

ANDREW TULLOCH

## 1. EXISTENCE

**Definition.** For  $C \subseteq \mathbb{R}^n$ , define  $\delta_C$  as

$$\delta_C(x) = \begin{cases} 0 & x \in C \\ \infty & x \notin C \end{cases} \quad (1.1)$$

Note  $x'$  minimizes  $f$  over  $C$  if and only if  $x'$  minimizes  $f + \delta_C$  over  $\mathbb{R}^n$ .

**Definition.** (i)  $\text{dom } f = \{x \in \mathbb{R}^n \mid f(x) < \infty\}$ ,  
 (ii)

$$\arg \min f = \begin{cases} \emptyset & f \equiv \infty \\ \{x \in \mathbb{R}^n \mid f(x) = \inf f\} & f < \infty \end{cases} \quad (1.2)$$

(iii)  $f$  is **proper** if and only if  $\text{dom } f \neq \emptyset$  and  $f(x) > -\infty$  for all  $x \in \mathbb{R}^n$ .

**Definition.** For  $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ , denote

$$\text{epi } f = \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \leq \alpha\} \quad (1.3)$$

**Theorem.** A set  $C$  is an epigraph if and only if for every  $x$  there is an  $\alpha \in \bar{\mathbb{R}}$  such that  $C \cap (x \times \mathbb{R}) = [\alpha, \infty]$  - so all vertical one-dimensional sections must be closed upper half-lines. If  $f$  is proper then  $\text{epi } f$  is not empty and does not include a complete vertical line.

**Definition.** For  $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ , define

$$\liminf_{x \rightarrow x'} f(x) = \lim_{\delta \downarrow 0} \inf_{\|x-x'\|_2 \leq \delta} f(x) = \lim_{k \rightarrow \infty} \inf_{\|x-x'\|_2 \leq \frac{1}{k}} f(x) \quad (1.4)$$

$f$  is **lower semicontinuous** at  $x'$  if and only if  $f(x') \leq \liminf_{x \rightarrow x'} f(x)$ .  
 $f$  is **lower semicontinuous** if  $f$  is lower semicontinuous at every  $x' \in \mathbb{R}^n$ .

**Theorem.**

$$\liminf_{x \rightarrow x'} f(x) = \min\{\alpha \in \bar{\mathbb{R}} \mid \exists (x^k) \rightarrow x' : f(x^k) \rightarrow \alpha\} \quad (1.5)$$

In particular,  $f$  is lower semi-continuous at  $x'$  if and only if  $f(x') \leq \liminf_{k \rightarrow \infty} f(x^k)$  for all convergence sequences  $x^k \rightarrow x'$ .

**Theorem.** Let  $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ . Then the following are equivalent:

- (i)  $f$  is lsc on  $\mathbb{R}^n$
- (ii)  $\text{epi } f$  is closed in  $\mathbb{R}^n \times \mathbb{R}$
- (iii) The sub level sets  $\text{lev}_{\leq \alpha} f = \{x \in \mathbb{R}^n \mid f(x) \leq \alpha\}$  are closed in  $\mathbb{R}^n$  for all  $\alpha \in \bar{\mathbb{R}}$ .

*Proof.*  $\text{epi } f$  can only be not closed along vertical lines.  $\square$

**Definition.**  $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  is **level bounded** if and only if  $\text{lev}_{\leq \alpha} f$  is bounded for all  $\alpha \in \bar{\mathbb{R}}$ .

**Theorem.**  $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  is level-bounded if and only if  $f(x^k) \rightarrow \infty$  for all sequences  $(x^k)$  satisfying  $\|x^k\|_2 \rightarrow \infty$ .

*Proof.*  $K(\alpha)$  such that  $x^k \notin \text{lev}_{\leq \alpha} f$  for  $k \geq K(\alpha)$  by boundedness. In reverse, find  $\alpha$  with  $\text{lev}_{\leq \alpha} f$  unbounded and choose  $x^k$  in this set with  $f(x^k) \leq \alpha$ .  $\square$

**Theorem.** Assume  $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  is lsc, level-bounded, and proper. Then  $\inf f(x) \in (-\infty, +\infty)$  and  $\arg \min f$  is nonempty and compact.

*Proof.* Consider  $\bigcap_{\alpha \in \mathbb{R}, \alpha > \inf f} \arg \min f$ , then this is countable intersection of nonempty, compact sets, so intersection is nonempty.

Finiteness of  $\inf$  is shown as for any  $x \in \arg \min f \neq \emptyset$ , must have  $f(x) = -\infty$  contradicting properness.  $\square$

**Theorem.** (i)  $f, g$  lsc, proper implies  $f + g$  is lsc.  
 (ii)  $f$  lsc,  $\lambda \geq 0$  implies  $\lambda f$  is lsc  
 (iii)  $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  is lsc and  $g : \mathbb{R}^m \rightarrow \mathbb{R}^n$  is continuous implies  $f \circ g$  is lsc.

**Definition.**  $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  is **convex** if and only if

$$f((1-\tau)x + \tau y) \leq (1-\tau)f(x) + \tau f(y) \quad (2.1)$$

for all  $x, y \in \mathbb{R}^n, \tau \in (0, 1)$ .

A set  $C \subseteq \mathbb{R}^n$  is **convex** if and only if  $\delta_C$  is convex if and only if  $(1-\tau)x + \tau y \in C$  for all  $x, y \in C, \tau \in (0, 1)$ .

$f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  is **strictly convex** if and only if  $f$  is convex and the inequality is strict for all  $x \neq y$  and  $\tau \in (0, 1)$ .

**Definition.** Let  $x_0, \dots, x_m \in \mathbb{R}^n$  and  $\lambda_0, \dots, \lambda_m \geq 0, \sum_{i=0}^m \lambda_i = 1$ . The linear combination  $\sum_{i=0}^m \lambda_i x_i$  is a **convex combination** of the points  $x_0, \dots, x_m$ .

**Theorem.** (i)  $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  is convex if and only if

$$f\left(\sum_{i=0}^m \lambda_i x_i\right) \leq \sum_{i=0}^m \lambda_i f(x_i) \quad (2.2)$$

for all  $m \geq 0, x_i \in \mathbb{R}^n, \lambda_i \geq 0, \sum_{i=0}^m \lambda_i = 1$ .

(ii)  $C \subseteq \mathbb{R}^n$  is convex if and only if  $C$  contains all convex combinations of its elements.

**Theorem.**  $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  is convex implies  $\text{dom } f$  is convex.

**Theorem.** (i)  $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  is convex if and only if  $\text{epi } f$  is convex in  $\mathbb{R}^n \times \mathbb{R}$ .

(ii)  $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  is strictly convex if and only if  $\{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) < \alpha\}$  is convex in  $\mathbb{R}^n \times \mathbb{R}$ .

(iii)  $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  is convex implies  $\text{lev}_{\leq \alpha} f$  is convex for all  $\alpha \in \bar{\mathbb{R}}$ .

**Theorem.** Assume  $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  is convex. Then

- (i)  $\arg \min f$  is convex.
- (ii)  $x$  is a local minimizer of  $f$  implies  $x$  is a global minimize of  $f$ .
- (iii)  $f$  is strictly convex and proper implies  $f$  has at most one global minimizer.

