subseteq dom(g) \subseteq ri(dom(f))$$

  and implies

  $$ri(dom(g)) \cap ri(dom(f)) = ri(dom(g)) \neq \emptyset$$

Lemma 5.18: L19-1 (Sufficient Decrease Lemma

$$F(y_+) \leqslant F(y) - \frac{L}{2}\|y - y_+\|^2$$

Proof. Use L18-8 with x replaced by y and recall that, because f is convex,

$$D_f(x,y) = f(x) - f(y) - \langle \nabla f(y), x - y\rangle \geqslant 0$$

□

The Proximal Gradient Method:
Given $y \in int(dom(f))$, update via

$$\begin{aligned} y_+ &:= Prox_{\frac{1}{L}g}\left(y - \frac{1}{L}\nabla f(y)\right) \\ &=: Ty \in dom(g) \subseteq int(dom(f)) = dom(\nabla f) \end{aligned}$$

The Algorithm:
Given $x_0 \in int(dom(f))$. $\forall n \in \mathbb{N}$, update via

$$x_{n+1} := Tx_n = Prox_{\frac{1}{L}g}\left(x_n - \frac{1}{L}\nabla f(x_n)\right)$$

Theorem 5.19: L19-2 $O(1/n)$ rate of convergence of function values

The following hold:

1. $\forall s \in S,\ n \in \mathbb{N},\ \|x_{n+1} - s\| \leqslant \|x_n - s\|$, *i.e.,* $(x_n)_{n\in\mathbb{N}}$ *is Fejér monotone w.r.t* S.
2. $(F(x_n))_{n\in\mathbb{N}}$ *decreases to* μ*. more precisely,*

$$0 \leqslant F(x_n) - \mu \leqslant \frac{L \cdot d_S^2(x_0)}{2n} = O\left(\frac{1}{n}\right)$$

Proof. Applying L19-1 with y replaced by x_n ($y_+ = x_{n+1}$) yields

$$F(x_{n+1}) \leqslant F(x_n) - \frac{L}{2}\|x_{n+1} - x_n\|^2 \leqslant F(x_n)$$

1. Recalling: Let $s \in S$, let $k \in \mathbb{N}$.
   Applying L18-8 with (x, y) replaced by (s, x_k) yields

   $$0 \geqslant \underbrace{F(s)}_{\mu} - F(x_{k+1}) \geqslant \frac{L}{2}\|s - x_{k+1}\|^2 - \frac{L}{2}\|s - x_k\|^2 \ldots (*)$$

   implies that

   $$(x_n)_{n\in\mathbb{N}} \text{ is Fejér monotone w.r.t. } S$$

2. Multiplying $(*)$ by $\frac{2}{L}$ and adding the resulting inequalities from $k = 0$ to $k = n - 1$ and telescoping yields:

   $$\frac{2}{L}\left(\sum_{k=0}^{n-1}(\mu - F(x_{k+1}))\right) \geqslant \|s - x_n\|^2 - \|s - x_0\|^2 \geqslant -\|s - x_0\|^2$$

   In particular, setting $s = P_S(x_0) \in S$, we obtain

   $$\begin{aligned} d_S^2(x_0) &= \|P_S(x_0) - x_0\|^2 \\ &\geqslant \frac{2}{L}\left(\sum_{k=0}^{n-1}(F(x_{k+1}) - \mu)\right) \\ &\geqslant \frac{2}{L}\left(\sum_{k=0}^{n-1}(F(x_n) - \mu)\right) \text{ by } F(x_{n+1}) \leqslant F(x_n) \\ &= \frac{2}{L}n(F(x_n) - \mu) \end{aligned}$$

   Equivalently,

   $$0 \leqslant F(x_n) - \mu \leqslant \frac{L \cdot d_S^2(x_0)}{2n}$$

   and

   $$F(x_n) \to \mu$$

□

Theorem 5.20: L19-3 Convergence of PGM

x_n converges to some solution in $S = \arg\min_{x\in\mathbb{R}^m} F(x)$

Proof. By the previous theorem we have $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone w.r.t S. Done if we can show that every cluster point of $(x_n)_{n\in\mathbb{N}}$ lies in S.
Suppose that $\overline{x}$ is a cluster point of $(x_n)_{n\in\mathbb{N}}$, say $x_{k_n} \to \overline{x}$.
Indeed,

$$\begin{aligned} &\mu \leqslant F(\overline{x}) \leqslant \liminf_{n\to\infty} F(x_{k_n}) = \mu \\ \implies &F(\overline{x}) = \mu \\ \iff &\overline{x} \in S \end{aligned}$$

□

Proposition 5.21: L19-4

The following hold:

1. $\frac{1}{L}\nabla f$ *is f.n.e.*
2. $I_d - \frac{1}{L}\nabla f$ *is f.n.e*
3. $T = Prox_{\frac{1}{L}g}(I_d - \nabla f)$ *is* $2/3$*-averaged*

Proof.

