ISyE 7683 (Winter 2018)

Advanced Nonlinear Optimization

Prof. R. D. C. Monteiro
Georgia Institute of Technology

LaTeXer: W. Kong
http://wwkong.github.io
Last Revision: May 22, 2018

Table of Contents

These notes are currently a work in progress, and as such may be incomplete or contain errors.

Acknowledgments:

Special thanks to *Michael Baker* and his LaTeX formatted notes. They were the inspiration for the structure of these notes.

Abstract

The purpose of these notes is to provide the reader with a secondary reference to the material covered in ISyE 7863.

1 Convex Sets

1.1 Basic Definitions

Definition 1.1. A **convex set** $C \subseteq \mathbb{R}^n$ is such that for $x, y \in C$ we have $[x, y] \subseteq C$ where

$$[x, y] = \{\alpha x + (1-\alpha)y : \alpha \in [0, 1]\}$$

Example 1.1. (a) The **hyperplane** $H_{s,\alpha}$ parameterized by $(s, \alpha) \in \mathbb{R}^n \times \mathbb{R}$ with $s \neq 0$ where

$$H_{s,\alpha} = \{x \in \mathbb{R}^n : s^T x = \alpha\}.$$

Note that if $x_0 \in H_{s,\alpha}$ then also

$$H_{s,\alpha} = \{x_0\} + \{s\}^{\perp}, \{s\}^{\perp} = \{x \in \mathbb{R}^n : s^T x = 0\}.$$

(b) The **affine manifold** which is a subset $\emptyset \neq V \subseteq \mathbb{R}^n$ such that if $x, y \in V$ then $\overleftrightarrow{xy} \subseteq V$ where

$$\overleftrightarrow{xy} = \{\alpha x + (1-\alpha)y : \alpha \in \mathbb{R}\}.$$

(*Exercise*. If $x^0 \in V$ then $V - \{x^0\}$ is a subspace that does not depend on x^0, i.e. $V = \{x^0\} + S$ where S is a subspace. We define the **dimension** of V as $\dim V = \dim S$.)

(c) The **convex cone**. A set $K \subseteq \mathbb{R}^n$ is a cone if $x \in K, \alpha > 0 \implies \alpha x \in K$ and if it is convex, it is a convex cone. An example is the recession cones of the form $\{x \in \mathbb{R}^n : A_1 x = 0, A_2 x \leq 0\}$.

Definition 1.2. $A : \mathbb{R}^n \mapsto \mathbb{R}^m$ is affine if $x, y \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ then

$$A(\alpha x + (1-\alpha)y) = \alpha Ax + (1-\alpha)Ay.$$

Example 1.2. A is affine if $A(x) = Tx + c$ where $T \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$.

Calculus of Convex Sets

Operations that preserve convexity:

(1) $\{C_i\}_{i \in I}$ family of convex sets $\implies \bigcap_{i \in I} C_i$ is convex

(2) $C_i \subseteq \mathbb{R}^{n_i}$ convex for $i = 1, ..., k \implies C_1 \times ... \times C_k \subseteq \mathbb{R}^{n_1} \times .... \times \mathbb{R}^{n_k}$ is convex

(3) $C_i \subseteq \mathbb{R}^n$ convex and $\alpha_i \in \mathbb{R}$ for $i = 1, ..., k \implies \alpha_1 C_1 + ... + \alpha_k C_k \subseteq \mathbb{R}^n$ is convex

(4) $C \subseteq \mathbb{R}^n$ convex and T affine $\implies T(C)$ is convex

(5) $D \subseteq \mathbb{R}^n$ convex and T affine $\implies T^{-1}(D)$ is convex

Applications

- We can prove (3) with (2), (4) using $T(x_1, ..., x_k) = \alpha_1 x_1 + ... + \alpha_k x_k$ and the fact that $T(C_1 \times ... \times C_k)$ is convex.

Definition 1.3. A **linear combination** of $x^1, ..., x^k \in \mathbb{R}^n$ is of the form $\alpha_1 x^1 + ... + \alpha_k x^k$ where $\alpha_i \in \mathbb{R}$ for $i = 1, ..., k$. An **affine combination** is a linear combination with $\sum \alpha_i = 1$ and a **convex combination** is an affine combination with α_i.

Proposition 1.1. *$C \subseteq \mathbb{R}^n$ is a convex set $\iff$ every convex combination of elements of C lies in C.*

Convexification

Given any $\emptyset \neq S \subseteq \mathbb{R}^n$, one can construct is smallest convex set co S containing S by

$$\text{co } S := \bigcap \{C \subseteq \mathbb{R}^n : C \text{ convex}, C \supseteq S\}$$

which we can the **convex hull** of S. Observe that if $S \subseteq C$ for convex C then co $S \subseteq C$. Similarly the **affine hull** of S is

$$\text{aff } S := \bigcap \{V \subseteq \mathbb{R}^n : V \text{ affine manifold}, V \supseteq S\}.$$

and also for the **linear hull** we have

$$\text{lin } S := \bigcap\{V \subseteq \mathbb{R}^n : V \text{ subspace}, V \supseteq S\}.$$

Proposition 1.2. *Let $\emptyset \neq S \subseteq \mathbb{R}^n$ be given. Then,*

$$\text{co } S = \textit{set of all convex combinations of elements of } S.$$

Proof. (Sketch) Let

$$\mathcal{C} = \text{set of all convex combination of elements of } S.$$

(1) $\mathcal{C}$ is convex (exercise) and $\mathcal{C} \supseteq S$. Thus, co $S \subseteq \mathcal{C}$.

(2) Since $S \subseteq$ co S and co S is convex, then co S contains all convex combinations of its elements and hence of S. Thus, co $S \supseteq \mathcal{C}$. □

Proposition 1.3. *(Caratheodory) Define*

$$\begin{aligned} S[k] = \{&\alpha_0 x^0 + ... + \alpha_k x^0 : \\ &x^i \in S, i = 0, ..., k, \\ &\alpha_i \geq 0, i = 1, ..., k, \\ &\sum_{i=0}^{n} \alpha_i = 1\} \end{aligned}$$

and suppose that S is dimension d (i.e. its affine hull is of dimension d). Then co $S = S[d]$ and also co $S = \bigcup_{k=0}^{\infty} S[k]$.

Definition 1.4. The **closed convex hull** is

$$\overline{\text{co}}\ S := \bigcap\{C \subseteq \mathbb{R}^n : C \text{ closed convex}, C \supseteq S\} = \text{cl }(\mathcal{C})$$

where cl S is the **closure** of S and

$$\mathcal{C} = \text{set of all convex combination of elements of } S.$$

1.2 Topology

<u>Relative Interior</u>

Given set $S \subseteq X \subseteq \mathbb{R}^n$.

Definition 1.5. $\bar{x} \in X$ is called an **interior point** of S wrt X if $\exists \delta > 0$ s.t. $\bar{B}(\bar{x};\delta) \cap X \subseteq S$ where $\bar{B}(\bar{x};\delta) := \{x \in \mathbb{R}^n : \|x - \bar{x}\| \leq \delta\}$. We remark that int$_X S \subseteq S$ where use the notation $\bar{x} \in$ int$_X S$ for the **interior with respect to** X**.**

Definition 1.6. $\bar{x} \in$ aff$S$ is a **relative interior point** of S if $\bar{x} \in$ int$_{\text{aff } S} S$. We remark that ri $S \subseteq S$ and use the notation $\bar{x} \in$ ri S for the **relative interior**.

Remark 1.1. Note that $S \subseteq T$ does not necessarily imply $\text{ri}\, S \subseteq \text{ri}\, T$, but is true if $\text{aff}\, S = \text{aff}\, T$.

Definition 1.7. $\bar{x} \in$ cl $S \backslash$ri S is a **relative boundary point** and use the notation $\bar{x} \in$ rbd S for the **relative boundary**. We define the **relative closure** cl$_X S$ as cl$_X S$ = (cl S) $\cap X$.

Example 1.3. (a) $C = \{x\}$, aff $C = \{x\}$, ri $C = \{x\}$ = cl C, rbd $C = \emptyset, \dim C = 0$.

(b) $C = [x, y], x \neq y, \text{aff}\, C = \overleftrightarrow{xy}, \text{ri}\, C = (x, y), \text{rbd}\, C = \{x, y\}, \dim C = 1$.

(c) $C = \bar{B}(x;\delta), \text{aff}\, C = \mathbb{R}^n, \text{ri}\, C = B(x;\delta), \text{rbd}\, C = \{x : \|x - \bar{x}\| = \delta\}, \dim C = n$.

Proposition 1.4. *Let $V \subseteq \mathbb{R}^n$ be an affine manifold and $S \subseteq V$ be given. Then:*

(a) int$_V S \neq \emptyset \implies V = \text{aff}\, S$ and hence $\text{ri}\, S \neq \emptyset$.

(a) int $S \neq \emptyset \implies \mathbb{R}^n = \text{aff}\, S$ and hence $\text{ri}\, S =$ int $S \neq \emptyset$.

Proposition 1.5. *If* $\emptyset \neq C \subseteq \mathbb{R}^n$ *convex, then* $\operatorname{ri} C \neq \emptyset$.

Proposition 1.6. *(resolution lemma) Let* $\emptyset \neq C \subseteq \mathbb{R}^n$ *convex,* $x \in \operatorname{cl} C$ *and* $y \in \operatorname{ri} C$. *Then* $[y, x) \subseteq C$.

Proof. Assume first that $x \in C$. Since $y \in \operatorname{ri} C$ then $\exists \delta > 0$ s.t.

$$\bar{B}(y; S) \cap \operatorname{aff} C \subseteq C.$$

For $t \in (0, 1]$ define $z_t = (1-t)x + ty$ and $\delta_t = t\delta$.

<u>Claim</u>. The set $\bar{B}(z_t; \delta_t) \cap \operatorname{aff} C \subseteq C$ for all $t \in [0, 1]$. In other words, $z_t \in \operatorname{ri} C$.

<u>Proof of Claim</u>. Remark that

$$\begin{aligned}
&\bar{B}(z_t; \delta_t) \cap \operatorname{aff} C \\
=&\bar{B}((1-t)x + ty; \delta_t) \cap [(1-t)x + t \cdot \operatorname{aff} C] \\
=&(1-t)x + t \cdot \bar{B}(y; \delta_t) \cap [(1-t)x + t \cdot \operatorname{aff} C] \\
=&(1-t)x + t \cdot \left[\bar{B}(y; \delta_t) \cap \operatorname{aff} C\right] \\
\subseteq&(1-t)x + tC \subseteq C
\end{aligned}$$

where the last inclusion is by the convexity of C. ☐

Corollary 1.1. *If* C *is convex then* $\operatorname{ri} C$ *is convex.*

Proposition 1.7. *Assume that* $\bar{x} \in \operatorname{ri} C$. *Then*

(a) $\exists \delta > 0$ *such that* $\bar{B}(\bar{x}; \delta) \cap \operatorname{aff} C \subseteq \operatorname{ri} C$

(b) Given any $x \in \operatorname{aff} C$, $\exists \varepsilon > 0$ *s.t.* $\bar{x} + t(x - \bar{x}) \in \operatorname{ri} C$, *for all* t *s.t.* $|t| \leq \varepsilon$

(c) Given any u *lying in the subspace parallel to* $\operatorname{aff} C$, $\exists \varepsilon > 0$ *s.t.* $\bar{x} + tu \in \operatorname{ri} C$, *for all* t *s.t.* $|t| < \varepsilon$.

Proposition 1.8. *Let* $\emptyset \neq C \subseteq \mathbb{R}^n$ *be convex. Then,*

(a) $\operatorname{aff}(\operatorname{ri} C) \overset{(1)}{=} \operatorname{aff} C \overset{o}{=} \operatorname{aff}(\operatorname{cl} C)$

(b) $\operatorname{ri}(\operatorname{ri} C) \overset{o}{=} \operatorname{ri} C \overset{(2)}{=} \operatorname{ri}(\operatorname{cl} C)$

(c) $\operatorname{cl}(\operatorname{ri} C) \overset{(3)}{=} \operatorname{cl} C \overset{o}{=} \operatorname{cl}(\operatorname{cl} C)$

where $\overset{o}{=}$ *means "obvious".*

Proof. (1) $\operatorname{ri} C \subseteq C \implies \operatorname{aff}(\operatorname{ri} C) \subseteq \operatorname{aff} C$. Conversely, let $x \in \operatorname{aff} C$. Since $C \neq \emptyset$ and convex, $\exists \bar{x} \in \operatorname{ri} C$ and $\exists \varepsilon > 0$ s.t. $\bar{x} + \varepsilon(x - \bar{x}) \in \operatorname{ri} C$. Now

$$\{x\} \subseteq \operatorname{aff}\left(\{\bar{x}, \bar{x} + \delta(x - \bar{x})\}\right) \subseteq \operatorname{aff}(\operatorname{ri} C)$$

and hence $\operatorname{aff}(C) \subseteq \operatorname{aff}(\operatorname{ri} C)$.

(2) Given $C \subseteq \operatorname{cl} C$, since $\operatorname{aff} C \overset{o}{=} \operatorname{aff}(\operatorname{cl} C)$ then $\operatorname{ri} C \subseteq \operatorname{ri}(\operatorname{cl} C)$. Conversely, let $y \in \operatorname{ri}(\operatorname{cl} C) \neq \emptyset$. Then $\exists \varepsilon > 0$ s.t. $z_\varepsilon = y + \varepsilon(y - \bar{x}) \in \operatorname{cl} C$. By the resolution lemma, $y \in \operatorname{ri} C$.

(3) *Exercise.* ☐

Proposition 1.9. *The sets* $\operatorname{ri} C$, C, *and* $\operatorname{cl} C$ *all have the same* ri,cl, *and* aff.

Proposition 1.10. *Let* C_1, C_2 *convex. Then the following are equivalent:*

(1) $\operatorname{ri} C_1 = \operatorname{ri} C_2$,

(2) $\operatorname{cl} C_1 = \operatorname{cl} C_2$,

(3) $\operatorname{ri} C_1 \subseteq C_2 \subseteq \operatorname{cl} C_1$.

Proof. (1) $\iff$ (2) follows from $\operatorname{ri}(\operatorname{cl} C) = \operatorname{ri} C$ and $\operatorname{cl}(\operatorname{ri} C) = \operatorname{cl} C$.

(3) $\implies$ (2) Directly,

$$\operatorname{cl}(C_1) = \operatorname{cl}(\operatorname{ri} C_1) \subseteq \operatorname{cl}(C_2) \subseteq \operatorname{cl}(C_1)$$

(2) $\implies$ (3) Directly,

$$\operatorname{ri} C_1 \overset{(1)}{=} \operatorname{ri} C_2 \subseteq C_2 \subseteq \operatorname{cl} C_2 \overset{(2)}{=} \operatorname{cl} C_1$$

□

Proposition 1.11. *If $C \subseteq \mathbb{R}^m$ is convex and $A : \mathbb{R}^m \mapsto \mathbb{R}^n$ is affine, then*

(1) $\operatorname{ri} A(C) = A(\operatorname{ri} C)$

(2) $\operatorname{cl} A(C) \supseteq A(\operatorname{cl} C)$ *(no need for convexity)*

(3) $\operatorname{aff} A(C) = \operatorname{aff}(A(\operatorname{ri} C)) = \operatorname{aff}(A(\operatorname{cl} C)) = A(\operatorname{aff} C)$

Example 1.4. Consider $C = \{(x_1, x_2) : x_2 \geq 1/x_1, x_1 > 0\}$ and $A(x_1, x_2) = x_2$ where $A(C) = \mathbb{R}_{++} = A(\operatorname{cl} C)$ and $\operatorname{cl} A(C) = \mathbb{R}_+$.

Corollary 1.2. *If $\alpha_1, ..., \alpha_k \in \mathbb{R}$ and $C_1, ..., C_k \in \mathbb{R}^n$ convex. Then,*

$$\operatorname{ri}(\alpha_1 C_1 + ... + \alpha_k C_k) = \alpha_1 \operatorname{ri} C_1 + ... + \alpha_k \operatorname{ri} C_k.$$

Lemma 1.1. *For $S_i \subseteq \mathbb{R}^n$, $i = 1, ..., k$,*

$$\operatorname{ri}(S_1 \times ... \times S_k) = \operatorname{ri} S_1 \times ... \times \operatorname{ri} S_k.$$

Proof. Exercise. □

Proof. (of Corollary) Define the linear map $A : \mathbb{R}^n \times \overbrace{...}^{k} \times \mathbb{R}^n \mapsto \mathbb{R}^n$ as

$$A(x_1, ..., x_k) = \alpha_1 x_1 + ... + \alpha_k x_k$$

where we have

$$A(C_1 \times ... \times C_k) = \alpha_1 C_1 + ... + \alpha_k C_k.$$

The result follows from the fact that $\operatorname{ri} A(C) = A(\operatorname{ri} C)$ where $C = C_1 \times ... \times C_k$. □

Proposition 1.12. *Let $A : \mathbb{R}^n \mapsto \mathbb{R}^n$ be affine and $D \subseteq \mathbb{R}^n$ be convex. If $A^{-1}(\operatorname{ri} D) \neq \emptyset$ then*

$$(*)\ \operatorname{ri} A^{-1}(D) = A^{-1}(\operatorname{ri} D)$$
$$(**)\ \operatorname{cl} A^{-1}(D) = A^{-1}(\operatorname{cl} D).$$

The sets $A^{-1}(\operatorname{ri} D), A^{-1}(D), A^{-1}(\operatorname{cl} D)$ have the same affine hull, namely $A^{-1}(\operatorname{aff} D)$.

Proposition 1.13. *If $C_1, ..., C_k \subseteq \mathbb{R}^n$ are convex and $\bigcap_{i=1}^k \operatorname{ri} C_i \neq \emptyset$ then*

$$\operatorname{ri}\left(\bigcap_{i=1}^k C_i\right) = \bigcap_{i=1}^k \operatorname{ri} C_i$$
$$\operatorname{cl}\left(\bigcap_{i=1}^k C_i\right) = \bigcap_{i=1}^k \operatorname{cl} C_i.$$

Proof. Consider the map $A(x) = (x, ..., x)$ where $A : \mathbb{R}^n \mapsto \mathbb{R}^n \times \overbrace{...}^{k} \times \mathbb{R}^n$. We have

$$A^{-1}(S_1 \times ... S_k) = \bigcap_{i=1}^k S_i$$

for any $S_i \subseteq \mathbb{R}^n$ and $i = 1, ..., k$. Since $D := C_1 \times ... \times C_k$ is convex, then by $(*)$ we have

$$\operatorname{ri}\left[\bigcap_{i=1}^{k} C_i\right] = \operatorname{ri} A^{-1}(D) = A^{-1}(\operatorname{ri} D) = A^{-1}(\operatorname{ri} D) = A^{-1}(\operatorname{ri} C_1 \times ... \times \operatorname{ri} C_k) = \bigcap_{i=1}^{k} \operatorname{ri} C_i.$$

and the same proof hold if we replace ri with cl and using $(**)$. ☐

1.3 Asymptotic or Recession Cone

For this section, let us always assume that $\emptyset \neq C \subseteq \mathbb{R}^n$ is closed convex.

Definition 1.8. A set $S \subseteq \mathbb{R}^n$ is **bounded** if $\exists R > 0$ such that $S \subseteq \bar{B}(0; R)$.

Definition 1.9. Let $\emptyset \neq C \subseteq \mathbb{R}^n$ be closed and convex. Its **asymptotic cone**, denoted by C_∞, is defined as

$$C_\infty := \{d \in \mathbb{R}^n : x + td \in C, \forall t > 0, \forall x \in C\}.$$

Proposition 1.14. *C_∞ is a closed convex cone containing 0.*

Proof. Exercise. *Hint*: We can show that

$$C_\infty = \bigcap_{\substack{t > 0 \\ x \in C}} \frac{C - x}{t}$$

and define the map $A(y) = (y - x)/t$. Use previous propositions to finish. ☐

Proposition 1.15. *If for source $x_0 \in C$ and $d \in \mathbb{R}^n$ we have*

$$\{x_0 + td : t > 0\} \subseteq C$$

then $d \in C_\infty$.

Proof. Let $x \in C$ be given. We claim that $x + td \in C$ for all $t > 0$. To show this, note that for $\epsilon \in (0, 1)$ we have

$$\begin{aligned} x_\epsilon^+ &:= x_0 + td + (1 - \epsilon)(x - x_0) \\ &= (1 - \epsilon)x + \epsilon\left(x_0 + \frac{t}{\epsilon}d\right) \in C \end{aligned}$$

which is a convex combination of two elements in C. Also (1) $x_\epsilon^+ \to x + td$ as $\epsilon \to 0$ and so $x + td \in \operatorname{cl} C = C$. ☐

Lemma 1.2. *If $d = \lim_{k\to\infty} \alpha_k x^k$ where $\{x^k\} \subseteq C$ and $\{\alpha_k\} \subseteq \mathbb{R}_{++} \to 0$ then $d \in C_\infty$.*

Proof. Let $x \in C, t > 0$ be given. Since $x^k, x \in C$ and C is convex, then

$$y^k := (1 - t\alpha_k)x + t\alpha^k x^k \in C$$

for sufficiently large enough k. We have $y^k \to x + td \in \operatorname{cl} C = C$. So $d \in C_\infty$. ☐

Proposition 1.16. *C is bounded $\iff C_\infty = \{0\}$.*

Proof. ($\implies$) Simple proof by contradiction.

($\impliedby$) Assume for contradiction that C is unbounded. So, $\exists$ a sequence $\{x_k\} \subseteq C$ such that $\|x_k\| \to \infty$. Consider $d_k = x_k/\|x_k\|$ where $d_k \overset{k\in K}{\to} d$ for some subsequence K by the Bolzano-Weierstrass theorem. Also note that $\|d\| = 1$ and so $d \neq 1$. By the previous lemma, using $\alpha_k = 1/\|x_k\|$, we have that $d \in C_\infty$. ☐

Proposition 1.17. *(a) If $\{C_j\}_{j\in J}$ is a family of closed convex sets such that $\bigcap_{j\in J} C_j \neq \emptyset$ then*

$$\left(\bigcap_{j\in J} C_j\right)_\infty = \bigcap_{j\in J} (C_j)_\infty$$

(b) If $C_i \subseteq \mathbb{R}^{n_i}$ is a non-empty closed convex set for $i = 1, 2, ..., k$ then

$$(C_1 \times ... \times C_k)_\infty = (C_1)_\infty \times ... \times (C_k)_\infty.$$

(c) Let $A : \mathbb{R}^n \mapsto \mathbb{R}^m$ be linear. Then,

(i) If $\emptyset \neq C$ is closed convex and $A(C)$ is closed then $A(C_\infty) \subseteq [A(C)]_\infty$.

(ii) If $\emptyset \neq D$ is closed convex and $A^{-1}(D) \neq \emptyset$ then $A^{-1}(D_\infty) = [A^{-1}(D)]_\infty$.

Proposition 1.18. *Let $A : \mathbb{R}^n \mapsto \mathbb{R}^m$ be linear, $\emptyset \neq C \subseteq \mathbb{R}^n$ closed convex such that $A^{-1}(0) \cap C_\infty = \{0\}$ (or $\subseteq -C_\infty$) then:*

(i) $A(C)$ is closed

(ii) $A(C_\infty) = [A(C)]_\infty$

Definition 1.10. The **linearity space** of C is defined as $C_\infty \cap (-C_\infty)$ which you can prove is the largest subspace contained in C_∞.