**Theorem.** Let  $I$  be an arbitrary index set. Then

- (i)  $f_i, i \in I$  convex implies  $\sup_{i \in I} f_i(x)$  is convex.
- (ii)  $f_i, i \in I$  strictly convex,  $I$  finite implies  $\sup_{i \in I} f_i(x)$  is strictly convex.
- (iii)  $C_i, i \in I$  is convex implies  $\bigcap_{i \in I} C_i$  is convex.
- (iv)  $f_k, k \in N$  is convex implies  $\limsup_{k \rightarrow \infty} f_k(x)$  is convex.

**Theorem.** Assume  $C \subseteq \mathbb{R}^n$  is open and convex, and  $f : C \rightarrow \mathbb{R}$  is differentiable. Then the following are equivalent:

- (i)  $f$  is [strictly] convex
- (ii)  $\langle y - x, \nabla f(y) - \nabla f(x) \rangle \geq 0$  for all  $x, y \in C$  [and  $> 0$  if  $x \neq y$ ]
- (iii)  $f(x) + \langle y - x, \nabla f(x) \rangle \leq f(y)$  for all  $x, y \in C$  [and  $< f(y)$  if  $x \neq y$ ]
- (iv) If  $f$  is additionally twice differentiable, then  $\nabla^2 f(x)$  is positive semidefinite for all  $x \in C$ .

*Proof.* Reduce to one-dimensional sections  $g(t) = f(x + t(y - x))$ . Take  $g(x+h)$ , use convexity to show  $g'(x) \leq \frac{g(y) - g(x)}{y-x}$ .

Use  $g(x) = \sup_{y \in \text{dom } g} g(y) + (y-x)g'(y)$  to show the supremum over convex functions is convex.  $\square$

**Theorem.** (i) Assume  $f_1, \dots, f_m : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  are convex,  $\lambda_1, \dots, \lambda_m \geq 0$ . Then  $f = \sum_{i=1}^m \lambda_i f_i$  is convex. If at least one of the  $f_i$  with  $\lambda_i > 0$  is strictly convex, then  $f$  is strictly convex.

(ii) Assume  $f_i : \mathbb{R}^{n_i} \rightarrow \bar{\mathbb{R}}$  are convex. Then

$$f : \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_m} \rightarrow \bar{\mathbb{R}} \quad f(x_1, \dots, x_m) = \sum_{i=1}^m f_i(x_i) \quad (2.3)$$

is convex. If all  $f_i$  are strictly convex, then  $f$  is strictly convex.

(iii) If  $f : \mathbb{R}^m \rightarrow \bar{\mathbb{R}}$  is convex,  $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$ . Then  $g(x) = f(Ax + b)$  is convex.

**Theorem.** (i)  $C_1, \dots, C_m$  convex implies  $C_1 \times \dots \times C_m$  is convex.  
 (ii)  $C \subseteq \mathbb{R}^n$  convex,  $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, L(x) = Ax + b$  implies  $L(C)$  is convex.  
 (iii)  $C \subseteq \mathbb{R}^m$  convex,  $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, L(x) = Ax + b$  implies  $L^{-1}(C)$  is convex.  
 (iv)  $C_1, C_2$  is convex implies  $C_1 + C_2$  is convex.  
 (v)  $C$  convex,  $\lambda \in \mathbb{R}$  implies  $\lambda C$  is convex.

**Definition.** For a set  $S \subseteq \mathbb{R}^n$  and  $x \in \mathbb{R}^n$ , define the projection of  $x$  onto  $S$  as

$$\Pi_S(y) = \arg \min_{x \in S} \|x - y\|_2. \quad (2.4)$$

**Theorem.** Assume  $C \subseteq \mathbb{R}^n$  is convex, closed, and  $C \neq \emptyset$ . Then  $\Pi_C$  is single-valued - that is, the projection of  $x$  onto  $C$  is unique for every  $x \in \mathbb{R}^n$ .

*Proof.*  $f$  is lsc, level-bounded, and proper, so  $\arg \min f \neq \emptyset$ .  $f$  is strictly convex so has at most one minimizer.  $\square$

**Definition.** For an arbitrary set  $S \subseteq \mathbb{R}^n$ ,

$$\text{con } S = \bigcap_{C \text{ convex}, S \subseteq C} C \quad (2.5)$$

is the convex hull of  $S$ . It is the smallest convex set that contains  $S$ .

**Theorem.**  $\text{con } S = \{\sum_{i=0}^p \lambda_i x_i \mid x_i \in S, \lambda_i \geq 0, \sum_{i=0}^p \lambda_i = 1, p \geq 0\} = D$

*Proof.*  $D \subseteq \text{con } S$  as a convex set contains all convex combinations of points in  $S$ , thus  $D \subseteq \text{con } S$ .

$\text{con } S \subseteq D$  as taking linear combination of  $x, y \in D$ , we have a convex combination of elements in  $D$ , which is in  $D$ , so  $\text{con } S \subseteq D$ .  $\square$

**Theorem.** We have

- (i)  $\text{cl } C = \{x \in \mathbb{R}^n \mid \forall \text{ open neighborhoods } N \text{ of } x, N \cap C \neq \emptyset\}$
- (ii)  $\text{int } C = \{x \in \mathbb{R}^n \mid \exists \text{ open neighborhood } N \text{ of } x \text{ such that } N \subseteq C\}$
- (iii)  $\text{bnd } C = \text{cl } C \setminus \text{int } C$ .

**Theorem.**  $\text{cl } C = \bigcap_{S \text{ closed}, C \subseteq S} S$ .

### 3. CONES AND GENERALIZED INEQUALITIES

**Definition.**  $K \subseteq \mathbb{R}^n$  is a cone if and only if  $0 \in K$  and  $\lambda x \in K$  for all  $x \in K, \lambda \geq 0$ .

A cone  $K$  is pointed if and only if  $\sum_{i=0}^m x_i = 0, x_i \in K$  implies  $x_i = 0$  for all  $i$ .

**Theorem.** Let  $K \subseteq \mathbb{R}^n$  be arbitrary. Then the following are equivalent:

- (i)  $K$  is a convex cone.
- (ii)  $K$  is a cone and  $K + K \subseteq K$ .
- (iii)  $K \neq \emptyset$  and  $\sum_{i=0}^m \alpha_i x_i \in K$  for all  $x_i \in K$  and  $\alpha_i \geq 0$  (not necessarily summing to 1).

*Proof.*  $x = \sum_{i=0}^m \alpha_i x_i \in K, \alpha_i \geq 0 \iff \sum_{i=0}^m \frac{\alpha_i}{\sum_j \alpha_j} x \in K$ , which is a convex combination.  $\square$

**Theorem.** Assume  $K$  is a convex cone. Then  $K$  is pointed if and only if  $K \cap -K = \{0\}$ .

*Proof.*  $x_1 + x_2 + \dots + x_m \in K, x_1 \neq 0, x_2 + \dots + x_m = -x_1, x_2 + \dots + x_m \in K$ , so  $x_1 \in K \cap -K$ .  $\square$

**Theorem.** For a closed convex cone  $K \subseteq \mathbb{R}^n$  we define the generalized inequality

$$x \leq_K y \iff x - y \in K \quad (3.1)$$

Then

- (i)  $x \leq_K x$
- (ii)  $x \leq_K y, y \leq_K z \Rightarrow x \leq_K z$
- (iii)  $x \leq_K y \Rightarrow -y \leq_K -x$
- (iv)  $x \leq_K y, \lambda \geq 0 \Rightarrow \lambda x \leq_K \lambda y$
- (v)  $x \geq_K y, x' \geq_K y' \Rightarrow x + x' \geq_K y + y'$
- (vi)  $x^k \rightarrow x, y^k \rightarrow y$  with  $x^k \geq_K y^k$  for all  $k \in \mathbb{N}$ , then  $x \geq_K y$ .
- (vii)  $x \geq_K y, y \geq_K x \Rightarrow x = y$  (antisymmetry) holds if and only if  $K$  is pointed.