1. Recall L9-8 4). Dividing both sides by $\frac{1}{L}$ yields

$$\left\langle \underbrace{\frac{1}{L}\nabla f(x)}_{T} - \underbrace{\frac{1}{L}\nabla f(y)}_{T}, x - y \right\rangle \geqslant \left\| \underbrace{\frac{1}{L}\nabla f(x)}_{T} - \underbrace{\frac{1}{L}\nabla f(y)}_{T} \right\|^2$$

   There fore 1) and 2) follows from A3 Problem 3 a),b),d)

Problem 3.

Let $T : \mathbb{R}^m \to \mathbb{R}^m$.

(i) Prove that the following are equivalent:

(a) T is firmly nonexpansive.

(b) $\mathrm{Id} - T$ is firmly nonexpansive.

(c) $2T - \mathrm{Id}$ is nonexpansive.

(d) $(\forall x \in \mathbb{R}^m)\ (\forall y \in \mathbb{R}^m)\ \|T(x) - T(y)\|^2 \leq \langle x - y, T(x) - T(y)\rangle$.

(e) $(\forall x \in \mathbb{R}^m)\ (\forall y \in \mathbb{R}^m)\ \langle (\mathrm{Id} - T)(x) - (\mathrm{Id} - T)(y), T(x) - T(y)\rangle \geq 0$.

2. as above

3. Recall that $Prox_{\frac{1}{L}g}$ is f.n.e. Hence, $Prox_{\frac{1}{L}g}$ and $I_d - \frac{1}{L}\nabla f$ are both $\frac{1}{2}$-average. Consequently, the composition $Prox_{\frac{1}{L}g}\left(I_d - \frac{1}{L}\nabla f\right)$ is averaged with constant $2/3$

$\square$

Remark. L19-5 Recall L14-4 1), 3). One can show that for $T = Prox_{\frac{1}{L}g}\left(I_d - \frac{1}{L}g\right)$ we have $\forall x, y$,

$$\frac{1}{2}\|(I_d - t)x - (I_d - T)y\|^2 \leqslant \|x - y\|^2 - \|Tx - Ty\|^2$$

Theorem 5.22: L19-6

Recalling the PGM iteration we have.

$$\|x_{n+1} - x_n\| \leqslant \frac{\sqrt{2} d_S(x_0)}{\sqrt{n}} = O\left(\frac{1}{\sqrt{n}}\right)$$

Proof. Using the previous remark we have, $\forall x, y$

$$\frac{1}{2}\|(I_d - T)x - (I_d - T)y\|^2 < \|x - y\|^2 - \|Tx - Ty\|^2 \ldots (*)$$

Let $s \in S$ and observe that $s = Ts$ by L18-6.
Applying $(*)$ with $x = x_k$, $y = s \in S$, we get

$$\frac{1}{2}\|(I_d - T)x_k - \underbrace{(I_d - T)s}_{0}\|^2 < \|x_k - s\|^2 - \|\underbrace{Tx_k}_{x_{k+1}} - \underbrace{Ts}_{s}\|^2$$

That is

$$\frac{1}{2}\|x_k - x_{k+1}\|^2 \leqslant \|x_k - s\|^2 - \|x_{k+1} - s\|^2 \ldots (\#)$$

Using the previous proposition T is $2/3$-averaged, hence T is nonexpansive. Therefore

$$\begin{aligned}\|\underbrace{x_k}_{Tx_{k-1}} - \underbrace{x_{k+1}}_{Tx_k}\| &\leqslant \|x_{k-1} - x_k\| \\ &\leqslant \dots \\ &\leqslant \|x_0 - x_1\|\end{aligned}$$

Summing $(\#)$ oVer $k = 0 \to n-1$,

$$\|x_0 - s\|^2 - \|x_n - s\|^2 \geqslant \frac{1}{2}\sum_{k=0}^{n-1}\|x_k - x_{k+1}\|^2 \geqslant \frac{1}{2}n\|x_{n-1} - x_n\|^2$$

In particular, for $s = P_S(x_0)$, we get

$$\begin{aligned}&\frac{1}{2}n\|x_{n-1} - x_n\|^2 \leqslant d_S^2(x_0) \\ \implies &\|x_{n-1} - x_n\| \leqslant \frac{\sqrt{2}}{\sqrt{n}}d_S(x_0) = O\left(\frac{1}{\sqrt{n}}\right)\end{aligned}$$

□

Corollary 5.23: L19-7 The Classical Proximal Point Algorithm 1970's Rockafeller

$g : \mathbb{R}^m \to (-\infty, \infty]$ convex lsc and proper, $c > 0$.

$$(P) \quad \min_{x \in \mathbb{R}^m} g(x)$$

Assume that $S := \arg\min_{x\in\mathbb{R}^m} g(x) \neq \emptyset$.
Let $x_0 \in \mathbb{R}^m$. Update via

$$x_{n+1} = Prox_{cg} x_n$$

Then

$$\begin{aligned}g(x_n) \searrow \mu &= \min g(\mathbb{R}^m) \\ 0 \leqslant g(x_n) - \mu &\leqslant \frac{d_S^2(x_0)}{2cn} \\ x_n &\to \text{some point in } S \\ \|x_{n-1} - x_n\| &\leqslant \frac{\sqrt{2}d_S(x_0)}{\sqrt{n}}\end{aligned}$$

