- post syllabus on gsthub

Cambridge Short Course, June 2018

Lecture 1: Convex analysis

- outline lectures - rooms ...
- ask for email comments.

▷ INTRO

1/18

covered

What's our goal? $\underset{x \in \mathbb{R}^n}{\text{minimize}}\ f(x)$ st $\underbrace{x \in C}_{\text{constraint}}$

$f(x)$ ← objective

$f: \mathbb{R}^n \to \mathbb{R}$

- We assume $n > 1$. Why? If $n = 1$, and has some basic smoothness properties, and we can bound domain of x, then just do grid search.
  Eg, $x \in [0,1]$, grid spacing h, evaluate f at $\frac{1}{h}$ point.
  But $x \in [0,1]^n$, spacing h, need $(\frac{1}{h})^n$ grid pts: "curse of dimensionality"
- Assume range is $\mathbb{R}$, not $\mathbb{R}^m$; otherwise vague
  ie, if output is work hours / get paid: $\begin{bmatrix} 10 \text{ hrs} \\ \$100 \end{bmatrix}$ vs $\begin{bmatrix} 20 \text{ hrs} \\ \$250 \end{bmatrix}$ which is better?
- Avoid integers

later for examples, p. 12

Ex Markowitz Portfolio Opt.

Demonstrates tricks. $x \in \mathbb{R}^n$ represents how much to invest in each of n investments. $x_i > 0$ = long position, $x_i < 0$ = short position.

We estimate a mean return rate $\mu \in \mathbb{R}^n$, μ_i = expected return (after, say, 1yr) on asset i.

(Naive: put all our \$ on one asset).

Also estimate variances $\hat{\Sigma}$.

For a portfolio $x \in \mathbb{R}^n$, expected return is $\mu^T x$ (maximize this)

expected variance is $x^T \hat{\Sigma} x$ (min. this)

"risk averse"

How to encode?

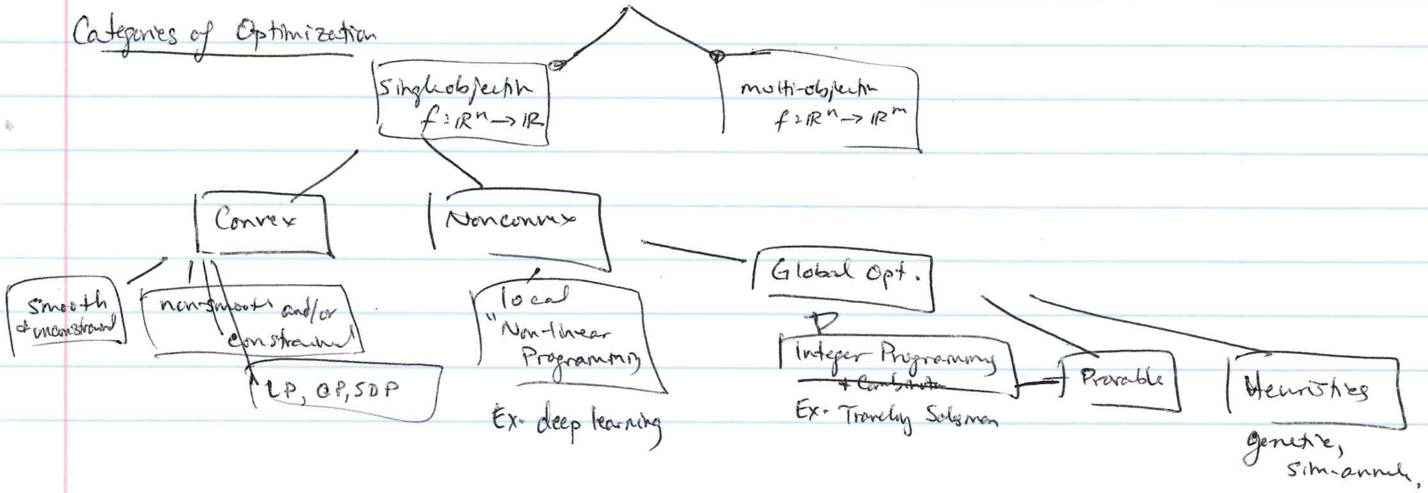
$$\min_{x \in \mathbb{R}^n} x^T \hat{\Sigma} x - \alpha \cdot \mu^T x \quad \text{s.t.} \quad \mathbb{1}^T x = 1$$

min/max trick

α is a tradeoff parameter.