**Definition** ( $K_n^{LP}$ ). For any pointed, closed, convex cone  $K \subseteq \mathbb{R}^m$ , a matrix  $A \in \mathbb{R}^{m \times n}$ , and vectors  $c \in \mathbb{R}^n, b \in \mathbb{R}^m$ , define the conic problem  $\inf_x c^T x$  s.t.  $Ax \geq_K b$ .

**Definition** ( $K_n^{SOCP}$ ).

$$K_n^{SOCP} = \{x \in \mathbb{R}^n \mid x_n \geq \sqrt{x_1^2 + \dots + x_{n-1}^2}\}. \quad (3.2)$$

### 4. SUBGRADIENTS

**Definition.** For any  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  and  $x \in \mathbb{R}^n$ ,

$$\partial f(x) = \{v \in \mathbb{R}^n \mid f(x) + \langle v, y - x \rangle \leq f(y) \forall y \in \mathbb{R}^n\} \quad (4.1)$$

is the set of subgradients of  $f$  at  $x$ .

**Theorem.** Assume  $f, g : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  are convex. Then

- (i) If  $f$  is differentiable at  $x$ , then  $\partial f(x) = \{\nabla f(x)\}$
- (ii) If  $f$  is differentiable at  $x$  and  $g(x) \in \mathbb{R}$ , then  $\partial(f+g)(x) = \partial g(x) + \nabla f(x)$ .

**Theorem.** Assume  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is proper. Then  $x \in \arg \min f \iff 0 \in \partial f(x)$ .

*Proof.*

$$0 \in \partial f(x) \iff f(x) \leq f(y) \forall y \iff x \in \arg \min f \quad (4.2)$$

where properness was required since  $\arg \min +\infty = \emptyset$  by definition.  $\square$

**Definition.** For a convex set  $C \subseteq \mathbb{R}^n$  and  $x \in C$ , the normal cone  $N_C(x)$  at  $x$  is

$$N_C(x) = \{v \in \mathbb{R}^n \mid \langle v, y - x \rangle \leq 0 \forall y \in C\} \quad (4.3)$$

$N_C(x) = \emptyset$  for  $x \notin C$ .

Note that  $N_C(x)$  is a cone if  $x \in C$ .

**Theorem.** Assume  $C \subseteq \mathbb{R}^n$  is convex with  $C \neq \emptyset$ . Then  $\partial \delta_C(x) = N_C(x)$ .

*Proof.* For  $x \in C$ , this follows easily, and for  $x \notin C$ , choosing  $y \in C$  shows that  $\partial \delta_C(x) = \emptyset$ .  $\square$

**Theorem.** Assume  $C \subseteq \mathbb{R}^n$  is closed and convex with  $C \neq \emptyset$  and  $x \in \mathbb{R}^n$ . Then  $y \in \Pi_C(x) \iff x - y \in N_C(y)$ .

*Proof.* Follows from  $y$  being the unique minimizer of  $f(y') = \frac{1}{2} \|y' - x\|_2^2 + \delta_C(y')$ .  $\square$

**Theorem.** Assume  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is proper and convex. Then

$$\partial f(x) = \begin{cases} \emptyset & x \notin \text{dom } f \\ \{v \in \mathbb{R}^n \mid (v, -1) \in N_{\text{epi } f}(x, f(x))\} & x \in \text{dom } f \end{cases} \quad (4.4)$$

If  $x \in \text{dom } f$  then  $N_{\text{dom } f}(x) = \{v \in \mathbb{R}^n \mid (v, 0) \in N_{\text{epi } f}(x, f(x))\}$ .

*Proof.*  $x \notin \text{dom } f$ : choose  $f(y) < \infty$ , then  $\partial f(x) = \emptyset$ .

$x \in \text{dom } f$ :  $v^T(y - x) + (-1)(-f(x)) \leq 0 \forall (y, \alpha) \in \text{epi } f \iff (v, -1) \in N_{\text{epi } f}(x, f(x))$ .  $\square$

**Definition.** For any set  $C \subseteq \mathbb{R}^n$ , define the affine hull and relative interior by

$$\text{aff } C = \bigcap_A \text{ affine}, C \subseteq A \quad (4.5)$$

$$\text{rint } C = \{x \in \mathbb{R}^n \mid \exists \text{ open neighborhood } N \text{ of } x \text{ with } N \cap \text{aff } C \subseteq C\}. \quad (4.6)$$

**Theorem.** (i) Assume  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is convex. Then

$$g(x) = f(x + y) \Rightarrow \partial g(x) = \partial f(x + y) \quad (4.7)$$

$$g(x) = f(\lambda x) \Rightarrow \partial g(x) = \lambda \partial f(\lambda x), \lambda \neq 0 \quad (4.8)$$

$$g(x) = \lambda f(x) \Rightarrow \partial g(x) = \lambda \partial f(x), \lambda > 0 \quad (4.9)$$

$$(4.10)$$

(ii) Assume  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is proper and convex, and  $A \in \mathbb{R}^{n \times m}$  is such that

$$\{Ay \mid y \in \mathbb{R}^m\} \cap \text{rint dom } f \quad (4.11)$$

If  $x \in \text{dom}(f \circ A) = \{y \in \mathbb{R}^m \mid Ay \in \text{dom } f\}$ , then  $\partial(f \circ A)(x) = A^T \partial f(Ax)$ .

(iii) Assume  $f_0, \dots, f_m : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  are proper and convex, and  $\text{rint dom } f_0 \cap \dots \cap \text{rint dom } f_m \neq \emptyset$ . If  $x \in \text{dom } f$ , then  $\partial(\sum_{i=0}^m f_i)(x) = \sum_{i=0}^m \partial f_i(x)$ .

### 5. CONJUGATE FUNCTIONS

**Definition.** For  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ ,  $\text{con } f(x) = \sup_{g \leq f, g \text{ convex}}$  is the convex hull of  $f$ .  $\text{con } f$  is the greatest convex function majorized by  $f$ .

**Definition.** For  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ , the lower closure  $\text{cl } f$  is defined as  $(\text{cl } f)(x) = \liminf_{y \rightarrow x} f(y)$ . Alternatively,  $(\text{cl } f)(x) = \sup_{g \leq f, g \text{ lsc}} g(x)$ .

**Theorem.** For  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ , we have  $\text{epi}(\text{cl } f) = \text{cl}(\text{epi } f)$ . Moreover, if  $f$  is convex then  $\text{cl } f$  is convex.

*Proof.*  $(x, \alpha) \in \text{cl}(\text{epi } f) \iff \exists x^k \rightarrow x, \alpha^k \rightarrow \alpha, f(x^k) \leq \alpha^k \iff \liminf_{y \rightarrow x} f(y) \leq \alpha \iff (x, \alpha) \in \text{epi}(\text{cl } f)$ .  $\square$

**Theorem.** Assume  $C \subseteq \mathbb{R}^n$  is closed and convex. Then

$$C = \bigcap_{(b,\beta), C \subseteq H_{b,\beta}} H_{b,\beta} \quad (5.1)$$

where  $B_{b,\beta} = \{x \in \mathbb{R}^n \mid \langle x, b \rangle - \beta \leq 0\}$

*Proof.* Separating hyperplane theorem -  $x \in C$  then  $x$  is the intersection. If  $x \notin C$  consider the separating hyperplane of  $C$  and  $\{x\}$ .  $\square$

**Theorem.** Assume  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is proper, lsc, and convex. Then  $f(x) = \sup_{g \text{ affine}, f \leq g} g(x)$ .