Proof. Set $\forall x \in \mathbb{R}^m$, $f(x) = 0$. Then $\forall x \in \mathbb{R}^m$, $\nabla f(x) = 0$

$$\implies \nabla f \equiv 0 \text{ is L-Lipschitz}$$

for any $L > 0$. In particular, for $L = \frac{1}{c} > 0$.
Observe that (P) can be written as

$$\begin{aligned}
&\min_{x \in \mathbb{R}^m} \underbrace{f(x) + g(x)}_{F(x)=g(x)} \\
\implies & S = \arg\min_{x \in \mathbb{R}^m} F(x) \\
&= \arg\min_{x \in \mathbb{R}^m} g(x)
\end{aligned}$$

$$\begin{aligned}
\nabla f \equiv 0 &\implies I_d - \frac{1}{L}\nabla f = I_d \\
\implies T &= Prox_{\frac{1}{L}g}\left(I_d - \frac{1}{L}\nabla f\right) \\
&= Prox_{cg} \circ (I_d) \\
&= Prox_{cg}
\end{aligned}$$

Done by the previous results. $\square$

5.7 Fast Iterative Shrinkage Thresholding Algorithm(FISTA)

$$(P)\ \min_{x\in\mathbb{R}^m} F(x) := f(x) + g(x)$$

Assumptions:
(P) has solutions

$$S := \arg\min_{x\in\mathbb{R}^m} F(x) \neq \emptyset$$

and

$$\mu = \min_{x\in\mathbb{R}^m} F(x)$$

- f is "nice": convex, lsc, proper and differentiable on $\mathbb{R}^m$. ∇f is L-Lipschitz on $\mathbb{R}^m$
- g is convex, lsc and proper.

FISTA:
$x_0 \in \mathbb{R}^m,\ t_0 = 1,\ y_0 = x_0$. Update via

$$\begin{aligned}
t_{n+1} =& \frac{1+\sqrt{1+4t_n^2}}{2} \\
\implies\ 2t_{n+1} - 1 =& \sqrt{1+4t_n^2} \\
\implies\ t_{n+1}^2 - t_{n+1} =& t_n^2 \\
x_{n+1} =& \mathrm{Prox}_{\frac{1}{L}g}(\left(\mathrm{I_d} - \frac{1}{L}\nabla \mathrm{f}\right)(\mathrm{y_n}) = \mathrm{Ty_n} \\
y_{n+1} =& x_{n+1} + \frac{t_n - 1}{t_{n+1}}(x_{n+1} - x_n) \\
=& \left(1 - \frac{1-t_n}{t_{n+1}}\right)x_{n+1} + \frac{1-t_n}{t_{n+1}}x_n \\
\in& \mathrm{aff}\{\mathrm{x_n}, \mathrm{x_{n+1}}\}
\end{aligned}$$

Remark. L20-1
The sequence $(t_n)_{n\in\mathbb{N}}$ satisfies $\forall n \in \mathbb{N},\ t_n \geqslant \frac{n+2}{2} \geqslant 1$. Verify using induction!
Indeed, base case:

$$t_0 = 1 = \frac{0+2}{2}$$

Now suppose for some $n \geqslant 0$,

$$t_n \geqslant \frac{n+2}{2}$$

Now

$$\begin{aligned}
t_{n+1} &= \frac{1+\sqrt{1+4t_n^2}}{2} \\
&\geqslant \frac{1+\sqrt{1+4\frac{(n+2)^2}{4}}}{2} \\
&= \frac{1+\sqrt{1+(n+2)^2}}{2} \\
&\geqslant \frac{1+\sqrt{(n+2)^2}}{2} \\
&= \frac{1+n+2}{2} \\
&= \frac{(n+1)+2}{2}
\end{aligned}$$

and the conclusion follows.

Theorem 5.24: L20-2 ($O(1/n^2)$) convergence for FISTA

$$0 \leqslant F(x_n) - \mu \leqslant \frac{2Ld_S^2(x_0)}{(n+1)^2} = O(1/n^2)$$

Proof. Set $s = P_S(x_0)$
By convexity of F (note $t_n \geqslant (n+2)/2 \geqslant 1$), we have

$$F\left(\frac{1}{t_n}S + \left(1 - \frac{1}{t_n}\right)x_n\right) \leqslant \frac{1}{t_n}F(s) + \left(1 - \frac{1}{t_n}\right)F(x_n)$$

Set $\forall n \in \mathbb{N}$,

$$\delta_n = F(x_n) - \mu \geqslant 0$$

Observe that,

$$\begin{aligned}
\left(1 - \frac{1}{t_n}\right)\delta_n - \delta_{n+1} &= \left(1 - \frac{1}{t_n}\right)(F(x_n) - \underbrace{F(s)}_{\mu}) - (F(x_{n+1}) - F(s)) \\
&= \left(1 - \frac{1}{t_n}\right)F(x_n) - \left(1 - \frac{1}{t_n}\right)F(s) - F(x_{n+1}) + F(s) \\
&= \left(1 - \frac{1}{t_n}\right)F(x_n) + \frac{1}{t_n}F(s) - F(x_{n+1}) \\
&\geqslant F\left(\frac{1}{t_n}s + \left(1 - \frac{1}{t_n}\right)x_n\right) - F(x_{n+1}) \ldots (1)
\end{aligned}$$

