ANDREW TULLOCH

CONVEX OPTIMIZATION

TRINITY COLLEGE
THE UNIVERSITY OF CAMBRIDGE

Contents

1

Introduction

1.1 Setup

A setup is

(i) A function $f : U \to \Omega$ (energy), an input image (i), and a reconstructed candidate image (u) and find the minimizer of the problem

$$\min_{u \in U} f(u,i) \tag{1.1}$$

where f is typically of the form

$$f(u,i) = \underbrace{l(u,i)}_{\text{cost}} + \underbrace{r(u)}_{\text{regulariser}} \tag{1.2}$$

Example 1.1 (Rudin-Osher-Fortem)**.**

$$min_u \frac{1}{2} \int_\Omega (u - I)^2 dx + \int_\Omega \|\nabla u\| du \tag{1.3}$$

- *Will reduce contrast*

- *Will not introduce new jumps*

Example 1.2 (Total variation (TV) L1)**.**

$$min_u \frac{1}{2} \int_\Omega |u - I| dx + \lambda \int_\Omega \|\nabla u\| du \tag{1.4}$$

- *In general does not cause contrast loss.*

- *Can show that if $I = B_r(0)$, then*

$$u = \begin{cases} I & r \geq \frac{\lambda}{2} \\ c & r < \frac{\lambda}{2} \end{cases} \tag{1.5}$$

Example 1.3 (Inpainting).

$$\min_u \frac{1}{2} \int_{\Omega/A} |u - I| dx + \lambda \int_{\Omega} \|\nabla u\| dx \tag{1.6}$$

1.2 *Convexity*

Theorem 1.4

If a function f is convex, every local optimizer is a global optimizer.

Theorem 1.5

If a function f is convex, it can (very often) be efficiently optimized (polynomial time in number of bits in the input)

In computer vision, problems are:

- Usually large-scale (10^5 - 10^7 variables)
- Usually non-differentiable

Definition 1.6. f is lower semicountinuous if

$$f(x') \leq \liminf_{x \to x'} f(x) = \min\{\alpha \in \bar{\mathbb{R}} | (x^k) \to x'_1, f(x^k) \to \alpha\} \tag{1.7}$$

Theorem 1.7. $f : R^n \to \bar{\mathbb{R}}$. *Then the following are equivalent*

(i) *f is lower semi continuous,*

(ii) $\operatorname{epi} f$ *is closed C in $\mathbb{R}^n \times \mathbb{R}$*

(iii) $\operatorname{lev}_\alpha f = \{x \in \mathbb{R}^n | f(x) \leq \alpha\}$ *are closed for all $\alpha \in \bar{\mathbb{R}}$*

Proof. ((*i*) $\to$ (*ii*)) Take $(x^k, \alpha^k) \in \operatorname{epi} f \to (x, \alpha) \in \mathbb{R}^n \times \mathbb{R}$. Then By 2.8 $f(x) \leq \liminf_{k \to \infty} f(x^k) \leq liminf_{k \to \infty} x^k = \alpha$ as $(x, \alpha) \in \operatorname{epi} f$.

((*ii*) $\to$ (*iii*)) $\operatorname{epi} f$ closed implies $\operatorname{epi} f \cap (\mathbb{R}^n \times \{a'\})$ closed for all $\alpha' \in \mathbb{R}$, if and only if

$$\{x \in \mathbb{R}^n | f(x) \leq \alpha'\} \tag{1.8}$$

is closed for all $\alpha' \in \mathbb{R}$ If $\alpha' = \infty$, $\text{lev}_{\leq +\infty} f = \mathbb{R}^n$. If $\alpha' = -\infty$, $\text{lev}_{\leq -\infty} f = \cap_{k \in \mathbb{N}} \text{lev}_{\leq -k} f$ which is the intersection of closed sets.

($(iii) \to (i)$) For $x^j \to x'$, take the subsequence $x^k \to x'$, with

$$f(x^k) \to \liminf_{x \to x'} f(x) = c \tag{1.9}$$

if $c \in \mathbb{R}$, for all $K(\epsilon) = K'$ such that $f(x^k) \leq C + \epsilon$ for all $k > K'$. Then

$\Rightarrow x^k \in \text{lev}_{\leq C+\epsilon} f \Rightarrow x' \in \text{lev}_{\leq C+\epsilon}$.

Equivalently, $x' \in \text{lev}_{\leq C} f \Rightarrow f(x') \leq C = \liminf x \to x' f(x)$.

If $c = +\infty$, done - $f(x') \leq +\infty = \liminf$. If $c = \infty$, same argument with $k \in \mathbb{N}$, $f(x^k) \leq -k, \ldots$ □

Example 1.8. *$f(x) = x$ is lower semi-continuous, but does not have a minimizer.*

Definition 1.9. $f : R^n \to \bar{\mathbb{R}}$ is **level-bounded** if and only if $\text{lev}_{\leq \alpha} f$ is bounded for all $\alpha \in \mathbb{R}$.

This is also known as **coercivity**.

Example 1.10. *$f(x) = x^2$ is level-bounded. $f(x) = \frac{1}{|x|}$ is not level bounded.*

Theorem 1.11. *$f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is level bounded if and only if $f(x^k) \to \infty$ if $\|x^k\| \to \infty$.*

Proof. Exercise. □

Theorem 1.12. *$f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is lower-semicontinuous, level-bounded, and proper. Then $\inf_x f(x) \in (-\infty, +\infty)$ and $argmin f = \{x | f(x) \leq inf f(x)\}$ is nonempty and compact.*

Proof.

$$\arg\min f = \{x | f(x) \leq \inf_x f(x)\} \tag{1.10}$$

$$= \{x | f(x) \leq \inf f(x) + \frac{1}{k}, \forall k \in \mathbb{N}\} \tag{1.11}$$

$$= \cap_{k \in \mathbb{N}} \text{lev}_{\inf f + \frac{1}{k}} f \tag{1.12}$$

If $\inf f$ is $-\infty$, just replace $\inf f + \frac{1}{k}$ by α with $\alpha > \infty$, and set $\alpha_k = -k, k \in \mathbb{N}$.

These are bounded (by level-boundedness), closed (by f being lower semicontinuous), and non-empty (since $\frac{1}{k} \geq 0$.). Then these limit sets are compact, we can just take the limit of the left boundaries of the level sets, and construct the convergent subsequence that is contained in every level set.

We need to show that $\inf f \neq -\infty$. If $\inf f = -\infty$, then there exists $x \in argminf$ with $f(x) = -\infty$. Since f is proper, this cannot exist. Thus, $\inf f \neq = -infty$. $\square$

Remark 1.13. *For Theorem 1.12, it suffices to have* $\operatorname{lev}_{\leq \alpha} f$ *bounded and non-empty for at least one* $\alpha \in \mathbb{R}$.

Proposition 1.14. *We have the following properties for lower semi continuity.*

(i) *If f, g is lower semicontinuous, then $f + g$ is lower semicontinuous*

(ii) *IF f is lower semicontinuous, $\lambda \geq 0$, then λf is lower semicontinuous*

(iii) *$f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is lower semicontinuous, $g : \mathbb{R}^m \to \mathbb{R}^n$ is continuous, then $f \circ g$ is lower semicontinuous.*

2

Convexity

Definition 2.1. (i) $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is **convex** if

$$f((1-\tau)x + \tau y) \leq (1-\tau)f(x) + \tau f(y) \tag{2.1}$$

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

(ii) A set $C \subseteq \mathbb{R}^n$ is convex if and only if $\mathbb{I}(C)$ is convex.

(iii) f is **strictly convex** if and only if (2.1) holds strictly whenever $x \neq y$ and $f(x), f(y) \in \mathbb{R}$.

Remark 2.2. *$C \subseteq \mathbb{R}^n$ is convex if and only if for all $x, y \in C$, the connecting line segment is contained in C.*

Exercise 2.3. *Show*

$$\{x | a^T x + b \geq 0\} \tag{2.2}$$

is convex.

$$f(x) = a^T x + b \tag{2.3}$$

is convex.