*Proof.*  $f$  lsc implies  $\text{epi } f$  is closed,  $f$  is convex implies  $\text{epi } f$  is convex, so  $\text{epi } f$  is the intersection of all half spaces containing it. Show  $g$  affine  $\iff$  there exists  $(b, c), \beta$  with  $c < -$ ,  $\text{epi } g = H_{(b,c),\beta}$ .

Then as  $\text{epi}(\sup_{g \leq f} g(x)) = \bigcap_{g \leq f} \text{epi } g$  where  $g$  is affine and  $g \leq f \iff \text{epi } f \subseteq \text{epi } g$ , we need only show  $I_1 = \bigcap_{(b,c), \beta \in S} H_{(b,c),\beta} = \bigcap_{(b,c), \beta \in S, c < 0} H_{(b,c),\beta} = I_2$ .  $\square$

**Definition.** Let  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ , then

$$f^* : \mathbb{R}^n \rightarrow \overline{\mathbb{R}} \quad (5.2)$$

$$f^*(v) = \sup_{x \in \mathbb{R}^n} \langle v, x \rangle - f(x) \quad (5.3)$$

is the *conjugate to  $f$* . The mapping  $f \mapsto f^*$  is the Legendre-Fenchel transform.

**Theorem.** Assume  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ . Then  $f^* = (\text{con } f)^* = (\text{cl } f)^* = (\text{cl con } f)^*$  and  $f^{**} = (f^*)^* \leq f$ . If  $\text{con } f$  is proper, then  $f^*$  and  $f^{**}$  are proper, lsc, and convex, and  $f^{**} = \text{cl con } f$ . If  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is proper, lsc, and convex, then  $f^{**} = f$ .

*Proof.*  $(v, \beta) \in \text{epi } f^* \iff (v, x) - \beta \leq f(x) \forall x$  so  $(v, \beta) \in \text{epi } f^*$  define all affine functions majorized by  $f$ . Thus for every affine function  $h$ ,  $h \leq f \iff h \leq \text{cl } f \iff h \leq \text{con } f \iff h \leq \text{cl con } f$ . Which gives our result.

For the inequality, expand  $f^{**}(y) = \sup_v \langle v, y \rangle + \inf_x (f(x) - \langle v, x \rangle) \leq f(y)$  at  $y = x$ .

As  $\text{con } f$  is proper, then  $\text{cl con } f$  is proper, lsc, and convex, so  $\text{cl con } f = \sup_{g \leq \text{cl con } f} g(x)$  where  $g$  is affine, which equals  $f^{**}$  by definition.

Finally, if  $f$  is convex, then  $\text{con } f = f$ , so  $\text{con } f$  is proper, and  $\text{con } f = f$  is lsc, so  $f^{**} = \text{cl con } f = \text{cl } f = f$ .  $\square$

**Theorem.** Assume  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ . Then

- (i)  $\text{con } f$  is not proper implies  $f^* \equiv +\infty$  or  $f^* \equiv -\infty$ .
- (ii) In particular,  $f^*$  proper implies  $\text{con } f$  is proper.

*Proof.* If  $\text{con } f = \infty$ , then  $f^*(v) = -\infty$ . If  $\text{con } f(x') = -\infty$ , then  $f^*(v) = (\text{con } f)^*(v) \geq +\infty$ .  $\square$

**Theorem.** Assume  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is proper, lsc, and convex. Then  $\partial f^* = (\partial f)^{-1}$ , specifically,

$$v \in \partial f(x) \iff f(x) + f^*(v) = \langle v, x \rangle \iff x \in \partial f^*(v). \quad (5.4)$$

Moreover,

$$\partial f(x) = \arg \max_{v'} \langle v', x \rangle - f^*(v') \iff \partial f^*(x) = \arg \max_{x'} \langle v, x' \rangle - f(x) \quad (5.5)$$

*Proof.*  $f(x) + f^*(v) = \langle v, x \rangle$  iff  $x \in \arg \max_{x'} \langle v, x' \rangle - f(x') \iff v \in \partial f(x)$ .  $\square$

**Theorem.** For a proper, lsc, convex  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ , we have

$$(f(\cdot) - \langle a, \cdot \rangle)^* = f^*(\cdot + a) \quad (5.6)$$

$$(f(\cdot + b))^* = f^*(\cdot) - \langle \cdot, b \rangle, \quad (5.7)$$

$$(f(\cdot) + c)^* = f^*(\cdot) - c \quad (5.8)$$

$$(\lambda f(\cdot))^* = \lambda f^{*star}(\frac{\cdot}{\lambda}), \lambda > 0 \quad (5.9)$$

$$(\lambda f(\frac{\cdot}{\lambda}))^* = \lambda f^*(\cdot), \lambda > 0 \quad (5.10)$$

**Theorem.** Let  $f_i : \mathbb{R}^{n_i} \rightarrow \overline{\mathbb{R}}, i = 0, \dots, m$  be proper and  $f(x_0, \dots, x_m) = \sum_{i=0}^m f_i(x_i)$ . Then  $f^*(v_1, \dots, v_m) = \sum_{i=0}^m f_i^*(v_i)$ .

*Proof.* For support functions,  $f^*(x) = \sigma_C(x)$ , so  $C$  nonempty closed convex implies  $f^*$  is proper lsc convex with  $f^*(0) = 0$  and  $f^*(\lambda v) = \lambda f^*(v)$ , so  $f^* = \sigma_C$  is positively homogeneous.

If  $g$  is positively homogeneous lower semicontinuous,  $g^*(x) \in \{0, \infty\}$ , so  $g^*$  is an indicator function, which must be convex and nonempty by properness lsc convexity of  $g^*$ .

One to one follows from  $\delta_C^{**} = \delta_C$ .

For cones, show  $g$  is positively homogeneous lsc convex proper indicator function  $\iff g = \delta_K$  for  $K$  a convex closed cone.

Taking any such indicator function  $\delta_K$ , then  $\delta_K^*$  is an indicator function, with  $\delta_K^*(v) < \infty \iff v \in K^*$ .  $\square$

**Definition.** For any set  $S \subseteq \mathbb{R}^n$  define the **support function**  $\text{supp}_S(v) = \sup_{x \in S} \langle v, x \rangle = (\delta_S^*)(v)$

**Definition.** A function  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is said to be **positively homogeneous** if and only if  $0 \in \text{dom } f$  and  $f(\lambda x) = \lambda f(x)$  for all  $x \in \mathbb{R}^n$  and  $\lambda > 0$ .

**Theorem.** The set of positively homogeneous proper lsc convex functions and the set of closed convex nonempty sets are in one-to-one correspondence through the Legendre-Fenchel transform:

$$\delta_C \leftrightarrow \text{supp}_C \quad (5.11)$$

$$x \in \partial \text{supp}_C(v) \iff x \in C \quad (5.12)$$

$$\text{supp}_C(v) = \langle v, x \rangle \iff v \in N_C(x). \quad (5.13)$$

In particular, the set of closed convex cones is in one-to-one correspondence with itself - for any cone  $K$  define the **polar cone** or **dual cone** as  $K^* = \{v \in \mathbb{R}^d \mid \langle v, x \rangle \leq 0 \forall x \in K\}$ . Then

$$\delta_K \leftrightarrow \delta_{K^*} \quad (5.14)$$

$$x \in N_{K^*}(v) \iff v \in N_K(x) \quad (5.15)$$

TODO: fill next condition/implication in.