Recall the FISTA updates, applying L18-8 with $x = \frac{1}{t_n}s + (1 - 1/t_n)x_n,\ y = y_n$, implies

$$y_+ = T_{y_n} = x_{n+1}$$

yields

$$
\begin{aligned}
& F\left(\frac{1}{t_n}s + (1-1/t_n)x_n\right) - F(x_{n+1}) \\
& \geqslant \frac{L}{2}\left\|\frac{1}{t_n}s + (1-1/t_n)x_n - x_{n+1}\right\|^2 - \frac{L}{2}\left\|\frac{1}{t_n}s + (1-1/t_n)x_n - y_n\right\|^2 \\
& \geqslant \frac{L}{2}\left\|\frac{1}{t_n}(s + (t_n-1)x_n - t_n x_{n+1})\right\|^2 - \frac{L}{2}\left\|\frac{1}{t_n}(s + (t_n-1)x_n - t_n y_n)\right\|^2 \\
& = \frac{L}{2t_n^2}\|t_n x_{n+1} - (s + (t_n-1)x_n)\|^2 - \frac{L}{2t_n^2}\|t_n y_n - (s + (t_n-1)x_n)\|^2 \ldots (2)
\end{aligned}
$$

and

$$
\begin{aligned}
& \|t_n y_n - (s + (t_n-1)x_n)\|^2 \\
= & \left\|t_n\left(x_n + \frac{t_{n-1}-1}{t_n}(x_n - x_{n-1})\right) - (s + (t_n-1)x_n)\right\|^2 \\
= & \|t_n x_n + (t_{n-1}-1)(x_n - x_{n-1}) - s - t_n x_n + x_n\|^2 \\
= & \|t_{n-1}x_n - t_{n-1}x_{n-1} + x_{n-1} - s\|^2 \\
= & \|t_{n-1}x_n - (s + (t_{n-1}-1)x_{n-1})\|^2 \ldots (3)
\end{aligned}
$$

Then using $t_{n+1}^2 - t_{n+1} = t_n^2$, we have

$$
\begin{aligned}
t_{n-1}^2\delta_n - t_n^2\delta_{n+1} &= (t_n^2 - t_n)\delta_n - t_n^2\delta_{n+1} \\
&= t_n^2\left(\left(1 - \frac{1}{t_n}\right)\delta_n - \delta_{n+1}\right) \\
&\underbrace{\geqslant}_{(1)} t_n^2\left(F\left(\frac{1}{t_n}s + \left(1 - \frac{1}{t_n}\right)x_n\right) - F(x_{n+1})\right) \\
&\underbrace{\geqslant}_{(2)} \frac{L}{2}\|t_n x_{n+1} - (s + (t_n-1)x_n)\|^2 - \frac{L}{2}\|t_n y_n - (s + (t_n-1)x_n)\|^2 \\
&\underbrace{=}_{(3)} \frac{L}{2}\|\underbrace{t_n x_{n+1} - (s + (t_n-1)x_n)}_{u_{n+1}}\|^2 - \frac{L}{2}\|\underbrace{t_{n-1}x_n - (s + (t_{n-1}-1)x_{n-1})}_{u_n}\|^2
\end{aligned}
$$

Multiplying by $\frac{2}{L}$ and rearranging yield

$$
\|u_{n+1}\|^2 + \frac{2}{L}t_n^2\delta_{n+1} \leqslant \|u_n\|^2 + \frac{2}{L}t_{n-1}^2\delta_n
$$

Therefore,

$$\begin{aligned}
\frac{2}{L}t_{n-1}^2\delta_n &\leqslant \|u_n\|^2 + \frac{2}{L}t_{n-1}^2\delta_n \\
&\leqslant \dots \\
&\leqslant \|u_1\|^2 + \frac{2}{L}t_0^2\delta_1 \\
&= \|\underbrace{t_0}_{=1} x_1 - (s + (t_0 - 1)x_0)\|^2 + \frac{2}{L}(1)(F(x_1) - \mu) \\
&= \|x_1 - s\|^2 + \frac{2}{L}(F(x_1) - \mu) \\
&\leqslant \|x_0 - s\|^2
\end{aligned}$$

where the last inequality follows from applying L18-8 with $x = s,\ y = y_0$, $y_+ = Ty_0 = x_1$ to obtain

$$\underbrace{F(s)}_{\mu} - F(x_1) \geqslant \frac{L}{2}\|s - x_1\|^2 - \frac{L}{2}\|x_0 - s\|^2$$