Capital constraints.

Categories of Optimization



skipped

▷ BACKGROUND

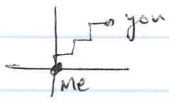
- Work in vector spaces, always $\mathbb{R}^n$: can add, scale vectors.
- To measure size, use a norm:

  Def A norm $\|\cdot\|$ acts on $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}_+$ s.t.

  ① $\|x\| = 0$ iff $x = 0$, ② $\forall \alpha \in \mathbb{R}$, $\|\alpha x\| = |\alpha| \cdot \|x\|$,

  ③ $\|x+y\| \le \|x\| + \|y\|$.

Hence, a possible distance between ~~x~~ x and y (a metric) is $\|x-y\|$



Ex: Eucl. distance or "ℓ^2":

$$\|x-y\|_2 = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$ "crow flies"

Ex: $$\|x-y\|_1 = \sum_{i=1}^{n} |x_i - y_i|$$ "Taxicab/Manhattan"

Ex: Mahalanobois

$$\|x-y\|_A = \langle x-y, A(x-y) \rangle \text{ if } A \succ 0$$

- To multiply vectors, inner-product "dot-product": $\langle x, y \rangle = x^T y$ (Euclidean)

- $A \succ 0$ means A is pos-def., i.e., $\forall x \neq 0$, $\langle x, Ax \rangle > 0$

  i.e., and (typically) $A = A^*$ ($A^* = \bar{A}^T$)

  in which case,

  ($A = VDV^T$ eigenvalues) $A \succ 0$ iff $\forall i$, $D_{ii} > 0$. (eigenvals.)

- Matrix norms:
  - $\|A\|_2 = \sup_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}$. If $A = U\Sigma V^T$, $\Sigma = \mathrm{diag}(\sigma)$, $\|A\|_2 = \|\sigma\|_\infty$.

    "operator norm":

    it's sub-multiplicative, $\|Ax\|_2 \le \|A\|_2 \cdot \|x\|_2$ — different means

  - $\|A\|_F = \|\mathrm{vec}(A)\|_2$ ⚠ MATLAB: norm(A) vs norm(A(:))

    $= \|\sigma\|_2$. Also sub-multiplicative, since $\|A\|_2 \le \|A\|_F$.

  - (In finite dim.),

    all norms are equiv: $\exists\, c, C > 0$ st.

    $\forall x \in \mathbb{R}^n$, $c\|x\|_a \le \|x\|_b \le C \cdot \|x\|_a$.

American-style, not $]0,1]$

- sup vs max, inf vs min : min of $[0,1]$ is 0 (smallest element)
  inf of $(0,1]$ is 0 (largest lower bound)

- Sequences $(x_k)_{k=1,2,\dots}$: $x_k \to x$ in $\mathbb{R}^n$ if $\lim_{k\to\infty} \|x_k - x\| = 0$ for any norm.
  i.e., $\forall \varepsilon > 0$, $\exists N$ s.t. $k > N \Rightarrow \|x_k - x\| < \varepsilon$.

Q: $x_k \to 0 \Rightarrow \sum_{k=1}^{\infty} x_k < \infty$?
A: no, eg., $x_k = \frac{1}{k}$.

[MISC] → Cauchy ..., Complete, ..., Banach, Hilbert (don't go into details)

- Open set: $\mathcal{U}$ is open if $\forall x \in \mathcal{U}$, I can draw a ball around x and stay inside $\mathcal{U}$.
- Closed set: contains limit pts.
- Compact: K is a compact set if $\forall (x_k) \subseteq K$, $\exists x \in K$ s.t. $x_k \to x$.

↑ skipped
↓ covered

Thm: Bolzano-Weierstrass / Heine-Borel
in Euclidean Space $\mathbb{R}^n$, a set is compact iff it is closed and bounded.

$x_n \to x$ then $f(x_n) \to f(x)$.

- $f: \mathbb{R}^n \to \mathbb{R}$ is cts at x, $\forall \varepsilon > 0$, $\exists \delta > 0$ s.t. $\|y - x\| < \delta \Rightarrow |f(y) - f(x)| < \varepsilon$.
  not cts.