## 6. DUALITY IN OPTIMIZATION

**Definition.** Assume  $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$  is proper, lsc, and convex. Define the **primal problem** as

$$\inf_{x \in \mathbb{R}^n} \phi(x), \phi(x) = f(x, 0), \quad (6.1)$$

the **dual problem** as

$$\sup_{y \in \mathbb{R}^m} \psi(y), \psi(y) = -f^*(0, y) \quad (6.2)$$

and the **inf-projections**

$$p(u) = \inf_x f(x, u) \quad (6.3)$$

$$q(v) = \inf_y f^*(v, y) = -\sup_y (-f^*(v, y)) \quad (6.4)$$

$f$  is sometimes called a **perturbation function** for  $\psi$ , and  $p$  the associated **marginal function**.

**Theorem.** Assume  $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$  is proper, lsc, and convex. Then

- (i)  $\phi$  and  $-\psi$  are lsc and convex.
- (ii)  $p, q$  are convex.
- (iii)  $p(0)$  and  $p^{**}(0)$  are the optimal values of the primal and dual problems -

$$p(0) = \inf_x \phi(x), p^{**}(0) = \sup_y \psi(y). \quad (6.5)$$

- (iv) The primal and dual problems are feasible if and only if their associated marginal function contains 0:

$$\inf_x \phi(x) < \infty \iff 0 \in \text{dom } p \quad (6.6)$$

$$\sup_y \psi(y) > -\infty \iff 0 \in \text{dom } q \quad (6.7)$$

*Proof.*  $f$  proper lsc convex implies  $f^*$  is proper lsc convex implies  $\pi, \psi$  lsc convex.

$p, q$  are convex from the strict epigraph set, with  $E = \{(u, \alpha) \in \mathbb{R}^m \times \mathbb{R} \mid p(u) = \inf_{x \in \mathbb{R}^n} f(x, u) < \alpha\} = A(E')$ , where  $A$  is the linear projection mapping  $A(x, u, \alpha) = (u, \alpha)$ , and  $E'$  is the strict epigraph of  $f$  and thus convex, so  $A(E')$  is convex.

$p^*(y) = -\psi(y)$ , so  $p^{**}(0) = \sup_y \psi(y)$ .

$0 \in \text{dom } p \iff p(0) < \infty \iff \inf_x f(x, 0) < \infty \iff \inf \psi < \infty$ .  $\square$

**Theorem.** Assume  $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$  is proper, lsc, and convex. Then **weak duality** always holds,

$$\inf_x \phi(x) \geq \sup_y \psi(y), \quad (6.8)$$

and under certain conditions the infimum and supremum are equal and finite - **strong duality**

$$p(0) \in \mathbb{R}, p \text{ lsc in } 0 \iff \inf_x \phi(x) = \sup_y \psi(y) \in \mathbb{R}. \quad (6.9)$$

The difference  $\inf \phi - \sup \psi$  is the **duality gap**.

*Proof.*  $\inf_x \phi(x) = p(0) \geq p^{**}(0) = \sup_y \psi(y)$  so must show  $p(0) \in \mathbb{R}$  and  $p$  lsc in 0 if and only if  $p(0) = p^{**}(0) \in \mathbb{R}$ .

( $\Rightarrow$ ) follows as  $p^{**}(0) \leq \text{cl}p(0) \leq p(0)$ , so  $\liminf_{y \rightarrow 0} p(y) = \text{cl}p(0) = p(0) \in \mathbb{R}$ , so  $p$  is lsc in 0.

( $\Leftarrow$ ) follows from the claim that if it holds then  $\text{cl}p$  is proper lsc convex. Convexity, lsc is clear, and an improper convex lsc function is always constant  $\infty$  or  $-\infty$ , contradiction  $p(0) \in \mathbb{R}$ . So  $(p^*)^*(0) = ((\text{cl}p)^*)^*(0) = \text{cl}p(0) = p(0)$ .  $\square$

**Theorem.** Assume  $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$  is proper, lsc, and convex. Then we have the **primal-dual optimality conditions**,

$$(0, y') \in \partial f(x', 0) \quad (6.10)$$

$$\iff \{x' \in \arg \min_x \phi(x), \quad y' \in \arg \max_y \psi(y), \inf_x \phi(x) = \sup_y \psi(y)\} \quad (6.11)$$

$$\iff (x', 0) \in \partial f^*(0, y'). \quad (6.12)$$

The set of **primal-dual optimal points**  $(x', y')$  satisfying this equation is either empty or equal to  $(\arg \min \phi) \times (\arg \max \psi)$ .

*Proof.* Follows from invertibility of subgradient in terms of conjugate functions, showing  $f(x', 0) = -f^*(0, y') \iff \phi(x') = \psi(y') \in \mathbb{R}$ , and equality with infinite value is explicitly excluded by definition of  $\arg \min$ .  $\square$

**Theorem.** Assume  $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$  is proper, lsc, and convex. Then

- (i)  $0 \in \text{int dom } p$  or  $0 \in \text{int dom } q$  implies  $\inf_x \phi(x) = \sup_y \psi(y)$ . ( $S'$ )
- (ii)  $0 \in \text{int dom } p$  and  $0 \in \text{int dom } q$  implies  $\inf_x \phi(x) = \sup_y \psi(y) \in \mathbb{R}$ . ( $S$ )
- (iii)  $0 \in \int \text{dom } p$  and  $\inf_x \phi(x) \in \mathbb{R}$  if and only if  $\arg \max_y \psi(y)$  is nonempty and bounded. ( $P$ )
- (iv)  $0 \in \int \text{dom } q$  and  $\sup_y \psi(y) \in \mathbb{R}$  if and only if  $\arg \min_x \phi(x)$  is nonempty and bounded. ( $D$ )

In particular, if any of  $S, P, D$  hold, then strong duality holds -  $\inf \phi = \sup \psi \in \mathbb{R}$ . If  $S$ , or ( $P$  and  $D$ ) hold, then there exists  $x', y'$  satisfying the primal-dual optimality conditions. Also,  $P$  implies  $\partial p(0) = \arg \max_y \psi(y)$ , and  $D$  implies  $\partial q(0) = \arg \min_x \phi(x)$ .

*Proof.* (i): If  $p(0) = -\infty$ , then  $p^{**}(0) \leq p(0) = -\infty$  Use  $\text{cl}p = p$  on  $\int \text{dom } p$ , so  $\sup \psi = \inf \psi = -\infty$ . Otherwise,  $p(0) \in \mathbb{R}$  so as  $\text{cl}p = p$  on  $\int \text{dom } p$ , we have  $p$  is lsc in 0 and so  $p^{**}(0) = p(0) \in \mathbb{R}$ .

This follows symmetrically on  $f'(x, y) = f^*(y, x)$ .

If both  $0 \in \int \text{dom } p, 0 \in \int \text{dom } q$ , then  $+\infty > p(0) \geq p^{**}(0) = \sup \psi = -q(0) > -\infty$ , which is finite.

Nonemptiness and boundedness follows from  $0 \in \int \text{dom } p$  if and only if  $\psi$  is proper lsc convex and level bounded.