That is,

$$\begin{aligned}
F(x_n) - \mu = \delta_n& \\
&\leqslant \frac{L}{2}\|x_0 - s\|^2 \frac{1}{t_{n-1}^2} \\
&\leqslant \frac{L}{2}\|x_0 - s\|^2 \frac{4}{(n+1)^2} \text{ by } t_n \geqslant \frac{n+2}{2} \\
&= \frac{2Ld_S^2(x_0)}{(n+1)^2} \text{ recall } s = P_S(x_0)
\end{aligned}$$

□

5.7.1 The Iterative Shrinkage Thresholding Algorithm (ISTA)

Special case of the PGM with

$$\begin{aligned}
&g(x) = \lambda\|x\|, \lambda > 0 \\
\Longrightarrow\ &\frac{1}{L}g(x) = \frac{\lambda}{L}\|x\|_1
\end{aligned}$$

$$\begin{aligned}
Prox_{\frac{1}{L}g}(x) &= \left(Prox_{\frac{\lambda}{L}\|\cdot\|}(x)\right)_{i=1}^n \\
&= \left(sign(x_i)\max\left\{0, |x_i| - \frac{\lambda}{L}\right\}\right)_{i=1}^n
\end{aligned}$$

5.8 The Fast Iterative Thresholding Algorithm (FISTA)

Is the accelerated version of ISTA?
$\|\cdot\|$ VS $\|\cdot\|_2$ Consider the two problems

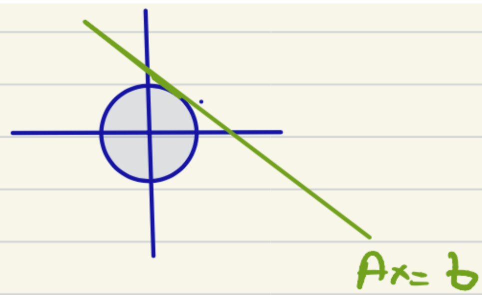
$$(P_1)\ \min \|x\|_2 \ s.t.\ Ax = b$$

and

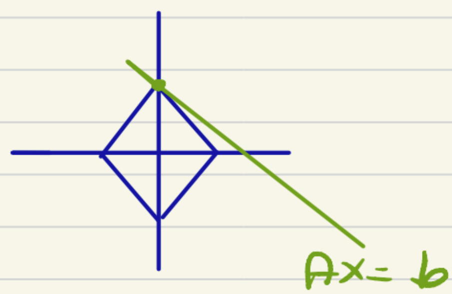
$$(P_2)\ \min \|x\|_1 \ s.t.\ Ax = b$$

$Ax = b$ is underdetermined system of equations.

solution in $\mathbb{R}^2_{++}$ P_1



solution in $\{0\} \times \mathbb{R}$ P_2



Example 5.25: L20-3

l_1 regularized least squares. Consider the problem

$$(P)\ \min_{x\in\mathbb{R}^m} \frac{1}{2}\|Ax-n\|_2^2+\lambda\|x\|_1$$

$\lambda > 0$, A is $n\times m$ matrix.

- *$g(x) = \lambda\|x\|_1$ convex, lsc, proper*
- *$f(x) = \frac{1}{2}\|Ax-b\|_2^2$ smooth, $\forall x\in\mathbb{R}^m$, $\nabla f(x) = A^T(Ax-b)$*
- *$dom(f) = dom(g) = \mathbb{R}^m$. Is ∇f Lipschitz? Recall L10-6,*

$$\begin{aligned} &\nabla f \text{ is L-Lipschitz continuous} \\ \iff &\lambda_{\max}(\nabla^2 f(x)) \leqslant L \\ \iff &\lambda_{\max}(A^TA) \leqslant L \end{aligned}$$

Take $L := \lambda_{\max}(A^TA)$

- *$S \neq \emptyset$. Indeed, $F(x) = f(x)+g(x) = \frac{1}{2}\|Ax-b\|_2^2+\lambda\|x\|_1$ is continuous, convex, coercive, $dom(F) = \mathbb{R}^m$, implies $S = \arg\min F \neq \emptyset$ (Here we used the fact that: F is convex lsc proper + coercive. C convex closed $\neq\emptyset$, $dom(F)\cap C\neq\emptyset$ Then F has a minimizer over C)*

Computational Tip Somtimes m is large and computing the eigenvalues of A^TA ($m\times m$ matrix) is not so easy.
In this case, you could use an upper bound on eigenvalues, e.g., the Frobenius norm:

$$\begin{aligned} \|A\|_F^2 &= \sum_{j=1}^m\sum_{i=1}^n a_{ij}^2 \\ &= tr(A^TA) \\ &= \sum_{i=1}^m \lambda_i(A^TA) \end{aligned}$$

Consider the problem

$$(P)\ \mathrm{minimize}_{\mathrm{x}\in\mathbb{R}^{\mathrm{m}}}\ \mathrm{F(x) = f(x) + g(x)}$$

- f is convex lsc and proper
- g is convex lsc and proper
- $S = \arg\min_{x\in\mathbb{R}^m} F(x) \neq \emptyset$