- $f: \mathbb{R}^n \to \mathbb{R}$ is lsc at x if $\forall (x_k)$ st $x_k \to x$, $f(x) \le \liminf f(x_n)$
  is lsc, not lsc.

$\boxed{\text{epi}(f) \text{ is closed}}$

- an operator $T: \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz cts, w/ constant L, if $\forall x, y \in \mathbb{R}^n$, $\|Tx - Ty\| \le L \cdot \|x - y\|$
  (depends on choice of norm)

- If K is compact and $f: K \to \mathbb{R}$ is cts, $f(K)$ is bounded and attains min, max
  ... is lsc, $f(K)$ bdd from below and attains its inf.

$f(x) = \frac{1}{x}$ on $K = \underbrace{(0, \infty)}_{\text{not compact}}$ } doesn't attain inf.

▷ DIFFERENTIATION IN $\mathbb{R}^n$, $n>1$

- Differentiability of $f:\mathbb{R}^n\to\mathbb{R}$ (for simplicity, think of $\mathbb{R}^2\to\mathbb{R}$

  1) Partial derivs. exist: $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$

     e.g., $f(x,y)=(xy)^{1/3}$, $\frac{\partial f}{\partial x}=\frac{1}{3}x^{-2/3}y^{1/3}\ \ldots$

     ⇑

     but along line $y=x$, define $g(x)=f(x,x)$. Is g differentiable? No.

     $g(x)=x^{2/3}$, $g'(x)=\frac{2}{3}x^{-1/3}$ so $g'(0)$ undefined!

C requires $f'(x;\mathbf{0})$ $\in B(d,K)$

$\boxed{\text{Gateaux v2}}$

  2) Directional derivs exist (in all directions): GATEAUX DIFF. — not agreed on!

     means $\forall$ directions $d\in\mathbb{R}^n$, $\underbrace{\lim_{h\to0}\frac{f(x+h\cdot d)-f(x)}{h}}_{=D(x;d)}$ exists.
     $f'(x;d)$

     ⇑

     (2')

  3) Directional deriv is a ~~etc~~ linear fcn in d, and a bit extra,

     $D(x;d)=\langle \underbrace{\nabla f(x)}_{\text{gradient}},d\rangle$ "FRÉCHET DIFF."

     meaning $\lim_{\|d\|\to0}\frac{\|f(x+d)-f(x)-\langle\nabla f(x),d\rangle\|}{\|h\|}=0$

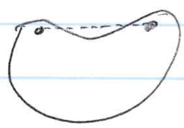
In 1D, ~~all~~ concepts (1), (2), (3) equivalent

* i.e., ~~Fréchet requires~~ the rate-of-convergence to be uniform. = FRÉCHET.

▷ CONVEXITY

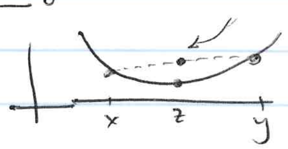
- A set $C\subseteq\mathbb{R}^n$ is convex if $\forall x,y\in C$ and $\forall t\in[0,1]$,

  $t\cdot x+(1-t)\cdot y\in C$. i.e.,



- A function $f:\mathbb{R}^n\to f$ is convex if $\forall x,y$, $\forall t\in[0,1]$,

  $f(\underbrace{tx+(1-t)y}_{z})\le t\cdot f(x)+(1-t)f(y)$



  , i.e., $\underbrace{\text{epi}(f)}$ is a convex set
  $=\{(x,s)\in\mathbb{R}^{n+1}: f(x)\le s\}$

- We're going to allow $f:\mathbb{R}^n\to\overline{\mathbb{R}}:=\mathbb{R}\cup\{\pm\infty\}$

and strictly cvx if $\forall x\ne y$, $t\in(0,1)$, have "<".

and concave if $-f$ is convex, and affine if both concave & convex.

Skipped

Why? simplifies things: $\min\limits_{x \in C} \overset{g(x)}{\cancel{\langle d, x \rangle}}$ $\Big\}$ $\min f(x) := \begin{cases} \overset{g(x)}{\cancel{\langle d, x \rangle}} & x \in C \\ +\infty & x \notin C \end{cases}$

Use $\delta_C(x) = \begin{cases} 0 & x \in C \\ +\infty & x \notin C \end{cases}$ as indicator fcn, NOT $\begin{cases} 1 \\ 0 \end{cases}$

Define: $\text{dom}(f) = \{x : f(x) < \infty\}$ (also called "effective domain")

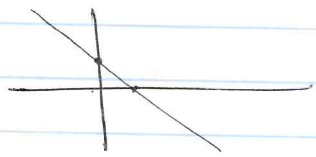
▷ Q: Differentiability of convex functions

First, domains → want to exclude boundaries.