2 Convex Functions

Notation 1. Let us denote $\bar{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\} = [-\infty, \infty]$ and for $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ we denote

$$\begin{aligned}
\operatorname{dom} f &= \{x \in \mathbb{R}^n : f(x) < \infty\} \\
\operatorname{epi} f &= \{(x, r) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq r\} \\
\operatorname{epi_S} f &= \{(x, r) \in \mathbb{R}^n \times \mathbb{R} : f(x) < r\} \\
f^{-1}(-\infty, r] &= \{x \in \mathbb{R}^n : f(x) \leq r\} \\
f^{-1}(-\infty, r) &= \{x \in \mathbb{R}^n : f(x) < r\}.
\end{aligned}$$

Definition 2.1. A **convex function** $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ is a function where its **epigraph** $\operatorname{epi} f$ is convex. We say such functions $f \in$ E-Conv $\mathbb{R}^n$.

Definition 2.2. $f : \mathbb{R}^n \mapsto \mathbb{R}$ is **proper convex** if $f \in$ E-Conv $\mathbb{R}^n$, $f(x) > -\infty$ for all $x \in \mathbb{R}^n$, and $f \neq \infty$ (or equivalently, $\exists x \in \mathbb{R}^n$ such that $f(x) < \infty$). We say that such that such functions $f \in$ Conv $\mathbb{R}^n$.

Definition 2.3. A function f is (proper) **concave** if $-f$ is (proper) convex.

Proposition 2.1. *Let $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ be given. Then the following are equivalent:*

(a) $f \in$ E-Conv $\mathbb{R}^n$

(b) $\operatorname{epi_S} f$ is a convex set

(c) $f(\alpha x_0 + (1-\alpha)x_1) \leq \alpha f(x_0) + (1-\alpha) f(x_1)$ for all $\alpha \in (0, 1)$ and $\forall x_0, x_1 \in \operatorname{dom} f$.

Proof. (a) $\implies$ (b) Let $(x_1, r_1), (x_2, r_2) \in \operatorname{epi}_S f$ and $\alpha \in (0, 1)$. We wish to show that

$$\alpha(x_1, r_1) + (1-\alpha)(x_2, r_2) \in \operatorname{epi_S} f.$$

Note that $\exists \epsilon > 0$ such that $(x_1, r_1 - \epsilon), (x_2, r_2 - \epsilon) \in \operatorname{epi} f$ by (a). Also by (a),

$$\begin{aligned}
&\alpha(x_1, r_1 - \epsilon) + (1-\alpha)(x_2, r_2 - \epsilon) \in \operatorname{epi} f \\
\implies &\alpha(x_1, r_1) + (1-\alpha)(x_2, r_2) - (0, \epsilon) \in \operatorname{epi} f
\end{aligned}$$

and hence the result follows.

(b) $\implies$ (c) Let $x_0, x_1 \in \operatorname{dom} f$ and $\alpha \in (0,1)$ be given. There exists $r_0, r_1 \in \mathbb{R}$ such that $f(x_i) < r_i$ for $i = 0,1$ or $(x_i, r_i) \in \operatorname{epi_S} f$. Since $\operatorname{epi_S} f$ is convex, then

$$(\alpha x_0 + (1-\alpha)x_1, \alpha r_0 + (1-\alpha)r_1) = \alpha(x_0, r_0) + (1-\alpha)(x_1, r_1) \in \operatorname{epi_S} f$$

or equivalently

$$f(\alpha x_0 + (1-\alpha)x_1) < \alpha r_0 + (1-\alpha) r_1$$

and as $r_i \downarrow f(x_i)$ for $i = 0,1$ the result follows.

(c) $\implies$ (a) Left as an exercise. □

Proposition 2.2. *Let $f \in$ E-Conv $\mathbb{R}^n$. Then*

(a) $f^{-1}[-\infty, r)$ is convex for all $r \in \bar{\mathbb{R}}$

(b) $f^{-1}[-\infty, r]$ is convex for all $r \in \bar{\mathbb{R}}$

So $\operatorname{dom} f$ is convex.

Proposition 2.3. *(Jensen's inequality) If $f \in$ E-Conv $\mathbb{R}^n$ then*

$$f(\alpha_0 x_0 + ... + \alpha_k x_k) \leq \sum_{i=1}^{k} \alpha_i f(x_i)$$

for all $(\alpha_0, ..., \alpha_k) \in \Delta_k$ the k-dimensional probability simplex and $x_i \in \operatorname{dom} f$ for $i = 0, 1, ..., k$.

Proof. Since $(x_i, r_i) \in \operatorname{epi} f$ for all $r_i > f(x_i)$ then $\sum_i \alpha_i (x_i, r_i) \in \operatorname{epi} f$ and $f(\sum_i \alpha_i x_i) \leq \sum_i \alpha_i r_i$. Letting $r_i \downarrow f(x_i)$ for all $i = 0, 1, ..., k$ gives the result. □

Definition 2.4. A function $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ is **strictly convex** if f is proper and

$$f(\alpha x_0 + (1-\alpha)x_1) < \alpha f(x_0) + (1-\alpha) f(x_1)$$

for all $\alpha \in (0,1)$ and $x_0 \neq x_1 \in \operatorname{dom} f$.

Definition 2.5. A function $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ is β**-strongly convex** if f is proper and

$$f(\alpha x_0 + (1-\alpha)x_1) \leq \alpha f(x_0) + (1-\alpha) f(x_1) - \frac{\beta}{2}\alpha(1-\alpha)\|x_0 - x_1\|^2$$

for all $\alpha \in (0,1)$ and $x_0 \neq x_1 \in \operatorname{dom} f$.

Remark 2.1. We have f is $\beta-$strongly convex $\implies$ f is strictly convex $\implies$ f convex

Proposition 2.4. *f is β-strongly convex $\iff$ $f - \frac{\beta}{2}\|\cdot\|^2$ is convex.*

Proposition 2.5. *(a) If $f_1, ..., f_k \in$ Conv $\mathbb{R}^n$ and $\alpha_1, ..., \alpha_n \geq 0$ then*

$$\alpha_1 f_1 + ... + \alpha_n f_k \in \begin{cases} \text{Conv } \mathbb{R}^n, & \text{if } \bigcap_{i=1}^k \operatorname{dom} f_i \neq \emptyset \\ \infty, & \text{otherwise.} \end{cases}$$

(b) If $\{f_i\}_{i \in I} \in$ E-Conv $\mathbb{R}^n$ then $\sup_{i \in I} f_i \in$ Conv $\mathbb{R}^n$ or $f = \infty$. Note that this can follow from $\operatorname{epi}(\sup_{i\in I} f_i) = \bigcap_{i \in I} \operatorname{epi} f$.

(c) If $f \in$ Conv $\mathbb{R}^n$ and $A : \mathbb{R}^n \mapsto \mathbb{R}^m$ is affine such that $A(\mathbb{R}^n) \cap \operatorname{dom} f \neq \emptyset$ then $f \circ A \in$ Conv $\mathbb{R}^n$.

Example 2.1. Let S^n be the space of real symmetric matrices. We have $\dim S^n = n(n+1)/2$. Define

$$\lambda_{\max}(A) = \max x^T A x \quad \text{s.t. } \|x\| = 1$$

$$\lambda_{\min}(A) = \min x^T A x \quad \text{s.t. } \|x\| = 1$$

Consider the linear map $\phi_x(A) = x^T Ax$. Then $\lambda_{\max}(A) = \sup_{x:\|x\|=1} \phi_x(A)$ which from (b) of the previous proposition implies that $\lambda_{\max}$ is convex. Since $\lambda_{\max}(A) = -\lambda_{\min}(-A)$ then $\lambda_{\min}$ is concave.

Example 2.2. Let $C \subseteq \mathbb{R}^n$ and define the **indicator function** $I_C : \mathbb{R}^n \mapsto [0, +\infty]$ by

$$x \mapsto I_C(x) = \begin{cases} 0 & \text{if } x \in C \\ \infty & \text{if } x \notin C. \end{cases}$$

It is not had to show that I_C is convex $\iff$ C is convex and $\operatorname{epi} I_C = C \times \mathbb{R}_+$.

Example 2.3. The **support function** $\sigma_c : \mathbb{R}^n \mapsto [-\infty, +\infty]$ where $\sigma_C(x) = \sup_{c \in C} x^T c$. Note that $\sigma_C = -\infty \iff C = \emptyset$ and $C \neq \emptyset \implies \sigma_C \in \text{Conv}\,\mathbb{R}^n$ (use $x = 0$).

2.1 Continuity

Proposition 2.6. *If $f \in$ Conv $\mathbb{R}^n$ then $\forall x_0 \in \operatorname{ri}(\operatorname{dom} f)$, $\exists L = L(x_0) \geq 0$ and neighbourhood $N(x_0)$ of x_0 such that*

$$|f(x) - f(\bar{x})| \leq L\|x - \tilde{x}\|$$

for all $x, \tilde{x} \in N(x_0) \cap \operatorname{aff}(\operatorname{dom} f)$. In particular, this result implies that f is continuous on $\operatorname{ri}(\operatorname{dom} f)$.

Proposition 2.7. *If $f \in$ Conv $\mathbb{R}^n$ then for all compact set $K \subseteq \operatorname{ri}(\operatorname{dom} f)$ there exists $L = L(K)$ such that*

$$|f(x) - f(\bar{x})| \leq L\|x - \tilde{x}\|.$$

Proof. Suppose that $\{x_k\}, \{\tilde{x}_k\}$ are sequences such that

$$|f(x_k) - f(\tilde{x}_k)| > k\|x_k - \tilde{x}_k\|.$$

By the continuity of f and the compactness of K, we have $x_k \to x$, $\tilde{x}_k \to \tilde{x}$ from Bolzano-Weierstrass and $f(x_k) \to f(x)$, $f(\tilde{x}_k) \to f(\tilde{x})$. Since $|f(x) - f(\tilde{x})|$ is then finite, then $x = \tilde{x}$.

From the previous proposition, $\exists \delta \geq 0$ such that f is L-Lipschitz on $B(x;\delta) \cap \operatorname{aff}(\operatorname{dom} f)$, i.e.

$$|f(x) - f(\tilde{x})| \leq L\|x - \tilde{x}\|, \forall x, \tilde{x} \in B(x;\delta) \cap \operatorname{aff}(\operatorname{dom} f)$$

which is a contradiction as $x_k, \tilde{x}_k$ enter in the neighbourhood $B(x;\delta) \cap \operatorname{aff}(\operatorname{dom} f)$ and k becomes large enough. □

Corollary 2.1. *If $f \in$ Conv $\mathbb{R}^n$ finite everywhere, then f is continuous on $\mathbb{R}^n$ and for every bounded set $C \subseteq \mathbb{R}^n$ there exists $L = L(C)$ such that*

$$|f(x) - f(\tilde{x})| \leq L\|x - \tilde{x}\|, \forall x, \tilde{x} \in C.$$

Definition 2.6. The **lower semi-continuous hull** of $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$, denoted by $\operatorname{lsc} f$ is defined as

$$\begin{aligned}(\operatorname{lsc} f)(x) &= \liminf_{y \to x} f(y) \\ &= \inf\left\{v : \exists \{y_k\} \to x \text{ s.t. } \lim_{n\to\infty} f(x_k) = v\right\} \leq f(x).\end{aligned}$$

Definition 2.7. A function $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ is lower semi-continuous (lsc) at $x \in \mathbb{R}^n$ if $f(x) = (\operatorname{lsc} f)(x)$. The function f is **lower semi-continuous** if $(\operatorname{lsc} f) = f$.

Proposition 2.8. *Let $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$. Then:*

(a) $\operatorname{epi}(\operatorname{lsc} f) = \operatorname{cl}(\operatorname{epi} f)$

(b) If $f \in$ E-Conv $\mathbb{R}^n$ then $\operatorname{lsc} f \in$ E-Conv $\mathbb{R}^n$

Proof. (a) $(x, r) \in \operatorname{cl}(\operatorname{epi} f) \iff \exists$ a sequence $\{(x_k, r_k)\} \subseteq \operatorname{epi} f$ such that $(x_k, r_k) \to (x, r) \iff \exists$ a sequence $\{x_k\} \subseteq \mathbb{R}^n$, $\{r_k\} \subseteq \mathbb{R}$ such that $x_k \to x, r_k \to r$ and $f(x_k) \leq k \overset{(*)}{\iff} \liminf_{y \to x} f(y) \leq r \iff (\operatorname{lsc} f)(x) \leq r \iff (x, r) \in \operatorname{epi}(\operatorname{lsc} f)$. Everything here but $(*)$ is obvious, so let us show $(*)$.

Proof of (*)

($\implies$) We have

$$\liminf_{y\to x} f(y) \le \liminf_{k\to\infty} f(x_k) \le \liminf_{k\to\infty}(r_k) = r.$$

($\impliedby$) There exists a sequence $\{x_k\} \to x$ such that $f(x_k) \to \liminf_{y\to x} f(y) \le r$. Let $r_k = \max\{r, f(x_k)\} \to r$. Then $f(x_k) \le r_k$. ☐

Proposition 2.9. *For $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ the following are equivalent:*

(a) $\operatorname{epi} f$ *is closed*

(b) $f^{-1}[-\infty, r]$ *is (possibly empty) closed for all* $\forall r \in \mathbb{R}$

(c) f *is lsc*

Proof. (a) $\implies$ (c) Comes from the fact that $\operatorname{epi} f = \operatorname{cl}(\operatorname{epi} f) = \operatorname{epi}(\operatorname{lsc} f)$ and so $f = \operatorname{lsc} f$. Similar idea from the converse. ☐

Proposition 2.10. *Let $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$. Then,*

(a) $\operatorname{lsc} f$ *is lsc and* $\operatorname{lsc} f \le f$.

(b) $\operatorname{lsc} f = \sup\{g : g \le f, g \text{ lsc}\} =: h$

(c) $\operatorname{lsc} f$ *is the largest lsc function minorizing f, i.e. if g is lsc with $g \le f$ then $g \le \operatorname{lsc} f$.*

Proof. (b) We have

$$\operatorname{epi} h = \bigcap_{\substack{g \le f \\ g \text{ lsc}}} \operatorname{epi} g \supseteq \operatorname{epi} f \implies \operatorname{cl}\left(\bigcap_{\substack{g \le f \\ g \text{ lsc}}} \operatorname{epi} g\right) \supseteq \operatorname{cl}(\operatorname{epi} f) = \operatorname{epi}(\operatorname{lsc} f)$$

and so $h \le \operatorname{lsc} f$. The reverse inequality follows from (a). ☐

Proposition 2.11. *Assume that $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ is lsc and $K \subseteq \mathbb{R}^n$ is compact and non-empty. Then $\exists x^* \in K$ such that*

$$f(x^*) = \inf\{f(x) : x \in K\}.$$

Proof. Let $f_* = \inf\{f(x) : x \in K\}$. If $f_* = \infty$ then $f(x) = \infty$ for all $x \in K$ and the result trivially follows by picking any $x^* \in K$. Assume instead that $f_* < \infty$. Then $\exists\{x_k\} \subseteq K$ such that $f(x_k) \downarrow f_*$ and $x_k \to x^* \in K$. Since f is lsc then

$$f(x^*) = \liminf_{y\to x^*} f(y) \le \liminf_{k\to\infty} f(x_k) = f_*$$

and so $f(x^*) = f_*$. ☐

Definition 2.8. A function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is **0-coercive** if $\lim_{\|x\|\to\infty} f(x) = \infty$ or equivalently $\forall r \in \mathbb{R}$, $\exists M > 0$ such that $\|x\| > M \implies f(x) > r$. Also equivalently, $\forall r \in \mathbb{R}, \exists M > 0$ such that $x \in f^{-1}[-\infty, r] \implies \|x\| \le M$ or equivalently $\forall r \in \mathbb{R}, \exists M > 0$ such that $f^{-1}[-\infty, r] \subseteq \bar{B}(0; M)$ or equivalently $\forall r \in \mathbb{R}$, $f^{-1}[-\infty, r]$ is bounded.

Proposition 2.12. *Assume $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ is lsc and 0-coercive. Then $\exists x^* \in \mathbb{R}^n$ such that*

$$f(x^*) = \inf\{f(x) : x \in \mathbb{R}^n\}.$$

Proof. If $f = \infty$ then it is obvious. Assume $\exists x_0 \in \mathbb{R}^n$ such that $f(x_0) < \infty$. Let $r \in \mathbb{R}$ be such that $r \ge f(x_0)$. Then $K = f^{-1}[-\infty, r] \neq \emptyset$ so it is compact since f is lsc. By the previous result, $\exists x^* \in K$ such that

$$f(x^*) = \inf f(x) \quad \text{s.t. } x \in K. \qquad = \inf f(x) \quad \text{s.t. } x \in \mathbb{R}^n.$$

☐

2.2 Closure of Convex Functions

Definition 2.9. For $f \in$ E-Conv $\mathbb{R}^n$ the closure of f, denoted by $\operatorname{cl} f$ is defined as

$$\operatorname{cl} f = \begin{cases} \operatorname{lsc} f, & \text{if } f \in \text{Conv } \mathbb{R}^n \text{ or } f = \infty \\ -\infty, & \text{otherwise.} \end{cases}$$

Definition 2.10. f is **closed** if $f = \operatorname{cl} f$.

Notation 2. E-$\overline{\text{Conv}}\ \mathbb{R}^n$ is the set of all **closed convex functions**. $\overline{\text{Conv}}\ \mathbb{R}^n$ is the set of all **proper closed convex functions**.

Lemma 2.1. *For $f \in$ E-Conv $\mathbb{R}^n$,*

$$\operatorname{ri}(\operatorname{epi} f) = \{(x, r) \in \mathbb{R}^n \times \mathbb{R} : x \in \operatorname{ri}(\operatorname{dom} f), r > f(x)\}$$

Proposition 2.13. *Suppose $f \in$ E-Conv $\mathbb{R}^n$ and $x_0 \in \operatorname{ri}(\operatorname{dom} f)$. Then $\forall x \in \mathbb{R}^n$ we have*

$$(\operatorname{lsc} f)(x) = \lim_{t \downarrow 0} f(x + t(x_0 - x)).$$

Proof. Let $x_0 \in \operatorname{ri}(\operatorname{dom} f)$ and $x \in \mathbb{R}^n$ be given. We have

$$(\operatorname{lsc} f)(x) = \liminf_{y \to x} f(y) \leq \liminf_{t \downarrow 0} f(x + t(x_0 - x)).$$

Claim: $(\operatorname{lsc} f)(x) \geq \limsup_{t \downarrow 0} f(x + t(x_0 - x))$

Proof of Claim: Let $r \geq (\operatorname{lsc} f)(x)$. Then $(x, r) \in \operatorname{epi}(\operatorname{lsc} f) = \operatorname{cl}(\operatorname{epi} f) = \operatorname{cl} C$. Since $x_0 \in \operatorname{ri}(\operatorname{dom} f) \subseteq \operatorname{dom} f$, $\exists r_0 \in \mathbb{R}^n$ such that $f(x_0) < r_0$. By the previous lemma, $(x_0, r_0) \in \operatorname{ri}(\operatorname{epi} f) = \operatorname{ri} C$. By the resolution theorem, for all $t \in (0, 1]$ we have

$$(x + t(x_0 - x), r + t(r_0 - r)) = (x, r) + t\left[(x_0, r_0) - (x, r)\right] \in \operatorname{ri} C \subseteq C = \operatorname{epi} f.$$

So for all $t \in (0, 1]$,

$$f(x + t(x_0 - x)) \leq r + t(r_0 - r)$$

and as $t \downarrow 0$,

$$\limsup_{t \downarrow 0} f(x + t(x_0 - x)) \leq r.$$

□

Proposition 2.14. *Suppose that $f \in$ E-Conv $\mathbb{R}^n$. Then:*

(a) $f(x) = (\operatorname{lsc} f)(x)$ *for all* $x \in \mathbb{R}^n \setminus \operatorname{rbd}(\operatorname{dom} f)$

(b) $\operatorname{dom} f \subseteq \operatorname{dom}(\operatorname{lsc} f) \subseteq \operatorname{cl}(\operatorname{dom} f)$

Proof. (a) Either $x \in \operatorname{ri}(\operatorname{dom} f)$ or $x \notin \operatorname{cl}(\operatorname{dom} f)$. First assume that $x \in \operatorname{ri}(\operatorname{dom} f)$. Then by the proposition with $x_0 = x$ we have

$$(\operatorname{lsc} f)(x) = \lim_{t \downarrow 0} f(x + t(x - x)) = \lim_{t \downarrow 0} f(x) = f(x).$$

(b) Second, assume $x \notin \operatorname{cl}(\operatorname{dom} f) = \operatorname{int}(\mathbb{R}^n \setminus \operatorname{dom} f)$. Then $\exists \delta > 0$ such that $\bar{B}(x; \delta) \subseteq \mathbb{R}^n \setminus \operatorname{dom} f$. So $\forall y \in \bar{B}(x; \delta)$, $f(y) = \infty$ and hence

$$(\operatorname{lsc} f)(x) = \liminf_{y \to x} f(y) = \infty.$$

Hence, $(\operatorname{lsc} f)(x) = \infty = f(x)$ for all $x \notin \operatorname{cl}(\operatorname{dom} f)$. □

Corollary 2.2. *If $f \in$ Conv $\mathbb{R}^n$ then*

(a) $f(x) = (\operatorname{cl} f)(x)$ *for all* $x \in \mathbb{R}^n \setminus \operatorname{rbd}(\operatorname{dom} f)$

(b) $\operatorname{dom} f \subseteq \operatorname{dom}(\operatorname{cl} f) \subseteq \operatorname{cl}(\operatorname{dom} f)$

Corollary 2.3. *If $f \in$ Conv $\mathbb{R}^n$ and $\operatorname{dom} f$ is an affine manifold then $f \in \overline{\text{Conv}}\ \mathbb{R}^n$.*

Proof. Have $\operatorname{rbd}(\operatorname{dom} f) = \emptyset$. So by previous proposition $f = \operatorname{lsc} f = \operatorname{cl} f$. So f is closed. □

Proposition 2.15. *Suppose $f \in$ E-Conv $\mathbb{R}^n$ and $(\operatorname{lsc} f)(x_0) = -\infty$ for some $x_0 \in \mathbb{R}^n$ (e.g. $f(x_0) = -\infty$ for some $x_0 \in \mathbb{R}^n$). Then,*

(a) $(\operatorname{lsc} f)(x) = -\infty$ for all $x \in \operatorname{cl}(\operatorname{dom} f)$ and $\operatorname{dom}(\operatorname{lsc} f) = \operatorname{cl}(\operatorname{dom} f)$

(b) $f(x) = -\infty$ for all $x \in \operatorname{ri}(\operatorname{dom} f)$

As a consequence of (a) and (b), $\operatorname{cl} f, \operatorname{lsc} f$ agree on $\operatorname{cl}(\operatorname{dom} f)$ and $f, \operatorname{cl} f$ agree on $\operatorname{ri}(\operatorname{dom} f)$.

Proof. Claim: $(\operatorname{lsc})(x) = -\infty$ for all $x \in \operatorname{dom}(\operatorname{lsc} f)$.

If we assume the claim, then $\operatorname{epi}(\operatorname{lsc} f) = \operatorname{dom}(\operatorname{lsc} f) \times \mathbb{R}$. Since $\operatorname{lsc} f$ is lsc it follows that $\operatorname{epi}(\operatorname{lsc} f)$ is closed. Hence $\operatorname{dom}(\operatorname{lsc} f)$ is closed. By Corollary 2.14, using the closure on all of the sets in the inclusion in (b) of the Corollary, we have

$$\operatorname{cl}(\operatorname{dom} f) = \operatorname{cl}(\operatorname{dom}(\operatorname{lsc} f)) = \operatorname{dom}(\operatorname{lsc} f).$$

Also by Corollary 2.14, $\forall x \in \operatorname{ri}(\operatorname{dom} f)$, we have

$$f(x) = (\operatorname{lsc} f)(x) = -\infty$$

so part (b) holds.

