Lecture 3 Lect III

(1). Recap. We saw some important convex functions and some ways of constructing them.

(2) Thm: Let $h(x,y)$ be convex in x, for every $y \in \mathcal{Y}$. Then,

$$f(x) := \sup_{y \in \mathcal{Y}} h(x,y) \quad \text{is convex.}$$

(3) Special case: Fenchel-Conjugate:

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

Observe f^* is convex even if f is not.

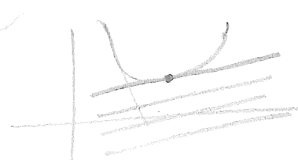
(4) Basic theorem: Let f be a closed convex function (i.e. a function whose epi f is a closed set, i.e. a lower semi cont. cvx fn)

then $f^{**} = f$;

(5). Axiomatic characterization of the Legendre-Fenchel transform

— Why study f^*: Intimately related to duality

— Ultimately grounded in the simple idea.



f being convex is globally underestimated by linear functions. (affine act.)

The tightest affine underestimator is

$$A(x) := \inf_y f(y) - \langle x, y \rangle$$

i.e. at a point x the value of tightest linear underestimator.

A encodes epi f via its supporting hyperplanes.

~~Theorem: (Artstein-Avidan, Milman).~~ Let $\Gamma_0(\mathbb{R}^n)$ denote the class of closed cvx fns on $\mathbb{R}^n$. The Legendre-Fenchel transform of a fn. $f \in \Gamma_0$, is defined by

$$\mathcal{L} : f \longmapsto \sup_y \langle \cdot, y \rangle - f(y)$$

i.e. $$(\mathcal{L}f)(x) = \sup_y \langle x, y \rangle - f(y)$$

Theorem: Let T be any transform that maps $\Gamma_0 \to \Gamma_0$ and satisfies

(1). $T(Tf) = f$ (involution)

(2) $f \leq g \Rightarrow Tf \geq Tg$ (order-reversing)

Then T must be "essentially" the Legendre-Fenchel transform.

That is, $\exists\, C \in \mathbb{R}, v \in \mathbb{R}^n, B \in GL_n$ s.t.

$$(Tf)(x) = (\mathcal{L}f)(Bx + v) + \langle z, v \rangle + C$$

Remarks: Gives axiomatic support to great success of LF-transform.

Playful: Let $\mathcal{S}$ be a distinguished class of your favorite functions on $\mathbb{R}^n$.

- we call T a duality transform on $\mathcal{S}$ if
  - for any $f \in \mathcal{S}$, $T(Tf) = f$
  - for any two fns $f, g \in \mathcal{S}$, st $f \leq g$, $Tf \geq Tg$.

Eg. Say $\mathcal{S}$ is class of log-concave fns

then $$(Tf)(x) := \inf_{y \in \mathbb{R}^n} \frac{e^{-\langle x, y \rangle}}{f(y)}$$ is such a transform.

(In fact, essentially unique).

\# Try your own class!! Eg try $(\mathbb{R}_{++}, |\log x - \log y|)$

(3)

Concept of subdifferentials, intimately related to ~~convex~~ Fenchel-conjugates

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be any function.

Define: $\partial f(x) := \{ g \in \mathbb{R}^n \mid f(y) \geq f(x) + \langle g, y-x \rangle \ \forall y \in \mathbb{R}^n \}$

Notice: if $x \notin \mathrm{dom} f$, then $\partial f(x) = \phi$.

Theorem: Let f be a proper (i.e. $\not\equiv +\infty$) convex function.

Then $g \in \partial f(x)$ iff

$$f^*(g) + f(x) = \langle g, x \rangle$$

Proof: $g \in \partial f(x) \iff f(y) \geq f(x) + \langle g, y-x \rangle$

$\iff$ ~~$f(x) - \langle g, x \rangle \geq$~~

$\langle g, y \rangle - f(y) \leq \langle g, x \rangle - f(x) \quad \forall y \in \mathrm{dom} f$

i.e.

$$f^*(g) = \sup_y \langle g, y \rangle - f(y) = \leq \langle g, x \rangle - f(x).$$

But in view of def of

$$f^*(g) \leq \langle g, x \rangle - f(x)$$

But by defn of f^* this is an equality.

Some basic theorems you should be aware of.