Def Interior of a set $\text{int}(C) = \{x \in C : \exists \varepsilon > 0, B_\varepsilon(x) \subseteq C\}$

Problem: Suppose I have $\min\limits_{\binom{x}{y} \in \mathbb{R}^2} x^2 + y^2$ s.t. $x + y = 1$ (i.e., $y = 1 - x$

Define $f(x) = x^2 + y^2 + \delta_{\{x+y=1\}}(x, y)$.



Then $\text{int}(\text{dom}(f)) = \ldots \phi$.

But we think of "parent space" as $\mathbb{R}^2$. However, a subspace of a Hilbert Space is still Hilbert (& ignore affine issue...) so work in this.

It is a Hilbert Space

Def $\text{aff}(C)$ = affine hull of C := smallest affine space containing C

= intersection of all...

Similarly,

Def $\text{conv}(C)$ = cvx hull of C = smallest convex set containing C

so

Def $\text{relint}(C) = \{x \in C : \exists \varepsilon > 0, B_\varepsilon(x) \cap \text{aff}(C) \subseteq C\}$

i.e.,

is $\frac{1}{2}$ an interior pt. of $C = [0,1]$? yes.

is $(\frac{1}{2}, 0)$ " _____ " $C = [0,1] \times \{0\}$ No, but it is in the relative interior.

Thm (cf corollary 8.39 B+C 2nd ed). $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ proper + convex, then f is cts on $\text{int}(\text{dom}(f))$

but in fact, ...

essential) Thm. Alexandrov '39 f convex, then a.e. for $x \in \text{int dom } f$,

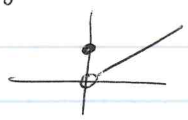
$\exists\ \nabla f(x),\ \nabla^2 f(x)$ s.t.

skipped

$$\forall d,\quad f(x+d) = f(x) + \langle \nabla f(x), d\rangle + \tfrac{1}{2}\langle Ad, d\rangle + o(\|d\|^2).$$

Stronger than C^2, see myth

Only on interior, eg.,



$\text{dom}(f) = \{x : x \geq 0\} \in \mathbb{R}$,
not cts even at $x=0$ (note: it is not lsc).

(DO THIS EARLIER)

▷ MORE CONVEX FCN:

covered.

If f is (Fréchet) differentiable, and convex, then ∇f is monotone

i.e., $\langle y-x, \nabla f(y) - \nabla f(x)\rangle \geq 0$ (and it's "iff" if dom(f) open & cvx)

(in 1D, this is saying $y > x \Rightarrow f'(y) > f'(x)$
$y < x \Rightarrow f'(y) < f'(x)$, i.e., f' is increasing)

If $f \in C^2$, so $\nabla^2 f(x)$ exists,

f is convex iff $\nabla^2 f(x)$ is pos. semi-def, meaning,

$\forall\ d,\quad \langle d, \nabla^2 f(x) d\rangle \geq 0.$

Written: $\nabla^2 f(x) \succeq 0$. (generalizes $f'' \geq 0$ in 1D)

have + examples

If $\exists\ \mu > 0$ st. $\nabla^2 f(x) \succeq \mu I$, f is strongly cvx

(more generally, if $f(x) - \mu/2 \|x\|^2$ is cvx).

If $\exists\ L$ st. $\nabla^2 f(x) \preceq L \cdot I$, then ∇f is L-Lipschitz cts.

Standard assumption: $0 \preceq \mu I \preceq \nabla^2 f(x) \preceq L I$, ie., bounded by quadratics.