Definition 2.4. $x_0, \ldots, x_m \in \mathbb{R}^n$, $\lambda_0, \ldots, \lambda_m \geq 0$ with $\sum_{i=0}^m \lambda_i = 1$, then $\sum_i \lambda_i x_i$ is a convex combination of the x_i.

Theorem 2.5. $\mathbb{R}^n \to \bar{\mathbb{R}}$ *is convex if and only if*

$$f\left(\sum_{i=1}^n \lambda_i x_i\right) \leq \sum_{i=1}^n \lambda_i f(x_i) \tag{2.4}$$

$C \subseteq \mathbb{R}^n$ is convex if and only if C contains all convex combinations of it's elements.

Proof.

$$\sum_{i=1}^{n} \lambda_i x_i = \lambda_m x_m + \left(1 - \lambda_m \frac{\lambda_i}{1-\lambda_m} x_i\right) \tag{2.5}$$

which is a convex combination of two points, and proceed by induction on m.

The set version is proven by application on $\mathbb{I}(C)$. □

Proposition 2.6. *$f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is convex implies that the domain of f is convex.*

Proposition 2.7. *$f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is convex if and only if $\operatorname{epi} f$ is convex in $\mathbb{R}^n \times \mathbb{R}$ if and only if*

$$\{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} | f(x) < \alpha\} \tag{2.6}$$

is convex in $\mathbb{R}^n \times \mathbb{R}$.

Proof. $\operatorname{epi} f$ is convex if and only if for all $\tau \in (0,1), \forall x, y, \alpha \geq f(x), \beta > f(y)$, $(x, \alpha), (y, \beta) \in \operatorname{epi} f$, we have

$$f((1-\tau)x + \tau y) \leq (1-\tau)\alpha + \tau\beta \tag{2.7}$$

$$\iff \forall \tau \in (0,1), \forall x, y, f((1-\tau)x + \tau y) \leq (1-\tau)f(x) + \tau f(y) \tag{2.8}$$

$$\iff f \text{ is convex.} \tag{2.9}$$

□

Proposition 2.8. *$f : R^n \to \bar{\mathbb{R}}$ is convex implies $lev_{\leq \alpha} f$ is convex for all $\alpha \in \overline{\mathbb{R}}$.*

Proof.

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

which implies $lev_{\leq \alpha}$ is convex.

$\alpha = \infty$ then the $lev_{\leq \alpha} f = \mathbb{R}^n$ which is convex. □

Theorem 2.9. *$f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is convex. Then* *(i)* $\arg\min f$ *is convex*

(ii) *x is a local minimizer of f implies x is a global minimizer of f*

(iii) *f is strictly convex and proper implies f has at most **one** global minimizer.*

Proof. (i) $f = \infty \Rightarrow \arg\min f = \emptyset$. $f \neq \infty \Rightarrow \arg\min f = lev_{\leq \inf f} f$ is convex by previous proposition.

(ii) Assume x is a local minimizer and there exists y with $f(y) < f(x)$. Then

$$f((1-\tau)x + \tau y) \leq (1-\tau)f(x) + \tau \underbrace{f(y)}_{<f(x)} < f(x)$$

Taking $\tau \to 0$ shows that x cannot be a local minimizer, and thus y cannot exist.

(iii) Assume x, y minimizes, which implies $f(x) = f(y)$. Then

$$f(\frac{1}{2}x + \frac{1}{2}y) < \frac{1}{2}f(x) + \frac{1}{2}f(y) = f(x) = f(y)$$

which implies x, y not global minimizers.

$\square$

Proposition 2.10. *To construct convex functions:*

(i) *Let $f_i, i \in \mathbb{I}$ convex, then*

$$f(x) = \sup_{i \in \mathbb{I}} f(x) \tag{2.10}$$

is convex.

(ii) *Let $f_i, i \in \mathbb{I}$ strictly convex, $\mathbb{I}$ finite, then*

$$\sup_{i \in \mathbb{I}} f(i) \tag{2.11}$$

is strictly convex.

(iii) *$C_i, i \in \mathbb{I}$ convex sets, then*

$$\cap_{i \in \mathbb{I}} C_i \tag{2.12}$$

is convex.

(iv) *$f_k, k \in \mathbb{N}$ convex,*

$$\limsup_{k\to\infty} f_k(x) \tag{2.13}$$

is convex.

Proof. Exercise. □

Example 2.11. *(i) C, D convex does not imply $C \cup D$ is convex (e.g. disjoint)*

(ii) *$f : R^n \to R$ is convex, C is convex implies $f + \mathbb{I}(C)$ is convex.*

(iii) *$f(x) = |x| = \max\{x' - x\}$ is convex.*

(iv) *$f(x) = \|x\|_p, p \geq 1$ is convex, as*

$$\| \cdot \|_p = \sup_{\|y\|_p = 1} \langle \cdot, y \rangle \tag{2.14}$$

Theorem 2.12. *Let $C \subseteq \mathbb{R}^n$ be open and convex. Let $f : C \to \mathbb{R}$ be differentiable. Then the following are equivalent:*

(i) *f is [strictly] convex*

(ii) *$\langle y - x, \nabla f(y) - \nabla f(x) \rangle \geq 00$ for all $x, y \in C$ [with > 0 for $x \neq y$]*

(iii) *$f(x) + \langle \nabla f(x), y - x \rangle \leq f(y)$ for all $x, y \in C$ [with $<$ for $x \neq y$]*

(iv) *If f is additionally twice differentiable, then $\nabla^2 f$ is positive semidefinite for all $x \in C$.*

(ii) is monotonicity of ∇f.

Proof. Exercise (reduce to $n = 1$, then extend by a convex function is convex on R^n iff it is convex on all R^{n-1} subsets.) □

Remark 2.13. *Note that the inverse of (iv) does not necessarily hold, for example $f(x) = x^4$.*

Proposition 2.14. *We have the following results hold for convex functions.*

(i) *$f_i, \ldots, f_m : \mathbb{R}^n \to \bar{\mathbb{R}}$ is convex, $\lambda_i, \ldots, \lambda_m \geq 0$, the n*

$$f = \sum_i \lambda_i f_i \tag{2.15}$$

*is convex, and **strictly** convex if there exists i such that $\lambda_i > 0$ and f_i is strictly convex.*

(ii) $f_i : \mathbb{R}^n \to \mathbb{R}$ *is convex implies* $f(x_1, \ldots, x_m) = \sum f_i(x_i)$ *is convex, strictly convex if* ***all*** f_i *are strictly convex.*

(iii) $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ *is convex,* $A \in \mathbb{R}^{n \times n}, b \in \mathbb{R}^n$ *implies* $g(x) = f(Ax + b)$ *is convex.*

Remark 2.15.

$$\|M\|_\epsilon = \sup_{x \in \mathbb{R}^n} (\|Mx\|)$$

is convex.

Proposition 2.16. *(i)* $c_1, \ldots, C_m$ *convex implies* $C_1 \times \cdots \times C_m$ *convex*

(ii) $C \subseteq \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ *implies* $L(C)$ *is convex with* $L(x) = Ax + b$.

(iii) $C \subseteq \mathbb{R}^m, ... \Rightarrow L^\star(C)$ *is convex.*

(iv) f, g *convex implies* $f + g$ *are convex*

(v) f *convex,* $\lambda \in \mathbb{R}$ *implies* λf *is convex*

Definition 2.17. For $S \subseteq \mathbb{R}^n, x \in \mathbb{R}^n$, define the projection of x onto S as

$$\Pi_S(y) = \arg\min_{x \in S} \|x - y\|_2 \tag{2.16}$$

Proposition 2.18. *If* $C \subseteq \mathbb{R}^n$ *is convex, closed, and non-empty, then* Π_C *is single-valued - that is, the projection is unique.*

Proof. Let

$$\Pi_S(y) = \arg\min_{x \in S} \frac{1}{2}\|x - y\|_2^2 + \mathbb{I}_C(x) \tag{2.17}$$

To show uniqueness, $\frac{1}{2}\|x - y\|^2$ is strictly convex, and $\nabla^2 \frac{1}{2}\|x - y\|^2 = I > 0$, so f is strictly convex.

f is proper ($C \neq \emptyset$), and thus f has at most one minimizer.