~~e.g. Let $f \in \Gamma_0(\mathbb{R}^n)$~~; (a) $\partial f(x) \neq \phi$ whenever $x \in \mathrm{ri}(\mathrm{dom} f)$

~~(b). Let f be a proper cvx fn.~~

(b) If $f \in \Gamma_0(\mathbb{R}^n)$, then

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

Thm: (Rockafeller) Let $f_1, \dots, f_m$ be proper convex functions on $\mathbb{R}^n$ and let $f = f_1 + \cdots + f_m$

Then

$$\partial f_1 + \cdots + \partial f_m \subset \partial f(x), \ \forall x$$

If the convex sets $\bigcap_{i=1}^m ri(\operatorname{dom} f_i) \neq \phi$,

then, actually

$$\partial (f_1 + \cdots + f_m) = \partial f_1 + \cdots + \partial f_m \quad \forall x$$

& Moreover, if say $f_1, \dots, f_k$ are polyhedral, then it is enough that

$$\left[\bigcap_{i=1}^{k} \operatorname{dom} f_i\right] \cap \left(\bigcap_{i=k+1}^{m} ri(\operatorname{dom} f_i)\right) \neq \phi.$$

Proof: '$\subset$' part is easy:

~~Assume $\bigcap_{i=1}^m ri(\operatorname{dom} f_i) \neq \phi$, we have~~

~~$$(f_1 + \cdots + f_m)^*(g) = \inf \{ f_1^*(g_1) + \cdots + f_m^*(g_m) \mid g_1 + \cdots + g_m = g \}$$~~

~~where for each g the inf is attained.~~

~~(I'll leave the proof in the notes)~~

say ① $g = g_1 + \cdots + g_m$ where $g_i \in \partial f_i(x)$.

(So $g \in \partial f(x)$)

② Then, $\forall z$, $f(z) = f_1(z) + \cdots + f_m(z) \geq \sum_i f_i(x) + \langle g_i, z - x \rangle$

$= f(x) + \langle g_1 + \cdots + g_m, z - x \rangle = f(x) + \langle g, z - x \rangle$

Example: Define f_1, f_2 by

$$f_1(x) := \begin{cases} -2\sqrt{x} & \text{if } x \geq 0 \\ +\infty & \text{if } x < 0 \end{cases}$$

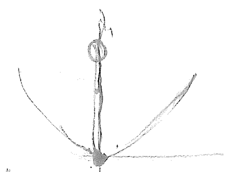
$$f_2(x) := \begin{cases} +\infty & \text{if } x > 0 \\ -2\sqrt{-x} & \text{if } x \leq 0 \end{cases}$$

$\sqrt{x}$ is concave on $[0, \infty)$ so $-2\sqrt{x}$ is cvx; $\sqrt{|x|}$ cve $-\sqrt{\cdot}$.

$$f \equiv f_1(x) + f_2(x) \equiv \begin{cases} 0, & x = 0 \\ +\infty, & \text{otherwise} \end{cases} \equiv \delta_{\{0\}}(x)$$

so $\partial f(0) \equiv \mathbb{R}$

But $\partial f_1(0) = \phi$
$\partial f_2(0) = \phi$ (Exercise: $\sqrt{x}$ is non diff at 0, in fact $\to \infty$)



---o---

Subgradient calculus

- Reading assignment: (Link posted on class website !)

. Determine $\partial f(x)$: Usually very hard

- Usually we need just one $g_x \in \partial f$.
- However testing optimality req. knowledge of ∂f.

Thm: Let $f(x) := \sup\{ f_i(x) \mid i \in I\} < +\infty$
$\forall x \in \mathbb{R}^n$.

I is an arbitrary index set.

Let $I(x) := \{ i \in I \mid f_i(x) = f(x) \}$ active set

then
$$\partial f(x) = \operatorname{conv} \bigcup_{i \in I(x)} \partial f_i(x)$$

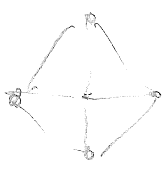
Question: what happens when $f(x)$ can be $\pm\infty$ for some values ?

Example: ~~$f(x)$~~ - what is $\partial \|x\|_\infty$?

clearly : ~~$\|x\|_\infty = \sup_u \{ \langle u, x \rangle \mid \|u\|_1 \leq 1 \}$~~

$$\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$$
$$= \max_{1 \leq i \leq n} |e_i^T x| \quad ,$$

$$\Rightarrow \partial \|\cdot\|(0) = \operatorname{conv}\{\pm e_1, \dots, \pm e_n\}$$



For other examples and details please refer to notes

IMPT: (If $g \in \partial f(x) \Rightarrow x \in \operatorname{dom} f \Rightarrow g \in \operatorname{dom} f^*$

But $\operatorname{dom} f^*$ can be larger than $\partial f(\mathbb{R}^n)$!)

Optimization Problems: (7)

(1) $\min_{x \in \mathbb{R}^n} f(x)$ ✱.