Ex: $f(x) = \langle c, x\rangle$ is cvx, not strictly.
$f(x) = e^{-x}$ is strictly, not strongly, cvx.

mention briefly

• PRESERVING CONVEXITY $\qquad f, g$ cvx. $\qquad (f(\cdot, y)$ cvx $\forall y$, $w \geq 0)$

• non-neg. weighted sums; or integrals: $f(x) + 3 \cdot g(x)$, or $\int w(y) f(x,y)\,dy$

• $f = h \circ g$, $g: \mathbb{R}^n \to \mathbb{R}^k$, $h: \mathbb{R}^k \to \mathbb{R}$, $\boxed{h(+\infty) \text{ defeq } +\infty}$, must be increasing/decreasing

then $k=1$, cvx if: $\begin{cases} h \text{ convex nondecreasing}, & g \text{ convex} \\ h \text{ convex nonincreasing}, & g \text{ concave} \end{cases}$ $\qquad \left(k > 1,\ h \text{ nondecreasing in each argument}, g_i \text{ cvx} \quad \ldots \quad g(x) = \begin{bmatrix} g_1(x) \\ g_k(x) \end{bmatrix} \right)$

or, h convex, g affine

Use: $f'' = h''(g(x))(g'(x))^2 + h'(g(x)) \cdot g''(x)$

(Preserving crxty, ct'd) • max: $f(x) = \max(g(x), h(x))$ or even $\sup_{y \in A} f_y(x)$ (proof: intersect epigraph)

~~min~~ ($\min(|x|, |x-1|)$) Not for min, even w/ 2,

• f jointly crx in (x,y), $C \neq \emptyset$ is crx, then $\inf_{y \in C} f(x,y)$ is crx (as long as $> -\infty$)

Ex. $\mathrm{dist}(x, C)$

(still, differentiability of crx fcn)

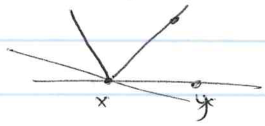
So if f is crx, and thus f is C^2 a.e., can we "pretend" it's differentiable everywhere? No. Sets of 0-measure often coincide w/ sol'n sets!

▷ SUBGRADIENTS

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be proper, then the subdifferential is

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

subgradient



May be empty.

Thm (Fermat's Rule) If f is proper, $\operatorname{argmin}_x f(x) = \{ x : 0 \in \partial f(x) \}$

Generalizes idea of $f'(x) = 0$, but more global.

Facts: f proper + convex, in $\mathbb{R}^n$, 1) $x \in$ relint (dom (f)) $\Rightarrow \partial f(x) \neq \emptyset$.

Prop 17.26 p.251 PLC ed 1 { 2) $x \in \mathrm{dom}(f)$ and differentiable at $x \Rightarrow \partial f(x) = \{\nabla f(x)\}$

3) $x \in$ relint (dom(f)) and f cts at x, then $\partial f(x) = \{u\}$

$\Rightarrow f$ is differentiable, $u = \nabla f(x)$.

$\nabla(f+g) = \nabla f + \nabla g$. True for subdiff?

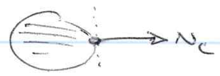
Not always: in $\mathbb{R}^n$, $f, g \in \Gamma_0(\mathbb{R}^n)$, relint (dom f) $\cap$ relint (dom g) $\neq \emptyset$,

or f or g has full domain, $\Rightarrow \partial(f+g) = \partial f + \partial g$.

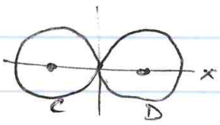
(cf. Cor 16.38 (iv) PLC ed 1)

"Constraint Qual.", ie., Slater

(Ex. when it fails: note $\partial(\delta_C) = N_C(x) = \begin{cases} \{u : \langle y-x, u \rangle \leq 0\ \forall y \in C\} & x \in C \\ \emptyset & x \notin C \end{cases}$ normal cone



= all vectors u defining a supporting hyperplane.



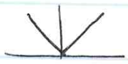
$f = \delta_C$, $g = \delta_D$, closed unit balls, int dom f $\cap$ int dom g $= \emptyset$

$\partial f(0) = N_C(0) = \mathbb{R}_+ \times \{0\}$

$\partial g = \mathbb{R}_- \times \{0\}$, $\partial f(0) + \partial g(0) = \mathbb{R} \times \{0\}$

But... $\partial(f+g)(0) = N_{C \cap D}(0) = N_{\{0\}}(0) = \mathbb{R}^2$ ← unequal.

FYI, tangent cone is polar cone to N_C, $\{u : \langle v, u \rangle \leq 0\ \forall v \in N_C\}$

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

but note directional deriv. is single-valued

(cf. Prop 17.2 PLC ed 2, i.e., f proper, cvx, $x \in \text{int}(\text{dom}(f))$ $\Rightarrow f'(x;d)$ exists, not $\pm\infty$).