Proof of Claim: By Proposition 2.15, $\{x_0\} \times \mathbb{R} \subseteq \operatorname{epi}(\operatorname{lsc} f)$. So $(0, -1) \in [\operatorname{epi}(\operatorname{lsc} f)]_\infty$. Let $x \in \operatorname{dom}(\operatorname{lsc} f)$. So $\exists r \in \mathbb{R}$ such that $(\operatorname{lsc} f)(x) < r$ or $(x, r) \in \operatorname{epi}(\operatorname{lsc} f)$. So $(x, r) + t(0, -1) \in \operatorname{epi}(\operatorname{lsc} f)$ for all $t > 0 \implies (\operatorname{lsc} f)(x) < r - t, \forall t > 0 \implies (\operatorname{lsc} f)(x) = -\infty$. □

Definition 2.11. The **convex hull** of denoted by $\operatorname{co} f$, is defined as

$$\operatorname{co} f = \sup\{g \in \text{E-Conv } \mathbb{R}^n : g \le f\}$$

Definition 2.12. The **closed convex hull** of $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$, denoted by $\overline{\operatorname{co}} f$, is defined as $\overline{\operatorname{co}} f = \operatorname{cl}(\operatorname{co} f)$.

Proposition 2.16. *(1) $\operatorname{co} f \in$ E-Conv $\mathbb{R}^n$, $\operatorname{co} f \le f$*

(2) if $g \in$ E-Conv $\mathbb{R}^n$, $g \le f$, then $g \le \operatorname{co} f$.

Proposition 2.17. *(1) $\overline{\operatorname{co}} f \in$ E-Conv $\mathbb{R}^n$, $\overline{\operatorname{co}} f \le f$*

(2) if $g \in$ E-$\overline{\text{Conv}}$ $\mathbb{R}^n$, $g \le f$, then $g \le \overline{\operatorname{co}} f$.

Proposition 2.18. *(1) $\operatorname{cl} f \in$ E-$\overline{\text{Conv}}$ $\mathbb{R}^n$*

(2) If $g \in$ E-$\overline{\text{Conv}}$ $\mathbb{R}^n$, $g \le f \implies g \le \operatorname{cl} f$.

2.3 Directional Derivatives

Definition 2.13. Let $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ and $\bar{x} \in \mathbb{R}^n$ such that $f(\bar{x}) \in \mathbb{R}$. The **directional derivative** of f at $\bar{x}$ along d is

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

whenever it exists where $\pm\infty$ is possible.

Definition 2.14. $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ is **differentiable** at $\bar{x}$ if $f(\bar{x}) \in \mathbb{R}$ and $\exists$ linear map $f'(\bar{x}) : \mathbb{R}^n \mapsto \mathbb{R}$ such that

$$\lim_{\substack{h \to 0 \\ h \in \mathbb{R}^n}} \frac{f(\bar{x} + h) - [f(\bar{x}) + f'(\bar{x})h]}{\|h\|} = 0.$$

Remark 2.2. (1) $f'(\bar{x})$ is unique

(2) f is differentiable at $\bar{x} \implies \bar{x} \in \operatorname{int}(\operatorname{dom} f)$.

(3) f is differentiable at $\bar{x} \implies f'(\bar{x}; d) = f'(\bar{x})d$.

Linear Algebra: The gradient is $T : \mathbb{R}^n \mapsto \mathbb{R}$ over inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^n$ where $\exists ! a \in \mathbb{R}^n$ such that $T(\cdot) = \langle a, \cdot \rangle$. In particular, $T = f'(\bar{x})$ and $f'(\bar{x})d = \langle a, d \rangle$.

Proposition 2.19. *Let $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ and $\bar{x} \in \mathbb{R}^n$ be such that $f(\bar{x}) \in \mathbb{R}$. If $\bar{x}$ is a local minimum of* $\inf\{f(x) : x \in \mathbb{R}^n\}$ *then*

$$f'(\bar{x}; d) \geq 0, \forall d \in \mathbb{R}^n$$

whenever it exists. As a consequence, if f is differentiable at $\bar{x}$ then $f'(\bar{x}) = 0$.

Proposition 2.20. *Assume $f \in$ E-Conv $\mathbb{R}^n$ and $\bar{x}, d \in \mathbb{R}^n$ are such that $f(\bar{x}) \in \mathbb{R}$. Define*

$$\Delta f(\cdot; x, d) : \mathbb{R}_{++} \mapsto \bar{\mathbb{R}}$$

as

$$\Delta f(t; \bar{x}, d) = \frac{f(\bar{x} + td) - f(\bar{x})}{t}.$$

Then,

(1) $\Delta f(\cdot; \bar{x}, d)$ is non-decreasing

(2) if $f(\cdot)$ is strictly convex and $d \neq 0$ then $\Delta f(\cdot; \bar{x}, d)$ is increasing

(3) if f is β-strongly convex, then for all $0 < t_1 < t_2$,

$$\Delta f(t_1) \leq \Delta f(t_2) - \frac{\beta}{2}(t_2 - t_1)\|d\|^2.$$

Proof. Let $0 < t_1 < t_2$ be given and suppose that f is β-strongly convex. Then

$$\bar{x} + t_1 d = \left(1 - \frac{t_1}{t_2}\right)\bar{x} + \frac{t_1}{t_2}(\bar{x} + td).$$

So,

$$f(\bar{x} + td_1) \leq \left(1 - \frac{t_1}{t_2}\right) f(\bar{x}) + \frac{t_1}{t_2} f(\bar{x} + t_2 d) - \frac{\beta}{2}\left(\frac{t_1}{t_2}\right)\left(1 - \frac{t_1}{t_2}\right)\|t_2 d\|^2$$
$$\implies \Delta f(t_1) \leq \Delta f(t_2) - \frac{\beta}{2}(t_2 - t_1)\|d\|^2$$

and similar arguments can be made for (a) and (b). □

Proposition 2.21. *Assume that $f \in$ E-Conv $\mathbb{R}^n$ and $\bar{x} \in \mathbb{R}^n$ such that $f(\bar{x}) \in \mathbb{R}$. Then,*

(a) $\forall d \in \mathbb{R}^n$, $f'(\bar{x}; d)$ exists and $f'(\bar{x}; d) = \inf_{t>0} \Delta f(t; \bar{x}, d)$

(b) $f(x) - f(\bar{x}) \geq f'(\bar{x}; x - \bar{x})$, $\forall x \in \mathbb{R}^n$

(c) $f(x) - f(\bar{x}) > f'(\bar{x}; x - \bar{x})$, $\forall x \in \mathbb{R}^n \backslash \{\bar{x}\}$ if f is strictly convex

(d) $f(x) - f(\bar{x}) \geq f'(\bar{x}; x - \bar{x}) + \frac{\beta}{2}\|x - \bar{x}\|^2, \forall x \in \operatorname{dom} f$ if f is β-strongly convex

Proof. (a) obvious

(b) $f(x) - f(\bar{x}) = \Delta f(1; \bar{x}, d) > \inf_{t>0} \Delta f(t; \bar{x}, d) = f'(\bar{x}; d)$ where $d = x - \bar{x}$

(c) $f(x) - f(\bar{x}) = \Delta f(1; \bar{x}, d) > \Delta f(t; \bar{x}, d) + \frac{\beta}{2}(1 - t)\|d\|^2$ and as $t \downarrow 0$ we get the RHS equal to $f'(\bar{x}; d) + \frac{\beta}{2}$. □

Note 1. $\lim_{t \to a^+} \phi(t) = \inf_{t>a} \phi(t)$ and $\lim_{t \to a^-} \phi(t) = \sup_{t<a} \phi(t)$ for $\phi(t)$ nondecreasing.

Proposition 2.22. *Assume that $f \in$ E-Conv $\mathbb{R}^n$ and $\bar{x} \in \mathbb{R}^n$ such that $f(\bar{x}) \in \mathbb{R}$. Then the following are equivalent:*

(a) $\bar{x}$ is a global min of $f(x)$ on $\mathbb{R}^n$

(b) $\bar{x}$ is a local min of $f(x)$ on $\mathbb{R}^n$

(c) $f'(\bar{x}; d) \geq 0$ for all $d \in \mathbb{R}^n$

(d) $f'(\bar{x}; x-\bar{x}) \geq 0$ *for all* $x \in \operatorname{dom} f$

If f *is differentiable at* $\bar{x}$ *then,*

(e) $f'(\bar{x}) = 0$

Proof. (a) $\implies$ (b) $\implies$ (c) $\implies$ (d) $\implies$ (a) by the previous proposition, part (b). □

Corollary 2.4. *Assume* f *is* β*-strongly convex and* $\bar{x}$ *is a global minimum of* f *over* $\mathbb{R}^n$*. Then:*

$$f(x) - f(\bar{x}) \geq \frac{\beta}{2}\|x-\bar{x}\|^2.$$

Definition 2.15. If $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ is proper and $\emptyset \neq C \subseteq \operatorname{dom} f$ is convex, we say f **is convex on** C if

$$f_C(x) = \begin{cases} f(x), & x \in C \\ +\infty, & \text{otherwise} \end{cases}$$

is convex.

Proposition 2.23. *Assume* $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ *is proper,* $\emptyset \neq C \subseteq \operatorname{dom} f$ *is convex, and* f *is convex on* C*. Then following are equivalent:*

(a) $\bar{x} \in C$ *is a global minimum of* f *over* C

(b) $\bar{x} \in C$ *is a local minimum of* f *over* C

(c) $f'(\bar{x}; d) \geq 0$ *for all* $d \in \mathbb{R}_+ \cdot (C - \bar{x})$

(d) $f'(\bar{x}; x-\bar{x}) \geq 0$ *for all* $x \in C$

Proof. (Outline) Assumptions imply f_C is convex.

Claim: We have

$$f_C'(\bar{x}; d) = \begin{cases} f'(\bar{x}; d), & d \in \mathbb{R}_+ \cdot (C - \bar{x}) \\ +\infty, & \text{otherwise.} \end{cases}$$

Proof of Claim: Exercise. □

Proposition 2.24. *Assume* $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ *is proper,* $\emptyset \neq C \subseteq \operatorname{dom} f$ *is convex, and* f *is strictly convex on* C*. Assume* $\bar{x}$ *is a global minimum of* f *over* C*. then* $\bar{x}$ *is the unique global minimum of* f *over* C*.*

Exercise. If $\emptyset \neq C$ is closed convex,

(1) $C_\infty = \{d \in \mathbb{R}^n : x + td \in C, \forall x \in C, \forall t > 0\}$

(2) $C_\infty = \{d \in \mathbb{R}^n : x + d \subseteq C, \forall x \in C\}$

(3) If $x_0 \in C$ then $C_\infty = \{d \in \mathbb{R}^n : x_0 + td \in C, \forall x_0 \in C, \forall t > 0\}$

2.4 Asymptotic Function

Definition 2.16. For $f \in \overline{\text{Conv}}\, \mathbb{R}^n$, its **asymptotic function** $f'_\infty : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ is defined as

$$f'_\infty(d) = \sup_{\substack{t>0 \\ x \in \operatorname{dom} f}} \frac{f(x+td) - f(x)}{t}.$$

Proposition 2.25. *For* $f \in \overline{\text{Conv}}\, \mathbb{R}^n$*, have:*

(a) $\operatorname{epi} f'_\infty = (\operatorname{epi} f)_\infty$

(b) If $x_0 \in \operatorname{dom} f$ *then*

$$f'_\infty(d) = \sup_{t>0} \underbrace{\frac{f'(x_0 + td) - f(x_0)}{t}}_{:=h_1(d)} \overset{(o)}{=} \sup_{x \in \operatorname{dom} f} \underbrace{f(x+d) - f(x)}_{:=h_2(d)}.$$

Proof. Claim: $\text{epi}\, f'_\infty = \text{epi}\, h_1 = \text{epi}\, h_2 = (\text{epi}\, f)_\infty$

Proof of Claim: If $(d,p) \in \text{epi}\, h$ then

$$\begin{aligned} & h_1(d) \le p \\ \iff & \frac{f(x_0+td)-f(x_0)}{t} \le p, \forall t>0 \\ \iff & f(x_0+td)-f(x_0) \le pt, \forall t>0 \\ \iff & (x_0+td, f(x_0)+pt) \in \text{epi}\, f, \forall t>0 \\ \iff & (x_0, f(x_0)) + t(d,p) \in \text{epi}\, f, \forall t>0 \\ \iff & (d,p) \in (\text{epi}\, f)_\infty \end{aligned}$$

The rest of the main proof is left as an exercise, using the different characterizations of C_∞. □

Proposition 2.26. *Let $f \in \overline{\mathit{Conv}}\, \mathbb{R}^n$. Then,*

(a) $f'_\infty \in \overline{\mathit{Conv}}\, \mathbb{R}^n$

(b) $f'_\infty(\alpha d) = \alpha f'_\infty(d)$ for all $\alpha \ge 0, d \in \mathbb{R}^n$

(c) $\forall r \in \mathbb{R}$ s.t. $f^{-1}[-\infty, r] \neq \emptyset$, we have $\left(f^{-1}[\infty, r]\right)_\infty = (f'_\infty)^{-1}[-\infty, 0]$.

Proof. (a) $\text{epi}(f'_\infty) = (\text{epi}\, f)_\infty$ was previously shown. Since the RHS is a closed convex cone, f'_∞ is a lsc convex function and also $f'_\infty(0) = 0$. So f'_∞ is proper.

Exercise. f is closed, $f(x_0) = -\infty$ for some $x_0 \in \mathbb{R} \implies f = -\infty$ on its closed domain.

(b) We have

$$f'_\infty(\alpha d) = \lim_{t\to\infty} \alpha \left[\frac{f(x_0+t\alpha d)-f(x_0)}{\alpha t}\right] = \alpha f'_\infty(d)$$

(c) $f^{-1}[-\infty, r] \times \{r\} \overset{\dagger}{=} (\text{epi}\, f) \cap \mathbb{R}^n \times \{r\}$ and hence

$$\begin{aligned} \left(f^{-1}[-\infty, r] \times \{r\}\right)_\infty &= (\text{epi}\, f)_\infty \cap (\mathbb{R}^n \times \{r\})_\infty \\ &= (\text{epi}\, f'_\infty) \cap (\mathbb{R}^n \times \{0\}) \\ &= f'_\infty[-\infty, 0] \times \{0\} \end{aligned}$$

from $(\dagger)$. Now the first expression is

$$\left(f^{-1}[-\infty, r] \times \{r\}\right)_\infty = \left(f^{-1}[-\infty, r]\right)_\infty \times \{0\}$$

and therefore (c) follows. □

Example 2.4. If $f = -\log x$ then

$$f'_\infty(d) = \begin{cases} 0, & \text{if } d \ge 0 \\ +\infty, & \text{if } d < 0. \end{cases}$$

Proposition 2.27. *Let $f \in \overline{\mathit{Conv}}\, \mathbb{R}^n$. Then the following are equivalent:*

(a) $\forall r \in \mathbb{R}$, $f^{-1}[-\infty, r]$ is bounded (i.e. f is coercive).

(b) $\exists r_0 \in \mathbb{R}$ s.t. $f^{-1}[-\infty, r_0] \neq \emptyset$ and bounded.

(c) the set of optimal solutions of $\min_{x\in\mathbb{R}^n} f(x) \neq \emptyset$ and bounded.

(d) $f'_\infty(d) > 0, \forall d \in \mathbb{R}^n \setminus \{0\}$.

Proof. (a) $\implies$ (c) Already done

(c) $\implies$ (b) Take $r_0 = \inf\{f(x) : x \in \mathbb{R}^n\}$. Then $f^{-1}[-\infty, r_0] =$ set of optimal solutions $\neq \emptyset$. Hence, $f^{-1}[-\infty, r_0]$ is $\neq \emptyset$ and bounded

(b) $\implies$ (d) Since (b), then $f^{-1}[-\infty, r]$ is $\neq \emptyset$ and bounded $\iff$ $(f^{-1}[-\infty, r_0])_\infty = \{0\}$ $\iff$ $(f'_\infty)^{-1}[-\infty, 0] = \{0\}$.

(d) $\implies$ (c) Exercise. □

Extension. The above can be generalized to optimal solutions over a closed and convex set $C \subseteq \mathbb{R}^n$ where for (d) we require $f'_\infty(d)$ for $d \in C_\infty \backslash \{0\}$. This can be done by considering $f_C(x) = f(x) + I_C(x)$.

Proposition 2.28. *(1) If $f_1, ..., f_k \in \overline{Conv}\, \mathbb{R}^n$ such that $\bigcap_{i=1}^k \operatorname{dom} f_i \neq \emptyset$ then for all $\alpha_1, ..., \alpha_k \geq 0$*

$$(\alpha_1 f_1 + ... + \alpha_k f_k)'_\infty = \alpha_1 (f_1)'_\infty + ... + \alpha_k (f_k)'_\infty$$

and $\alpha_1 f_1 + ... + \alpha_k f_k \in \overline{Conv}\, \mathbb{R}^n$.

(2) If $\{f_i\}_{i\in I} \subseteq \overline{Conv}\, \mathbb{R}^n$ such that $\sup_{i\in I} f_i(x_0) < \infty$ for some $x_0 \in \mathbb{R}^n$ then $f := \sup_{i\in I} f_i \in \overline{Conv}\, \mathbb{R}^n$ and $f'_\infty = \sup_{i\in I} (f_i)'_\infty$.

(3) If $f \in \overline{Conv}\, \mathbb{R}^n$, $A : \mathbb{R}^n \mapsto \mathbb{R}^m$ affine such that $A(\mathbb{R}^n) \cap \operatorname{dom} f \neq \emptyset$ then $f \circ A \in \overline{Conv}\, \mathbb{R}^n$ and

$$(f \circ A)'_\infty = f'_\infty \circ (A_0) \text{ where } A_0(\cdot) = A(\cdot) - A(0).$$

Exercise 2.1. $(I_C)'_\infty = I_{C_\infty}$.

Corollary 2.5. *We have*

$$(f_C)'_\infty(d) = (f + I_C)'_\infty(d) = f'_\infty(d) + (I_C)'_\infty = f'_\infty(d) + I_{C_\infty}(d).$$

2.5 Differentiable Functions

Proposition 2.29. *Let $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ be differentiable on a nonempty convex set $C \subseteq \operatorname{dom} f$. Then the following are equivalent:*

(a) f is convex on C, i.e.

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y), \forall x, y \in C, \alpha \in (0,1)$$

(b) $f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle$, $\forall x, y \in C$

(c) $[f'(y) - f'(x)]\,(y - x) \geq 0$, $\forall x, y \in C$.

Corollary 2.6. *Assume $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ is differentiable on a nonempty convex set $C \subseteq \operatorname{dom} f$. Then for all $\forall \beta \in \mathbb{R}$, the following are equivalent,*

(a) $\forall x, y \in C, \forall \alpha \in (0,1)$ we have

$$f(\alpha x + (1-\alpha)y) + \frac{\beta}{2}\alpha(1-\alpha) \leq \alpha f(x) + (1-\alpha) f(y)$$

(b) $f - \frac{\beta}{2}\|\cdot\|^2$ is convex.

(c) $\forall x, y \in C$, $f(y) \geq f(x) + f'(x)(y-x) + \frac{\beta}{2}\|y - x\|^2$

(d) $\forall x, y \in C$, $[f'(y) - f'(x)]\,(y - x) \geq \beta\|y - x\|^2$.

Corollary 2.7. *Assume $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ is differentiable on a nonempty convex set $C \subseteq \operatorname{dom} f$. Then $\forall L \in \mathbb{R}$ the following are equivalent:*

(a) $\forall x, y \in C, \forall \alpha \in (0,1)$ we have

$$f(\alpha x + (1-\alpha)y) + \frac{L}{2}\alpha(1-\alpha) \geq \alpha f(x) + (1-\alpha) f(y)$$

(b) $\frac{L}{2}\|\cdot\|^2 - f$ is convex.

(c) $\forall x, y \in C$, $f(y) \leq f(x) + f'(x)(y-x) + \frac{L}{2}\|y - x\|^2$

(d) $\forall x, y \in C$, $[f'(y) - f'(x)]\,(y - x) \leq L\|y - x\|^2$.

3 Separation Theory

Definition 3.1. Let $\emptyset \neq C \subseteq \mathbb{R}^n$ closed convex and inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^n$. Define the **projection operator** $\Pi_C : \mathbb{R}^n \mapsto C$ onto C as

$$x \mapsto \Pi_C(x) := \operatorname*{argmin}_{c \in C} \|c - x\| = \operatorname*{argmin}_{c \in C} \underbrace{\frac{1}{2}\|c - x\|^2}_{:= f_x(c)}$$

which is 1-strongly convex.

Proposition 3.1. $\bar{c} = \Pi_C(x) \iff \langle c - \bar{c}, x - \bar{c} \rangle \leq 0$ *for all* $c \in C$.

Proof. $\bar{c} = \Pi_C(x) \iff \bar{c}$ is an optimal solution of $\operatorname{argmin}_{c \in C} \frac{1}{2}\|c - x\|^2 \iff \langle \nabla f_x(\bar{c}), c - \bar{c} \rangle \geq 0$ for all $c \in C \iff \langle \bar{c} - x, c - \bar{c} \rangle \geq 0$ for all $c \in C$. □

Proposition 3.2. *For every* $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$*,*

$$\|\Pi_C(x) - \Pi_C(y)\|^2 \leq \langle x - y, \Pi_C(x) - \Pi_C(y) \rangle$$

and as a consequence,

$$\|\Pi_C(x) - \Pi(y)\| \leq \|x - y\|.$$

Proof. We have

$$\langle \Pi_C(y) - \Pi_C(x), x - \Pi_C(x) \rangle \leq 0$$
$$\langle \Pi_C(x) - \Pi_C(y), y - \Pi_C(y) \rangle \leq 0$$

and summing the two gives

$$\|\Pi_C(x) - \Pi_C(y)\|^2 \leq \langle x - y, \Pi_C(x) - \Pi_C(y) \rangle .$$

□

3.1 Hyperplanes

Definition 3.2. Let $C_1, C_2 \subseteq \mathbb{R}^n$ be nonempty an H be a hyperplane.

(a) H **separates** C_1, C_2 if $C_1 \subseteq H^{\leq}$ and $C_2 \subseteq H^{\geq}$.

(b) H **properly separates** C_1, C_2 if H separates them and $C_1 \cup C_2 \not\subseteq H$.

(c) H **strongly separates** C_1, C_2 if H separates $C_1 + \bar{B}(0; \delta_1), C_2 + \bar{B}(0; \delta_2)$ for some $\delta_1, \delta_2 > 0$.

Proposition 3.3. *Let* $\emptyset \neq C_1, C_2 \subseteq \mathbb{R}^n$ *be given*

(a) $\exists$ *hyperplane separating* $C_1, C_2 \iff \exists 0 \neq s \in \mathbb{R}^n$ *s.t.* $\sup_{x_1 \in C} \langle s, x_1 \rangle \leq \inf_{x_2 \in C} \langle s, x_2 \rangle \qquad (*)$.

(b) $\exists$ *hyperplane properly separating* $C_1, C_2 \iff \exists s \in \mathbb{R}^n$ *s.t. (*) holds and* $\inf_{x_1 \in C} \langle s, x_1 \rangle < \sup_{x_2 \in C} \langle s, x_2 \rangle$.

(c) $\exists$ *hyperplane strongly separating* $C_1, C_2 \iff \exists s \in \mathbb{R}^n$ *s.t. (*) holds strictly.*

Proposition 3.4. *Let* $\emptyset \neq C_1, C_2 \subseteq \mathbb{R}^n$ *be given. Then* C_1, C_2 *can be separated* $\iff \{0\}, C = C_1 - C_2$ *can be separated.*

Proposition 3.5. *Let* $\emptyset \neq C \subseteq \mathbb{R}^n$ *be a convex set and* $x \in \mathbb{R}^n$*. Then,*

(a) x, C *(*C_1, C_2*) can be strongly separated* $\iff x \notin \operatorname{cl} C$ $(0 \notin \operatorname{cl}(C_1 - C_2))$

(b) x, C *(*C_1, C_2*) can be properly separated* $\iff x \notin \operatorname{ri} C$ $(0 \notin \operatorname{ri}(C_1 - C_2))$.