Subdifferential: if (iii) then  $p(0) \in \mathbb{R}$ , so  $\text{cl}p(0) = p(0) \in \mathbb{R}$ ,  $\text{cl}p$  is then proper and lsc convex, so  $\partial(\text{cl}p)(0) = \arg \max \psi$ , but  $\partial p(0) = \partial(\text{cl}p)(0)$ .  $\square$

**Theorem.** Assume  $k : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  and  $h : \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$  are both proper, lsc, convex, and  $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, c \in \mathbb{R}^n$ . For  $f(x, u) = \langle c, x \rangle + k(x) + h(Ax - b + u)$ , the primal and dual problems are of the form

$$\inf \phi(x), \phi(x) = \langle c, x \rangle + k(x) + h(Ax - b) \quad (6.13)$$

$$\sup_y \psi(y), \psi(y) = -\langle b, y \rangle - h^*(y) - k^*(-A^T y - c) \quad (6.14)$$

with

$$\int \text{dom } p = \int (\text{dom } h - A \text{dom } k) + b \quad (6.15)$$

$$\int \text{dom } q = \int (\text{dom } k^* - (-A^T) \text{dom } h^*) + c \quad (6.16)$$

and optimality conditions

$$\{-A^T y' - c \in \partial k(x'), y' \in \partial h(Ax' - b)\} \quad (6.17)$$

$$\iff \{x' \in \arg \min_x \phi(x), y' \in \arg \max_y \psi(y), \inf_x \phi(x) = \sup_y \psi(y)\} \quad (6.18)$$

$$\iff \{Ax' - b \in \partial h^*(y'), x' \in \partial k^*(-A^T y' - c)\} \quad (6.19)$$

**Theorem.** Assume  $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$  is proper, lsc, convex. Define the associated **Lagrangian** as  $l(x, y) = -f(x, \cdot)^*(y)$ , so  $l(x, y) = \inf_u (f(x, u) - \langle y, u \rangle)$ . Then  $l(\cdot, y)$  is convex for every  $y$ ,  $-l(x, \cdot)$  is lsc and convex for every  $x, y$  and  $f(x, \cdot) = (-l(x, \cdot))^*$ , and  $(v, y) \in \partial f(x, u) \iff v \in \partial_x l(x, y)$  and  $u \in \partial_y(-l)(x, y)$ .

*Proof.* Consider  $g(x, y, u) = f(x, u) - \langle y, u \rangle$ , which is proper lsc convex. Then  $l(\cdot, y) = \inf_u g(\cdot, y, u)$  is convex. Then  $f_x(y) = f(x, y)$ , then  $-l(x, \cdot) = f_x^*(\cdot)$  so  $f_x$  is either  $+\infty$  or proper lsc convex, and  $-l(x, \cdot)$  is either  $-\infty$  or proper lsc convex, but always lsc convex.

By subgradient definition,  $(v, y) \in \partial f(x, u) \iff \inf_{u'} f(x', u') - \langle y, u' \rangle \geq f(x, u) - \langle y, u \rangle + \langle v, x' - x \rangle \forall x'$ , which evaluated at  $x' = x$  implies  $\inf_{u'} f(x', u') - \langle y, u' \rangle = f(x, u) - \langle y, u \rangle$ .

Continuing, we obtain  $(v, y) \in \partial f(x, u) \iff y \in \partial_u f(x, u), \in \partial_x l(x, y)$ , and the first condition is equivalent to  $u \in \partial f_x^*(y)$ , or  $u \in \partial(-l(x, \cdot))^* = \partial_y(-l)(x, y)$  as required.  $\square$

**Definition.** For any function  $l : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$  we say that  $(x', y')$  is a **saddle point** of  $l$  if  $l(x, y') \geq l(x', y') \geq l(x', y)$  for all  $x, y$ . The set of all saddle points is denoted by  $\text{spl}$ .

Equivalently,  $(x', y') \in \text{spl}$  if  $\inf_x l(x, y') = l(x', y') = \sup_y l(x', y)$ .

**Theorem.** Assume  $f$  is proper, lsc, and convex with associated Lagrangian  $l$ . Then  $\phi(x) = \sup_y l(x, y)$ , and  $\psi(y) = \inf_x l(x, y)$ , and the primal problem is  $\inf_x \phi(x) = \inf_x \sup_y l(x, y)$ , and the dual problem is  $\sup_y \psi(y) = \sup_y \inf_x l(x, y)$ . Moreover, the optimality condition

$$\{x' \in \arg \min_x \phi(x), y' \in \arg \max_y \psi(y), \inf_x \phi(x) = \sup_y \psi(y)\} \quad (6.20)$$

$$\iff (x', y') \in \text{spl} \quad (6.21)$$

$$\iff \{0 \in \partial_x l(x', y'), 0 \in \partial_y(-l)(x', y')\} \quad (6.22)$$

**Theorem.** Assume  $X \subseteq \mathbb{R}^n, Y \subseteq \mathbb{R}^m$  are nonempty, closed, convex, and  $L : X \times Y \rightarrow \mathbb{R}$  is a continuous function with  $L(\cdot, y)$  is convex for every  $y$  and  $-L(x, \cdot)$  convex for every  $x$ . Then  $l(x, y) = L(x, y) + \delta_X(x) - \delta_Y(y)$  with the convention  $+\infty - \infty = +\infty$  on the right, is the Lagrangian to  $f(x, u) = \sup_y l(x, y) + \langle u, y \rangle = (-l(x, \cdot))^*(u)$ .

$f$  is proper, lsc, and convex, so the previous result applies with primal and dual problems  $\phi(x) = \delta_X(x) + \sup_{y \in Y} L(x, y)$ ,  $\psi(y) = -\delta_Y(y) + \int_{x \in X} L(x, y)$ . Moreover, if  $X$  and  $Y$  are bounded, then  $\text{spl}$  is nonempty and bounded.

*Proof.* For a fixed  $y$ ,  $\inf_x l(x, y) = -f^*(0, y) = \psi(y)$ , and for a fixed  $x$ ,  $\sup_y l(x, y) = f(x, 0) = f_x^*(0) = f(x, 0) = \phi(x)$  as  $f_x$  is either proper lsc convex or  $+\infty$ .

The optimality condition is equivalent to  $(0, y') \in f(x', 0) \iff 0 \in \partial_x l(x', y'), 0 \in \partial_y(-l)(x', y')$ , which is the saddle point condition.  $\square$

## 7. NUMERICAL OPTIMALITY

**Definition.** For  $\phi : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ , a point  $x$  is an  $\epsilon$ -optimal solution if  $\phi(x) - \inf \phi \leq \epsilon$ .

**Definition.** Assume  $(x^k, y^k)$  is a primal-dual feasible pair - so  $x^k \in \text{dom } \phi$  and  $y^k \in \text{dom } \psi$ . Then  $\phi(x^k) \geq \psi(y^k)$ , and  $0 \leq \phi(x^k) - \inf \phi \leq \phi(x^k) - \psi(y^k) = \gamma(x^k, y^k) := \gamma$ .  $\gamma$  is the **numerical primal-dual gap**. If  $\gamma < \epsilon$  then  $x^k$  is an  $\epsilon$ -optimal solution with **optimality certificate**  $y^k$ .

The normalized gap is  $\bar{\gamma} = \bar{\gamma}(x^k, y^k) = \frac{\phi(x^k) - \psi(y^k)}{\psi(y^k)}$ .