▷ FIRST INEQUALITIES. let $f \in \Gamma_0(\mathbb{R}^n)$

lower bound: $\forall y$,

or any subgradient if not diff-able

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$$

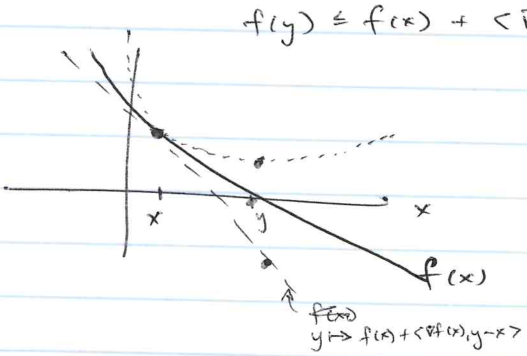
If f is μ-strongly cvx,

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \mu/2 \, \|x - y\|^2$$

upper-bnd. If ∇f is L-Lipschitz,

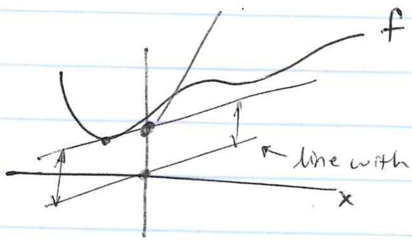
$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + L/2 \, \|x - y\|^2$$

"DESCENT LEMMA"



▷ FENCHEL-LEGENDRE CONJUGATE

$$f^*(y) = \sup_x \langle y, x \rangle - f(x)$$



Convex even if f isn't!

Ex: If $f(x) = \delta_C$,

$f^*(y) = \sup_{x \in C} \langle y, x \rangle$ is "support fcn"

i.e., $f(x) = \delta_C$ w/ $C = \{x : \|x\| \leq 1\}$,

then $f^*(y) = \sup_{\|x\| \leq 1} \langle y, x \rangle = \|y\|_*$ "dual norm"

($\|\cdot\| = \|\cdot\|_p$ then $\|\cdot\|_* = \|\cdot\|_q$, $\frac{1}{p} + \frac{1}{q} = 1$)

Ex: $f(x) = \frac{1}{2}\|x\|^2 = f^*(x)$

Ex: $f(x) = f(u,v) = g(u) + h(v)$ "separable"

then ~~$f^*(x,y) = g^*(x) + h^*(y)$~~

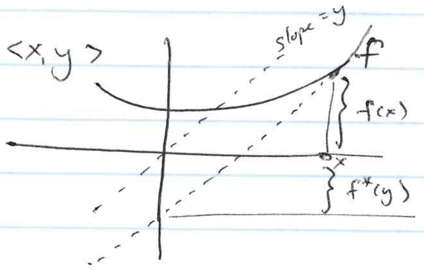
$f^*(u',v') = g^*(u') + h^*(v')$

Fenchel-Young Inequality: $\forall x, y,\ f(x) + f^*(y) \geq \langle x, y \rangle$

Proof by def'n of sup

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

or, $x \in \partial f^*(y)$



i.e. THM (Cor. 16.24 PLC 1st ed)

If $f \in \Gamma_0(\mathbb{R}^n)$, $(\partial f)^{-1} = \partial f^*$. (and $f, g \in \Gamma_0(\mathbb{R}^n)$, $f = g^*$ iff $g = f^*$, i.e., $f = f^{**}$)

▷ CONVEX RELAXATIONS.

(Prop. 13.14 PLC 1st ed) (no assumptions on f)

1) $f^{**} \leq f$

Proof: $f^{**}(x) = \sup_y \langle x, y \rangle - \underbrace{\left(\sup_z \langle y, z \rangle - f(z) \right)}_{f^*(y)}$

$= \sup_y \langle x, y \rangle + \inf_z -\langle y, z \rangle + f(z)$, choose $z = x$

$\leq \sup_y \cancel{\langle x, y \rangle} - \cancel{\langle y, x \rangle} + f(x) = f(x)$. ✓.

2) $f \leq g \Rightarrow f^* \geq g^*$ Proof: similar.