Proof. (a) ($\implies$) Easy. Left as an exercise.

$(\Longleftarrow)$ Assume $x \notin \operatorname{cl} C = \tilde{C}$. Take $s = x - \Pi_{\tilde{C}}(x) \neq 0$. We have

$$\begin{aligned} & \langle c - \Pi_{\tilde{C}}(x), x - \Pi_{\tilde{C}}(x) \rangle \leq 0, \forall c \\ \iff & \langle c - x + s, s \rangle \leq 0, \forall c \\ \iff & \langle c, s \rangle \leq \langle x, s \rangle - \|s\|^2, \forall c \\ \iff & \langle c, s \rangle < \langle x, s \rangle, \forall c \\ \implies & \sup_{c \in C} \langle c, s \rangle \leq \langle x, s \rangle . \end{aligned}$$

(b) $(\Longrightarrow)$ Exercise.

$(\Longleftarrow)$ Assume $x \notin \operatorname{ri} C$. Either $x \notin \operatorname{cl} C$ (strongly separable) or $x \in \operatorname{rbd} C$. We want to check the latter case. Note that $\operatorname{rbd}(C) = \operatorname{rbd}(\operatorname{cl} C)$. Now, we have

$$x \in \operatorname{rbd}(\operatorname{cl} C) = \operatorname{rbd}(\operatorname{aff} C \setminus \operatorname{cl} C) \subseteq \operatorname{cl}(\operatorname{aff} C \setminus \operatorname{cl} C).$$

and so $\exists$ a sequence $\{x_k\} \subseteq \operatorname{aff} C \setminus \operatorname{cl} C$ where $x_k \to x$. So $\exists s_k \neq 0$ s.t. $\langle s_k, c \rangle \leq \langle s_k, x_k \rangle$ and $\|s_k\| = 1$ for all $c \in C$. By Bolzano-Weierstrass, $\exists K$ such that $\{s_k\}_{k \in K}$. So as $k \in K \to \infty$ we conclude that

$$\langle s, c \rangle \leq \langle s, x \rangle, \forall c \in C \iff \sup_{c \in C} \langle s, c \rangle \leq \langle s, x \rangle .$$

Assume for contradiction that

$$\begin{aligned} & \inf_{c \in C} \langle s, c \rangle = \langle s, x \rangle \\ \iff & \langle s, c - x \rangle = 0, \forall c \in C \\ \iff & \langle s, u \rangle = 0, \forall u \in C - x \\ \iff & \langle s, u \rangle = 0, \forall u \in \text{lin } (C - x) = \operatorname{aff}(C - x) =: L \end{aligned}$$

where L is a subspace parallel to $\operatorname{aff} C$. However, $s_k = x_k - \Pi_C(x_k) \in L$ where L is closed. Hence $s \in L$. So $s = 0$ since $s \in L^\perp$ as well which is impossible as $\|s\| = 1$. □

Proposition 3.6. *Let $\emptyset \neq C \subseteq \mathbb{R}^n$ be a convex set and $x \in \mathbb{R}^n$. Then,*

$$\operatorname{cl} C = \bigcap \left\{ H^{\leq} : H \textit{ is a hyperplane}, C \subseteq H^{\leq} \right\} .$$

Proof. $(\subseteq)$ Obvious

$(\supseteq)$ Assume $x \notin \operatorname{cl} C$. Then exists hyperplane H such that $C \subseteq H^{\leq}$ and $x \in H^{\geq} = \mathbb{R}^n \setminus H^{\leq}$. So $x \in \bigcap \left\{ H^{\leq} : H \text{ is a hyperplane}, C \subseteq H^{\leq}\right\}$ □

Corollary 3.1. *If $f \in$ E-Conv $\mathbb{R}^n$ then*

$$\operatorname{epi}(\operatorname{lsc} f) = \operatorname{cl}(\operatorname{epi} f) = \bigcap \left\{ H^{\leq} : H \textit{ is a hyperplane}, \operatorname{epi} f \subseteq H^{\leq} \right\} .$$

Remark 3.1. A closed halfspace has one of the following representations:

(1) $H^+(s, \beta) = \{(x, t) : \langle s, t \rangle + t \leq \beta\}$

(2) $H^-(s, \beta) = \{(x, t) : \langle s, t \rangle - t \leq \beta\}$

(3) $H^0(s, \beta) = \{(x, t) : \langle s, t \rangle \leq \beta\}$

Observe that

(1) $H^+(s, \beta)$ is **not** an epigraph

(2) $H^-(s, \beta) = \operatorname{epi}(\langle s, \cdot \rangle - \beta)$

(3) $H^0(s, \beta) = H^{\leq}_{s,\beta} \times \mathbb{R}$

Proposition 3.7. *If $f \in$ E-Conv $\mathbb{R}^n$ then*

$$\begin{aligned}\operatorname{cl} f &= \sup\{A : A \text{ is affine}, A \le f\}\\ &= \sup_{(s,\beta)}\{\langle s,\cdot\rangle - \beta : \langle s,\cdot\rangle - \beta \le f\}.\end{aligned}$$

Also if $f \in$ Conv $\mathbb{R}^n$ then $\exists$ affine function minorizing f.

Proof. If $f = +\infty$ or $f(x^0) = -\infty$ for some $x^0 \in \mathbb{R}$ then the claim clearly holds. Assume now that $f \in \operatorname{Conv} \mathbb{R}^n$. Define

$$\begin{aligned}\Sigma^- &= \left\{(s,\beta) : H^-(s,\beta) \supseteq \operatorname{epi} f\right\}\\ \Sigma^0 &= \left\{(s,\beta) : H^0(s,\beta) \supseteq \operatorname{epi} f\right\}.\end{aligned}$$

So by the previous result

$$\begin{aligned}\operatorname{epi}(\operatorname{cl} f) = \operatorname{epi}(\operatorname{lsc} f) = \operatorname{cl}(\operatorname{epi} f) &= \bigcap\left\{H^{\le} : H \text{ is a hyperplane}, \operatorname{epi} f \subseteq H^{\le}\right\}\\ &= \left(\bigcap_{(s,\beta)\in\Sigma^-} H^-(s,\beta)\right) \cap \left(\bigcap_{(s,\beta)\in\Sigma^0} H^0(s,\beta)\right).\end{aligned}$$

<u>Claim</u>. $\bigcap_{(s,\beta)\in\Sigma^-} H^-(s,\beta) \subseteq \bigcap_{(s,\beta)\in\Sigma^0} H^0(s,\beta)$.

By the claim,

$$\begin{aligned}\operatorname{epi}(\operatorname{cl} f) = \bigcap_{(s,\beta)\in\Sigma^-} H^-(s,\beta) &= \bigcap_{(s,\beta)\in\Sigma^-} \operatorname{epi}\left(\langle s,\cdot\rangle - \beta\right)\\ &= \operatorname{epi}\left(\sup_{(s,\beta)\in\Sigma^-} \langle s,\cdot\rangle - \beta\right)\\ &= \operatorname{epi}\left(\sup_{\substack{(s,\beta)\\ \langle s,\cdot\rangle - \beta \le f}} \langle s,\cdot\rangle - \beta\right)\end{aligned}$$

and the conclusion follows from the fact that $\left(\bigcap_{(s,\beta)\in\Sigma^-} H^-(s,\beta)\right) \ne \emptyset$ (*exercise*). Note that if $\left(\bigcap_{(s,\beta)\in\Sigma^-} H^-(s,\beta)\right) = \emptyset$ then $\operatorname{cl}(\operatorname{epi} f) = \left(\bigcap_{(s,\beta)\in\Sigma^0} H^0(s,\beta)\right) = D \times \mathbb{R}$ for some set D.

<u>Proof of the claim</u>. (Sketch) Rotate the vertical hyperplane so that it becomes non-vertical using convex combinations. □

3.2 Conjugate Functions

Definition 3.3. The **conjugate** of $f : \mathbb{R}^n \mapsto \mathbb{R}$, denoted by f^*, is defined as $f^* : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ where

$$s \mapsto f^*(s) = \sup_{x\in\mathbb{R}^n} \langle x, s\rangle - f(x).$$

Observe that $\forall s \in \mathbb{R}^n$ we have

$$f^*(s) = \sup_{x\in\operatorname{dom} f} \langle x, s\rangle - f(x) = \sup_{(x,t)\in\operatorname{epi} f} \langle x, s\rangle - t.$$

Proposition 3.8. *We have:*

(a) if $f = \infty$ then $f^* = -\infty$

(b) if $f(x_0) = -\infty$ for some x_0 then $f^* = \infty$

(c) $\operatorname{epi} f^* = \{(s,\beta) : \langle s,\cdot\rangle - \beta \le f\}$

(d) $f^*(s) = \inf\{\beta : \langle s,\cdot\rangle - \beta \le f\}$

(e) $-f^*(0) = \inf\{f(x) : x \in \mathbb{R}^n\}$

(f) $\forall x, s \in \mathbb{R}^n$, $f^*(s) \geq \langle x, s\rangle - f(x)$

Proof. (a) obvious

(b) obvious

(c) directly,

$$\begin{aligned}(s,\beta) \in \operatorname{epi} f^* &\iff f^*(s) \leq \beta \\ &\iff \sup_x \langle x, s\rangle - f(x) \leq \beta \\ &\iff \langle x, s\rangle - \beta \leq f(x), \forall x \\ &\iff \langle \cdot, s\rangle - \beta \leq f.\end{aligned}$$

(d) directly,

$$f^*(s) = \inf\{\beta : (s,\beta) \in \operatorname{epi} f^*\} \overset{(c)}{=} \inf\{\beta : \langle s, \cdot\rangle - \beta \leq f\}.$$

(e) obvious

(f) obvious ☐

Proposition 3.9. *For any* $f \in$ *E-Conv* $\mathbb{R}^n$,

$$f^* = (\operatorname{cl} f)^* = (\operatorname{lsc} f)^*.$$

Proof. Let $A = \langle s, \cdot\rangle - \beta$. Then $A \leq f \iff A \leq \operatorname{lsc} f \iff A \leq \operatorname{cl} f$. ☐

Definition 3.4. Fenchel's inequality is

$$f^*(s) \geq \langle x, s\rangle - f(x).$$

Proposition 3.10. *Let* $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ *be such that*

(1) $f \neq \infty$

(2) f *is minorized by an affine function*

Then, $f^* \in \overline{\textit{Conv}}\, \mathbb{R}^n$. *As a consequence, if* $f \in$ *Conv* $\mathbb{R}^n$ *then* $f^* \in \overline{\textit{Conv}}\, \mathbb{R}^n$.

Proof. Since

$$\operatorname{epi} f^* = \{(s,\beta) : \langle s, \cdot\rangle - \beta \leq f\}$$

it follows from (2) that $\operatorname{epi} f^* \neq \emptyset$. Also, since

$$f^* = \sup_{x \in \operatorname{dom} f} \langle x, \cdot\rangle - f(x)$$

and $\operatorname{dom} f \neq \emptyset$ due to (1), then $f^* \neq \infty$ and $f^* \in \overline{\text{Conv}}\, \mathbb{R}^n$. ☐

Proposition 3.11. *Assume that* $f \in$ *E-Conv* $\mathbb{R}^n$. *Then*

$$\operatorname{cl} f = f^{**} = (f^*)^*.$$

Proof. Here,

$$\begin{aligned}\operatorname{cl} f &= \sup\{A : A \text{ affine}, A \leq f\} \\ &= \sup_{(s,\beta)} \{\langle s, \cdot\rangle - \beta : \langle s, \cdot\rangle - \beta \leq f\} \\ &= \sup_{(s,\beta)} \{\langle s, \cdot\rangle - \beta : \langle s, \beta\rangle \leq \operatorname{epi} f^*\} \\ &= f^{**}.\end{aligned}$$

☐

3.3 Subgradients

Definition 3.5. We say $s \in \partial f(\bar{x})$ where ∂f is the subgradient of f if and only if

$$f(x) \geq f(\bar{x}) + \langle s, x - \bar{x} \rangle, \forall x \in \mathbb{R}^n.$$

Remark 3.2. We have

- $f(\bar{x}) = +\infty \implies \partial f(\bar{x}) = \mathbb{R}^n$
- $f(\bar{x}) = +\infty$ then $\partial f(\bar{x}) \neq \emptyset \iff f = +\infty$ in which case $\partial f(\bar{x}) = \mathbb{R}^n$.

Assumption. *(A) $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ and $\bar{x} \in \mathbb{R}^n$ such that $f(\bar{x}) \in \mathbb{R}$.*

Proposition 3.12. *If (A) holds then*

(a) $\bar{x}$ is a global minimum of f over $\mathbb{R}^n \iff 0 \in \partial f(\bar{x})$.

(b) $\partial f(\bar{x})$ is a (possibly empty) closed convex set.

Proposition 3.13. *Assume that $f \in$ E-Conv $\mathbb{R}^n$ and $\bar{x} \in \mathbb{R}^n$ such that $f(\bar{x}) \in \mathbb{R}$. Then,*

$$\partial f(\bar{x}) = \{s \in \mathbb{R}^n : \langle s, \cdot \rangle \leq f'(\bar{x}; \cdot)\}$$

and also

$$\operatorname{cl} f'(\bar{x}; \cdot) = \sigma_{\partial f(\bar{x})} = \sup_{s \in \partial f(\bar{x})} \langle s, \cdot \rangle.$$

Proof. Directly

$$\begin{aligned} s \in \partial f(\bar{x}) &\iff f(x) \geq f(\bar{x}) + \langle s, x - \bar{x} \rangle, \forall x \\ &\iff f(\bar{x} + td) \geq f(\bar{x}) + t \langle s, d \rangle, \forall d, \forall t > 0 \\ &\iff \inf_{t>0} \frac{f(\bar{x} + td) - f(\bar{x})}{t} \geq \langle s, d \rangle, \forall d \\ &\iff f'(\bar{x}; d) \geq \langle s, d \rangle, \forall d. \end{aligned}$$

□

Proposition 3.14. *Let $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ and $\bar{x} \in \mathbb{R}^n$ be given. Then $s \in \partial f(x) \iff f^*(s) \leq \langle x, s \rangle - f(x)$.*

Proof. Directly,

$$\begin{aligned} s \in \partial f(x) &\iff f(\tilde{x}) \geq f(x) + \langle s, \tilde{x} - x \rangle, \forall \tilde{x} \\ &\iff \langle s, x \rangle - f(x) \geq \langle s, \tilde{x} \rangle - f(\tilde{x}), \forall \tilde{x} \\ &\iff \langle s, x \rangle - f(x) \geq \sup_{\tilde{x}} \langle s, \tilde{x} \rangle - f(\tilde{x}) \\ &\iff \langle s, x \rangle - f(x) \geq f^*(s). \end{aligned}$$

□

Definition 3.6. For a multivalued map $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, define $A^{-1}(y) = \{x : y \in A(x)\}$.

Lemma 3.1. *Let $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ and $\bar{x} \in \mathbb{R}^n$ such that $\partial f(\bar{x}) \neq \emptyset$ be given. Then:*

(a) $(\operatorname{lsc} f)(\bar{x}) = f(\bar{x})$*, i.e. f is lsc at $\bar{x}$*

(b) If $f \in$ E-Conv $\mathbb{R}^n$ then $(\operatorname{cl} f)(\bar{x}) = f(\bar{x})$.

Proof. (a) Assume $f(\bar{x}) \in \mathbb{R}$ and let $s \in \partial f(\bar{x}) \neq \emptyset$. Then $A(\cdot) = \langle s, \cdot - \bar{x} \rangle + f(\bar{x})$ minorizes f. Since $A \leq f$, A is lsc, have $A \leq \operatorname{lsc} f \leq f$. So

$$f(\bar{x}) = A(\bar{x}) \leq (\operatorname{lsc} f)(\bar{x}) \leq f(\bar{x}).$$

(b) Exercise. □

Proposition 3.15. *Let $f \in$ E-Conv $\mathbb{R}^n$ and $x \in \mathbb{R}^n$ be given. Then for $s \in \mathbb{R}^n$ the following are equivalent*

(a) $s \in \partial f(x)$

(b) $s \in \partial(\mathrm{cl}\, f)(x)$ *and* $(\mathrm{cl}\, f)(x) = f(x)$

(c) $x \in \partial f^*(s)$ *and* $(\mathrm{cl}\, f)(x) = f(x)$.

Proof. We have

$$(a) \iff \begin{cases} f^*(s) \le \langle x, s\rangle - f(x) \\ (\mathrm{cl}\, f)(x) = f(x) \end{cases}$$
$$\iff \begin{cases} (\mathrm{cl}\, f)^*(s) = f^*(s) \le \langle x, s\rangle - (\mathrm{cl}\, f)(x) \\ (\mathrm{cl}\, f)(x) = f(x) \end{cases}$$
$$\iff \begin{cases} s \in \partial(\mathrm{cl}\, f)(x) \\ (\mathrm{cl}\, f)(x) = f(x) \end{cases} \iff (b)$$

Next,

$$(a) \iff \begin{cases} (\mathrm{cl}\, f)^*(s) = f^*(s) \le \langle x, s\rangle - (\mathrm{cl}\, f)(x) \\ (\mathrm{cl}\, f)(x) = f(x) \end{cases}$$
$$\iff \begin{cases} f^*(s) \le \langle x, s\rangle - f^{**}(x) \\ (\mathrm{cl}\, f)(x) = f(x) \end{cases}$$
$$\iff \begin{cases} f^{**}(x) \le \langle x, s\rangle - f^*(s) \\ (\mathrm{cl}\, f)(x) = f(x) \end{cases}$$
$$\iff \begin{cases} x \in \partial f^*(s) \\ (\mathrm{cl}\, f)(x) = f(x) \end{cases} \iff (c)$$

□

Corollary 3.2. *If $f \in$ E-Conv $\mathbb{R}^n$ then $s \in \partial f(x) \iff x \in \partial f^*(s)$.*

Corollary 3.3. *If $f \in \overline{\text{Conv}}\, \mathbb{R}^n$ then $\partial f^*(0) = \operatorname{argmin}_{x \in \mathbb{R}^n} f(x)$.*

3.4 Sublinear Functions

Definition 3.7. $\sigma : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ is **sublinear** if $\mathrm{epi}\, \sigma$ is a convex cone.

Definition 3.8. $\sigma : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ is **subadditive** if $\sigma(x_0 + x_1) \le \sigma(x_0) + \sigma(x_1)$ and is **positively homogeneous** (of degree 1) if $\sigma(tx) = t\sigma(x)$ for all $t > 0$ and for all $x \in \mathbb{R}^n$.

Proposition 3.16. *Let $\sigma : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$. Then the following are equivalent:*

(a) σ is sublinear

(b) σ is convex and positively homogeneous

(c) σ is subadditive and positively homogeneous

(d) $\sigma(t_0 x_0 + t_1 x_1) \le t_0 \sigma(x_0) + t_1 \sigma(t_1)$ for all $t_0, t_1 > 0$ and for all $x_0, x_1 \in \mathrm{dom}\, \sigma$

Proposition 3.17. *Let $\sigma : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ be sublinear. Then,*

(a) $\mathrm{dom}\, \sigma$ is a convex cone

(b) $\sigma(0) \in \{-\infty, 0, +\infty\}$

(c) if σ is proper then $\sigma(x) + \sigma(-x) \ge \sigma(0) \ge 0$

(d) if σ is proper closed then $\sigma(0) = 0$

Proof. (a) Exercise

(b) Follows from the fact that $t\sigma(0) = \sigma(t \cdot 0) = \sigma(0)$ for all $t > 0$

(c) Obvious

(d) Let $x_0 \in \operatorname{dom} \sigma \neq \emptyset$. Then

$$\begin{aligned}\sigma(0) = (\operatorname{cl} \sigma)(0) = (\operatorname{lsc} \sigma)(0) &= \liminf_{x \to 0} \sigma(x) \\ &\leq \liminf_{t \to 0} \sigma(tx_0) = \liminf_{t \to 0} t \underbrace{\sigma(x_0)}_{\in \mathbb{R}} = 0.\end{aligned}$$

□

Example 3.1. For $C \subseteq \mathbb{R}^n$, the support function $\sigma_C \in \overline{\text{Conv}}\, \mathbb{R}^n$ for $C \neq \emptyset$ is a closed sublinear function.

Remark 3.3. $\sigma_C(s) = \sup_{x \in C} \langle s, x \rangle = (I_C)^*$.

For the following results, let Σ be a map from the set of closed convex sets to the set of closed sublinear functions σ such that $\sigma \neq \infty$.

Claim: Σ is a bijection

Proposition 3.18. *For any $C \subseteq \mathbb{R}^n$ we have*

$$\begin{aligned}\operatorname{lsc} I_C &= I_{\operatorname{cl} C} \\ \operatorname{co} I_C &= I_{\operatorname{co} C} \\ \overline{\operatorname{co}} I_C &= I_{\overline{\operatorname{co}} C}.\end{aligned}$$

Proof. Exercise. □

Proposition 3.19. *For any $C \subseteq \mathbb{R}^n$ we have*

$$\sigma_C = \sigma_{\operatorname{cl} C} = \sigma_{\operatorname{co} C} = \sigma_{\overline{\operatorname{co}} C}.$$

Proof. Fact. $f^* = (\operatorname{lsc} f)^* = (\operatorname{co} f)^* = (\overline{\operatorname{co}} f)^*$. Take $f = I_C$ to get

$$(I_C)^* = (I_{\operatorname{cl} C})^* = (I_{\operatorname{co} C})^* = (I_{\overline{\operatorname{co}} C})^*.$$

□

Proposition 3.20. *Let $C_1, C_2 \subseteq \mathbb{R}^n$ be closed convex. Then,*

$$C_1 \subseteq C_2 \iff \sigma_{C_1} \leq \sigma_{C_2}$$

and in particular, $C_1 = C_2 \iff \sigma_{C_1} = \sigma_{C_2}$.

Proof. (Fact: $f \leq g \implies f^* \geq g^*$) ($\implies$) $C_1 \subseteq C_2 \implies I_{C_1} \geq I_{C_2} \implies (I_{C_1})^* \leq (I_{C_2})^* \implies \sigma_{C_1} \leq \sigma_{C_2}$.
($\impliedby$) $\sigma_{C_1} \leq \sigma_{C_2} \iff (I_{C_1})^* \leq (I_{C_2})^* \implies (I_{C_1})^{**} \geq (I_{C_2})^{**}$ and now

$$\begin{aligned}&(I_{C_1})^{**} \geq (I_{C_2})^{**} \\ \iff& \overline{\operatorname{co}} I_{C_1} \geq \overline{\operatorname{co}} I_{C_2} \\ \iff& I_{\overline{\operatorname{co}} C_1} \geq I_{\overline{\operatorname{co}} C_2} \\ \iff& \overline{\operatorname{co}} C_1 \subseteq \overline{\operatorname{co}} C_2 \\ \iff& C_1 \subseteq C_2.\end{aligned}$$

□

Corollary 3.4. *Σ is one-to-one.*

Corollary 3.5. *For any $C \subseteq \mathbb{R}^n$,*

$$\overline{\operatorname{co}} C = \{x \in \mathbb{R}^n : \langle x, \cdot \rangle \leq \sigma_C(\cdot)\}.$$

Proof. $\sigma_{\{x\}} = \langle x, \cdot \rangle$. Hence,

$$\begin{aligned} x \in \overline{\mathrm{co}} C &\iff \{x\} \subseteq \overline{\mathrm{co}} C \\ &\iff \sigma_{\{x\}} \le \sigma_{\overline{\mathrm{co}} C} \\ &\iff \langle x, \cdot \rangle \le \sigma_C. \end{aligned}$$

□

Proposition 3.21. *(Σ is onto) If σ is a closed sublinear function such that $\sigma \neq \infty$ then $\sigma = \sigma_C$ where*

$$C = C(\sigma) = \{x \in \mathbb{R}^n : \langle x, \cdot \rangle \le \sigma\}.$$

By the previous result,

$$C = \overline{\mathrm{co}} C = \{x \in \mathbb{R}^n : \langle x, \cdot \rangle \le \sigma_C\} = \{x \in \mathbb{R}^n : \langle x, \cdot \rangle \le \sigma\} = C(\sigma).$$

Proof. If $\sigma = -\infty$ then $C = C(\sigma) = \emptyset$ and hence $\sigma_C = -\infty = \sigma$. Assume now $\sigma \neq -\infty$. Then σ is proper.