No further assumptions of smoothness or domain inclusions.
Suppose that $\exists s \in S$ such that $0 \in \partial f(s) + \partial g(s) \subseteq \partial(f+g)(s)$
One situation is when

$$\mathrm{ri(dom(f))} \cap \mathrm{ri(dom(g))} \neq \emptyset$$

then sum rules applies, i.e.

$$\partial(f+g) = \partial f + \partial g$$

Recall that in A4 you proved that

$$\mathrm{Prox_f} = (I_d + \partial f)^{-1}$$
$$\mathrm{Prox_g} = (I_d + \partial g)^{-1}$$

Set

$$R_f := 2\mathrm{Prox_f} - \mathrm{I_d}$$
$$Rg := 2\mathrm{Prox_g} - \mathrm{I_d}$$

Define the Douglas-Rachford (DR) operator as follows:

$$\begin{aligned} T &= I_d - \mathrm{Prox_f} + \mathrm{Prox_g}(2\mathrm{Prox_f} - \mathrm{I_d}) \\ &= I_d - \mathrm{Prox_f} + \mathrm{Prox_g R_f} \end{aligned}$$

Lemma 5.26: L22-1

The following hold:

1. R_f *and* R_g *are nonexpansive*
2. $T = \frac{1}{2}(I_d + R_g R_f)$
3. T *is firmly nonexpansive*

Proof.

1. Recall that $\mathrm{Prox_f}$ is f.n.e by L14-2
   Now combine with A3, T is f.n.e $\iff$ $2T - I_d$ is nonexpansive.

2. Indeed,

$$
\begin{aligned}
&\frac{1}{2}(I_d + R_g R_f)\\
=&\frac{1}{2}(I_d + (2\mathrm{Prox_g} - \mathrm{I_d})\mathrm{R_f})\\
=&\frac{1}{2}(I_d + 2\mathrm{Prox_g R_f} - \mathrm{R_f})\\
=&\frac{1}{2}(I_d + 2\mathrm{Prox_g R_f} - (2\mathrm{Prox_f} - \mathrm{I_d}))\\
=&\frac{1}{2}(I_d + 2\mathrm{Prox_g R_f} - 2\mathrm{Prox_f} + \mathrm{I_d})\\
=&\frac{1}{2}(2I_d - 2\mathrm{Prox_f} + 2\mathrm{Prox_g R_f})\\
=&I_d - \mathrm{Prox_f} + \mathrm{Prox_g R_f}\\
=:&T
\end{aligned}
$$

3. Observe that $R_g R_f (= R_g \circ R_f)$ is a composition of two nonexpansive mappings, hence $R_g R_f$ is nonexpansive.
   Therefore,

$$
\begin{aligned}
T &= \frac{1}{2}(I_d + R_g R_f)\\
&= \frac{1}{2}I_d + \frac{1}{2}\underbrace{R_g R_f}_{=:N}
\end{aligned}
$$

   That is T is $1/2$-averaged, equivalently, T is f.n.e by L12-7

□

Useful if we plan to iterate T. Shall we?

Remark. L22-2

$$\mathrm{FixT} = \mathrm{FixR_g R_f}$$

Indeed, let $x \in \mathbb{R}^m$. Then

$$
\begin{aligned}
x \in \mathrm{FixT} &\iff x = Tx\\
&\iff x = \frac{1}{2}(I_d + R_g R_f)(x) = \frac{1}{2}(x + R_g R_f x)\\
&\iff 2x = x + R_g R_f x\\
&\iff x = R_g R_f x\\
&\iff x \in \mathrm{FixR_g R_f}
\end{aligned}
$$

Proposition 5.27: L22-3

$$\mathrm{Prox_f(FixT) \subseteq S}$$

Proof. Let $x \in \mathbb{R}^m$, and set $s = \mathrm{Prox_f x}$. On the one hand.

$$
\begin{aligned}
& s = \mathrm{Prox_f(x)}\ (\mathrm{Prox_f} = (\mathrm{I_d} + \partial \mathrm{f})^{-1} \\
\iff & x \in (I_d + \partial f)(s) = s + \partial f(s) \\
\iff & 2 \underbrace{\mathrm{Prox_f x}}_{=s} - (2 \underbrace{\mathrm{Prox_f x} - x)}_{R_f x} \in s + \partial f(s) \\
\iff & 2s - R_f(x) \in s + \partial f(s) \\
\iff & 2s - R_f(x) - s \in \partial f(s) \\
\iff & s - R_f(x) \in \partial f(s)
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
x \in \mathrm{Fix(T)} \iff & x = Tx,\ (T = I_d - \mathrm{Prox_f} + \mathrm{Prox_g R_f}) \\
\iff & x = x - \mathrm{Prox_f(x)} + \mathrm{Prox_g R_f(x)} \\
\iff & \mathrm{Prox_f(x)} = \mathrm{Prox_g R_f(x)} \\
\iff & s = \mathrm{Prox_g R_f(x)} \\
\iff & R_f(x) \in s + \partial g(s),\ (\mathrm{Prox_g} = (\mathrm{I_d} + \partial \mathrm{g})^{-1}) \\
\iff & 0 \in s - R_f(x) + \partial g(s) \\
\iff & R_f(x) - s \in \partial g(s)
\end{aligned}
$$