To show existence, we have that f is proper, lower semicontinuous (left part from continuous, right part from C closed). Level bounded as $\|x - y\|_2^2 \to \infty$ as $\|x\|_2 \to \infty$, and $\mathbb{I}(C) \geq 0$.

Thus, the $\arg\min f \neq \emptyset$. $\square$

Definition 2.19. Let $S \subseteq \mathbb{R}^n$ is arbitrary. Then

$$\operatorname{con} S = \cap_{C \text{ convex}, \, C \supseteq S} C \tag{2.18}$$

is the convex ball of S.

Remark 2.20. $\operatorname{con} S$ *is the* ***smallest*** *convex set that contains* S*.*

Theorem 2.21. *Let* $S \subseteq \mathbb{R}^n$*, then*

$$\operatorname{con} S = \left\{ \sum_{i=0}^{n} \lambda_i x_i | x_i \in S, \lambda_i \geq 0, \sum_{i=0}^{n} \lambda_i = 1 \right\} \tag{2.19}$$

Proof. D is convex and contains S - if $x, y \in D$, then $(1-\tau)x + (\tau)y \in D$. Thus $\operatorname{con} S \subseteq D$.

From a previous theorem, $\operatorname{con} S$ convex implies $\operatorname{con} S$ contains all convex combinations of points in S. Thus $\operatorname{con} S \supseteq D$

Thus $\operatorname{con} S = D$. $\square$

Definition 2.22. For a set $C \subseteq \mathbb{R}^n$, define $\operatorname{cl} C$ as

$$\operatorname{cl} C = \{x \in \mathbb{R}^n | \text{for all open neighborhoods } N \text{ of } x, \, N \cap C \neq \emptyset\} \tag{2.20}$$

$\operatorname{int} C$ as

$$\operatorname{int} C = \{x \in \mathbb{R}^n | \text{there exists an open neighborhood } N \text{ of } x \text{ with } N \subseteq C\} \tag{2.21}$$

∂C (the boundary) as

$$\partial C = \operatorname{cl} C \setminus \operatorname{int} C \tag{2.22}$$

Remark 2.23.

$$\operatorname{cl} G = \bigcap_{S \text{ closed}, \, S \supseteq G} S \tag{2.23}$$

3

Cones and Generalized Inequalities

Definition 3.1. $K \subseteq \mathbb{R}^n$ is a cone if and only if $0 \in K$ and $\lambda x \in K$ for all $x \in K, \lambda \geq 0$.

Definition 3.2. A cone in **pointed** if and only if

$$x_1 + \cdots + x_n = 0 \iff x_1 = \cdots = x_n = 0 \tag{3.1}$$

Example 3.3. *(i)* R^n *is a pointed cone.*

(ii) $\{(x_1, x_2) \in \mathbb{R}^2 | x_2 > 0\}$ *is not a cone.*

(iii) $\{(x_1, x_2) \in \mathbb{R}^2 | x_2 \geq 0\}$ *is a cone, but is not pointed.*

(iv) $\{x \in \mathbb{R}^n | x_i \geq 0, i \in 1, \ldots, n\}$ *is a pointed cone.*

Proposition 3.4. $K \subseteq \mathbb{R}^n$ *be any set. 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* $\sum_{i=0}^n \alpha_i x_i \in K$ *for all* $x_i \in K$, $\alpha_i \geq 0$.

Proof. Exercise. □

Proposition 3.5. *If* K *is a convex cone, then* K *is pointed if and only if* $K \cap (-K) = \{0\}$.

Proof. Exercise. □

Definition 3.6.

$$\partial f(x) = \{v \in \mathbb{R}^n | f(x) + \langle v, y - x \rangle \leq f(y) \forall y\} \tag{3.2}$$

$$f'(x) = \lim_{t \to 0} \frac{f(x+t) - f(x)}{t} \tag{3.3}$$

Proposition 3.7. *Let $f, g : \mathbb{R}^n \to \bar{\mathbb{R}}$ be convex. Then*

(i) f differentiable at x implies $\partial f(x) = \{\nabla f(x)\}$.

(ii) f differentiable at x, $g(x) \in \mathbb{R}$ implies

$$\partial(f+g)(x) = \nabla f(x) + \partial g(x) \tag{3.4}$$

Proof. We have

(i) Equivalent to (ii) with $g = 0$.

(ii) To show $LHS \supseteq RHS$, if $v \in \partial g(x)$, then $\langle v, y - x \rangle \leq g(y)$, which by Theorem 3.12 in notes gives

$$\langle \nabla f(x), y - x \rangle + f(x) \leq f(y) \tag{3.5}$$

$$\langle v + \nabla f(x), y - x \rangle + (f+g)(x) \leq (f+g)(y) \tag{3.6}$$

$$\iff v + \nabla f(x) \in \partial(f+g) \tag{3.7}$$

The other direction is given as

$$\liminf z \to x \frac{f(z) + g(z) - f(x) - g(x) - \langle v, z - x \rangle}{\|z - x\|} \geq 0 \tag{3.8}$$

$$\Rightarrow \liminf_{z \to x} \frac{g(z) - g(x) - \langle v - \nabla f(x), z - x \rangle}{\|z - x\|} \geq 0 \tag{3.9}$$

$$\Rightarrow \liminf_{t \downarrow 0} \frac{g(z(t)) - g(x) - \langle v - \nabla f(x), (1-t)x + ty - y \rangle}{t\|y - x\|} \geq 0 \tag{3.10}$$

$$\Rightarrow \liminf_{t \downarrow 0} \frac{t(g(y) - g(x)) - t\langle v - \nabla f(x), y - x \rangle}{t\|y - x\|} \geq 0 \tag{3.11}$$

$$\Rightarrow g(y) - g(x) - \langle v - \nabla f(x), y - x \rangle \geq 0 \Rightarrow v - \nabla f(x) \quad \in \partial g(x) \tag{3.12}$$

$$v \in \nabla f(x) - \partial g(x) \tag{3.13}$$

□

Theorem 3.8. *Let $f : \mathbb{R}^n \to \bar{R}$ be proper. Then*

$$x \in \arg\min f \iff 0 \in \partial f(x) \tag{3.14}$$

Proof.

$$0 \in \partial f(x) \iff \underbrace{\langle 0, y - x\rangle}_{0} + f(x) \leq f(y) \tag{3.15}$$

□

Definition 3.9. For a convex set $C \subseteq \mathbb{R}^n$ and point $x \in C$,

$$N_C(x) = \{v \in \mathbb{R}^n | \langle v, y - x\rangle \leq 0 \forall y \in G\} \tag{3.16}$$

is the "normal cone" of x. By convention, $N_C(x) = \emptyset$ for all $x \notin C$.