Claim. $\sigma^* = I_{C(\sigma)}$.

If the claim is true, then $\sigma = \sigma^{**} = (I_{C(\sigma)})^* = \sigma_{C(\sigma)}$.

Proof of claim. Assume $x \in C(\sigma)$. Then, $\langle x, \cdot \rangle \le \sigma$ so $\sigma^*(x) = \sup_{s \in \mathbb{R}^n} \langle x, s \rangle - \sigma(s) \le 0$. Also, $\sigma(0) = 0$ since σ is a proper closed sublinear function and hence $\sigma^*(x) \ge 0$ for all $x \in \mathbb{R}^n$. Thus, $\sigma^*(x) = 0$ if $x \in C(\sigma)$.

Assume now $x \notin C(\sigma)$. Then, $\langle x, \cdot \rangle \not\le \sigma$. So $\exists s_0 \in \mathbb{R}^n$ such that $\langle x, s_0 \rangle > \sigma(s_0)$. So

$$\sigma^*(x) \ge \sup_{t>0} \langle x, ts_0 \rangle - \sigma(ts_0) = \sup_{t>0} t \cdot [\langle x, s_0 \rangle - \sigma(s_0)] = \infty$$

and hence $\sigma^*(x) = \infty$ if $x \notin C(\sigma)$. Therefore $\sigma^* = I_{C(\sigma)}$. □

Proposition 3.22. *Assume $f \in$ E-Conv $\mathbb{R}^n$ and $\bar{x} \in \mathbb{R}^n$ is such that $f(\bar{x}) \in \mathbb{R}$. Then,*

(a) $\mathrm{dom}\,(f'(\bar{x}; \cdot)) = \mathbb{R}_{++} \cdot (\mathrm{dom}\, f - \bar{x})$

(b) $f'(\bar{x}; \cdot)$ *is sublinear.*

Proof. (a) $d \in \mathrm{dom}(f'(\bar{x}; \cdot)) \iff f'(\bar{x}; d) < \infty \iff \inf_{t>0} \frac{f(\bar{x}+td)-f(\bar{x})}{t} < \infty \iff \exists t > 0$ s.t. $f(\bar{x} + td) < \infty \iff \exists t > 0$ s.t. $\bar{x} + td \in \mathrm{dom}\, f \iff \exists t > 0$ s.t. $d \in \frac{\mathrm{dom}\, f - \bar{x}}{t} \iff d \in \mathbb{R}_{++} \cdot (\mathrm{dom}\, f - \bar{x})$.

(b) $\forall t > 0$ and $\forall d \in \mathbb{R}^n$, obviously have $f'(\bar{x}; td) = tf'(\bar{x}; d)$. Let $d_0, d_1 \in \mathrm{dom}\, f'(\bar{x}; \cdot)$ and $(\alpha_0, \alpha_1) \in \Delta_1$. Then $\exists t_i > 0$ s.t. $\bar{x} + t_i d_i \in \mathrm{dom}\, f$. Let $\bar{t} = \min\{t_1, t_2\}$. Then,

$$\bar{x} + td \in \mathrm{dom}\, f, \forall t \in [0, \bar{t}]$$

and hence

$$\frac{f(\bar{x} + t(\alpha_0 d_0 + \alpha_1 d_1)) - f(\bar{x})}{t} \le \alpha_0 \frac{f(\bar{x} + td_0) - f(\bar{x})}{t} + \alpha_1 \frac{f(\bar{x} + td_1) - f(\bar{x})}{t}.$$

Taking $t \to \infty$ we arrive at

$$f'(\bar{x}; \alpha_0 d_0 + \alpha_1 d_1) \le \alpha_0 f'(\bar{x}; d_0) + \alpha_1 f'(\bar{x}; d_1).$$

□

Proposition 3.23. *Assume $f \in$ E-Conv $\mathbb{R}^n$ and $\bar{x} \in \mathbb{R}^n$ is such that $f(\bar{x}) \in \mathbb{R}$. Then,*

$$\mathrm{cl}\, f'(\bar{x}; \cdot) = \sigma_{\partial f(\bar{x})}.$$

Proof. Take $\sigma = f'(\bar{x}; \cdot)$. From previous two results, have $\mathrm{cl}\, f'(\bar{x}; \cdot) = \sigma_{\partial f(\bar{x})}$. □

Proposition 3.24. *Assume $f \in$ E-Conv $\mathbb{R}^n$ and $\bar{x} \in \mathbb{R}^n$ is such that $f(\bar{x}) \in \mathbb{R}$. Then,*

$$\partial f(\bar{x}) = \emptyset \iff \exists d_0 \in \mathbb{R}^n \text{ s.t. } f'(\bar{x}; d_0) = -\infty$$

in which case

$$f'(\bar{x}; d) = -\infty, \forall d \in \mathrm{ri}(\mathrm{dom}\, f - \bar{x}).$$

Proof. $\partial f(\bar{x}) = \emptyset \iff \sigma_{\partial f(\bar{x})} = -\infty = \operatorname{cl} f'(\bar{x};\cdot) = -\infty \iff \exists d_0$ s.t. $f'(\bar{x};d_0) = -\infty$. Recall that if $f(x_0) = -\infty$ for some x_0 then $f(x) = -\infty$ for all $x \in \operatorname{ri}(\operatorname{dom} f)$. Using this fact,

$$f'(\bar{x};d) = -\infty, \forall d \in \operatorname{ri}(\operatorname{dom} f'(\bar{x};d)) = \operatorname{ri}(\mathbb{R}_{++}\cdot(\operatorname{dom} f - \bar{x})).$$

Since $\operatorname{aff}(\operatorname{dom} f - \bar{x}) = \operatorname{aff}(\mathbb{R}_{++}\cdot(\operatorname{dom} f - \bar{x}))$ [exercise], we have $\operatorname{ri}(\operatorname{dom} f - \bar{x}) \subseteq \operatorname{ri}(\mathbb{R}_{++}\cdot(\operatorname{dom} f - \bar{x}))$. □

Proposition 3.25. *Assume $f \in$ E-Conv $\mathbb{R}^n$ and $\bar{x} \in \mathbb{R}^n$ is s.t. $f(\bar{x}) \in \mathbb{R}$. Then:*

(a) if $\bar{x} \in \operatorname{ri}(\operatorname{dom} f)$, then $\partial f(\bar{x}) \neq \emptyset$ and $f'(\bar{x};\cdot) = \sigma_{\partial f(\bar{x})}$.

(b) $\bar{x} \in$ int$(\operatorname{dom} f)$ iff $\partial f(\bar{x}) \neq \emptyset$ and bounded, in which case $f'(\bar{x};d) = \max\{\langle d,s\rangle : s \in \partial f(\bar{x})\}$.

Proof. (a) $\operatorname{dom} f'(\bar{x};\cdot) = \mathbb{R}_{++}\cdot(\operatorname{dom} f - \bar{x})$. Since $\bar{x} \in \operatorname{ri}(\operatorname{dom} f)$, have $\mathbb{R}_{++}\cdot(\operatorname{dom} f - \bar{x})$ is a subspace, $f'(\bar{x};\cdot)$ convex and $f'(\bar{x};0) = 0$, then $f'(\bar{x};\cdot) \in \overline{\text{Conv}}\,\mathbb{R}^n$ [exercise]. Then $f'(\bar{x};\cdot) = \operatorname{cl} f'(\bar{x};\cdot) = \sigma_{\partial f(\bar{x})}$ and hence $\partial f(\bar{x}) \neq \emptyset$.

(b) We have

$$\begin{aligned} \bar{x} \in \text{int}(\operatorname{dom} f) &\iff \mathbb{R}_{++}\cdot(\operatorname{dom} f - \bar{x}) = \mathbb{R}^n \\ &\iff f'(\bar{x};\cdot) \text{ finite everywhere} \\ &\overset{(*)}{\iff} \operatorname{cl} f'(\bar{x};\cdot) \text{ finite everywhere} \\ &\iff \sigma_{\partial f(\bar{x})} \text{ finite everywhere} \\ &\overset{(\dagger)}{\iff} \partial f(\bar{x}) \neq \emptyset \text{ and bounded} \end{aligned}$$

$(*)$ is left as an exercise and the forward direction of $(\dagger)$ is from the fact that $\exists M \in \mathbb{R}$ s.t. $\sigma_{\partial f(\bar{x})}(d) \leq M\|d\|, \forall d \in \mathbb{R}^n \implies \sigma_{\partial f(\bar{x})} \leq \sigma_{\bar{B}(0;M)} \iff \partial f(\bar{x}) \subseteq \bar{B}(0;M)$. □

4 Duality

4.1 Equality Constrained Problems

Consider the (ECP) optimization problem

$$\text{(ECP)} \qquad \begin{aligned} f_* = \inf\ & f(x) \\ \text{s.t. } & g_i(x) = 0, i \in E \\ & x \in X \end{aligned}$$

for some finite index set E where we will denote $(-\infty,+\infty]^E \ni g_E(x) := (g_i(x))_{i\in E}$. Assume that:

(a) $f, g_i : \mathbb{R}^n \mapsto (-\infty,+\infty]$

(b) $\emptyset \neq X \subseteq \operatorname{dom} f \cap \left(\bigcap_{i\in E} \operatorname{dom} g_i\right)$.

Observe that $f_* < \infty \iff \exists x \in X$ s.t. $g_E(x) = 0$.

Definition 4.1. Define the **Lagrangian function** for (ECP) $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^E \mapsto (-\infty,+\infty]$ by

$$(x,\lambda) \mapsto \begin{cases} f(x) + \sum_{i\in E} \lambda_i g_i(x), & \text{if } x \in X \\ +\infty, & \text{otherwise} \end{cases} = \begin{cases} f(x) + \langle \lambda, g_E(x)\rangle, & \text{if } x \in X \\ +\infty, & \text{otherwise.} \end{cases}$$

Note that (ECP) $\iff \inf_x \sup_\lambda \mathcal{L}(x,\lambda) \geq \sup_\lambda \inf_x \mathcal{L}(x,\lambda)$ which we call the dual. Also,

$$\sup_{\lambda\in\mathbb{R}^E} \mathcal{L}(x,\lambda) = \begin{cases} f(x), & \text{if } g_E(x) = 0, x \in X \\ +\infty, & \text{otherwise} \end{cases}$$

and so (ECP) $\leftrightarrow \inf_{x\in\mathbb{R}^n} \sup_{\lambda\in\mathbb{R}^E} \mathcal{L}(x,\lambda)$.

Definition 4.2. The **dual function** $\theta : \mathbb{R}^E \mapsto [-\infty, \infty)$ is defined as $\theta(\lambda) = \inf_{x\in\mathbb{R}^n} \mathcal{L}(x, \lambda)$. The dual (ECP) is

$$\text{(DECP)} \qquad \theta^* = \sup_{\lambda\in\mathbb{R}^E} \theta(\lambda) = \sup_{\lambda\in\mathbb{R}^E} \inf_{x\in\mathbb{R}^n} \mathcal{L}(x, \lambda).$$

Note that $-\theta \in \overline{\text{Conv}}\,\mathbb{R}^n$.

Notation 3. For $\lambda \in \mathbb{R}^E$, denote $X(\lambda) = \{x \in \mathbb{R}^n : \mathcal{L}(x, \lambda) = \theta(\lambda)\}$. Observe that:

(1) if $\theta(\lambda) = -\infty$ then $X(\lambda) = \emptyset$

(2) $\theta(\lambda) < \infty$ for all $\lambda \in \mathbb{R}^E$

(3) $X(\lambda) = \{x \in X : \theta(\lambda) = f(x) + \langle \lambda, g_E(x)\rangle$

Proposition 4.1. *(Everett) Assume $x_\lambda \in X(\lambda)$ for some $\lambda \in \mathbb{R}^E$. Then x_λ is an optimal solution of*

$$\begin{aligned} (P_\lambda) \qquad & \inf f(x) \\ & \textit{s.t. } g_E(x) = g_E(x_\lambda) \\ & \quad x \in X. \end{aligned}$$

Proof. x_λ is clearly feasible for (P_λ). Also, $\mathcal{L}(x_\lambda, \lambda) \leq \mathcal{L}(x, \lambda)$ for all $x \in X$, so

$$f(x_\lambda) + \langle \lambda, g_E(x_\lambda)\rangle \leq f(x) + \langle \lambda, g_E(x)\rangle$$

for all $x \in X$. In particular, if x is feasible for P_λ then $f(x_\lambda) \leq f(x)$. □

Definition 4.3. $\lambda^* \in \mathbb{R}^E$ is a **Lagrange multiplier** (LM) of (ECP) if $f_* \in \mathbb{R}$ and $f_* = \theta(\lambda^*)$ ($\iff$ $f_* = \inf_{x\in X} f(x) + \langle \lambda^*, g_E(x)\rangle$).

Remark 4.1. Consider the set

$$S = \left\{ \begin{pmatrix} g_E(x) \\ f(x) \end{pmatrix} \in \mathbb{R}^E \times \mathbb{R} : x \in X \right\}$$

and let $\eta^* = \begin{pmatrix} \lambda^* \\ 1 \end{pmatrix}$, $s^* = \begin{pmatrix} 0 \\ f^* \end{pmatrix}$. Let $H^{\geq} = \left\{ s : (\eta^*)^T(s - s^*) \geq 0 \right\}$. Then $S \subseteq H^{\geq}$ since $f_* \leq f(x) + \langle \lambda^*, g_E(x)\rangle$ for all $x \in X$ or equivalently,

$$\begin{pmatrix} \lambda^* \\ 1 \end{pmatrix}^T \begin{pmatrix} g_E(x) - 0 \\ f(x) - f_* \end{pmatrix} \geq 0.$$

Proposition 4.2. *For a given $(x^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^E$, then following are equivalent:*

(a) x^* is an optimal solution and λ^* is a Lagrange multiplier for (ECP)

(b) $x^* \in X(\lambda^*)$, $g_E(x^*) = 0$.

Proof. (a) $\implies$ $g_E(x^*) = 0, x^* \in X$, and $f(x^*) = f_* = \theta(\lambda^*) \in \mathbb{R}$. Then,

$$f(x^*) + \langle \lambda^*, g_E(x^*)\rangle = f(x^*) = f_* = \theta(\lambda^*).$$

Since $x^* \in X$, we have by the definition of $X(\lambda)$ that $x^* \in X(\lambda^*)$. So (b) holds.

(b) $\implies$ x^* is an optimal solution of (ECP) due to Everette's proposition. Hence,

$$f_* = f(x^*) = f(x^*) + \langle \lambda, g_E(x^*)\rangle = \theta(\lambda^*) \in \mathbb{R}.$$

□

Proposition 4.3. *(Weak Duality) For every feasible x of (ECP) and $\lambda \in \mathbb{R}^E$, we have $f(x) \geq \theta(\lambda)$. As a consequence, $f_* \geq \theta_*$.*

Proof. $f(x) = \mathcal{L}(x, \lambda) \geq \inf_u \mathcal{L}(u, \lambda) = \theta$. □

Proposition 4.4. *λ^* is a LM of (ECP) $\iff$ $f_* = \theta_* \in \mathbb{R}$ and λ^* is an optimal solution of (DECP).*

Proof. Follows from $f_* \geq \theta_* \geq \theta(\lambda^*)$. So

$$\mathbb{R} \ni f_* = \theta(\lambda^*) \iff f_* = \theta_* \text{ and } \theta_* = \theta(\lambda^*).$$

□

Corollary 4.1. *Assume $f_* = \theta_* \in \mathbb{R}$. Then the set of LM's is equal to the set of dual optimal solutions.*

Definition 4.4. The **value function** for (ECP) is defined as

$$\begin{aligned} v(b) = \inf\ & f(x) \\ \text{s.t. } & g_E(x) + b = 0 \ (\iff g_E(x) = -b) \\ & x \in X. \end{aligned}$$

Observe that $f_* = v(0)$.

Proposition 4.5. *For all $\lambda \in \mathbb{R}^E$, $v^*(\lambda) = (-\theta)(\lambda)$.*

Proof. We have

$$\begin{aligned} -v^*(\lambda) &= \sup_b \langle \lambda, b\rangle - v(b) \\ &= \inf_b v(b) - \langle \lambda, b\rangle \\ &= \inf_b \begin{pmatrix} \inf & f(x) - \langle \lambda, b\rangle \\ \text{s.t.} & g_E(x) = -b \\ & x \in X \end{pmatrix} \\ &= \begin{matrix} \inf f(x) + \langle \lambda, g_E(x)\rangle = \theta(\lambda) \\ \text{s.t. } x \in X. \end{matrix} \end{aligned}$$

□

Corollary 4.2. $(-\theta)^* = \overline{\text{co}}v$ *using the fact that* $v^{**} = \overline{\text{co}}v$.

Proposition 4.6. $\theta_* = (\overline{\text{co}}v)(0)$.

Proof. Directly,

$$\begin{aligned} \theta_* &= \sup_\lambda \theta(\lambda) = \sup_\lambda \langle 0, \lambda\rangle - (-\theta)(\lambda) \\ &= (-\theta)^*(0) = (\overline{\text{co}}v)(0). \end{aligned}$$

□

Corollary 4.3. $f_* = \theta_* \iff v(0) = (\overline{\text{co}}v)(0)$.

Proposition 4.7. *The set of dual optimal solutions is equal to $\partial(\overline{\text{co}}v)(0)$.*

Proof. Recall that

$$-\theta \in \begin{cases} \overline{\text{Conv}}\, \mathbb{R}^n \\ +\infty. \end{cases}$$

We then have $\lambda^* \in \text{argmax}_\lambda\, \theta(\lambda) = \text{argmin}_\lambda (-\theta)(\lambda) \iff 0 \in \partial(-\theta)(\lambda^*) \iff \lambda^* \in \partial(-\theta)^*(0) \iff \lambda^* \in \partial(\overline{\text{co}}v)(0)$ by the previous corollary. □

Remark 4.2. Observe that $(-\theta)^*(0) = \theta_*$. Also if Λ^* is the set of optimal solutions of (DECP), then $\Lambda^* = \partial(-\theta)^*(0)$.

Corollary 4.4. $\overline{\text{co}}v(0) = \theta_*$ *and* $\partial(\overline{\text{co}}v)(0) = \Lambda^*$.

Proposition 4.8. λ^* *is a Lagrange multiplier (L.M.) of (ECP)* $\iff v(0) \in \mathbb{R}$ *and* $\lambda^* \in \partial v(0)$ *(or $f_* \in \mathbb{R}$).*

Proof. $\lambda^* \in \mathbb{R}^E$ is a L.M. of (ECP) $\iff$ $f_* = \theta_* \in \mathbb{R}$ and $\lambda^* \in \Lambda^*$ $\iff$ $v(0) = (\overline{\text{co}}v)(0)$ and $\lambda^* \in \partial(\overline{\text{co}}v)(0)$ $\iff$ $v(0) \in \mathbb{R}$ and $\lambda^* \in \partial v(0)$. The last one follows from the fact that $v(0) = (\operatorname{cl} v)(0)$ and $0 \in \partial v^*(\lambda^*)$ $\iff$ $v(0) = (\operatorname{cl} v)(0)$ and $\lambda^* \in \partial(\operatorname{cl} v)(0)$ $\iff$ $\lambda^* \in \partial v(0)$. □

Proposition 4.9. *Assume $f_* \in \mathbb{R}$, $v \in$ E-Conv $\mathbb{R}^n$, and $0 \in \operatorname{ri}(\operatorname{dom} v)$. Then (ECP) has a LM.*

4.2 Inequality Constrained Problems

Consider the (ICP) optimization problem

$$\text{(ICP)} \qquad \begin{aligned} f_* = \inf\ & f(x) \\ \text{s.t. } & g_i(x) \le 0, i \in I \\ & x \in X \end{aligned}$$

for some finite index set I, i.e. $g_I(x) \le 0$.

Assumption 1. *Assume that:*

(a) $f, g_i : \mathbb{R}^n \mapsto (-\infty, +\infty]$

(b) $\emptyset \neq X \subseteq \operatorname{dom} f \cap \left(\bigcap_{i \in E} \operatorname{dom} g_i\right)$.

Definition 4.5. The **Lagrangian function** for (ICP) is defined as

$$\mathcal{L}(x, \lambda) = \begin{cases} f(x) + \langle \lambda, g_I(x) \rangle, & \text{if } x \in X, \lambda \ge 0 \\ -\infty & \text{if } x \in X, \lambda \not\ge 0 \\ +\infty & \text{if } x \notin X. \end{cases}$$

Then,

$$\sup_\lambda \mathcal{L}(x, \lambda) = \begin{cases} f(x), & \text{if } x \in X, g_I(x) \le 0 \\ +\infty, & \text{otherwise.} \end{cases}$$

So we then have (ICP) $\iff$ $\inf_x \sup_\lambda \mathcal{L}(x, \lambda)$. The dual problem (DICP) is $\sup_\lambda \inf_x \mathcal{L}(x, \lambda)$ and we call $\theta(\lambda) = \inf_x \mathcal{L}(x, \lambda)$ the **dual function** of (ICP). Explicitly,

$$\theta(\lambda) = \begin{cases} \inf_{x \in X} f(x) + \langle \lambda, g_I(x) \rangle, & \text{if } \lambda \ge 0 \\ -\infty, & \text{if } \lambda \not\ge 0. \end{cases}$$

So,

$$\theta_* = \sup_{\lambda \in \mathbb{R}^I} \theta(\lambda) = \sup_{\lambda \ge 0} \inf_{x \in X} f(x) + \langle \lambda, g_I(x) \rangle.$$

and let Λ^* be the set of optimal solutions. Let us define

$$\begin{aligned} X(\lambda) &= \{x \in \mathbb{R}^n : \mathcal{L}(x, \lambda) = \theta(\lambda)\} \\ &= \begin{cases} \{x \in \mathbb{R}^n : f(x) + \langle \lambda, g_I(x) \rangle = \theta(\lambda)\}, & \text{if } \lambda \ge 0 \\ X, & \text{if } \lambda \not\ge 0 \end{cases} \end{aligned}$$

and the **value function**

$$\begin{aligned} v(b) = \inf\ & f(x) \\ \text{s.t. } & g_I(x) + b \le 0 \\ & x \in X. \end{aligned}$$

This can be related to (ECP) as follows. Define

$$(\widetilde{ECP}) \qquad f_* = \inf\ f(x) \quad \text{s.t. } g_i(x) + s = 0, i \in I \quad x \in X, s \in \mathbb{R}^I_+.$$

The Lagrangian function for this problem is

$$\tilde{L}(x, s; \lambda) = \begin{cases} f(x) + \langle \lambda, g_I(x) + s \rangle, & \text{if } x \in X, s \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

and also

$$\tilde{\theta}(\lambda) = \inf_{(x,s) \in X \times \mathbb{R}^I_+} \tilde{\mathcal{L}}(x, s; \lambda).$$

Let us define $\tilde{XS}(\lambda)$ as the set of optimal solution of the previous problem, $\tilde{\theta}_* = \sup_{\lambda \in \mathbb{R}^I} \theta(\lambda)$ and Λ^* is the set of optimal solutions, and

$$\tilde{v}(b) = \inf\ f(x) \quad \text{s.t. } g_I(x) + s + b \leq 0 \quad x \in X, s \geq 0.$$

Proposition 4.10. *We relate $(\widetilde{ECP})$ to (ICP):*

(a) $f_* = \tilde{f}_*$ and $v = \tilde{v}$ (i.e. x^* is an optimal solution of (ICP) $\iff$ $(x^*, -g_I(x^*))$ is an optimal solution of $(\widetilde{ECP})$)

(b) $\theta = \tilde{\theta}$ and

$$\tilde{XS}(\lambda) = \begin{cases} X(\lambda) \times \{s \in \mathbb{R}^I, s \geq 0, \langle s, \lambda \rangle \geq 0\}, & \textit{if } \lambda \geq 0 \\ \emptyset, & \textit{otherwise.} \end{cases}$$

(c) $\theta^* = \tilde{\theta}^*$ and $\Lambda^* = \tilde{\Lambda}^*$.