Altogether, the last inclusions imply that

$$
\begin{aligned}
0 \in & \partial f(s) + \partial g(s) \\
\subseteq & \partial (f+g)(s) \\
\implies s \in & S = \arg\min_{x \in \mathbb{R}^m} F(x)
\end{aligned}
$$

□

Theorem 5.28: L22-4

Let $x_0 \in \mathbb{R}^m$. Update via

$$x_{n+1} := x_n - \mathrm{Prox_f(x_n)} + \mathrm{Prox_g(2Prox_f(x_n) - x_n)}$$

Then

$$\mathrm{Prox_f(x_n)} \longrightarrow \textit{ a minimizer of } \mathrm{f + g}$$

Proof. Rewrite x_{n+1} as

$$
\begin{aligned}
x_{n+1} &= (I_d - \mathrm{Prox_f} + \mathrm{Prox_g(2Prox_f - I_d))x_n} \\
&= Tx_n \\
&= T^{n+1} x_0
\end{aligned}
$$

Then by L14-1 $x_{n+1} \to \overline{x} \in \mathrm{FixT}$. Observe that $\mathrm{Prox_f}$ is (firmly) nonexpansive by L14-2, hence continuous by L12-11. Consequently, $\mathrm{Prox_f x_n}$ will converge to $\mathrm{Prox_f\overline{x}} =: \mathrm{s}$
Finally, observe that

$$s \in \mathrm{Prox_f(FixT)} \underbrace{\subseteq}_{\text{Prop L22}-3} \mathrm{S}$$

The proof is complete. □

Consider the problem

$$(P) \text{ minimizer f(x) s.t x} \in \text{C}$$

Assumptions:

- f is convex lsc and proper
- $\emptyset \neq C$ closed and convex $\subseteq \text{int}(\text{dom}(\text{f}))$
- $S := \arg\min_{x \in C} f(x) \neq \emptyset$

Set $\mu := \min f(C)$.
Stochastic Projected subgradient Method:
Given $x_0 \in C$, update via:

$$x_{n+1} := P_C(x_n - t_n g_n)$$

Assumptions on t_n:
A0: $t_n > 0$,

$$\sum_{n=0}^{\infty} t_n \to \infty, \frac{\sum_{k=0}^{n} t_k^2}{\sum_{k=0}^{n} t_k} \to 0 \text{ as } k \to \infty$$

e.g., $t_n = \frac{\alpha}{n+1},\ \alpha > 0$
What about g_n?
Choose g_n to be a random vector, such that the following assumptions are satisfied.
A1: (”unbiased subgradient”)
$\forall n \in \mathbb{N}$,

$$E(g_n|x_n) \in \partial f(x_n)$$

(means expectation of g_n given x_n is a subgradient), equivalently,
$\forall y \in \mathbb{R}^m$,

$$f(x_n) + \langle E(g_n|x_n), y - x_n\rangle \leqslant f(y)$$

A2: (”boundedness”)
$\exists L > 0, \forall n \in \mathbb{N}$,

$$E(\|g_n\|^2|x_n) \leqslant L^2$$

Theorem 5.29: L23-1

Assuming the previous assumptions hold. Then

$$E(\mu_k) \to \mu \text{ as } k \to \infty$$

where

$$\mu_k := \min\{f(x_0), \ldots, f(x_k)\} \geqslant \mu$$

Proof. Let $s \in S$ and let $n \in \mathbb{N}$. Then

$$\begin{aligned} 0 &\leqslant \|x_{n+1} - s\|^2 \\ &= \|P_C(x_n - t_n g_n) - P_C(s)\|^2 \\ &\leqslant \|(x_n - t_n g_n) - s\|^2 \\ &= \|(x_n - s) - t_n g_n\|^2 \\ &= \|x_n - s\|^2 - 2t_n \langle g_n, x_n - s\rangle + t_n^2 \|g_n\|^2 \end{aligned}$$

Now taking the conditional expectation, given x_n, yields,

$$\begin{aligned} E(\|x_{n+1} - s\|^2 | x_n) \leqslant & \|x_n - s\|^2 + 2t_n \langle E(g_n|x_n), s - x_n\rangle \\ & + t_n^2 E(\|g_n\|^2 | x_n) \\ \underbrace{\leqslant}_{A1, A2} & \|x_n - s\|^2 + 2t_n(f(s) - f(x_n)) + t_n^2 L^2 \\ = & \|x_n - s\|^2 + 2t_n(\mu - f(x_n)) + t_n^2 L^2 \end{aligned}$$

Now taking the expectation w.r.t. x_n yields

$$E(\|x_{n+1} - s\|^2) \leqslant E(\|x_n - s\|^2) + 2t_n(\mu - E(f(x_n))) + t_n^2 L^2 \ldots (*)$$