**Definition.** Assume  $\phi(x) = \phi_0(x) = \sum_{i=1}^{n_p} \delta_{g_i(x) \leq 0}$ ,  $\psi(y) = \psi_0(y) = \sum_{i=1}^{n_d} \delta_{h_i(y) \leq 0}$  where  $\text{dom } \phi_0 = \text{dom } \psi_0 = \mathbb{R}^n$  and  $g_i : \mathbb{R}^n \rightarrow \mathbb{R}, h_i : \mathbb{R}^m \rightarrow \mathbb{R}$  are suitable continuous real-valued convex functions, so the primal and dual constraints are of the form  $g_i(x) \leq 0, h_i(y) \leq 0$ . Then the primal and dual infeasibilities are defined as  $\eta_p = \max\{0, g_1(x^k), \dots, g_{n_p}(x^k)\}$  and  $\eta_d = \max\{0, h_1(y^k), \dots, h_{n_d}(y^k)\}$ .

## 8. FIRST-ORDER METHODS

**Definition.** For  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ , we define

- (i) The **forward step**,  $F_{\tau_k f}(x^k) = (I - \tau_k \partial f)x^k$
- (ii) The **backward step**,  $B_{\tau_k f}(x^k) = (I + \tau_k \partial f)^{-1}x^k$ .

**Theorem.** If  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is proper lsc convex with  $\tau > 0$ , then the backward step is  $B_{\tau f}(x) = \arg \min_y \{\frac{1}{2}\|y - x\|_2^2 + \tau f(y)\}$  and is therefore unique.

*Proof.*  $y \in B_{\tau f}(x) \iff 0 \in y - x + \tau \partial f(y) \iff y \in \arg \min_{y'} \{\frac{1}{2}\|y' - x\|_2^2 + \tau f(y')\}$   $\square$

**Theorem.** Assume  $f$  is proper lsc convex and  $\arg \min f \neq \emptyset$ . The **forward step** is  $x^{k+1} \in F_{\tau_k f}(x^k)$ . The sequence is not unique, can get stuck if  $x^k$  is infeasible.

The **backward step** is  $x^{k+1} = B_{\tau_k f}(x^k)$  - which is a **unique** sequence, and cannot get stuck. Sub-steps are as hard as the original problem (but strictly convex).

- (i) Forward stepping:  $x^{k+1} \in F_{\tau_k f}(x^k)$ ;
- (ii) Backward stepping:  $x^{k+1} = B_{\tau_k f}(x^k)$ .

If  $f = g + h$ ,  $\partial f = \partial g + \partial h$  with  $f, g, h$  proper lsc convex,  $\arg \min f \neq \emptyset$ , we can do:

- (i) Backward-Backward stepping:  $x^{k+1} = B_{\tau_k h} B_{\tau_k g}(x^k)$ .
- (ii) Forward-Backward stepping:  $x^{k+1} \in B_{\tau_k h} F_{\tau_k g}(x^k)$ . If  $f(x) = g(x) + \delta_C(x)$ ,  $g$  differentiable,  $C \neq \emptyset$  closed and convex, then  $x^{k+1} \in \arg \min_x \{\frac{1}{2} \|y - (x^k - \tau_k \Delta g(x^k))\|_2^2 + \delta_C(x)\} = \Pi_C(x^k - \tau_k \Delta g(x^k))$ , which is a gradient projection.

## 9. INTERIOR-POINT METHODS

**Definition.** For a cone  $K$  we define the **canonical barriers**  $F = F_K$  and associated parameters  $\theta_F$ .

- (i)  $K = K_n^{LP} = \{x \in \mathbb{R}^n | x_1, \dots, x_n \geq 0\}$ ,  $F(x) = \sum_{i=1}^n -\log x_i$ ,  $\theta_F = n$ .
- (ii)  $K = K_n^{SOCP} = \{x \in \mathbb{R}^n | x_n \geq \sqrt{x_1^2 + \dots + x_{n-1}^2}\}$ ,  $F(x) = -\log(x_n^2 - x_1^2 - \dots - x_{n-1}^2)$ ,  $\theta_F = 2$ .
- (iii)  $K = K_n^{SDP} = \{X \in \mathbb{R}^{n \times n} | X \text{ symmetric positive semidefinite}\}$ ,  $F(x) = -\log \det X$ ,  $\theta_F = n$ .
- (iv)  $K = K^1 \times K^2$ , then  $F_K(x^1, x^2) = F_{K^1}(x^1) + F_{K^2}(x^2)$ , with  $\theta_F = \theta_{F^1} + \theta_{F^2}$ .

**Theorem.** If  $F$  is a canonical barrier for  $K$ , then  $F$  is smooth on  $\text{dom } F = \text{int } K$  and strictly convex,  $F(tx) = F(x) - \theta_F \log t$  for all  $x \in \text{dom } F$ , and for  $x \in \text{dom } F$ , we have

- (i)  $-\nabla F(x) \in \text{dom } F$
- (ii)  $\langle \nabla F(x), x \rangle = -\theta_F$ ,
- (iii)  $-\nabla F(-\nabla F(x)) = x$ ,
- (iv)  $-\nabla F(tx) = -\frac{1}{t} \nabla F(x)$

*Proof.* Differentiate with respect to  $t$  and let  $t = 1$ .  $\square$

**Theorem.** Consider the problem  $\inf \langle c, x \rangle$  s.t.  $Ax - b \geq_K 0$ . The dual problem is  $\sup \langle -b, y \rangle$  s.t.  $-A^T y = c$ ,  $y \geq_{K^*} 0$ . Replacing  $y$  by  $-y$  and assuming  $K$  is self-dual  $K^* = K$  we obtain the dual as  $\sup \langle b, y \rangle$  such that  $A^T y = c$ ,  $y \geq_K 0$ .

The **primal central path** is the mapping

$$t \mapsto x(t) = \arg \min \{-t \langle b, y \rangle + F(y) + \delta_{A^T y = c}\} \quad (9.1)$$

The **dual central path** is the mapping  $t \mapsto y(t) = \arg \min \{-t \langle b, y \rangle + F(y) + \delta_{A^T y = c}\}$

The **primal-dual central path** is the mapping  $t \mapsto z(t) = (x(t), y(t))$  for some  $t > 0$ , if and only if

$$Ax - b \in \text{dom } F \quad (9.2)$$

$$A^T y = c \quad (9.3)$$

$$ty + \nabla F(Ax - b) = 0 \quad (9.4)$$

*Proof.*  $y = -\frac{1}{y} \Delta f(Ax - b) \in \text{dom } F$  as  $Ax - b \in \text{dom } F$  and  $\text{dom } F$  is a cone. We also have that if  $x$  is on the primal path, then  $A^T y = c$ , so  $y$  is feasible.

For dual optimality, we need  $0 \in -tb + \nabla F(y) + N_{A^T y = c}$ . As  $-tb + \nabla F(y) = -tAx \in \text{range } A$ ,  $y$  is the unique dual solution. Multiplying the dual optimality result by  $A^T$  gives the optimality condition for the primal central point.  $\square$

**Theorem.** For feasible  $x, y$  (so  $Ax - b \in K$ ,  $A^T y = c$ ,  $y \in K$ ), the duality gap is  $\phi(x) - \psi(y) = \langle y, Ax - b \rangle$ . Moreover, for points  $(x(t), y(t))$  on the central path, the duality gap is  $\phi(x(t)) - \psi(y(t)) = \frac{\theta_F}{t}$ .