Proposition 4.11. *For $(x^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^I$, we have:*

$$\begin{array}{rcr} x^* \textit{ is an optimal solution of (ICP)} & \iff & \lambda^* \geq 0, g(x^*) \leq 0 \\ \lambda^* \textit{ is a LM of (ICP)} & & \langle \lambda^*, g(x^*) \rangle = 0 \\ & & x \in X(\lambda^*) \end{array}$$

and $x^* \in X(\lambda^*) \iff x^* \in \operatorname{argmin}_{x \in X} f(x) + \langle \lambda^*, g_I(x) \rangle$.

Proof. $(x^*, s^*) = (x^*, -g(x^*))$ is an optimal solution of $\widetilde{ECP}$ and λ^* is a LM of $\widetilde{ECP}$ $\iff$ $g(\lambda^*) + s^* = 0, (x^*, s^*) \in \tilde{XS}(\lambda^*)$ $\iff$ $g(x^*) + s^* = 0$, $\lambda^* \geq 0, s^* \geq 0$, $\langle \lambda^*, s^* \rangle = 0$, $x^* \in X(\lambda^*)$ $\iff$ RHS of the proposition. □

Proposition 4.12. *The following are equivalent:*

(a) $f_* = \theta_* \in \mathbb{R}$ and $\lambda^* \in \Lambda^*$

(b) λ^* is a LM of ICP

(c) $v(0) \in \mathbb{R}$ and $\lambda^* \in \partial v(0)$

Proposition 4.13. *Assume that $f_* \in \mathbb{R}$, $v \in$ E-Conv $\mathbb{R}^n$ and $0 \in \operatorname{ri}(\operatorname{dom} v)$. Then (ICP) has a LM.*

Assumption 2. *Suppose X is convex and f, g_i are convex for $i \in I$.*

Proposition 4.14. *Under assumption 1 and assumption 2, the value function v is convex and*

$$\operatorname{ri}(\operatorname{dom} v) = \left\{ b \in \mathbb{R}^I : \begin{array}{c} \exists x \in \operatorname{ri}(X) \textit{ s.t.} \\ g_I(x) + b < 0 \end{array} \right\}.$$

Proof. Let $\Pi(x,b) = -b$ and
$$U = \{(x,r) \in X \times \mathbb{R}^I : g_i(x) \le r_i\}.$$
Then $\operatorname{dom} v = \Pi(U)$ and $\operatorname{ri}(\operatorname{dom} v) = \Pi(\operatorname{ri} U)$. ☐

Proposition 4.15. *Let $f_1, ..., f_m : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ and convex set $X \subseteq \mathbb{R}^n$ such that $\emptyset \neq X \subseteq \bigcap_{i=1}^m \operatorname{dom} f_i$ be given. If each f_i is convex on X then*
$$U = \{(x,r) \in X \times \mathbb{R}^m : f_i(x) \le r_i, i = 1,2,...,m\}$$
is convex and
$$\operatorname{ri} U = \{(x,r) \in \operatorname{ri} X \times \mathbb{R}^m : f_i(x) < r_i, i = 1,2,...,m\}.$$

Theorem 4.1. *Consider the problem*
$$\begin{aligned} (NLP) \qquad f_* = \inf\ & f(x) \\ \text{s.t. } & g_I(x) \le 0, \quad g_i, i \in I \text{ convex} \\ & g_E(x) = 0, \quad g_i, i \in E \text{ affine} \\ & x \in X. \end{aligned}$$
and define
$$\begin{aligned} I_a &= \{i \in : g_i \text{ is affine}\} \\ I_c &= I \backslash I_a. \end{aligned}$$
If $f_* \in \mathbb{R}$ and $\exists x^0 \in \operatorname{ri} X$ such that $g_E(x^0) = 0$, $g_{I_a}(x^0) \le 0$, $g_{I_c}(x^0) < 0$ then (NLP) has a LM.

4.3 Calculus of Conjugate Functions

Note 2. We have
$$g_1(x) + g_2(x) + ... + g_m(x) = g \circ A$$
where $A(x) = (x, x, ..., x)$ and
$$g(x_1, .., .x_m) = g_1(x_1) + g_2(x_2) + ... + g_m(x_m).$$

Definition 4.6. Let $A : \mathbb{R}^n \mapsto \mathbb{R}^m$ affine and $f : \mathbb{R}^n \mapsto [-\infty, +\infty]$. Define $Af : \mathbb{R}^m \mapsto [-\infty, +\infty]$ as
$$\begin{aligned} y \mapsto (Af)(y) = \inf\ & f(x) \\ \text{s.t. } & Ax = y \end{aligned}$$

Proposition. *(1) $f \in$ E-Conv $\mathbb{R}^n \implies Af \in$ E-Conv $\mathbb{R}^n$*

(2) $\operatorname{dom}(Af) = A(\operatorname{dom} f)$

Proposition 4.16. $(Af)^* = f^* \circ A^*$

Proof. Directly,
$$\begin{aligned} (Af)^*(s) &= \sup_{y \in \mathbb{R}^m} \langle y, s \rangle - (Af)(y) \\ &= \sup_{y \in \mathbb{R}^m} \langle y, s \rangle - \inf\{f(x) : Ax = y\} \\ &= \sup_{(x,y)} \{\langle y, s \rangle - f(x) : Ax = y\} \\ &= \sup_x \langle Ax, s \rangle - f(x) \\ &= \sup_x \langle A^* s, x \rangle - f(x) \\ &= f^*(A^* s) \\ &= (f^* \circ A^*)(s). \end{aligned}$$
☐

Proposition 4.17. *For any $g \in$ E-Conv $\mathbb{R}^n$ and $B : \mathbb{R}^n \mapsto \mathbb{R}^m$ linear, we have*

$$(\operatorname{cl} g \circ B)^* = \operatorname{cl}(B^* g^*).$$

Proof. By the previous result,

$$(B^* g^*) = g^{**} \circ B = (\operatorname{cl} g) \circ B$$

and hence

$$(\operatorname{cl} g \circ B)^* = (B^* g^*)^{**} = \operatorname{cl}(B^* g^*).$$

□

Proposition 4.18. *Let $g \in$ E-Conv $\mathbb{R}^m$ and $B : \mathbb{R}^n \mapsto \mathbb{R}^m$ linear be such that*

$$(*) \qquad \textit{Im } B \cap \operatorname{ri}(\operatorname{dom} g) \neq \emptyset.$$

Then $(g \circ B)^* = B^* g^*$ and for every $s \in \mathbb{R}^n$ such that $B^* g^*(s)$ is finite, the infimum

$$\begin{aligned}(B^* g^*)(s) = \inf\ & g^*(y) \\ \textit{s.t. } & B^* y = s\end{aligned}$$

is achieved.

Proof. From $(*)$ we have $g \neq \infty$.
(1) $\exists y \in \mathbb{R}^n$ such that $g(y) = -\infty$. Then $g^*(y) = \infty$ and hence $B^* g^* = \infty$. By assumption, $\exists x^0 \in \operatorname{ri}(\operatorname{dom} g)$. So, $(g \circ B)(x^0) = g(Bx^0) = -\infty$ since $g(y') = -\infty$ for all $y' \in \operatorname{ri}(\operatorname{dom} g)$. So $(g \circ B)^* = \infty$.
(2) $g \in \operatorname{Conv} \mathbb{R}^n$. Then,

$$\begin{aligned}
-(g \circ B)^*(s) &= -\sup_x \langle s, x\rangle - (g \circ B)(x) \\
&= \inf_x g(Bx) - \langle s, x\rangle \\
&= \begin{bmatrix} \inf_{x,y} g(y) - \langle s, x\rangle \\ \text{s.t. } Bx - y = 0 \\ \quad y \in \operatorname{dom} g, x \in \mathbb{R}^n \end{bmatrix} \\
&\geq \sup_\lambda \begin{bmatrix} \inf_{x,y} g(y) - \langle s, x\rangle + \langle \lambda, Bx - y\rangle \\ \text{s.t. } y \in \operatorname{dom} g, x \in \mathbb{R}^n \end{bmatrix} \\
&= \sup_{B^*\lambda = s} \begin{bmatrix} \inf_y g(y) - \langle \lambda, y\rangle \\ \text{s.t. } y \in \operatorname{dom} g \end{bmatrix} \\
&= \sup_{B^*\lambda = s} \begin{bmatrix} -\sup_y \langle \lambda, y\rangle - g(y) \\ \text{s.t. } y \in \operatorname{dom} g \end{bmatrix} \\
&= \begin{bmatrix} \sup_\lambda -g^*(\lambda) \\ \text{s.t. } B^*\lambda = s \end{bmatrix} \\
&= \begin{bmatrix} -\inf_\lambda g^*(\lambda) \\ \text{s.t. } B^*\lambda = s \end{bmatrix} \\
&= -(B^* g^*)(s).
\end{aligned}$$

where under duality theory, the infimum must be achieved in the second to last expression. The inequality in the above expressions is made to be equality using duality arguments as well. □

Proposition 4.19. *Let $g \in$ E-Conv $\mathbb{R}^m$ and $B : \mathbb{R}^n \mapsto \mathbb{R}^m$ be linear. Then,*

$$B^*(\partial g(Bx)) \subseteq \partial(g \circ B)(x), \forall x.$$

If, in addition, Im $B \cap \mathrm{ri}(\mathrm{dom}\, g) \neq \emptyset$ *then equality holds.*

Proof. We have

$$\begin{aligned} s \in \partial(g \circ B)(x) &\iff (g \circ B)^*(x) \leq \langle s, x\rangle - (g \circ B)(s) \\ &\iff (B^* g^*)(s) \leq \langle s, x\rangle - g(Bx) \\ &\Longleftarrow \exists y \text{ s.t. } B^* y = s \text{ and } g^*(y) \leq \langle s, x\rangle - g(Bx) \\ &\iff \exists y \text{ s.t. } g^*(y) \leq \langle B^* y, x\rangle - g(Bx) \text{ and } s = B^* y \\ &\iff \exists y \text{ s.t. } s = B^* y \text{ and } g^*(y) \leq \langle y, Bx\rangle - g(Bx) \\ &\iff \exists y \text{ s.t. } s = B^* y \text{ and } y \in \partial g(Bx) \\ &\iff s \in B^*(\partial g(Bx)). \end{aligned}$$

□

Definition 4.7. The **ϵ-subgradient** is defined as

$$s \in \partial_\epsilon f(x) \iff f(x') \geq f(x) + \langle s, x' - x\rangle - \epsilon, \forall x'.$$

An equivalent characterization is

$$s \in \partial_\epsilon f(x) \iff f^*(s) \leq \langle x, s\rangle - f(x) + \epsilon.$$

Corollary 4.5. *Let* $\epsilon > 0$, $g \in$ *E-Conv* $\mathbb{R}^m$*, and* $B : \mathbb{R}^n \mapsto \mathbb{R}^m$ *be linear. Then,*

$$B^*(\partial g_\epsilon(Bx)) \subseteq \partial_\epsilon(g \circ B)(x), \forall x.$$

If, in addition, Im $B \cap \mathrm{ri}(\mathrm{dom}\, g) \neq \emptyset$ *then equality holds.*

<u>Infimal Convolution</u>

Definition 4.8. For $f_1, ..., f_m : \mathbb{R}^n \mapsto (-\infty, +\infty]$, their **infimal convolution** is defined as

$$(f_1 \square ... \square f_m)(x) = \begin{bmatrix} \inf\, f_1(x_1) + ... + f_2(x_m) \\ \text{s.t. } x_1 + ... + x_m = x \end{bmatrix}.$$

Proposition 4.20. $f_1, ..., f_m \in$ *Conv* $\mathbb{R}^n$ *implies that* $f_1 \square ... \square f_m \in$ *E-Conv* $\mathbb{R}^n$ *and*

$$\mathrm{dom}(f_1 \square ... \square f_m) = \mathrm{dom}\, f_1 + ... + \mathrm{dom}\, f_m.$$

Remark 4.3. Let $f(x_1, ..., x_m) = f_1(x_1) + ... + f_2(x_m)$ and $A(x_1, ..., x_m) = x$. Then $f_1 \square ... \square f_m = Af$ and $f \circ A^* = (f_1 + ... + f_m)$.

Proposition 4.21. *Let* $f_i : \mathbb{R}^m \mapsto (-\infty, \infty]$, $i = 1, 2, ..., m$ *be given. Then:*

(i) $(f_1 \square ... \square f_m)^* = f_1^* + ... + f_m^*$

(ii) If $f_i \in$ *Conv* $\mathbb{R}^n$ *for* $i = 1, 2, ..., m$ *then* $(\mathrm{cl}\,[f_1 + ... + f_m])^* = \mathrm{cl}\,(f_1^* \square ... \square f_m^*)$.

(iii) If $f_i \in$ *Conv* $\mathbb{R}^n$ *for* $i = 1, 2, ..., m$ *and*

$$\bigcap_{i=1}^m \mathrm{ri}(\mathrm{dom}\, f_i) \neq \emptyset$$

then

$$(f_1 + ... + f_m)^* = (f_1^* \square ... \square f_m^*).$$

Proof. Note that

$$\begin{aligned} f^*(s_1, ..., s_m) &= f_1^*(s_1) + ... + f_m^*(s_m) \qquad \text{[exercise]} \\ \mathrm{cl}\, f(s_1, ..., s_m) &= \mathrm{cl}\, f_1(s_1) + ... + \mathrm{cl}\, f_m(s_m) \\ A^*(x) &= (x, ..., x) \end{aligned}$$

(i) We have $(f_1 \square ... \square f_m)^* = (Af)^* = (f^* \circ A^*) = f_1^* + ... + f_m^*$.

(ii) We have $(\mathrm{cl}\, f_1 + ... + \mathrm{cl}\, f_m)^* = (\mathrm{cl}\, f \circ A^*)^* = \mathrm{cl}(Af^*) = \mathrm{cl}\,(f_1^* \square ... \square f_m^*)$.

(iii) Note that $\mathrm{ri}(\mathrm{dom}\, f) = \mathrm{ri}(\mathrm{dom}\, f_1) \times ... \times \mathrm{ri}(\mathrm{dom}\, f_m)$ and this is equivalent to

$$\bigcap_{i=1}^{m} \mathrm{ri}(\mathrm{dom}\, f_i) \neq \emptyset$$

and so by (ii) the closure can be removed to get $(f_1 + ... + f_m)^* = (f_1^* \square ... \square f_m^*)$. ☐

Corollary 4.6. *We have*

$$\begin{aligned} \partial(f_1 + ... + f_m)(x) &= \partial(f \circ A^*)(x) \\ &= A\left[\partial f(A^* x)\right] \\ &= A\left(\partial f_1(x) \times ... \times \partial f_m(x)\right) \\ &\overset{(*)}{=} \partial f_1(x) + ... + \partial f_m(x) \end{aligned}$$

if the standard constraint qualification holds, where $(*)$ is left as an exercise. Note that $\supseteq$ always holds regardless of the constraint set.

Corollary 4.7. *If $0 \leq \epsilon_1 + ... + \epsilon_m \leq \epsilon$ then*

$$\partial_\epsilon(f_1 + ... + f_m)(x) = \partial_{\epsilon_1} f_1(x) + ... + \partial_{\epsilon_m} f_m(x)$$

when the standard constraint qualification holds. Note that $\supseteq$ always holds regardless of the constraint set.

Applications

(1) Consider the problem

$$\begin{aligned} &\min\, f(x) \\ &\text{s.t. } x \in C \end{aligned}$$

where $f : \mathbb{R}^n \mapsto (-\infty, \infty]$ and $C \subseteq \mathbb{R}$. This is equivalent to

$$(*) \qquad \begin{aligned} &\min\, f(x) + I_C(x) = (f + I_C)(x) \\ &\text{s.t. } x \in \mathbb{R}^n. \end{aligned}$$

Now x^* is a global min of $(*)$ $\iff$ $0 \in \partial(f + I_C)(x^*)$ $\Longleftarrow$ $0 \in \partial f(x^*) + \partial I_C(x^*)$ $\iff$ $0 \in \partial f(x^*) + N_C(x^*)$ $\iff$ $-\partial f(x^*) \cap N_C(x^*) \neq \emptyset$. All the statements are equivalent if f is convex, C is convex, $\mathrm{ri}(\mathrm{dom}\, f) \cap \mathrm{ri}\, C \neq \emptyset$. The last expression is a generalization of the requirement $-\nabla f(x^*) \in N_C(x^*)$.

Proposition 4.22. *Consider ICP with $\emptyset \neq X \subseteq \mathrm{dom}\, f \cap \bigcap_{i \in I} \mathrm{dom}\, g_i$. Let $\bar{x}$ be a feasible point of $(*)$, i.e. $g_I(x) \leq 0,\ x \in X$. If $\exists \bar{\lambda} \in \mathbb{R}_+^m$ s.t.*

$$\begin{cases} \partial f(\bar{x}) + \sum_{i \in I} \bar{\lambda}_i \partial g_i(\bar{x}) \in N_X(\bar{x}), & \\ \bar{\lambda}^T g_I(\bar{x}) = 0 & (a) \end{cases} (*)$$

then $\bar{x}$ is an optimal solution and $\bar{\lambda}$ is a Lagrange multiplier of $(**)$.

Conversely suppose that $f, \{g_i\}_{i \in I}$ are convex, X is convex and $\exists x^0 \in \mathrm{ri}(\mathrm{dom}\, f) \cap \bigcap_{i \in I} \mathrm{ri}(\mathrm{dom}\, g_i) \cap \mathrm{ri}\, X$ such that $g_I(x_0) < 0$. Then if $\bar{x}$ is a global minimum of (2), $\exists \bar{\lambda} \in \mathbb{R}_+^m$ satisfying $(*)$.

Proof. ($\Longrightarrow$) We have

$$\begin{aligned} (*) &\implies 0 \in \partial(f + \sum_{i \in I} \bar{\lambda}_i g_i + I_X)(\bar{x}) \\ &\implies \bar{x} \in \operatorname*{argmin}_{x \in \mathbb{R}^n} L(x, \bar{\lambda}) \qquad (b) \end{aligned}$$

So (a), (b) $\iff$ $\bar{x}$ is an optimal solution, $\bar{\lambda}$ is a LM.

($\Longleftarrow$) Assume $\bar{x}$ is a global minimum $\implies$ $f_* \in \mathbb{R}$. Also, a LM $\bar{\lambda}$ exists. Then, (a), (b) holds. □

5 Algorithms

5.1 Composite Gradient Method

Consider the problem

$$(*) \qquad \min \psi(x) := f(x) + h(x) \\ \text{s.t. } x \in \mathbb{R}^n$$

where

(1) $g \in \overline{\text{Conv}}\, \mathbb{R}^n$

(2) $f : \mathbb{R}^n \mapsto (-\infty, +\infty]$ is differentiable on $\operatorname{dom} h$ and $\exists L$ such that

$$\|\nabla f(x) - \nabla f(\tilde{x})\| \leq L\|x - \tilde{x}\|, \forall x, \tilde{x} \in \operatorname{dom} h$$

(3) f is convex on $\operatorname{dom} h$

(4) $(*)$ has an optimal solution

Observe that:

(2) $\implies |f(\tilde{x}) - \ell_f(\tilde{x}; x)| \leq \frac{L}{2}\|\tilde{x} - x\|^2$ for all $x, \tilde{x} \in \operatorname{dom} h$.

(2,3) $\implies 0 \leq f(\tilde{x}) - \ell_f(\tilde{x}; x) \leq \frac{L}{2}\|\tilde{x} - x\|^2$

Composite Gradient Method

(0) Let $x_0 \in \operatorname{dom} h$ and $\lambda > 0$ be given. Set $k = 1$.

(1) Compute

$$x_k := \operatorname*{argmin}_{x \in \mathbb{R}^n} \left\{ \ell_f(x; x_{k-1}) + h(x) + \frac{1}{2\lambda}\|x - x_{k-1}\|^2 \right\}. \tag{5.1}$$

e.g. if $h = I_X$ then $x_k = \Pi_X(x_{k-1} - \lambda \nabla f(x_{k-1}))$.

(2) Set $k \hookleftarrow k + 1$ and go to (1).

Lemma 5.1. *For all $x \in \operatorname{dom} h$, $\forall k \geq 1$,*

$$\begin{aligned} &\ell_f(x; x_{k-1}) + h(x) + \frac{1}{2\lambda}\|x - x_{k-1}\|^2 \\ \geq &\ell_f(x_k; x_{k-1}) + h(x_k) + \frac{1}{2\lambda}\|x_k - x_{k-1}\|^2 + \frac{1}{2\lambda}\|x - x_k\|^2 \end{aligned}$$

Lemma 5.2. *For all $x \in \operatorname{dom} h$, $\forall k \geq 1$,*

$$(f + h)(x_k) - (f + h)(x) \leq \frac{1}{2}\left(L - \frac{1}{\lambda}\right)\|x_k - x_{k-1}\|^2 + \frac{1}{2\lambda}\left(\|x_{k-1} - x^*\|^2 - \|x_k - x^*\|^2\right)$$

Proof. From the previous lemma,

$$\begin{aligned} &(f + h)(x) \\ \geq &\ell_f(x; x_{k-1}) + h(x) \\ \geq &\ell_f(x_k; x_{k-1}) + h(x_k) + \frac{1}{2\lambda}\|x_k - x_{k-1}\|^2 + \frac{1}{2\lambda}\left(\|x - x_k\|^2 - \|x - x_{k-1}\|^2\right) \\ \geq &(f + h)(x) - \frac{L}{2}\|x_k - x_{k-1}\|^2 + \frac{1}{2\lambda}\|x_k - x_{k-1}\|^2 + \frac{1}{2\lambda}\left(\|x - x_k\|^2 - \|x - x_{k-1}\|^2\right) \end{aligned}$$

□

Lemma 5.3. *For all $k \geq 1$,*

$$(f+h)(x_{k-1}) - (f+h)(x_k) \geq \left(\frac{1}{\lambda} - \frac{L}{2}\right)\|x_k - x_{k-1}\|^2.$$

Hence, if $\lambda < \frac{2}{L}$ then $\{(f+h)(x_k)\}$ is decreasing.

Proof. Use the previous lemma with $x = x_k$. □

Lemma 5.4. *For every optimal solution x^* and $k \geq 1$,*

$$\psi(x_k) - \psi_* \leq \frac{1}{2}\left(L - \frac{1}{\lambda}\right)\|x_k - x_{k-1}\|^2 + \frac{1}{2}\left(\|x_{k-1} - x^*\|^2 - \|x_k - x^*\|^2\right).$$

As a consequence,

$$k\left[\psi(x_k) - \psi_*\right] \leq \sum_{\ell=1}^{k}\left[\psi(x_\ell) - \psi_*\right] \leq \frac{1}{2}\left(L - \frac{1}{\lambda}\right)\sum_{\ell=1}^{k}\|x_\ell - x_{\ell-1}\|^2 + \frac{1}{2\lambda}\left(\|x_0 - x^*\|^2 - \|x_k - x^*\|^2\right).$$

Proposition 5.1. *Assume $\lambda < 2/L$. Then $\forall k \geq 1$:*

$$k\left[\psi(x_k) - \psi_*\right] \leq \frac{1}{2}\left(L - \frac{1}{\lambda}\right)^+\left(\frac{1}{\lambda} - \frac{L}{2}\right)^{-1}\left[\psi(x_0) - \psi_*\right] + \frac{1}{2\lambda}\left(\|x_0 - x^*\|^2 - \|x_k - x^*\|^2\right).$$

Corollary 5.1. *For all $k \geq 1$,*

$$(f+h)(x_{k-1}) - (f+h)(x_k) \geq \left(\frac{1}{\lambda} - \frac{L}{2}\right)\|x_k - x_{k-1}\|^2.$$

Hence, if $\lambda < 2/L$ then $\{(f+h)(x_k)\}$ is decreasing.