- say f is differentiable we know that

$\nabla f(x^*) = 0$ is nec. for x^* to be a local min

- If $f \in C^2$, $\nabla^2 f(x^*) \succ 0$ suff for local min.

- ~~Nonconvex world~~

- For cvx fns $\nabla f(x^*) = 0$ also sufficient.

- Nonconvex world: No known "easy" necess + suff conditions. for optimality

(2) ~~$\min_{x \in \mathbb{R}^n} f(x)$ s.t. $x \in \mathcal{X}$~~.

~~write as unconstrained problem.~~

Theorem: (Fermat's law): Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. Then

$$\operatorname{Argmin} f = \operatorname{Zer}(\partial f) := \{x \in \mathbb{R}^n \mid 0 \in \partial f(x)\}.$$

Proof: $x \in \operatorname{Argmin} f \Rightarrow f(x) \le f(y) \ \forall y \in \mathbb{R}^n$.

So $f(y) \ge f(x) + \langle 0, y-x \rangle \ \forall y \Rightarrow 0 \in \partial f(x)$.

(Converse is obvious!)

—o—

Example: $\min f(x)$ st $x \in \mathcal{X}$

$$\iff \min_{x \in \mathbb{R}^n} f(x) + \delta_{\mathcal{X}}(x)$$

So $x^* \in \operatorname{Argmin}$ iff

$$0 \in \partial(f + \delta_{\mathcal{X}})(x^*)$$

Recall Rockafellar's theorem: If $\operatorname{ri}(\operatorname{dom} f) \cap \operatorname{ri}(\operatorname{dom} \delta_{\mathcal{X}}) \neq \emptyset$

then $\partial(f + \delta_{\mathcal{X}}) = \partial f + \partial\delta_{\mathcal{X}}$

(So assuming the above "constraint qualification")

$$0 \in \partial f(x^*) + \partial \delta_{\mathcal{X}}(x^*)$$

What is $\partial \delta_{\mathcal{X}}$?

$$g \in \partial \delta_{\mathcal{X}} \iff \delta_{\mathcal{X}}(y) \geq \delta_{\mathcal{X}}(x) + \langle g, y-x \rangle \quad \forall y$$

$x \in \mathcal{X}$ so $\delta_{\mathcal{X}}(x) = 0$

$\implies 0 \geq \langle g, y-x \rangle \quad \forall y \in \mathcal{X}$ must hold!

Def: (Normal cone)

$$\mathcal{N}_{\mathcal{X}}(x) := \{ g \in \mathbb{R}^n \mid 0 \geq \langle g, y-x \rangle, y \in \mathcal{X} \}$$

$$\implies \boxed{0 \in \partial f(x^*) + \mathcal{N}_{\mathcal{X}}(x^*)}$$

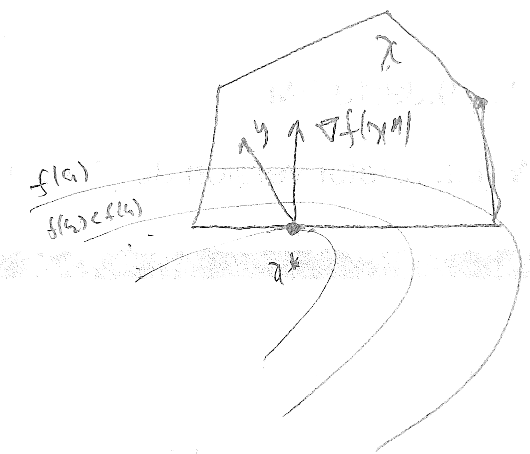
If f is differentiable, $\partial f(x^*) = \nabla f(x^*)$

So above condition translates into:

$$\langle \nabla f(x^*), y - x^* \rangle \geq 0 \quad \forall y \in \mathcal{X}$$

(So-called Kolmogorov condition on optimality)

Picture.



$\nabla f(x^*)$ defines supporting hyperplane at x^*

Example: Projection Lemma.

$$\min \frac{1}{2}\|x - x_0\|^2 \quad \text{s.t. } x \in \mathcal{X}$$

iff. $$\langle x^* - x_0, y - x^* \rangle \geq 0 \quad \forall y \in \mathcal{X}$$

(Classical result)

Question: What about optimality conditions in general convex, geodesic metric spaces $(\mathcal{X}, d)$

HW: Show that if f is g.c. fn on g.c. set $\mathcal{X}$ then local optimality $\Longrightarrow$ global optimality

Specific Examples of CP+ problems

① Linear program: $\min\ c^T x$, $Ax = b$, $x \geq 0$ — Standard form

WX!