Proposition 3.10. *Let C be convex and $C \neq \emptyset$. Then*

$$\partial \mathbb{I}(C)(x) = N_C(x) \tag{3.17}$$

Proof. For $x \in C$, we have

$$\partial \mathbb{I}(C) = \{v | \mathbb{I}(C)(x) + \langle v, y - x\rangle \leq \mathbb{I}(C)(y) \forall y \in C\} = N_C(x) \tag{3.18}$$

For $x \notin C$, . □ Fill in proof here

Proposition 3.11. *C, closed, $C \neq \emptyset$ and convex, $x \in \mathbb{R}^n$. Then*

$$y = \Pi_C(x) \iff x - y \in N_C(y) \tag{3.19}$$

Proof.

$$y \in \Pi_C(x) \iff y \text{ minimizes } \underbrace{\frac{1}{2}\|y - x\|^2 + \mathbb{I}(C)(y')}_{g(y)}. \tag{3.20}$$

If and only if $0 \in \partial g(y)$ if and only if $0 \in y - x + \partial \mathbb{I}(C)(y)$ □

Proposition 3.12. *Let $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ be convex proper. Then*

$$\partial f(x) = \begin{cases} \emptyset & x \notin dom f \\ \{v \in \mathbb{R}^n | (u, -1) \in N_{\text{epi}\, f}(x, f(x))\} & x \in dom f \end{cases} \tag{3.21}$$

If $x \in dom f$,

$$N_{dom f} = \{v \in \mathbb{R}^n | (x, 0) \in N_{\text{epi}\, f}(x, f(x))\} \tag{3.22}$$

Example 3.13. *Let subdifferential of $f : \mathbb{R} \to \mathbb{R}, f(x) = |x|$ is*

$$\partial f(x) = \begin{cases} \text{sign}\, x & x \neq 0 \\ [-1, 1] & x = 0 \end{cases} \tag{3.23}$$

Definition 3.14. For $C \subseteq \mathbb{R}^n$, define the **affine hull** as

$$\text{aff}(C) = \cap_{A \text{ affine}, C \subseteq A} A \tag{3.24}$$

$$\text{rint}(C) = \{x \in \mathbb{R}^n | \text{there exists an open neighborhood } N \text{ of } x \text{ such that } N \cap \text{aff}(C) \subseteq C\} \tag{3.25}$$

Example 3.15.

$$\text{aff}(\mathbb{R}^n) = \mathbb{R}^n \tag{3.26}$$

$$\text{aff}(\mathbb{R}^n) = \mathbb{R}^n \tag{3.27}$$

$$\text{rint}([0, 1]^2) = (0, 1)^2 \tag{3.28}$$

$$\text{rint}([0, 1] \times \{0\}) = (0, 1) \times \{0\} \tag{3.29}$$

Exercise 3.16. *We know*

$$int A \cap B \subseteq \int A \cap \int B \tag{3.30}$$

Does

$$\text{rint}(A \cap B) \subseteq \text{rint}\, A \cap \text{rint}\, B \tag{3.31}$$

Proposition 3.17. *Let $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ be convex. Then*

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

(ii) $g(x) = f(\lambda x) \Rightarrow \partial g(x) = \lambda \partial f(x)$

(iii) $g(x) = \lambda f(x) \Rightarrow \partial g(x) = \lambda \partial f(x)$

(ii) $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ *proper, convex,* $A \in \mathbb{R}^m \times n$ *such that*

$$\{Ay | y \in \mathbb{R}^m\} \cap \text{rint}\, \text{dom}\, f \neq \emptyset \tag{3.32}$$

Then for $x \in \operatorname{dom}(f \circ A)$ *we have*

$$\partial(f \circ A)(x) = A^T \partial f(Ax) \tag{3.33}$$

(iii) *Let* $f_1, \ldots f_m : \mathbb{R}^n \to \bar{\mathbb{R}}$ *be proper, convex, and* $\operatorname{rint}\operatorname{dom} f_1 \cap \cdots \cap \operatorname{rint}\operatorname{dom} f_m \neq \emptyset$*. Then*

$$\partial f(x) = \partial f_1(x) + \cdots + \partial f_m(x) \tag{3.34}$$

4

Conjugate Functions

4.1 The Legendre-Fenchel Transform

Definition 4.1. For $f : \mathbb{R}^n \to \bar{\mathbb{R}}$, define

$$\operatorname{con} f(x) = \sup_{g \le f, g \text{ convex}} g(x) \tag{4.1}$$

is the convex hull of f.

Proposition 4.2. *$\operatorname{con} f$ is the greatest convex function majorized by f.*

Definition 4.3. For $f : \mathbb{R}^n \to \bar{\mathbb{R}}$, the (lower) closure of f is defined as

$$\operatorname{cl} f(x) = \liminf_{y \to x} f(y) \tag{4.2}$$

Proposition 4.4. *For $f : \mathbb{R}^n \to \bar{\mathbb{R}}$,*

$$\operatorname{epi}(\operatorname{cl} f) = \operatorname{cl}(\operatorname{epi} f) \tag{4.3}$$

In particular, if f is convex, then $\operatorname{cl} f$ is convex.

Proof. Exercise. □

Proposition 4.5. *If $f : \mathbb{R}^n \to \bar{\mathbb{R}}$, then*

$$(\operatorname{cl} f)(x) = \sup_{g \le f, g \text{ lsc}} g(x) \tag{4.4}$$

Proof. □ Fill in

Theorem 4.6. *Let $C \subseteq \mathbb{R}^n$ be closed and convex. Then*

$$C = \bigcap_{(b,\beta) s.t. c \subseteq H_{b,\beta}} H_{b,\beta} \tag{4.5}$$

where

$$H_{b,\beta} = \{x \in \mathbb{R}^n | \langle x, b\rangle - \beta \leq 0\} \tag{4.6}$$

Proof. □

Theorem 4.7. *Let $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ be proper, lsc, and convex. Then*

$$f(x) = \sup_{g \leq f, g \text{ affine}} g(x) \tag{4.7}$$

Proof. □

This is quite an involved proof.

Definition 4.8. For $f : \mathbb{R}^n \to \bar{\mathbb{R}}$, let $f^\star : \mathbb{R}^n \to \bar{\mathbb{R}}$ be defined by

$$f^\star(v) = \sup_{x \in \mathbb{R}^n} \{\langle v, x\rangle - f(x)\} \tag{4.8}$$

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

Remark 4.9. *For $v \in \mathbb{R}^n$,*

$$f^\star(v) = \sup\{x \in \mathbb{R}^n\{\langle v, x\rangle - f(x)\} \tag{4.9}$$

$$\Rightarrow f^\star(v) \geq \langle v, x\rangle - f(x) \forall x \in \mathbb{R}^n \iff f(x) \geq \langle v, x\rangle - f^\star(x) \forall x \in \mathbb{R}^n. \tag{4.10}$$

Thus $f^\star$ is the largest affine function with gradient v majorized by f.

Theorem 4.10. *Assume $f : \mathbb{R}^n \to \bar{\mathbb{R}}$. Then*

(i) $f^\star = (\operatorname{con} f)^\star = (\operatorname{cl} f)^\star = (\operatorname{cl}\operatorname{con} f)^\star$ *and* $f \geq f^{\star\star} = (f^\star)^\star$*, the* ***biconjugate*** *of f.*

(ii) If $\operatorname{con} f$ *is proper, then $f^\star, f^{\star\star}$ as proper, lower semicontinuous, convex and* $f^{\star\star} = \operatorname{cl}\operatorname{con} f$.

(iii) If $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is proper, lower semicontinuous, and proper, then

$$f^{\star\star} = f \tag{4.11}$$

Proof. We have

$$(v, \beta) \in \operatorname{epi} f^\star \iff \beta \geq \langle v, x\rangle - f(x) \forall x \in \mathbb{R}^n \iff f(x) \geq \langle v, x\rangle - \beta \forall x \in \mathbb{R}^n \tag{4.12}$$

We claim that for an affine function h, we have $h \leq f \iff h \leq conf \iff h \leq \operatorname{cl} f \iff h \leq \operatorname{cl} \operatorname{con} f$. This is shown as $\operatorname{con} f$ is the largest convex function less than or equal to f, and h is convex. Same for cl, cl con, etc.

Thus in (4.12) we can replace f by $\operatorname{con} f$, $\operatorname{cl} f$, $\operatorname{cl} \operatorname{con} f$, which gives our required result.

We also have

$$\begin{aligned}
f^{\star\star}(y) &= \sup_{v \in \mathbb{R}^n} \{\langle v, y\rangle - \sup_{x \in \mathbb{R}^n} \{\langle v, x\rangle - f(x)\}\} && (4.13)\\
&\leq \sup_{v \in \mathbb{R}^n} \{\langle v, y\rangle - \langle v, y\rangle + f(y)\} && (4.14)\\
&= f(y) && (4.15)
\end{aligned}$$

For the second part, we have $\operatorname{con} f$ is proper, and claim that $\operatorname{cl} \operatorname{con} f$ is proper, lower semicontinuous, and convex.