Corollary 5.2. *We have*

$$\psi(x_0) - \psi_* \geq \psi(x_0) - \psi(x_k) \geq \left(\frac{1}{\lambda} - \frac{L}{2}\right)\sum_{\ell=1}^{k}\|x_\ell - x_{\ell-1}\|^2.$$

Remark 5.1. The optimality condition of (5.1) is

$$\begin{aligned}0 \in& \partial\left(\ell_f(\cdot; x_{k-1}) + h(\cdot) + \frac{1}{2\lambda}\|\cdot - x_{k-1}\|^2\right)(x_k)\\ =& \nabla f(x_{k-1}) + \partial h(x_k) + \frac{1}{\lambda}(x_k - x_{k-1}).\end{aligned}$$

An approximate solution is one where $v \in \nabla f(x) + \partial h(x)$ and $\|v\| \leq \bar{\rho}$ for small $\bar{\rho} > 0$. We can observe the previous inclusion is equivalent to $v_k \in \nabla f(x_k) + \partial h(x_k)$ where

$$v_k := \nabla f(x_k) - \nabla f(x_{k-1}) + \frac{1}{\lambda}(x_{k-1} - x_k).$$

Note that

$$\|v_k\| \leq \|\nabla f(x_k) - \nabla f(x_{k-1})\| + \frac{1}{\lambda}\|x_{k-1} - x_k\| \leq \left(L + \frac{1}{\lambda}\right)\|x_k - x_{k-1}\|.$$

From a previous lemma, we can get

$$\begin{aligned}&(f+h)(x_{k-1}) \geq (f+h)(x_k) - \frac{L}{2}\|x_k - x_{k-1}\|^2 + \frac{1}{\lambda}\|x_k - x_{k-1}\|^2\\ \Longrightarrow\ &\psi(x_{k-1}) - \psi(x_k) \geq \left(\frac{1}{\lambda} - \frac{L}{2}\right)\|x_k - x_{k-1}\|^2\\ \Longrightarrow\ &\psi(x_0) - \psi_* \geq \psi(x_0) - \psi(x_k) \geq \left(\frac{1}{\lambda} - \frac{L}{2}\right)\sum_{\ell=1}^{k}\|x_k - x_{k-1}\|^2.\end{aligned}$$

Let $\theta_k^2 = \min_{\ell=1,\dots,k} \|x_\ell - x_{\ell-1}\|^2$.Then the above implies

$$\psi(x_0) - \psi_* \geq \left(\frac{1}{\lambda} - \frac{L}{2}\right)\theta_k^2 k \implies \theta_k^2 \leq \frac{\psi(x_0) - \psi_*}{k}\left(\frac{1}{\lambda} - \frac{L}{2}\right)^{-1}.$$

Hence,

$$\min_{\ell=1,2,\dots,k} \|v_\ell\|^2 \leq \left(L + \frac{1}{\lambda}\right)^2 \theta_k^2 \leq \bar{\rho}^2 \implies k \sim O(1/\bar{\rho}^2).$$

5.2 Composite Subgradient Method

Consider the problem

$$\begin{aligned}\psi_* = \min\ & f(x) + h(x) \\ \text{s.t. } & x \in \mathbb{R}^n\end{aligned}$$

where:

- $h \in \overline{\text{Conv}}\,\mathbb{R}^n$
- f is convex on $\operatorname{dom} h$
- $\partial f(x) \neq \emptyset, \forall x \in \operatorname{dom} h$
- $\exists M \geq 0$ s.t. $|s| \leq M$ for all $s \in \partial f(x)$ for all $x \in \operatorname{dom} h$.
- $\psi_* \in \mathbb{R}$ is achieved.

Suppose it is easy to find a solution to

$$\begin{aligned}\min\ & c^T x + h(x) + \frac{1}{2\lambda} \\ \text{s.t. } & x \in \mathbb{R}^n.\end{aligned}$$

The optimal point to the easy subproblem is $x = (I + \lambda\partial h)^{-1}(\lambda c)$ which is called the **resolvent**.

Lemma 5.5. *f is M-Lipschitz on $\operatorname{dom} h$, i.e.*

$$|f(x) - f(x')| \leq M\|x - x'\|, \forall x, x' \in \operatorname{dom} h.$$

Proof. Directly

$$f(x') - f(x) \geq \langle s, x' - x\rangle \geq -\|s\|\|x' - x\| \geq -M\|x' - x\|$$

and symmetrically $f(x) - (x') \geq -M\|x' - x\|$. □

Remark 5.2. Let $\bar{s} \in \partial f(\bar{x})$ and $\ell_f(\cdot;\bar{x}) = f(\bar{x}) + \langle \bar{s}, \cdot - \bar{x}\rangle$. Then $\ell_f(\cdot;\bar{x}) \leq f$ and

$$\begin{aligned}0 \leq f(x) - \ell_f(x;\bar{x}) &\leq f(x) - f(\bar{x}) - \langle \bar{s}, x - \bar{x}\rangle \\ &\leq M\|x - \bar{x}\| + \|\bar{s}\|\|x - \bar{x}\| \\ &\leq (M + \|\bar{s}\|)\|x - \bar{x}\| \\ &\leq 2M\|x - \bar{x}\|.\end{aligned}$$

<u>Composite Subgradient Method</u>

(0) Let $x_0 \in \operatorname{dom} h$ be given and set $k = 1$.

(1) Choose $s_{k-1} \in \partial f(x_{k-1})$ and $\lambda_k > 0$. Compute

$$x_k = \operatorname{argmin} \left\{ \ell_f(x; x_{k-1}) + h(x) + \frac{1}{2\lambda_k}\|x - x_k\|^2 \right\}$$

where

$$\ell_f(x; x_{k-1}) = f(x_{k-1}) + \langle s_{k-1}, x - x_{k-1} \rangle .$$

(2) Set $k \hookleftarrow k+1$ and go to 1.

Lemma 5.6. *For all $x \in \operatorname{dom} h$,*

$$\begin{aligned}&\ell_f(x; x_{k-1}) + h(x) + \frac{1}{2\lambda_k}\|x - x_{k-1}\|^2 \\ \geq &\ell_f(x_k; x_{k-1}) + h(x_k) + \frac{1}{2\lambda_k}\|x_k - x_{k-1}\|^2 + \frac{1}{2\lambda_k}\|x - x_k\|^2.\end{aligned}$$

Lemma 5.7. *For all $x \in \operatorname{dom} h$,*

$$\lambda_k \left[\psi(x_k) - \psi(x)\right] \leq \frac{1}{2}\left(\|x - x_{k-1}\|^2 - \|x - x_k\|^2\right) + 2M^2\lambda_k^2$$

Proof. From the previous lemma,

$$\begin{aligned}&\frac{1}{2\lambda_k}\left(\|x - x_{k-1}\|^2 - \|x - x_k\|^2\right) \\ \geq &\ell_f(x_k; x_{k-1}) - \ell_f(x; x_{k-1}) + h(x_k) - h(x) + \frac{1}{2\lambda_k}\|x_k - x_{k-1}\|^2 \\ \geq &f(x_k) - 2M\|x_k - x_{k-1}\| - f(x) + h(x_k) - h(x) + \frac{1}{2\lambda_k}\|x_k - x_{k-1}\|^2 \\ = &\psi(x_k) - \psi(x) - (2M\sqrt{\lambda_k})\left(\frac{\|x_k - x_{k-1}\|}{\sqrt{\lambda_k}}\right) + \frac{1}{2\lambda_k}\|x_k - x_{k-1}\|^2 \\ \geq &\psi(x_k) - \psi(x) - 2M^2\lambda_k.\end{aligned}$$

□

Lemma 5.8. *For all $x \in \operatorname{dom} h$,*

$$\sum_{\ell=1}^{k} \lambda_\ell \left[\psi(x_\ell) - \psi(x)\right] \leq \frac{1}{2}\left(\|x - x_0\|^2 - \|x - x_k\|^2\right) + 2M^2 \sum_{\ell=1}^{k} \lambda_\ell^2.$$

Definition 5.1. Define

$$\theta_k = \max\left\{ \min_{1 \leq \ell \leq k} \psi(x_\ell), \psi\left(\frac{\sum_{\ell=1}^k \lambda_\ell x_\ell}{\sum_{\ell=1}^k \lambda_\ell}\right)\right\} - \psi_*$$

as a measure of optimality.

Proposition 5.2. *We will have*

$$\theta_k \leq \frac{1}{2\Lambda_k} \underbrace{\|x^* - x_0\|^2}_{d_0^2} + \frac{2M^2 \sum_{\ell=1}^k \lambda_\ell^2}{\Lambda_k}$$

where $\Lambda_k = \sum_{\ell=1}^k$ for any optimal solution x^*.

Proof. Follows from 1.4 with $x = x^*$ and the convexity of ψ. □

Remark 5.3. Consider taking $\lambda_\ell = \varepsilon/(4M^2) = \lambda$. Then

$$\begin{aligned}\theta_k &= d_0^2/(2k\lambda) + 2M^2\lambda \\ &= 2d_0^2M^2/(k\varepsilon) + \varepsilon/2.\end{aligned}$$

If we want $\theta_k \le \varepsilon$ then we need

$$\frac{2d_0^2 M^2}{k\varepsilon} \le \frac{\varepsilon}{2} \iff k \ge \frac{2d_0^2 M^2}{\varepsilon^2}.$$

Question: How do we improve if ψ is μ-strongly convex?

5.3 Nesterov's Method

For the problem $\min(f+h)(x) = \psi(x)$, Nesterov's method tries to keep the invariant

$$\min_{x\in\mathbb{R}^n} \left\{ A_k\Gamma_k(x) + \frac{1}{2}\|x - x_0\|^2 \right\} \ge A_k\psi(y_k)$$

where $A_k\Gamma_k \le A_k f$, $A_k \ge 0$, $\Gamma_k \in \overline{\text{Conv}}\,\mathbb{R}^n$. The idea is to send $A_k \to \infty$. At an optimal point x^* in the left subproblem, we have

$$A_k\psi(y_k) \le A_k\Gamma(x^*) + \frac{1}{2}\|x^* - x_0\|^2 \le A_k\psi_* + \frac{1}{2}\|x^* - x_0\|^2.$$

So

$$\psi(y_k) - \psi_* \le \frac{1}{2A_k}\|x_0 - x_*\|^2$$

and if $A_k \approx O(k^2)$ then convergence is fast. Let us first consider the problem

$$\begin{aligned} \phi_* = \min\ &\phi(u) \\ \text{s.t. } &u \in \mathbb{R}^n. \end{aligned}$$

where $\phi \in \overline{\text{Conv}}\,\mathbb{R}^n$ and let $X_* \neq \emptyset$ be the set of optimal solutions, i.e. $\phi_* \in \mathbb{R}$. The idea of the algorithm, again, is to choose $u_0 \in \mathbb{R}^n$, $\{y_n\} \subseteq \operatorname{dom}\phi$, $\{A_k\} \subseteq \mathbb{R}_+ \to \infty$, $\{\Gamma_k\} \subseteq \overline{\text{Conv}}\,\mathbb{R}^n$ such that

$$\begin{aligned} A_n\Gamma_n &\le A_n\phi \qquad\qquad (*)_k \\ A_k\phi(y_k) &\le \min\left\{ A_k\Gamma_k(u) + \frac{1}{2}\|u - u_0\|^2 \right\}. \end{aligned}$$

Proposition 5.3. *For all $x^* \in X^*$,*

$$\phi(y_k) - \phi_* \le \frac{\|x^* - x_0\|^2}{A_k}$$

and as a consequence,

$$\phi(y_n) - \phi_* \le \frac{d_0^2}{2A_n} \text{ where } d_0 = \min\{\|x^* - u_0\| : x^* \in X^*\}.$$

Accelerated Framework

0) Let $A_0 \ge 0$, $u_0, y_0 \in \mathbb{R}^n$, $\Gamma_0 \in \overline{\text{Conv}}\,\mathbb{R}^n$ (e.g. $A_0 = 0, y_0 = u_0 = x_0$, and $\Gamma_0 = 0$) be such that (Γ_0, A_0, y_0) satisfies $(*)_0$. Set $k = 0$ and

$$x_0 = \operatorname*{argmin}_u \left\{ A_0\Gamma_0(u) + \frac{1}{2}\|u - u_0\|^2 \right\}.$$

1) Compute $\lambda_k > 0$, $y_{k+1} \in \mathbb{R}^n$ and $\gamma_{k+1} \in \overline{\text{Conv}}\, \mathbb{R}^n$ such that $\gamma_{k+1} \leq \phi$ and compute

$$\begin{aligned}
a_{k+1} &= \frac{\lambda_k + \sqrt{\lambda_k^2 + 4\lambda_k A_k}}{2} \qquad (1)\\
A_{k+1} &= A_k + a_k\\
\Gamma_{k+1} &= \frac{A_k \Gamma_k + a_k \gamma_k}{A_k + a_k} \qquad (2')\\
x_{k+1} &= \operatorname*{argmin}_u \left\{ A_{k+1}\Gamma_{k+1}(u) + \frac{1}{2}\|u - u_0\|^2 \right\} \qquad (2)\\
\tilde{x}_{k+1} &= \frac{A_k}{A_{k+1}} y_k + \frac{a_{k+1}}{A_{k+1}} x_k\\
\tilde{y}_k &= \frac{A_k}{A_{k+1}} y_k + \frac{a_{k+1}}{A_{k+1}} x_{k+1}. \qquad (3)
\end{aligned}$$

Find y_{k+1} that satisfies

$$\phi(y_{k+1}) \leq \gamma_k(\tilde{y}_{k+1}) + \frac{1}{2\lambda_k}\|\tilde{y}_{k+1} - \tilde{x}_{k+1}\|^2. \qquad (**)$$

2) Set $k \leftarrow k+1$ and go to 1).

Observations

(1) A sufficient condition for $(**)$ is

$$\phi(y_{k+1}) \leq \min_u \left\{ \gamma_k(u) + \frac{1}{2\lambda_k}\|u - \tilde{x}_{k+1}\|^2 \right\}$$

(2) Choose γ_k such that $\phi(u) \leq \gamma_k(u) + \frac{1}{2\lambda_k}\|u - \tilde{x}_{k+1}\|$ for all u and

$$y_{k+1} = \operatorname*{argmin}_u \left\{ \gamma_k(u) + \frac{1}{2\lambda_k}\|u - \tilde{x}_{k+1}\|^2 \right\}.$$

Alternatively choose γ_k such that $\phi(u) \leq \gamma_k(u) + \frac{1}{2\lambda_k}\|u - \tilde{x}_{k+1}\|$ for all u and $y_{k+1} = \tilde{y}_{k+1}$.

Lemma 5.9. *If $(*)_k$ holds, then $A_{k+1}\Gamma_{k+1} \leq A_{k+1}\phi$ and for all $u \in \mathbb{R}^n$,*

$$A_{k+1}\Gamma_{k+1}(u) + \frac{1}{2}\|u - u_0\|^2 \geq A_{k+1}\left[\gamma_k(\tilde{u}_k(u)) + \frac{1}{2\lambda_k}\|\tilde{u}_k(u) - \tilde{x}_{k+1}\|^2\right]$$

where

$$\tilde{u}_k(u) = \frac{A_k y_k + a_k u}{A_{k+1}}.$$

Proof. We have

$$\begin{aligned}
&A_{k+1}\Gamma_{k+1}(u) + \frac{1}{2}\|u - u_0\|^2\\
&\overset{(2')}{=} A_k\Gamma_k(u) + a_k\gamma_k + \frac{1}{2}\|u - u_0\|^2\\
&\geq A_k\phi(y_k) + \frac{1}{2}\|u - u_k\|^2 + a_k\gamma_k(u)\\
&\overset{(*)_k}{\geq} A_k\gamma_k(y_k) + a_k\gamma_k(u) + \frac{1}{2}\|u - x_k\|^2\\
&\overset{\text{convexity}}{\geq} (A_k + a_k)\gamma_k\left(\frac{A_k y_k + a_k u}{A_k + a_k}\right) + \frac{1}{2}\|u - x_k\|^2\\
&= (A_k + a_k)\gamma_k\left(\tilde{u}_k(u)\right) + \frac{(A_k + a_k)^2}{2a_k^2}\|\tilde{u}_k(u) - \tilde{u}_k(x_k)\|^2\\
&= (A_k + a_k)\left[\gamma_k\left(\tilde{u}_k(u)\right) + \frac{A_k + a_k}{2a_k^2}\|\tilde{u}_k(u) - \tilde{u}_k(x_k)\|^2\right].
\end{aligned}$$

Noting that $(A_k + a_k)/a_k^2 = 1/\lambda_k$ we obtain:

$$
\begin{aligned}
&(A_k + a_k)\left[\gamma_k\left(\tilde{u}_k(u)\right) + \frac{A_k + a_k}{2a_k^2}\|\tilde{u}_k(u) - \tilde{u}_k(x_k)\|^2\right]. \\
=&A_{k+1}\left[\gamma_k\left(\tilde{u}_k(u)\right) + \frac{1}{2\lambda_k}\|\tilde{u}_k(u) - \tilde{u}_k(x_k)\|^2\right].
\end{aligned}
$$

□

Proposition 5.4. *If $(*)_k$ holds then $(*)_{k+1}$ holds.*

Proof. Directly,

$$
\begin{aligned}
&\min_u \left\{A_{k+1}\Gamma_{k+1}(u) + \frac{1}{2}\|u - u_0\|^2\right\} \\
&\overset{(2)}{=} A_{k+1}\Gamma_{k+1}(x_{k+1}) + \frac{1}{2}\|x_{k+1} - u_0\|^2 \\
&\overset{\text{lemma}}{\geq} A_{k+1}\left[\gamma_k(\tilde{u}_k(x_{k+1})) + \frac{1}{2\lambda_k}\|\tilde{u}_k(x_{k+1}) - x_{k+1}\|^2\right] \\
&\overset{(3)}{=} A_{k+1}\left[\gamma_k(\tilde{y}_{k+1}) + \frac{1}{2\lambda_k}\|\tilde{y}_{k+1} - \tilde{x}_{k+1}\|^2\right] \\
&\overset{(**)}{=} A_{k+1}\phi(y_{k+1})
\end{aligned}
$$

□

Proposition 5.5. *For every $k \geq 0$,*

(a) $x_k = \operatorname{argmin}_u \left\{A_k\Gamma_k(u) + \frac{1}{2}\|u - x_0\|^2\right\}$

(b) $A_k\Gamma_k = A_0\Gamma_0 + \sum_{i=1}^{k-1} a_i\gamma_i$

(c) $A_k = A_0 + \sum_{i=1}^{k-1} a_i$

(d) $A_k\Gamma_k \leq A_k\phi$

(e) $A_k\phi(y_k) \leq \min_u \left\{A_k\Gamma_k(u) + \frac{1}{2}\|u - x_0\|^2\right\}$

(1) $\phi \neq \operatorname{dom}(A_k\phi) \subseteq \operatorname{dom} A_k\Gamma_k$

(2) x_k is well-defined and $x_k \in \operatorname{dom}(A_k\Gamma_k)$

(3) $y_{k+1} \in \operatorname{dom}\phi$

(4) if $\{x_k\} \subseteq \operatorname{dom}\phi$ then $\tilde{x}_{k+1}, \tilde{y}_{k+1} \in \operatorname{dom}\phi$.

Proposition 5.6. *For every $k \geq 0$,*

(a) $A_k\Gamma_k = A_0\Gamma_0 + \sum_{i=1}^{k-1} a_i\gamma_i$ *and* $A_k = A_0 + \sum_{i=1}^{k-1} a_i$

(b) $(*)_k$ holds

(c) $A_k \geq \left(\sqrt{A_0} + \frac{1}{2}\sum_{i=0}^{k-1}\sqrt{\lambda_i}\right)^2$ *and for $\lambda_i = 1/L$ constant across i, we have $A_k \geq k^2/(4L)$. In particular,*

$$
\phi(y_k) - \phi_* \leq \frac{2Ld_0^2}{k^2}.
$$

Proof. For (c), we have $a_k \geq (\lambda_k/2) + \sqrt{\lambda_k A_k}$ and

$$
\begin{aligned}
&A_{k+1} = A_k + a_k \geq A_k + \sqrt{\lambda_k A_k} + (\lambda_k/2) \geq (\sqrt{A_k} + \sqrt{\lambda_k}/2)^2 \\
\implies &\sqrt{A_k} \geq \sqrt{A_k} + \sqrt{\lambda_k}/2.
\end{aligned}
$$

So $\sqrt{A_k} \geq \sqrt{A_0} + \frac{1}{2}\sum_{i=0}^{k-1}\sqrt{\lambda_i}$.

Composite Optimization (Again)

Consider the problem

$$\min_{x\in\mathbb{R}^n} \{\phi(x) = f(x) + h(x)\}$$

where: □

- $h \in \overline{\text{Conv}}\,\mathbb{R}^n$
- f is convex on $\operatorname{dom} h$
- f is differentiable on $\operatorname{dom} h$ and $\exists L \geq 0$ such that

$$\|\nabla f(x) - f(x')\| \leq L\|x - x'\|, \forall x, x' \in \operatorname{dom} h.$$

Variants I to III

$\tilde{\gamma}_k(\cdot) = \ell_f(\cdot; \tilde{x}_{k+1}) + h(\cdot)$ and $\lambda_k = 1/L$.

(I) $y_{k+1} = \tilde{y}_{k+1}$

(II) $y_{k+1} = \operatorname{argmin}\left\{\tilde{\gamma}_k(x) + \frac{L}{2}\|x - \tilde{x}_{k+1}\|^2\right\}$. Note $\partial\tilde{\gamma}_k(y_{k+1}) + L(y_{k+1} - \tilde{x}_{k+1}) = 0$.

(III) Consider $\gamma_k(x) = \tilde{\gamma}_k(y_{k+1}) + \left\langle \underbrace{L(\tilde{x}_{k+1} - y_{k+1})}_{\in\partial\tilde{\gamma}_k(y_{k+1})}, x - y_{k+1} \right\rangle$ and y_{k+1} in (II). We have

(1) $\gamma_k \leq \tilde{\gamma}_k$

(2) $\gamma_k(y_{k+1}) = \tilde{\gamma}_k(y_{k+1})$

(3) $y_{k+1} = \operatorname{argmin}\left\{\gamma_k(x) + \frac{L}{2}\|x - \tilde{x}_{k+1}\|^2\right\}$.

Proposition 5.7. *Given (x_k, y_k), if we find (x_{k+1}, y_{k+1}, a_k) such that*

$$A_k\phi(y_k) + a_k\gamma_k(x) + \frac{1}{2}\|x - x_k\|^2 \geq (A_k + a_k)\phi(y_{k+1}) + \frac{1}{2}\|x - x_{k+1}\|^2$$

then this implies the invariant

$$A_k\phi(y_k) \leq \min\left\{A_k\Gamma_k(x) + \frac{1}{2}\|x - u_0\|^2\right\}.$$

Proof. Summing from $i = 0, ..., k-1$ with $u_0 = x_0$, $A_0 = 0$, and $\Gamma_0 = 0$ we get

$$A_0\phi(y_0) + \frac{1}{2}\|x - x_0\|^2 + \sum_{i=0}^{k-1} a_i\gamma_i(x) \geq A_k\phi(y_k) + \frac{1}{2}\|x - x_1\|^2$$

or equivalently,

$$A_k\phi(y_k) \leq \sum_{i=0}^{k-1} a_i\gamma_i(x) + \frac{1}{2}\|x - u_0\|^2.$$

□

5.4 Conditional Gradient Method

Consider the problem

$$\begin{aligned} &\min\ f(x) \\ &\text{s.t. } x \in X \end{aligned}$$

where $X \neq \emptyset$ is compact convex. Suppose that:

- f is convex on X
- f is differentiable on X
- $\exists L \geq 0$ such that
$$\|\nabla f(x') - \nabla f(x)\| \leq L\|x' - x\|, \forall x, x' \in X.$$

Conditional Gradient Method

(0) Let $x_0 \in X$ and set $k = 0$.