3) $f^{***} = f^*$

Proof: ◦ $f^{***} \leq f^*$ via (1) applied to f^*.

◦ also, $f^{**} \leq f$, so (2) $\Rightarrow f^{***} \geq f^*$. ◻.

Take-away:

$f^{**} \leq f$ always, $f^{**} = f$ if $f \in \Gamma_0(\mathbb{R}^n)$.

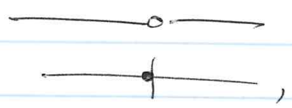
Prop (13.39 PLC 1st ed) f proper but not nec. convex, then

unless $f^* \equiv +\infty$ (so $f^{**} \equiv -\infty$), f^{**} is the lsc crx envelope of f,

i.e., $\mathrm{epi}(f^{**}) = \overline{\mathrm{conv}(\mathrm{epi}(f))}$

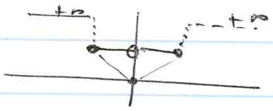
i.e., "crx relaxation" i.e., $f^{**} \leq f$, $f^{**} \in \Gamma_0(\mathbb{R}^n)$ and $\nexists\, g \in \Gamma_0(\mathbb{R}^n)$ w/ $g > f^{**}$ anywhere

Ex: $f(x) = \|x\|_0$,



, $f^{**} \equiv 0$.

Or, $f(x) = \|x\|_0 + \delta_{[-1,1]}(x)$,



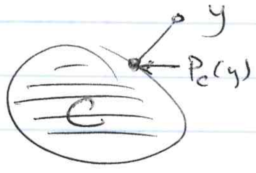
, geom. clear $f^{**}(x) = |x|$.

▷ PROXIMITY OPERATORS

Def for $f \in \Gamma_0(\mathbb{R}^n)$, $t > 0$, $\text{prox}_{t \cdot f}(y) = \underset{x}{\text{argmin}} \; t \cdot f(x) + \frac{1}{2}\|x-y\|^2$

Generalizes projection: if $f(x) = \delta_C$,

$$\text{prox}_{t \cdot f}(y) = \underset{x \in C}{\text{argmin}} \; \frac{1}{2}\|x-y\|^2 = P_C(y).$$



Ex $f(x) = \|x\|_1$, separable, do componentwise

$$\text{prox}_{t|\cdot|}(y) = \underset{x \in \mathbb{R}}{\text{argmin}} \; t|x| + \frac{1}{2}(x-y)^2$$

How to compute? Fermat's Rule: $0 \in t \cdot \partial|x| + (x-y)$, $x = \text{prox}$

i.e., $y \in (x + t \cdot \partial|x|)$

or more generally,

generally (set $t=1$ wlog)

Fermat's: if $x = \text{prox}_f(y)$,

$0 \in \partial f(x) + (x-y)$, i.e., $y \in (I + \partial f)(x)$

i.e., $\boxed{x = (I + \partial f)^{-1}(y)}$

So, $(t=1)$

$0 \in \partial|x| + (x-y)$.

Case $x = 0$, $\partial|x| = [-1,1]$, so $y \in [-1,1]$, ~~and the value is~~ $\cancel{\frac{1}{2}y^2}$

Case $x \neq 0$, $y > 0$

Pick $x > 0$ too, $\partial|x| = 1$, $x = y - 1$, so need $y \geq 1$,

~~value is~~ $\cancel{y - 1/2}$. $0 = -1 + x - y$

otherwise, $\partial|x| = -1$, $x = y + 1$, $y > 0$. contradiction.

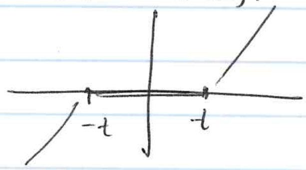
Case $x \neq 0$, $y < 0$

$x < 0$, $x = y + 1$, $y \leq -1$.

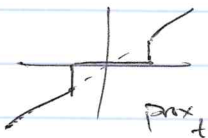
altogether, w/ $t > 0$,

$$\text{prox}_{t|\cdot|}(y) = \text{sign}(y) \cdot \max(0, |y| - t)$$

soft-thresholding.



vs hard-thresh.