Lower semicontinuity is a give. Convexity is given by the previous proposition that f is convex implies $\operatorname{cl} f$ is convex. Properness is to be shown in an exercise.

Applying the previous theorem,

$$\begin{aligned}
\operatorname{cl} \operatorname{con} f(x) &= \sup_{g \in \operatorname{cl} \operatorname{con} f, g \text{ affine}} g(x) && (4.16)\\
&= \sup_{(v,\beta) \in \operatorname{epi} f^\star} \{\langle v, x\rangle - \beta\} && (4.17)\\
&= \sup_{(v,\beta) \in \operatorname{epi} f^\star} \{\langle v, x\rangle - f^\star(v)\} && (4.18)\\
&= \sup_{v \in \operatorname{dom} f^\star} \{\langle v, x\rangle - f^\star(v)\} && (4.19)\\
&= \sup_{v \in \mathbb{R}^n} \{\langle v, x\rangle - f^\star(v)\} - f^{\star\star}(x) && (4.20)
\end{aligned}$$

with $g(x) \leq \langle v, x\rangle - \beta$, $(v, \beta) \in \operatorname{epi}(\operatorname{cl} \operatorname{con} f)^\star \iff (v, \beta) \in \operatorname{epi} f^\star$.

To show $f^\star$ is proper, lower semicontinuous, and convex, we have $\operatorname{epi} f^\star$ is the intersection of closed convex sets, and therefore closed and convex, and hence $f^\star$ is lower semicontinuous and convex.

To show properness, we have $\operatorname{con} f$ is proper implies there exists $x \in \mathbb{R}^n$ with $\operatorname{con} f(x) < \infty$. Then $f^\star(v) = \sup_{x \in \mathbb{R}^n}\{\langle v, x\rangle - f(x)\}$, which is greater than $-\infty$.

If $f^\star \equiv +\infty$, then $\operatorname{cl}\operatorname{con} f = f^{\star\star} = \sup_v \langle v, x - f^\star(x) \equiv -\infty$, and so $\operatorname{cl}\operatorname{con} f$ is proper, which implies $f^\star$ is proper, lower semicontinuous, and convex. Applying to $f^\star$ - we need $\operatorname{con} f^\star$ proper (which is proper by previous result),, and thus $f^{\star\star}$ is proper, lower semicontinuous, and convex.

For part 3, apply 2 - f is convex, which implies $\operatorname{con} f = f$ and $\operatorname{con} f$ is proper (as f is proper), and f is lsc and convex, and thus $f^{\star\star} = \operatorname{cl}\operatorname{con} f = f$. □

4.2 *Duality Correspondences*

Theorem 4.11. *Let $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ be proper, lower semicontinuous, and convex. Then:*

(i) $\partial f^\star = (\partial f)^{-1}$

(ii) $v \in \partial f(x) \iff f(x) + f^\star(v) = \langle v, x\rangle \iff x \in \partial(f^\star)(v)$.

(iii)

$$\partial f(x) = \arg\max_{v'}\{\langle v', x\rangle - f^\star(v')\} \tag{4.21}$$

$$\partial f^\star(x) = \arg\max_{x'}\{\langle v, x'\rangle - f(x')\} \tag{4.22}$$

$$\tag{4.23}$$

Proof. (i) This is obvious from (2)

(ii)

$$
\begin{aligned}
f(x) + f^\star(v) = \langle v, x\rangle \qquad\qquad & \qquad\qquad && (4.24)\\
\iff & \{f^\star(v) = \langle v, x\rangle - f(x) && (4.25)\\
\iff & x \in \arg\max\{\langle v, x'\rangle - f(x')\} && (4.26)\\
\ldots & && (4.27)
\end{aligned}
$$

□

Finish off proof

Proposition 4.12. *Let $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ be proper, lower semicontinuous, and convex. Then*

$$(f(\cdot) - \langle a, \cdot \rangle)^\star = f^{\star(\cdot + a)} \tag{4.28}$$

$$(f(\cdot + b))^\star = f^\star(\cdot) - \langle \cdot, b \rangle \tag{4.29}$$

$$(f(\cdot) + c)^\star = f^\star(\cdot) - c \tag{4.30}$$

$$(\lambda f(\cdot))^\star = \lambda f^\star(\frac{\cdot}{\lambda}), \lambda > 0 \tag{4.31}$$

$$(\lambda f(\frac{\cdot}{\lambda}))^\star = \lambda f^\star(\cdot) \tag{4.32}$$

Proof. Exercise □

Proposition 4.13. *$f_i : \mathbb{R}^n \to \bar{\mathbb{R}}$, $i : 1, \ldots, m$ proper, $f(x_1, \ldots, x_m) = \sum_i f_i(x_i)$. Then $f^\star(v_1, \ldots, v_m) = \sum_i f^\star(v_i)$*

Definition 4.14. For any set $S \subseteq \mathbb{R}^n$, define the support function

$$G_S(v) = \sup_{x \in S} \langle v, x \rangle = (\delta_S)^\star(v) \tag{4.33}$$

Definition 4.15. A function $h : \mathbb{R}^n \to \bar{\mathbb{R}}$ is **positively homogeneous** if $0 \in \operatorname{dom} h$ and $h(\lambda x) = \lambda h(x)$ for all $\lambda > 0, x \in \mathbb{R}^n$.

Proposition 4.16. *The set of positive homogeneous proper lower semicontinuous convex functions and the set of closed convex nonempty sets are in one-to-one correspondence through the Legendre-Fenchel transform.*

$$\delta_C \leftrightarrow G_C \tag{4.34}$$

and

$$x \in \partial G_C(v) \iff x \in C \tag{4.35}$$

and

$$G_C(v) = \langle v, x \rangle \iff v \in N_C(x) = \partial \delta_C(x) \tag{4.36}$$

The set of closed convex cones is in one-to-one correspondence with itself:

$$\delta_K \leftrightarrow \delta_{K^\star} \tag{4.37}$$

$$K^\star = \{v \in \mathbb{R}^n | \langle v, x \rangle \leq 0 \forall x \in K\} \tag{4.38}$$

and

$$x \in N_{K^\star}(v) \iff x \in K, v \in K^\star, \tag{4.39}$$

$$\langle x, v \rangle = 0 \iff v \in N_K(x) \tag{4.40}$$

5

Duality in Optimization

Definition 5.1. For $f : \mathbb{R}^n \times \mathbb{R}^m \to \bar{\mathbb{R}}$ proper, lower semicontinuous, convex, we define the primal and dual problems

$$\inf_{x \in \mathbb{R}^n} \phi(x), \phi(x) = f(x, 0) \tag{5.1}$$

$$\sup_{y \in \mathbb{R}^m} \psi(y), \psi(y) = f^\star(0, y) \tag{5.2}$$

and the inf-**projections**

$$p(u) = \inf_{x \in \mathbb{R}^n} f(x, u) \tag{5.3}$$

$$q(v) = \inf_{y \in \mathbb{R}^m} f^\star(v, y) \tag{5.4}$$

f is the perturbation function for ϕ, p is the associated projection function.

Consider the problem

$$\inf_x \frac{1}{2}\|x - z\|^2 + \delta_{\geq 0}(Ax - b) = \inf \phi(x) \tag{5.5}$$

Consider the perturbed problem

$$f(x, u) = \frac{1}{2}\|x - z\|^2 + \delta_{\geq 0}(Ax - b + u) \tag{5.6}$$

Proposition 5.2. *Assume f satisfying the assumptions in Definition 5.1. Then*

(i) $\phi, -psi$ *are convex and lower semicontinuous*

(ii) p, q *are convex*

(iii) $p(0) = \inf_x \phi(x)$

(iv) $p^{\star\star}(0) = \sup_y \psi(y)$

(v) $\inf_x \phi(x) < \infty \iff 0 \in \operatorname{dom} p$

(vi) $\sup \psi(y) > -\infty \iff 0 \in \operatorname{dom} q$

Proof. (i) ϕ is clearly convex. For ψ, $f^\star$ is lower semicontinuous and convex, which implies $-\psi$ is lower semicontinuous and convex.