*Proof.*

$$\phi(x) - \psi(y) = \langle c, x \rangle - \langle b, y \rangle \quad (9.5)$$

$$= \langle y, Ax - b \rangle \quad (9.6)$$

$$= \left\langle -\frac{1}{t} \nabla F(Ax(t) - b), Ax(t) - b \right\rangle \quad (9.7)$$

$$= \frac{\theta_F}{t}. \quad (9.8)$$

$\square$

**Theorem.** We define  $\|v\|_x^* = (v^T \nabla^2 F(Ax - b)^{-1} v)^{\frac{1}{2}}$ ,  $z = (x, y)$ , so  $z(t)$  is the primal-dual central path, and  $\text{dist}(z, z(t)) = \|ty + \nabla F(Ax - b)\|_x^*$ . Then for  $Ax - b \in \text{dom } F$ ,  $y \in \text{dom } F$ ,  $A^T y = c$ , we have  $\text{dist}(z, z(t)) \leq 1$  implies  $\phi(x) - \psi(y) \leq 2(\phi(x(t)) - \psi(y(t))) = 2\frac{\theta_F}{t}$ .

*Proof.* If we linearize  $\nabla F(Ax^{k+1} - b) = \nabla F(Ax^k - b + A\Delta x) \approx \nabla F(Ax^k - b) + \nabla^2 F(Ax^k - b)A\Delta x$ , and solve the constraints  $A^T(y) = c$ ,  $ty + \nabla F(Ax - b) = 0$ .  $\square$

**Theorem.** Assume  $0 < \rho \leq \kappa < \frac{1}{10}$ ,  $t^k > 0$  fixed, and  $z^k = (x^k, y^k)$  strictly feasible, so  $Ax^k - b \in \text{dom } F$ ,  $y^k \in \text{dom } f$ , such that  $\text{dist}(z^k, z(t^k)) < \kappa$ .

If we apply a full Newton step with  $\tau_k = 1$  and  $t^{k+1} = (1 + \frac{\rho}{\sqrt{\theta_F}})t^k$  to generate  $z^{k+1}$ , then  $x^{k+1}, y^{k+1}$  are strictly primal and dual feasible, and  $\text{dist}(z^{k+1}, z(t^{k+1})) > \kappa$  as well.

## 10. SUPPORT VECTOR MACHINES

**Definition.** The primal formulation of an SVM is

$$\inf_{w, b} \frac{1}{2} \|w\|_2^2 \quad (10.1)$$

such that  $1 \leq y^i (\langle w, x^i \rangle + b)$  for all  $1 \leq i \leq n$ .

The dual formulation is

$$\inf_{z \in \mathbb{R}^n} \frac{1}{2} \left\| \sum_{i=1}^n y^i x^i z_i \right\|_2^2 + e^T z \quad (10.2)$$

such that  $z_i \leq 0$ ,  $\sum_{i=1}^n y^i z_i = 0$ .

## 11. TOTAL VARIATION AND APPLICATIONS

**Definition.** For  $u \in L^1(\Omega, \mathbb{R}^m)$ , the **total variation** of  $u$  is defined as

$$TV(u) = \sup_{v \in C_c^1(\Omega, \mathbb{R}^m \times \mathbb{R}^n), \|v\|_\infty \leq 1} \int_\Omega \langle u, \text{div } v \rangle dx \quad (11.1)$$

**Theorem.** Assume  $A \subseteq \Omega$  is a set so that its boundary is  $C^1$  and satisfies  $\mathcal{H}^{n-1}(\Omega \cap \partial A) < \infty$ . Define

$$1_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases} \quad (11.2)$$

then  $TV(1_A) = \mathcal{H}^{n-1}(\Omega \cap \partial A)$ .

*Proof.* The lower bound follows from

$$TV(1_A) = \sup_{v \in C_c^1(\Omega, \mathbb{R}^n), \|v\|_\infty \leq 1} \int_{\partial A} \langle v, n \rangle ds \quad (11.3)$$

by Gauss's theorem.  $\square$

**Theorem** (Coarea formula). If  $u \in BV(\Omega)$ , then  $TV(1_{u(x) > t}) < \infty$  for  $\mathcal{L}^1$ -a.e.  $t \in \mathbb{R}$ , and  $TV(u) = \int_{\mathbb{R}} TV(1_{u > t}) dt$ .

**Definition.** For  $\Omega \subseteq \mathbb{R}^d$  and  $k \geq 1$ , define the space  $BV^k(\Omega)$  as  $BV^k = \{u \in W^{k-1,1} | \nabla^{k-1} u \in BV(\Omega, \mathbb{R}^{d^{k-1}})\}$  and the higher-order total variation as

$$TV^k(u) = \sup_{v \in C_c^k(\Omega, \text{Sym}^k(\mathbb{R}^d)), \|v\|_\infty \leq 1} \int_\Omega u \text{div}^k v dx = TV(\nabla^{k-1} u) \quad (11.4)$$

## 12. RELAXATION

**Definition.** Consider the **Chan-Vese model**

$$f_{CV}(C, c_1, c_2) = \int_C (g - c_1^2) dx + \int_{\Omega \setminus C} (g - c_2^2) dx + \lambda \mathcal{H}^{d-1}(C). \quad (12.1)$$

**Theorem.** Let  $c_1, c_2$  be fixed, and consider  $\inf_{u: \Omega \rightarrow [0,1], u \in BV(\Omega)} f(u)$ ,  $f(u) = \langle u, s \rangle_{L^1} + \lambda TV(u)$ . Then if  $u$  is a minimizer of  $f$ , and  $u(x) \in \{0, 1\}$  a.e., then  $C$  is a minimizer of  $f_{CV}(\cdot, c_1, c_2)$ .

*Proof.* Follows by definitions - must have  $u = 1_C$ .  $\square$

**Definition.** Let  $\mathcal{C} = BV(\Omega, [0, 1]) = \{u \in BV(\Omega) | u(x) \in [0, 1] \text{ a.e.}\}$ . Then for  $u \in \mathcal{C}$ ,  $\alpha \in [0, 1]$ . Define  $\bar{u}_\alpha = 1_{\{u > \alpha\}}$ . Then  $f: \mathcal{C} \rightarrow \mathbb{R}$  satisfies the **generalized coarea condition** if and only if

$$f(u) = \int_0^1 f(\bar{u}_\alpha) d\alpha \quad (12.2)$$

for all  $u \in \mathcal{C}$ .

**Theorem.** Let  $f^* u = TV(u)$ , the condition is the coarea formula. As the condition is additive, we need only show  $\int_\Omega s(x) u(x) = \int_0^1 \int_\Omega s(x) 1_{\{u(x) > \alpha\}} dx$ , where we use Fubini due to  $s \in L^\infty(\Omega)$ .

**Theorem.** Assume  $f: \mathcal{C} \rightarrow \mathbb{R}$  satisfies the generalized coarea condition, and  $u^*$  satisfies  $u^* \in \arg \min_{u \in \mathcal{C}} f(u)$ . Then for almost every  $\alpha \in [0, 1]$ , the thresholded function satisfies  $\bar{u}_\alpha^* \in \arg \min_{u \in BV(\Omega, \{0,1\})} f(u)$ .

*Proof.* Follows by considering the set  $S_\epsilon = \{\alpha \in [0, 1] \mid f(u^\star) \leq f(u_\alpha^\star) - \epsilon\}$  for some  $\epsilon > 0$ , and showing that this implies  $f(u^\star) \leq \int_0^1 f(\bar{u}_\alpha^\star) d\alpha - \epsilon L^1(S_\epsilon)$  which contradicts the generalized coarea formula.  $\square$

## REFERENCES