$f, g \in \Gamma_0$

Thm ~~$\forall y$: $f(x)$~~ • MOREAU IDENTITY

$$x = \underbrace{\text{prox}_f(x)}_{x_1} + \underbrace{\text{prox}_{f^*}(x)}_{x_2}$$

and $f(x_1) + f^*(x_2) = \langle x_1, x_2 \rangle$

didn't cover

(Prox. Operators)

⚠ $\operatorname{prox}_{f+g}(x) \neq \operatorname{prox}_f(x) + \operatorname{prox}_g(x)$

$\operatorname{prox}_{f\circ g}(x)$ also not simple, with a few exceptions : $g: \mathbb{R}^n \to \mathbb{R}$, $f: \mathbb{R} \to \mathbb{R}$ see Combettes

Major Issue in Algorithms.

: $g(x) = Q\cdot x + b$, $Q\cdot Q^T = I$, i.e. if $Q^* = Q^{-1}$, $\operatorname{prox}_{f\circ Q}(x) = Q^* \operatorname{prox}_f(Qx)$

But... simpler formulas work, e.g., reflection, translation, adding linear/quadratic, ... See Combettes, Beck.

- MATRIX PROX.

Fact: (24.67 PLC 2nd ed, cf [235 Thm 4.6])

von Neumann Trace Ineq.: $A, B \in \mathbb{R}^{m\times n}$,

$$|\operatorname{trace}(A^T B)| \leq \sum_{i=1}^{\min(m,n)} \sigma_i(A)\,\sigma_i(B).$$

Prop (24.68)

Define $f: \mathbb{R}^{m\times n} \to \mathbb{R}$ by $f(A) = \sum_{i=1}^{\min(m,n)} \varphi(\sigma_i(A))$, a spectral fn

e.g. norms : spectral, nuclear, Frob

where ~~φ:R→R is~~ $\varphi \in \Gamma_0(\mathbb{R})$ is an even fcn.

Let $A = U\Sigma V^T$ be SVD of A, $\Sigma = \begin{bmatrix} \sigma_1(A) & & 0 \\ & \ddots & \\ 0 & & \sigma_{\min(m,n)}(A) \end{bmatrix}$

Then $\operatorname{Prox}_f(A) = U\tilde{\Sigma}V^T$, $\tilde{\Sigma} = \begin{bmatrix} \operatorname{prox}_\varphi(\sigma_1(A)) & & 0 \\ & \ddots & \\ 0 & & \operatorname{prox}_\varphi(\sigma_{\min(m,n)}(A)) \end{bmatrix}$.

▷ EXAMPLES OF OPTIMIZATION (high-dim., convex, maybe not so obvious)

Imaging

observed data

high-dim, convex.

$$\min_x \tfrac{1}{2}\|Ax-b\|^2 \quad \text{s.t. } x \geq 0$$

"non-neg. LS", simplest non-trivial opt. problem.

intensities of fluorophores

or, better loss fcn to approximate log-lik of Poisson:

$$\tfrac{1}{2}\|Ax-b\|^2_{\Sigma^2}, \quad \Sigma \text{ is a diagonal matrix.}$$

Difficulty:

- sz(X) = # pixels, about 1 to 10,000 to 100,000 $\boxed{\text{in 3D}}$
- Solve about 20,000 of these
- in seconds
- A is ill-conditioned.

High-d. Statistics

Lasso: $\min_x \tfrac{1}{2}\|Ax-b\|^2 + \lambda\|x\|_1$ for some $\lambda > 0$

ie., $\underbrace{\min_{\beta\in\mathbb{R}^p} \tfrac{1}{2}\|X\beta-y\|_2^2}_{\text{ill-posed, since either no sol'n or } \infty \text{ sol'n.}} + \lambda\|\beta\|_1$, y_i = magnitude of response ~~#~~ in person i, $i=1,\dots,n$

X_{ij} might be ~~g~~ expression/presence of gene j in person i

"regularizer"

Matrix Completion

Ex, Netflix data. Y_{ij} = rating person i gave to movie j

$$\min_{X\in\mathbb{R}^{n_1\times n_2}} \left(\tfrac{1}{2}\sum_{(i,j)\in\Omega} (X_{ij}-Y_{ij})^2 = \tfrac{1}{2}\|\mathcal{A}_\Omega(X-Y)\|_2^2 \right) + \lambda \cdot \|X\|_*$$

$\mathcal{A}_\Omega : \mathbb{R}^{n_1\times n_2} \to \mathbb{R}^{|\Omega|}$

trace norm, nuclear norm.