(ii) Look at the strict epigraph of p:

$$E = \{(u, \alpha) \in \mathbb{R}^m \times \mathbb{R} | p(u) < \alpha\} \tag{5.7}$$

$$= \{(u, \alpha) \in \mathbb{R}^m \times \mathbb{R} | \exists x : f(x, u) < \alpha\} \tag{5.8}$$

$$= A\{(u, \alpha, x) \in \mathbb{R}^m \times \mathbb{R} \times \mathbb{R}^n | f(x, u) \leq \alpha\} \tag{5.9}$$

$$= A(u, \alpha, x) \mapsto (u, \alpha) \tag{5.10}$$

as a linear map over a convex set. As E is convex, p must be convex. Similarly with q.

(iii) For $p(0)$, this proceeds by definition. For $p^{\star\star}(0)$,

$$p^\star(y) = \sup_u \{\langle y, u \rangle - p(u)\} \tag{5.11}$$

$$= \sup_{u,x} \langle y, u \rangle - f(x, u) \tag{5.12}$$

$$= f^\star(0, y) \tag{5.13}$$

and

$$p^{\star\star}(0) = \sup_y \langle 0, y \rangle - p^\star(y) \tag{5.14}$$

$$= \sup_y -f^\star(0, y) \tag{5.15}$$

$$= \sup_y \psi(y) \tag{5.16}$$

(iv) By definition, $0 \in \operatorname{dom} p \iff p(0) = \inf_x f(x) < \infty$.

□

complete proof

Theorem 5.3. *Let f as in Definition 7.1. Then weak duality holds*

$$p(0) = \inf_x \phi(x) \geq \sup_y \psi(y) = p^{\star\star}(0) \tag{5.17}$$

and under certain conditions the inf, sup *are equal and finite (strong duality).*

$p(0) \in \mathbb{R}$ *and p lower-semicontinuous at 0 if and only if* $\inf \phi(x) = \sup \psi(y) \in \mathbb{R}$.

Definition 5.4.

$$\inf \phi - \sup \psi \tag{5.18}$$

is the **duality gap**

Proof. ($\Leftarrow$) $p^{\star\star} \leq \operatorname{cl} p \leq p \Rightarrow \operatorname{cl} p(0) = p(0)$.

($\Rightarrow$) $\operatorname{cl} p$ is lower semicontinuous, convex $\Rightarrow$ $\operatorname{cl} p(x)$ is proper, $\sup \psi = (p^\star)^\star(0) = (\operatorname{cl} p)^{\star\star}(0) = \operatorname{cl} p(0) = p(0) = \inf \phi$ $\square$

Proposition 5.5.

Fill in notes from lecture

b

6

First-order Methods

Idea is that we do gradient descent on our objective function f, with step size τ_k.

Definition 6.1. Let $f : \mathbb{R}^n \to \bar{\mathbb{R}}$, then

(i)
$$F_{\tau_k f}(x^k) = x^k - \tau_k \partial f(x^k) = (I - \tau_k \partial f)(x^k) \tag{6.1}$$

(ii)
$$B_{\tau_k f}(x^k) = (I + \tau_k \partial f)^{-1}(x^k) \tag{6.2}$$

so

$$\tag{6.3}$$

Fill in

Proposition 6.2. *If $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is proper, lower semicontinuous, and convex, for $\tau > 0$,*

$$B_{\tau_k f}(x) = \arg\min_y \{\frac{1}{2\tau_k}\|y - x\|_2^2 + f(y)\} \tag{6.4}$$

Proof.

$$y \in \arg\min\{...\} \iff 0 \in \frac{1}{\tau_k}(y - x)\partial f(y) \tag{6.5}$$

$$\tag{6.6}$$

□

Missed lecture material

7
Interior Point Methods

Theorem 7.1 (Problem)**.** *Consider the problem*

$$\inf_x \langle c, x\rangle + \delta_K(Ax - b) \tag{7.1}$$

with K a closed convex cone - thus $Ax \geq_K b$.

Notation.

The idea is to replace δ_K by a smooth approximation F, with $F(x) \to \infty$ as $x \to$ boundary K, and solve $\inf \langle c, x\rangle + \frac{1}{t}F(x)$, equivalent to solving $t\langle c, x\rangle + F(x)$.

Proposition 7.2. *F is canonical barrier for K. Then F is smooth on $\operatorname{dom} F = \int K$ and strictly convex, with*

$$F(tx) = F(x) - \Theta_F \lambda t, \tag{7.2}$$

and

(i) $-\nabla F(x) \in \operatorname{dom} F$

(ii) $\langle \nabla F(x), x\rangle = -\Theta_F$

(iii) $-\nabla F(-\nabla F(x)) = x$

(iv) $-\nabla F(tx) = -\frac{1}{t}\nabla F(x)$.

For the problem

$$\inf_x \langle c, x\rangle \tag{7.3}$$

such that $Ax \geq_K b$ with K closed, convex, self-dual cone, we have

$$\sup_x \inf_x \langle c, x \rangle + \delta_K(Ax - b) \tag{7.4}$$

$$\Rightarrow -\langle b, y \rangle - h^\star(-A^T y - c) - k^\star(y) \tag{7.5}$$

$$\iff \sup_y \langle b, y \rangle s.t. y \geq_K 0, A^T y = c \tag{7.6}$$

$$\iff \inf_x \langle c, x \rangle + F(ax - b) \tag{7.7}$$

$$\iff \sup_y \langle b, y \rangle - F(y) - \delta_{A^T y = c} \tag{7.8}$$

More conditions on, etc

Proposition 7.3. *For $t > 0$, define*

$$x(t) = \arg\min\{t\langle c, x \rangle + F(Ax - b)\} \tag{7.9}$$

$$y(t) = \arg\min\{-t\langle b, y \rangle + F(y) + \delta_{A^T y = c}\} \tag{7.10}$$

$$z(t) = (x(t), y(t)) \tag{7.11}$$

These paths exist, are unique, and

$$(x, y) = (x(t), y(t)) \iff \begin{cases} A^T y = c \\ ty + \nabla F(Ax - b) = 0 \end{cases} \tag{7.12}$$

Proof. The first part follows by definition.

For (7.12), the primal optimality condition is that $0 = tc + A^{T \nabla F(Ax-b)}$. The dual optimality condition is that $0 \in -tb + \nabla F(y) + N_{z | A^T z = c}(y)$, which is

$$\iff 0 \in -tb + \nabla F(y) + \begin{cases} \emptyset A^T y \neq c \\ \text{range } A \qquad A^T y = c \end{cases} \tag{7.13}$$

$$\iff 0 \in -tb + \nabla F(y) + \text{range } A, A^T y = c \tag{7.14}$$

Assume x satisfies the primal optimality condition. Then

$$tc + A^T \nabla F(Ax - b) = 0 \iff tc + A^T(-ty) = 0 \tag{7.15}$$

$$\iff A^T y = c \tag{7.16}$$

$\square$

A few more steps...

Proposition 7.4. *If x, y feasible, then*

$$\phi(x) - \psi(y) = \langle y, Ax - b \rangle \tag{7.17}$$

Moreover, for $(x(t), y(t))$ on the central path,

$$\phi(x(t)) - \psi(y(t)) = \frac{\Theta_F}{t} \tag{7.18}$$

Proof.

$$\begin{aligned}
\phi(x) - \psi(y) &= \langle c, x \rangle - \langle b, y \rangle && (7.19)\\
&= \left\langle A^T y, x \right\rangle - \langle b, y \rangle && (7.20)\\
&= \langle Ax - b, y \rangle && (7.21)
\end{aligned}$$

For the second part, we have

$$\begin{aligned}
\phi(x(t)) - \psi(y(t)) &= \langle y(t), Ax(t) - b \rangle && (7.22)\\
&= \left\langle -\frac{1}{t} \nabla F(Ax(t) - b), Ax(t) - b \right\rangle && (7.23)\\
&= \frac{\Theta_F}{t} && (7.24)
\end{aligned}$$

$\square$

Fill in from lecture notes

8

Support Vector Machines

8.1 Machine Learning

Given $X = \{x^1, \dots, x^n\}$ a sample set, $x^i \in \mathcal{F} \subseteq \mathbb{R}^m$, with $\mathcal{F}$ our feature space, and C a set of classes. We seek to find

$$h_\theta : \mathcal{F} \to C, \theta \in \Theta \tag{8.1}$$

with Θ our parameter space.