(1) Let $y_k \in \text{argmin}_{x \in X} \ell_f(x, x_k)$.

(2) Set $x_{k+1} = x_k + \theta_k(y_k - x_k)$ where

$$\theta_k = \underset{\theta \in [0,1]}{\text{argmin}} \left[\ell_f(x_k(\theta); x_k) + \frac{L}{2}\|x_k(\theta) - x_k\|^2\right]$$
$$x_k(\theta) = x_k + \theta(y_k - x_k).$$

(3) Set $k \leftarrow k + 1$ and go to (1).

Observations

(1) $\theta_k = \text{argmin}_{\theta \in [0,1]} \left[\theta \underbrace{\langle \nabla f(x_k), y_k - x_k \rangle}_{\leq 0} + \frac{\theta^2 L}{2}\|y_k - x_k\|^2\right]$

(2) $\langle \nabla f(x_k), y_k - x_k \rangle \leq 0$ for all $k \geq 0$.

Proof. Optimality conditions for step (1) of the conditional gradient method imply that

$$\nabla f(x_k)^T (x - y_k) \geq 0, \forall x.$$

For $x = x_k$ the result for (2) holds. □

(3) $\langle \nabla f(x_k), x_k - y_k \rangle = 0$ implies y_k is optimal for (1)

Proof. We have
$$\langle \nabla f(x_k), x - x_k \rangle = \langle \nabla f(x_k), x - y_k \rangle \geq 0, \forall x \in X.$$
So x_k satisfies the optimality condition for (1) and hence it is optimal for (1). □

Remark 5.4. We have

$$\begin{aligned} f(x_{k+1}) = f(x_k(\theta_k)) &\leq \ell_f(x_k(\theta_k); x_k) + \frac{L}{2}\|x_k(\theta_k) - x_k\|^2 \\ &= f(x_k) + \nabla f(x_k)^T (x_k(\theta_k) - x_k) + \frac{L}{2}\theta_k\|y - x_k\|^2 \end{aligned}$$

and also
$$\theta_k = \max\left\{1, \frac{\langle \nabla f(x_k), x_k - y_k \rangle}{L\|y_k - x_k\|^2}\right\}.$$

So
$$\theta_k \leq \frac{\langle \nabla f(x_k), x_k - y_k \rangle}{L\|y_k - x_k\|^2}$$

and

$$\begin{aligned} f(x_{k+1}) &\leq f(x_k) + \theta_k \nabla f(x_k)^T (y_k - x_k) + \frac{\theta_k}{2} \langle \nabla f(x_k), x_k - y_k \rangle \\ &= f(x_k) - \frac{\theta_k}{2} \langle \nabla f(x_k), x_k - y_k \rangle < f(x_k). \end{aligned}$$

So

$$\underbrace{f(x_{k+1}) - f_*}_{\alpha_{k+1}} \le \underbrace{f(x_k) - f_*}_{\alpha_k} - \frac{\theta_k}{2} \langle \nabla f(x_k), x_k - y_k \rangle$$
$$\le f(x_k) - f_* - \frac{\theta_k}{2} \langle \nabla f(x_k), x_k - x^* \rangle, x^* \in X^*$$
$$\le f(x_k) - f_* - \frac{\theta_k}{2} (f(x_k) - f_*)$$

and we have two cases:

(1) $\theta_k = 1 \implies f(x_{k+1}) - f_* \le \frac{1}{2} (f(x_k) - f_*)$

(2) $\theta_k < 1 \implies \alpha_{k+1} \le \alpha_k - \frac{(\langle \nabla f(x_k), x_k - y_k \rangle)^2}{\|x_k - y_k\|^2} \le \alpha_k - \frac{\alpha_k^2}{2L\|x_k - y_k\|^2}$.

Remark 5.5. Let $D = \max\{\|x' - x\| : x, x' \in X\}$. So $\alpha_{k+1} \le \alpha_k - \tau\alpha_k^2$ where $\tau = 1/(2LD^2)$ and

$$\frac{1}{\alpha_k} \ge \frac{1}{\alpha(1 - \tau\alpha_k)} \ge \frac{1 + \tau\alpha_k}{\alpha_k} = \frac{1}{\alpha_k} + \tau$$

and hence

$$\frac{1}{\alpha_k} \ge \frac{1}{\alpha_0} + k\tau \implies \alpha_k \le \frac{\alpha_0}{1 + k\alpha_0\tau} \le \frac{1}{k\tau} = \frac{2LD^2}{k}$$

and the reduction is $f(x_{k+1}) - f_* \le 2LD^2/k$. It can be shown that the rate of convergence for $f(x_k) - \eta_k$ for $\eta_k = \max_{\ell=1,..,k} \ell_f(y_\ell, x_\ell) \le f_*$ is similar.

6 Monotone Operators

Definition 6.1. For a multivalued map $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, i.e. $z \mapsto T(z) \subseteq \mathbb{R}^n$, e.g. $\partial f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ for $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ where ∂f maps to a set $\subseteq \mathbb{R}^n$, the **graph** of T is $\operatorname{gr} T = \{(z, v) : v \in T(z)\}$. T is said to be **monotone** if for $(z, v), (z', v') \in \operatorname{gr} T$ we have $\langle z - z', x - x' \rangle \ge 0$. T is said to be **maximal monotone** if $\nexists$ a monotone operator T' such that $\operatorname{gr} T' \supseteq \operatorname{gr} T$ and $\operatorname{gr} T' \ne \operatorname{gr} T$.

Examples

(1) The problem

$$\min f(x)$$
$$\text{s.t. } x \in \mathbb{R}^n$$

for $f \in \overline{\text{Conv}}\, \mathbb{R}^n$ is equivalent to finding x such that $0 \in \partial f(x)$ where $T = \partial f$ is maximal monotone.

(2) $0 \ne C \subseteq \mathbb{R}^n$ closed convex $\implies N_C(z) = \partial(I_X)(z) = \{n : \langle n, z' - z \rangle \le, \forall z' \in C\}$

(3) $C \subseteq \mathbb{R}^n$, $D \subseteq \mathbb{R}^m$ nonempty convex sets. The function $K : \mathbb{R}^n \times \mathbb{R}^m$ is **convex-concave** if

- $K(x, y) = \begin{cases} \in \mathbb{R}, & \text{if } (x, y) \in C \times D \\ -\infty, & \text{if } x \in C, y \notin D \\ +\infty, & \text{if } x \notin C \end{cases}$
- $\forall (x, y) \in C \times D$, $K(\cdot, y) - K(x, \cdot) \in \overline{\text{Conv}}(\mathbb{R}^n \times \mathbb{R}^m)$.

Proposition 6.1. *$T(x, y) = \partial_x K(x, y) \times \partial_y(-K)(x, y)$ is maximal monotone.*

Remark 6.1. Note that

$$\begin{aligned}(0,0)\in T(\bar{x},\bar{y}) &\iff \begin{cases}0\in\partial_x K(\bar{x},\bar{y})\\ 0\in\partial_y K(\bar{x},\bar{y})\end{cases}\\ &\iff \begin{cases}\bar{x}\in \operatorname{argmin}_{x\in\mathbb{R}^n} K(x,\bar{y}) = \operatorname{argmin}_{x\in C} K(x,\bar{y})\\ \bar{y}\in \operatorname{argmin}_{y\in\mathbb{R}^m} -K(\bar{x},y) = \operatorname{argmax}_{y\in D} K(\bar{x},y)\end{cases}\\ &\iff K(\bar{x},y)\le K(\bar{x},\bar{y})\le K(x,\bar{y}), \forall x,y\in C\times D.\end{aligned}$$

Min-Max Interpretation

Remark that

$$\inf_{x\in C}\underbrace{\sup_{y\in D} K(x,y)}_{p(x)} = \inf_{x\in C} p(x) \ge \sup_{y\in D}\underbrace{\inf_{x\in C} K(x,y)}_{d(y)} = \sup_{y\in D} d(y)$$

and let X^*, Y^* be the set of optimal points for $p(x)$ and $d(y)$ respectively. Then $(\bar{x},\bar{y})$ is a **saddle point** $\iff \bar{x}\in X^*, \bar{y}\in Y^*$ and $p(\bar{x}) = d(\bar{y})$.

(4) The problem

$$\begin{aligned}\min\ & f(x)\\ \text{s.t. } & g(x)\le 0\\ & x\in X\end{aligned}$$

which is equivalent to

$$\min_{x\in X}\max_{y\ge 0}\ f(x)+\langle y, g(x)\rangle \equiv \max_{y\ge 0}\min_{x\in X}\ f(x)+\langle y, g(x)\rangle\,.$$

If $f,g\in\overline{\text{Conv}}\,\mathbb{R}^n$ and X is closed then

$$K(x,y)=\begin{cases}f(x)+\langle y,g(x)\rangle, & \text{if } (x,y)\in X\times\mathbb{R}^m_+\\ +\infty, & \text{if } x\notin X\\ -\infty, & \text{if } x\in X, y\not\geq 0\end{cases}$$

is a convex-concave function. Now

$$\begin{aligned}(0,0)\in T(x,y) &= \partial_x K(x,y)\times\partial_y(-K)(x,y)\\ &= \begin{pmatrix}\partial f(x)+\sum_{i=1}^m y_i\partial g_i(x)+N_X(x)\\ -g(x)+N_{\mathbb{R}^m_+}(y)\end{pmatrix}.\end{aligned}$$

(5) [**Variational Inequalities**] $C\subseteq\mathbb{R}^n$ closed convex set, $F: C\mapsto\mathbb{R}^n$ is continuous monotone. Then,

$$T(x)=(F+N_C)(x)=\begin{cases}F(x)+N_C(x), & \text{if } x\in C\\ \emptyset, & \text{otherwise}\end{cases}$$

is maximal monotone. Now

$$\begin{aligned}& 0\in T(x)=(F+N_C)(x)\\ \iff & -F(x)\in N_C(x)\\ \iff & \langle F(x), x'-x\rangle \ge, \forall x'\in C.\end{aligned}$$

If C is also a cone then $\langle F(x), x\rangle\ge 0, F(x)\ge 0, x\ge 0$ and $F(x)=0$.

6.1 Proximal Point Methods

We have

$$\begin{aligned} 0 \in T(z) &\iff 0 \in \lambda T(z), \lambda > 0 \\ &\iff z \in z + (\lambda T)(z), \lambda > 0 \\ &\iff z \in (I + \lambda T)(z) \\ &\iff z \in \underbrace{(I + \lambda T)^{-1}}_{F_\lambda}(z) \end{aligned}$$

where $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, $T^{-1} : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$, and $T^{-1}(w) = \{z : T(z) \ni w\}$.

Lemma 6.1. *If T is monotone and $\lambda > 0$ then for all $z \in \mathbb{R}^n$ there is at most one $w \in \mathbb{R}^n$ such that $(I + \lambda T)(w) = z$.*

Lemma 6.2. *Assume $(I + \lambda T)w_i = z_i$, $i = 1, 2$. Then*

$$\langle z_1 - z_2, w_1 - w_2 \rangle \geq \|w_1 - w_2\|^2 \implies \|z_1 - z_2\| \geq \|w_1 - w_2\|.$$

Proof. We have $z_i - w_i \in (\lambda T)(w_i)$ and $\langle (z_1 - w_1) - (z_2 - w_2), w_1 - w_2 \rangle \geq 0$. The rest is left as an exercise. □

Proposition 6.2. *(Minty) Assume T is monotone and $\lambda > 0$. Then the following are equivalent:*

(a) T is maximal monotone

(b) Range$(I + \lambda T) = \mathbb{R}^n$

(c) $\operatorname{dom}(I + \lambda T)^{-1} = \mathbb{R}^n$

Remark 6.2. T maximal monotone implies that:

- $F_\lambda = (I + \lambda T)^{-1}$ is a point-to-point operator so that $\operatorname{dom}(F_\lambda) = \mathbb{R}^n$
- F_λ is non-expansive: $\|F_\lambda(z) - F_\lambda(z')\| \leq \|z - z'\|$ for all $z, z' \in \mathbb{R}^n$
- $\|F_\lambda(z) - F_\lambda(z')\|^2 \leq \langle F_\lambda(z) - F_\lambda(z'), z - z' \rangle$

Recall the optimality condition is $z = F_\lambda(z)$. The **prox-point method** (PPM) is to iterate:

(1) $z_{k+1} = F(z_k)$

(2) $k \hookleftarrow k + 1$.

for $F = F_\lambda$. Assume $\exists z^*$ such that $0 \in T(z^*)$ or $F(z^*) = z^*$. Note that the PPM method has conditions

$$\begin{aligned} \langle z_{k+1} - z^*, z_k - z^* \rangle &= \langle F(z_{k+1}) - F(z^*), z_k - z^* \rangle \\ &\geq \|F(z_k) - F(z^*)\|^2 \\ &= \|z_{k+1} - z^*\|^2 \geq 0 \end{aligned}$$

and hence

$$\langle z_{k+1} - z^*, z_k - z_{k+1} \rangle \geq 0.$$

Now,

$$\begin{aligned} \|z_k - z^*\| &= \|z_k - z_{k+1} + z_{k+1} - z^*\|^2 \\ &\geq \|z_k - z_{k+1}\|^2 + \|z_{k+1} - z^*\|^2 \end{aligned}$$

and so

$$\begin{aligned} &\|z_k - z^*\|^2 - \|z_{k+1} - z^*\|^2 \geq \|z_k - z_{k+1}\|^2 \\ \implies &\|z_0 - z^*\|^2 - \|z_{k+1} - z^*\|^2 \geq \sum_{i=1}^{k} \|z_{i-1} - z_i\|^2 = \sum_{i=1}^{k} \|F(z_{i-1}) - z_{i-1}\|^2 \end{aligned}$$

which says that $\{z_k\}$ is bounded. From Bolzano-Weierstrass, if $\{z_k\} \overset{k\in K}{\to} \bar{z}$ and $z_{k+1} = F(z_k) \to F(\bar{z})$ then

$$\|z_k - F(z_k)\| = \|z_k - z_{k+1}\| \overset{k\to\infty}{\to} 0$$

and since $\|z_k - F(z_k)\| \overset{k\in K}{\to} \|\bar{z} - F(\bar{z})\|$ then $\bar{z} = F(\bar{z})$. Now $\|z_k - \bar{z}\|$ is non-increasing and has a subsequence going to 0. Hence $z_k \to \bar{z}$ and $\bar{z} \in T^{-1}(0)$. the complexity is

$$\min_{i=0,\ldots,k-1} \|F(z_i) - z_i\|^2 \leq \frac{d_0^2}{k}.$$

6.2 Inexact Proximal Methods

From the proximal point methods, recall the equivalent definitions of the update

$$\begin{aligned} & x_k = (I + \lambda_k T)^{-1}(x_{k-1}) \\ \iff & x_{k-1} \in (I + \lambda_k)(x_k) \\ \iff & x_k \text{ is the unique solution of } \lambda_k T(x) + x - x_{k-1} = 0. \end{aligned}$$

This can also be rewritten as

$$\begin{cases} \lambda_k T(\tilde{x}) + x - x_{k-1} \ni 0 \\ \tilde{x} - x. \end{cases}$$

The **inexact case** is described as follows. For some $\lambda_k > 0$, find $(\tilde{x}, x, \varepsilon)$ such that

$$\begin{cases} \lambda_k T^{\varepsilon}(\tilde{x}) + \tilde{x} - x_{k-1} \ni 0 \\ \|\tilde{x} - x\|^2 + 2\lambda_k \varepsilon \leq \sigma^2 \|\tilde{x} - x_{k-1}\|^2, \sigma \in [0,1]. \quad (*) \end{cases}$$

Set $(x_k, \tilde{x}_k, \varepsilon_k) = (x, \tilde{x}, \varepsilon)$ and $k \hookleftarrow k+1$. Here we define

$$T^{\varepsilon}(\tilde{x}) = \{\tilde{v} : \langle \tilde{v} - v, \tilde{x} - x \rangle \geq -\varepsilon, \forall (x, v) \in \operatorname{gr} T\}$$

Properties

1) T monotone $\implies T \subseteq T^0$, i.e. $\operatorname{gr} T \subseteq \operatorname{gr} T^0$

2) T maximal monotone $\iff T = T^0$

3) $(A^{\varepsilon_1} + B^{\varepsilon_2}) \subseteq (A+B)^{\varepsilon_1 + \varepsilon_2}$

4) If $T = \partial f$, $f \in \overline{\text{Conv}}\, \mathbb{R}^n$ then $\partial_\varepsilon f \in T^{\varepsilon}$

5) If $T(x,y) = \partial \left[K(\cdot, y) - K(x, \cdot)\right](x,y)$ where K is a closed convex-concave map then

$$\partial_\varepsilon \left[K(\cdot, y) - K(x, \cdot)\right] \subseteq K^{\varepsilon}.$$

Goal

Given $(\bar{\rho}, \bar{\varepsilon}) \in \mathbb{R}^2_{++}$ find $(\tilde{x}, \tilde{v}, \tilde{\varepsilon})$ such that

$$\begin{aligned} & \tilde{v} \in T^{\tilde{\varepsilon}}(\tilde{x}) \\ & \|\tilde{v}\| \leq \bar{\rho}, \tilde{\varepsilon} \leq \bar{\varepsilon}. \end{aligned}$$

This is the **hybrid proximal extragradient** (HPE) method originally proposed by Solodov and Svaiter.

Pointwise

Assume $\lambda_k = \lambda$ for all k and $\sigma < 1$. Then $\exists i \leq k$ such that

$$\Sigma_i \leq \frac{\sigma^2 d_0^2}{2k(1-\sigma^2)\lambda}, \|\tilde{v}_i\|^2 \leq \frac{(1+\sigma)^2 d_0^2}{(1-\sigma^2)k\lambda^2}$$

Ergodic

Define the ergodic iterates

$$v_k^a = \frac{\sum_{i=1}^k \tilde{v}_i}{k}, \tilde{x}^a = \frac{\sum_{i=1}^k \tilde{x}_i}{k}, \tilde{\varepsilon}^a = \frac{1}{k}\sum_{i=1}^k \left[\varepsilon_i + \langle v_i, \tilde{x}_i - \tilde{x}^0 \rangle\right].$$

The equivalent formulas are

$$\begin{aligned} \tilde{v}_k^a &\in T^{\tilde{\varepsilon}_k^a}(\tilde{x}_k^a) \\ \tilde{\varepsilon}_k^a &\le \frac{1}{2}\left(2 + \frac{\sigma}{(1-\sigma^2)^{1/2}}\right)^2 \frac{d_0^2}{\lambda k} \\ \|\tilde{v}_k^a\| &\le \frac{2d_0}{\lambda k}. \end{aligned}$$

Generalizations

Consider the problem of finding x such that

$$0 \in F(x) + N_X(x) = (F + N_X)(x)$$

where $F : X \mapsto \mathbb{R}^n$ monotone and

$$\|F(x) - F(x')\| \le L\|x' - x\|$$

and X closed convex. The exact problem is

$$\lambda\left[F(x) + N_X(x)\right] + x - x_0 \ni 0, \forall x' \in X.$$

Tseng's Forward-Back Splitting method is to solve the inexact problem

$$\begin{cases} \lambda(F(x_0) + N_X(\tilde{x})) + \tilde{x} - x_0 \ni 0 & (1) \\ \lambda\left[F(\tilde{x}) + n\right] + x - x_0 \ni 0. & (2) \end{cases}$$

Here,

$$\begin{aligned} \tilde{x} &= \mathcal{P}_x\left(x_0 - \lambda F(x_0)\right) \\ n &= \frac{x_0 - \tilde{x}}{\lambda} - F(x_0) \\ x &= x_0 - \lambda\left[F(\tilde{x}) + n\right] \end{aligned}$$

and we will examine the convergence for when $\lambda = \sigma/L$ and $\sigma \in (0, 1)$.

Special case of HPE framework

The parameters are $T = F + N_X$, $\varepsilon = 0$, $\lambda_k = \lambda$ for all k. Also remark that

$$\begin{aligned} \lambda_k T^{\varepsilon}(\tilde{x}) + x - x_0 &= \lambda(F + N_X)(\tilde{x}) + x - x_0 \\ &\ni \lambda(F(\tilde{x}) + n) + x - x_0 \overset{(2)}{=} 0. \end{aligned}$$

If $\varepsilon = 0$ then $(*)$ is equivalent to

$$\|\tilde{x} - x\| \le \sigma\|\tilde{x} - x_0\|.$$

Subtracting (2) from (1) gives us

$$x - \tilde{x} + \lambda\left[F(\tilde{x}) - F(x_0)\right] = 0.$$

So,

$$\|x - \tilde{x}\| = \lambda\|F(\tilde{x}) - F(x_0)\| \le \lambda L\|\tilde{x} - x_0\| = \sigma\|\tilde{x} - x_0\|$$

where the last equality comes from $\lambda = \sigma/L$. The **Korpelevich method** is similar to Tseng's method except that x is obtained with another projection

$$x = \mathcal{P}_X(x_0 - \lambda F(\tilde{x})).$$

6.3 ADMM

Consider the problem

$$\begin{aligned} \min\ & f(x) + g(y) \\ \text{s.t. } & Ax + By = 0 \end{aligned}$$

and $f, g \in \overline{\text{Conv}}\, \mathbb{R}^n$. The method considers solving

$$\min_{x,y} f(x) + g(y) + \langle \lambda, Ax + By \rangle + \frac{\rho}{2}\|Ax + By\|^2.$$

The idea is, given y_0:

1. Set $x \hookleftarrow \min_u\ f(u) + g(y_0) + \langle \lambda_0, Au + By_0 \rangle + \frac{\rho}{2}\|Au + By_0\|^2$.
2. Set $y \hookleftarrow \min_v\ f(x) + g(v) + \langle \lambda_0, Ax + Bv \rangle + \frac{\rho}{2}\|Ax + Bv\|^2$.

The respective optimality conditions are

$$0 \in \partial f(x) + A^*(\lambda_0 + \rho(Ax + By_0)) \qquad (1)$$
$$0 \in \partial g(y) + B^*(\lambda_0 + \rho(Ax + By)) \qquad (2)$$

and for λ we have

$$0 = \lambda - \lambda_0 - \rho(Ax + By). \qquad (3)$$

Let $\tilde{\lambda} = \lambda_0 + \rho(Ax + By)$. Then

$$\begin{aligned} (1) &\iff \partial f(x) + A^*\tilde{\lambda} \ni 0 \\ (2) &\iff \partial g(x) + B^*\tilde{\lambda} + \rho(B^*B)(y - y_0) \ni 0 \\ (3) &\iff -\rho(Ax + By) + \lambda - \lambda_0 = 0. \end{aligned}$$

This can be written as the system

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \in \begin{bmatrix} 0 & 0 & A^* \\ 0 & 0 & B^* \\ -A & -B & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ \tilde{\lambda} \end{bmatrix} + \begin{bmatrix} \partial f(x) \\ \partial g(x) \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \rho(B^*B)(y - y_0) \\ (\lambda - \lambda_0)/\rho \end{bmatrix}$$

which can be written in the HPE framework with $\sigma = 1$ (see a paper by Monteiro).