② Quadratic: $\min\ x^T A x + b^T x$ s.t. $Ax \leq b$

● SOCP, and SDPs:

Ex: Write: $\min\ f(x) = \left(\max\limits_{1 \leq i \leq m} a_i^T x + b_i\right)$ as an LP

Q: Does QP always have a solution?

Semidefinite Programming:

$$\min_{x \in \mathbb{R}^n}\ c^T x$$

$$\text{s.t. } \mathcal{A}(x) := A_0 + x_1 A_1 + \cdots + x_n A_n \succeq 0$$

$A_0, \ldots, A_n \in S^{n \times n}$.

Exercise: SDP $\supset$ SOCP $\supset$ QP $\supset$ LP

Equivalent / Std. form.

$$\min\ \langle C, X \rangle$$
$$\langle A_i, X \rangle = b_i$$
$$X \succeq 0$$

Using duality one converts between the two. (to see later)

—

Where does the idea come from?

LP: $\min c^T x$, $Ax = b$, $x \geq 0$ | Super successful (WWII, Dantzig, Smale, CSR, etc. etc.)

How to generalize: $c^T x$: $\Rightarrow$ replace by nonlinear fn.

$Ax = b$ replace A by a nonlinear map? || use $x_i = 1$ $\Downarrow$ $x_i^2 = 1$!!!

Both quickly become non-tractable.

\# Generalize structural constraint. $\mathbb{R}^n_+$

$\mathbb{R}^n_+$ is a cone (polyhedral cone)

Δ Replace by general convex cone!

Δ Importantly: replace $x \geq 0$ by $x \succeq_K 0$

$\star$ Nesterov-Nemirovskii — 80's Conic Opt. theory

Seeking to gen. $x \geq 0$

$$K := \{x \in \mathbb{R}^n \mid x \geq 0\}$$

$$x \geq y \Leftrightarrow x - y \geq 0 \Leftrightarrow x - y \in K$$

Lemma: (Necc. & suff. for a set $K \subset \mathbb{R}^n$ to define a useful vector ineq. is that K should be a nonempty pointed cone: i.e.

- $K \neq \phi$
- $x, y \in K \Rightarrow x + y \in K$
- $\alpha K \subset K$,
- $x, -x \in K \Rightarrow x = 0$

Two other impt properties

Recall for $\mathbb{R}_n^+$: Closed with nonempty interior

($\{x_k\} \subset \mathbb{R}_+^n \to x \Rightarrow x \in \mathbb{R}_+^n$)
(int $\mathbb{R}_+^n$: contains Ball with +ve radius)

Eg. Conic problem: min $c^T x$
$Ax = b$
$x \succeq_K 0$

LP: $K = \mathbb{R}_+^n$;
SOCP: $K = Q^n := \{(x,t) \in \mathbb{R}^n \mid \|x\|_2 \le t\}$
SDP: $K = S_+^n := \{X = X^T \succeq 0,\ X \in S^n\}$
or cartesian products of these

Other cones could be used, but these three are "easiest" → class project could investigate some other norm cone.

(Not all cones are nice: Recall CP_n)

SDP (cone program).
min $c^T x$, $Ax = b$, $x \in K$

Think of x as a matrix
" so $c^T x = \mathrm{trace}(C^* X) = C(:)' X(:)$ matlab notation

wlog assume $K = S_+^n$

$Ax = b \Leftrightarrow a_i^T x = b_i \equiv \langle A_i, X\rangle = b_i$
$x \in K \equiv X \succeq 0$

min $\langle C, X\rangle$
$\langle A_i, X\rangle = b_i$
$X \succeq 0$

HW: try solving a few SDPs in CVX

(~~Don't forget~~.

Important Topic: Which convex (and nonconvex) optimization problems can be formulated as SDPs?

↳ Related to major open problems in convex geometry (We'll come to that later in the course)

Ex: $\min x^T A x$ s.t. $x_i^2 = 1$ — Notice $A \succeq 0$. But constraint is nonlinear!

$x^T A x = \operatorname{tr}(A x x^T)$

$$= \left\{ \begin{array}{l} \min \langle A, X \rangle \\ \operatorname{Diag}(X) = e \\ X \succeq 0, \quad \boxed{\cancel{\operatorname{rank}(X) = 1}} \end{array} \right.$$

$\Rightarrow \cancel{X = x x^T}$

$\boxed{X \succeq x x^T}$ convex!

$$X \succeq x x^T \iff \begin{bmatrix} X & x \\ x^T & 1 \end{bmatrix} \succeq 0$$