Our task is to find θ such that h_θ is the "best" mapping from X into C.

(i) Unsupervised learning (only X is known, usually $|C|$ not known)

- Clustering
- Outlier detection
- Mapping to lower dimensional subspace

(ii) Supervised learning ($|C|$ known). We have training date $T = \{(x^1, y^1), \dots, (x^n, y^n)\}$, with $y^i \in C$.

We seek to find

$$\theta = \arg\min_{\theta \in \Theta} f(\theta, T) = \sum_{i=1}^{n} g(y^i, h_\theta(x^i)) + R(\theta) \tag{8.2}$$

8.2 Linear Classifiers

The idea is to consider

$$h_\theta : \mathbb{R}^m \to \{-1, 1\}, \theta = \begin{pmatrix} w \\ b \end{pmatrix} \tag{8.3}$$

$$h_\theta(x) = \operatorname{sign}(\langle w, x \rangle + b) \tag{8.4}$$

We want to consider maximum margin classifiers, satisfying

$$\max_{w,b,w\neq 0} \min_i y^i (\left\langle \frac{w}{\|w\|}, x \right\rangle + \frac{b}{\|w\|}) \tag{8.5}$$

which can be rewritten as

$$\max_{w,b} c \tag{8.6}$$

such that

$$c \leq y^i \left(\left\langle \frac{w}{\|w\|}, x^i \right\rangle + \frac{b}{\|w\|} \right) \tag{8.7}$$

$$\|w\| c \leq y^i \left(\left\langle w, x^i \right\rangle + b \right) \tag{8.8}$$

$$\tag{8.9}$$

or just

$$\min_{w,b} \frac{1}{2} \|w\|^2 \tag{8.10}$$

such that

$$1 \leq y^i (\langle w, x \rangle + b) \tag{8.11}$$

Definition 8.1. In standard form,

$$\inf_{w,b} k(w, b) + h(M \begin{pmatrix} w \\ b \end{pmatrix} - e) \tag{8.12}$$

The conjugates are

$$k(w,b) = \frac{1}{2}\|w\|^2 \tag{8.13}$$

$$k^\star(u,c) = \frac{1}{2}\|u\|^2 + \delta_{\{0\}}(c) \tag{8.14}$$

$$h(z) = \delta_{\geq 0}(z) h^\star(v) = \delta_{\leq v}(v) \tag{8.15}$$

In saddle point form,

$$\inf_{w,b}\sup_{z} \frac{1}{2}\|w\|_2^2 + \left\langle M\begin{pmatrix} w \\ b \end{pmatrix} - e, z \right\rangle - \delta_{\leq 0}(z) \tag{8.16}$$

The dual problem is

$$\sup_{z} -\langle e,z\rangle - \delta_{\leq 0}(z) - \frac{1}{2}\| - \sum_{i=1}^{n} y^i x^i z_i\|_2^2 - \delta_{\{0\}}(\langle y,z\rangle) \tag{8.17}$$

and thus

$$\inf_{z} \frac{1}{2}\| \sum_{i} y^i x^i z_i\|_2^2 + \langle e,z\rangle \tag{8.18}$$

such that $z \leq 0, \langle y,z\rangle = 0$.

The optimality conditions are

$$\tag{8.19}$$

Fill in rest of optimality conditions

We use the fact that if $k(x) + h(Ax+b)$ is our primal, then the dual is $-\langle b,y\rangle - k^\star(-A^T y) - h^\star(z)$.

Explanation of support vectors

8.3 Kernel Trick

The idea is to embed our features into a a higher dimensional space, mapping function ϕ. Then our decision function takes the form

$$h_\theta(x) = \text{sign}(\langle \phi(x), w\rangle + b) \tag{8.20}$$

9

Total Variation

9.1 Meyer's G-norm

Idea - regulariser that favors **textured** regions

$$\|u\|_G = \{\inf \|v\|_\infty | \div v = u, v \in L^\infty(\mathbb{R}^d, \mathbb{R}^d)\} \tag{9.1}$$

Discretized, we have

$$\|u\|_G = \inf\{\delta^{\star C}(v) + \delta_{-G^T v = u}\} \tag{9.2}$$

$$C = \{v | \sum_x \|v(x)\|_2 \leq 1\} \tag{9.3}$$

$$= \inf v \sup_v \sup_w \{\delta_C^\star(v) - \left\langle w, G^T v + u \right\rangle\} \tag{9.4}$$

$$= \sup_w - \sup_v \{\langle v, -Gw \rangle + \delta_C^\star(v) - \langle w, u \rangle\} \tag{9.5}$$

$$= \sup_w \{\delta_C(-Gw) + \langle w, u \rangle\} \tag{9.6}$$

$$= \sup_w \{\langle w, u \rangle - \delta_C(-Gw)\} \tag{9.7}$$

$$= \sup\{\langle w, u \rangle | TV(w) \leq 1\} \tag{9.8}$$

and so $\|\cdot\|_G$ is the **dual** to TV.

$$\|\cdot\|_G = \delta_{B_{TV}}^\star \tag{9.9}$$

where $B_{TV} = \{u | TV(u) \leq 1\}$.

Similarly,

$$\sup\{\langle u, w\rangle | \|w\|_G \leq 1\} \tag{9.10}$$

$$= \sup\{\langle u, w\rangle | \exists v : w = -G^T v, \|v\|_\infty \leq 1\} \tag{9.11}$$

$$= \sup\{\left\langle u, -G^{Tv}\right\rangle | \|v\|_\infty \leq 1\} \tag{9.12}$$

$$= TV(u) \tag{9.13}$$

Why is $\| \cdot \|$ good in separating noise?

$$\arg\min \frac{1}{2}\|u - g\|_2^2 s.t. \|u\|_G \leq \lambda \tag{9.14}$$

$$= \prod_{\lambda B_{\|\cdot\|_G}}(g) = B_{\lambda B_{\|\cdot\|_G}}(g) = g - B_{\lambda B_{\|\cdot\|_G}} = g - B_{\lambda TV}(g) \tag{9.15}$$

9.2 *Non-local Regularization*

In real-world images, large $\| \nabla u\|$ does not always mean noise.

Definition 9.1. $\Omega = \{1, \dots, n\}$, given $u \in \mathbb{R}^n, x, y \in \Omega$, $w \in \Omega^2 \to R_{\geq 0}$, then

$$\partial_y u(x) = (u(y) - u(x))w(x, y) \tag{9.16}$$

$$\nabla_w u(x = (\partial_y u(x)))_{y \in \Omega} \tag{9.17}$$

A suitable divergence $\div_w u(x) = \sum_{y \in \Omega}(w(x, y) - v(y, x))w(x, y)$ adjoint to ∇_w with respect to Euclidean inner products.

$$\langle - \div_w v, u\rangle = \left\langle v, \nabla_w u\right\rangle \tag{9.18}$$

Non-local regularizers are

$$J(u) = \sum_{x \in \Omega} g(\nabla_w u(x)) \tag{9.19}$$

with

$$TV^g_{NL}(u) = \sum_{x\in\Omega} \| \underset{w}{\nabla} u(x) \|_2 \tag{9.20}$$

$$TV^d_{NL}(u) = \sum_{x\in\Omega} \sum_{y\in\Omega} |\partial_y u(x) \tag{9.21}$$

This reduces to classical TV if the weights are chosen as constant ($w(x,y) = \frac{1}{h}$).