Let $k \in \mathbb{N}$.
Summing $\sum_{n=0}^{k}$ over $(*)$ and cancelling duplicate terms yields

$$0 \leqslant E(\|x_{k+1} - s\|^2) \leqslant \|x_0 - s\|^2 - 2\sum_{n=0}^{k} t_n(E(f(x_n)) - \mu) + L^2 \sum_{n=0}^{k} t_n^2$$

Hence,

$$\begin{aligned} \frac{1}{2}\left(\|x_0 - s\|^2 + L^2 \sum_{n=0}^{k} t_n^2\right) &\geqslant \sum_{n=0}^{k} t_n(E(f(x_n)) - \mu) \\ &\geqslant \sum_{n=0}^{k} t_n(E(\mu_k) - \mu) \\ &\geqslant 0, \text{ by} f(x_n) \geqslant \mu_k \geqslant \mu \end{aligned}$$

Therefore,

$$0 \leqslant E(\mu_k) - \mu \leqslant \frac{\|x_0 - s\|^2 + L^2 \sum_{n=0}^{k} t_n^2}{2\sum_{n=0}^{k} t_n} \to 0, \text{ as } k \to \infty \text{ by} A0$$

The proof is complete. □

5.8.1 Key Application:

Minimizing a sum of functions

$$f_1, \ldots, f_r : \mathbb{R}^m \to (-\infty, \infty]$$

are convex, lsc proper
Set $I = \{1, \ldots, r\}$ and assume

$$\forall i \in I,\ \mathrm{int(dom(f_i)) \supseteq C \text{ is convex closed}, \neq \emptyset}$$

Also assume that

$$\forall i \in I, \exists L_i \geqslant 0,\ \sup\|\partial f_i(C)\|\leqslant L_i$$

Fact: $\sup\|\partial f_i(C)\|\leqslant L_i \iff f_i|_C$ is L_i-Lipschitz. True if, e.g., C is bounded.

Set

$$f = \sum_{i \in I} f_i$$

<u>Goal</u>

$$(P)\ \mathrm{minimizer_{x \in C} f}$$

We will apply SPGM to (P).
To do that, we verify

- f is convex lsc and proper
- $\emptyset \neq C$ closed and convex $\subseteq \mathrm{int(dom(f))}$
- $S := \arg\min_{x \in C} f(x) \neq \emptyset$

and we have

- $f = \sum_{i \in I} f_i$ is convex lsc. by f_i all convex and lsc proper.
- $\mathrm{dom(f) = \cap_{i \in I} dom(f_i) \supseteq C \neq \emptyset \implies f \text{ is proper.}}$
- $$\begin{aligned}\mathrm{int(dom(f))} &= \mathrm{int} \cap_{\mathrm{i \in I}} \mathrm{dom(f_i)} \\ &= \cap_{i \in I} \mathrm{int(dom(f_i))} \text{ by } I \text{ finite} \\ &\supseteq C \text{ by the previous point}\end{aligned}$$
- Now assume $\mu := \min f(C)$ is attained, i.e., P has a solution. We now will show that $A1, A2$ can be satisfied,
  By the fact above, we have each $f_i|_C$ is L_i-Lipschitz.
  Therefore, using the triangle inequality

  $$f|_C = \sum_{i \in I} f_i|_C \text{ is } \sum_{i \in I} L_i \text{ Lipschitz}$$

  Therefore, once again, by the fact, we learn that

  $$\sup\|\partial f(C)\|\leqslant \sum_{i \in I} L_i$$

Let $x_0 \in C$. Given $x_n \in C$, $x_{n+1} = P_C(x_n - t_n g_n)$, we pick an index $i_n \in I = \{1, \dots, r\}$ randomly using uniform distribution and we set

$$\begin{aligned} g_n &= r \cdot f'_{i_n}(x_n) \\ &\in r \cdot \partial f_{i_n}(x_n) \end{aligned}$$

Now,

$$\begin{aligned} E(g_n | x_n) &= \sum_{i=1}^{r} \frac{1}{r} \cdot r f'_i(x_n) \\ &= \sum_{i=1}^{r} \underbrace{f'_i(x_n)}_{\in \partial f_i(x_n)} \\ &\in \partial f_1(x_n) + \ldots + \partial f_r(x_n) \\ &= \partial(f_1 + \ldots + f_r)(x_n) \text{ sum rule} \\ &= \partial f(x_n) \end{aligned}$$

so A1 holds.
Next:

$$\begin{aligned} E(\|g_n\|^2 | x_n) &= \sum_{i=1}^{r} \frac{1}{r} \|r f'_i(x_n)\|^2 \\ &= \sum_{i=1}^{r} r \|f'_i(x_n)\|^2 \\ &\leqslant r \sum L_i^2 \\ &=: L^2 \end{aligned}$$

Therefore, A2 holds with $L := \sqrt{r \sum_{i=1}^{r} L_i^2}$.
Consequently,

$$x_{n+1} := P_C(x_n - t_n g_n)$$

generates a sequence such that

$$E(\mu_n) \to \mu$$

$\mu_n = \min_{i \in \{1, \ldots, n\}} \{f(x_0), \ldots, f(x_n)\}$