9.2.1 How to choose $w(x,y)$?

(i) Large if neighborhoods of x, y are similar with respect to a distance metric.

(ii) Sparse, otherwise we have n^2 terms in the regularized (n is number of pixels).

Possible choice

$$d_u(x,y) = \int_\Omega K_G(t)(u(y+t) - u(x+t))^2 dt \tag{9.22}$$

$$A(x) = \arg\min_A \{ \sum_{y\in A} d_u(x,y) | A \subseteq S(x), |A| = k \} \tag{9.23}$$

$$w(x,y) = \begin{cases} 1 & y \in A(x), x \in A(y) \\ 0 & \text{otherwise} \end{cases} \tag{9.24}$$

with K_G a Gaussian kernel of variance G^2.

10

Relaxation

Example 10.1 (Segmentation). *Find $C \subseteq \Omega$ that fits the given data and prior knowledge about typical shapes.*

Theorem 10.2 (Chan-Vese). *Given $g : \Omega \to \mathbb{R}, \Omega \subseteq \mathbb{R}^d$, consider*

$$\inf_{C \subseteq \mathbb{R}, c_1, c_2 \in \mathbb{R}} f_{CV}(C, c_1, c_2) \tag{10.1}$$

$$f_{CV}(C, c_1, c_2) = \int_C (g - c_1)^2 dx + \int_{\Omega \setminus C} (g - c_2)^2 dx + \lambda \mathcal{H}^{d-1}(\partial C) \tag{10.2}$$

.

Confirm form of the regu-
lariser

Thus fit c_1 to shape, c_2 to outside of shape, and C is the region of the shape.

10.1 Mumford-Shah model

$$\inf_{K \subseteq \Omega \text{ closed}, u \in C^\infty(\Omega \setminus K)} f(K, u) \tag{10.3}$$

$$f(K, u) = \int_\Omega (g - u)^2 dx + \lambda \int_{\Omega \setminus K} \| \nabla u \|_2^2 dx + \nu \mathcal{H}^{d-1}(\partial K). \tag{10.4}$$

Chan-Vese is a special case of this, with forcing $u = c_1 \mathbb{I}(C) + c_2(1 - \mathbb{I}(C))$.

$$\inf_{C\subseteq\Omega, c_1, c_2\in\mathbb{R}} \int_C (g-c_1)^2 dx + \int_{\Omega\setminus C} (g-c_2)^2 dx + \lambda \mathcal{H}^{d-1}(\partial C) \tag{10.5}$$

$$\inf_{u\in BV(D,\{0,1\}), c_1, c_2\in\mathbb{R}} \int_\Omega u(g-c_1)^2 + (1-u)(g-c_2)^2 dx + \lambda TV(u) \tag{10.6}$$

Fix c_1, c_2. Then

$$\inf_{u\in BV(\Omega,\{0,1\})} \int_\Omega u \underbrace{\left((g-c_1)^2 - (g-c_2)^2\right)}_{S(\cdot)} dx + \lambda TV(u) \tag{C}$$

Replacing $\{0,1\}$ by it's convex hull, we obtain

$$\inf_{u\in BV(\Omega,[0,1])} \int_\Omega u \cdot S dx + \lambda TV(u) \tag{R}$$

This is a "convex relaxation" -

(i) Replace non-convex energy by convex approximation

(ii) Replace non-convex f by con f.

Can the minima of (R) have values $\notin \{0,1\}$. Assume $u_1, u_2 \in BV(\Omega, \{0,1\})$ that u_2 minimized (R) and (C), and $u_1 \neq u_2$. Then

$$\frac{1}{2}u_1 + \frac{1}{2}u_2 \notin BV(\Omega, \{0,1\}) \tag{10.7}$$

is a minimizer of (R).

Proposition 10.3. *Let c_1, c_2 be fixed. If u minimizes (R) and $u \in BV(\Omega, \{0,1\})$, then u minimizes (C) and ($u = 1_C$) C minimizes $f_{CV}(\cdot, c_1, c_2)$.*

Proof. Let $C' \subseteq \Omega$. Then $1_{C'} \in BV(\Omega, [0,1])$. Then

$$f(1_{C'}) \geq f(1_C) \tag{10.8}$$

$\forall C' \subseteq \Omega$. □

Definition 10.4. $\mathcal{L} = BV(\Omega, [0,1])$. For $u \in \mathcal{L}$, $\alpha \in [0,1]$,

$$\overline{u}_\alpha(x) = \mathbb{I}(\{u > \alpha\})(x) = \begin{cases} 1 & u(x) > \alpha \\ 0 & u(x) \leq \alpha \end{cases} \tag{10.9}$$

Then $f : \mathcal{L} \to \mathbb{R}$ satisfies general coarea condition (GCC) if and only if

$$f(u) = \int_0^1 f(\overline{u}_\alpha) d\alpha \tag{10.10}$$

Upper bound?

Proposition 10.5. *Let $s \in L^\infty(\Omega)$, Ω bounded. Then f as in (R) satisfies* GCC.

Proof. If f_1, f_2 satisfy GCC, then $f_1 + f_2$ satisfy GCC. This follows as λTV satisfies GCC by the coarea formula for total variation.

Then we have

$$\int_\Omega u(x)S(x)dx = \int_\Omega (\int_0^1 \mathbb{I}(\{u(x) > \alpha\})\, S(x)d\alpha)dx \tag{10.11}$$

$$= \int_0^1 \int_\Omega \mathbb{I}(\{u(x) > \alpha\})\, dxd\alpha. \tag{10.12}$$

□

Theorem 10.6. *Assume $f : BV(\Omega, [0,1]) \to \mathbb{R}$ satisfies* GCC *and*

$$u^\star \in \operatorname*{arg\,min}_{u \in BV(\Omega,[0,1])} f(u). \tag{10.13}$$

Then for almost any $\alpha \in [0,1]$, $\overline{u}^\star_\alpha$ is a minimizer of f over $BV(\Omega, \{0,1\})$, with

$$\overline{u}^\star_\alpha \in \operatorname*{arg\,min}_{u \in BV(\Omega,\{0,1\})} f(u). \tag{10.14}$$

Proof.

$$S = \{\alpha | f(\overline{u}^\star_\alpha) \neq f(u^\star)\} \tag{10.15}$$

If $\mathcal{L}(S) = \emptyset$, then for a.e. α,

$$f(u^\star_a) = f(u^\star) = \inf_{u \in BV(\Omega,[0,1])} f(u) \tag{10.16}$$

If $\mathcal{L}(S) > 0$, then there exists $\epsilon > 0$ such that

$$L(S_\epsilon) > 0, S_\epsilon = \{\alpha | \overline{u}^\star_\alpha \geq f(u^\star) + \epsilon\} \tag{10.17}$$

which implies

$$f(u^\star) = \int_{[0,1]\setminus S_\epsilon} f(u^\star) d\alpha + \int_{S_\epsilon} f(u^\star) d\alpha \tag{10.18}$$

$$\leq \int_0^1 f(\overline{u}_\alpha^\star) d\alpha - \epsilon L(S_\epsilon) \tag{10.19}$$

$$< \int_0^1 f(\overline{u}_\alpha^\star) d\alpha \tag{10.20}$$

which contradicts GCC. □

Remark 10.7. *Discretization problem, consider*

$$\min_{u \in \mathbb{R}^n} f(u), f(u) = \langle s, u \rangle + \lambda \sum_i \|Gu\|_2 \tag{10.21}$$

such that $0 \leq u \leq 1$ *vs*

$$\min_{u \in \mathbb{R}^n} f(u), f(u) = \langle s, u \rangle + \lambda \sum_i \|Gu\|_2. \tag{10.22}$$

such that $u \in \{0, 1\}^n$.

If $u^\star$ *solves the relaxation, does* $\overline{u}_\alpha^\star$ *solve the combinatorial problem?*

Only if the ***discretized energy*** f *satisfies* GCC.

11

Bibliography