Agenda: prox... (last time)
- Convex Problems
- Existence, uniqueness of min
- Lagr. Duality
- LP
- SDP
- STRONG DUAL, SLATER
- OPTIMALITY, KKT.
- optim

Lecture 2: Convex Minimization Problems

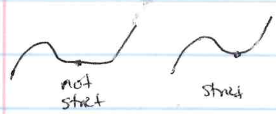
▷ Missing from last-time: joint convexity/continuity /linearity, Fenchel-Legendre conjugate, prox of ℓ_1 and $\|\cdot\|_*$, preserving convexity: non-neg. addition

▷ CONVEX OPT.

(P) $\min_{x \in C} f(x)$

1) "Feasible" means $x \in C$

2) "Global min." or "Solution" if ~~$x \in C$~~
x^* feasible and $f(x^*) \le f(x)$ $\forall x \in C$ ("nothing bigger")



3) "Local min" if $\exists\, \varepsilon > 0$ st. $\forall x \in B_\varepsilon(x^*) \cap C$, $f(x^*) \le f(x)$
(and x^* feasible)

4) "strict local min" if ... $f(x^*) < f(x)$ $\forall x \ne x^*$, $x \in B_\varepsilon(x^*) \cap C$

5) "isolated" ($\Rightarrow$ strict) if no other local min nearby.

6) (if $C = \mathbb{R}^n$), $\nabla f(x^*) = 0 \Rightarrow$ "critical" or "stationary" pt.
and "saddle-pt" if not a local min or max

Ex: $\min x^2(x-1)^2$ 2 global min 

Ex: $\min x$ no minimizer

Ex: $\min_{x \in (0,1)} x$ no minimizer: require C to be closed and non-empty

For global minimizers: suppose f is a convex fcn, C is a convex set,
then **Thm**: all local min. are actually global min.
(and set of all minimizers is cvx)

proof:
let $x \in C$ be a local min. If not global, $\exists\, y \in C$ st $f(y) < f(x)$.
Then $\forall t \in [0,1]$, $z = tx + (1-t)y \in C$ is feasible.
By local min, for t sufficiently close to 1, $f(z) \ge f(x)$.
Yet $f(z) \le t \cdot f(x) + (1-t) f(y) < f(x)$ by convexity of f. Contradiction.

To simplify, allow $f: \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$

$\boxed{f \in \Gamma_0(\mathbb{R}^n)}$ means f is
1) convex
2) proper (not always $+\infty$; no $-\infty$)
3) lsc

Ex: $f = \delta_C$, indicator of C
1) C a convex set
2) $C \ne \emptyset$
3) C is closed set.

▷ EXISTENCE, UNIQUENESS OF MIN

-11.8 p157 PLC 1st ed.

① convexity $\Rightarrow$ all local min are global, but do they exist? unique?
② $f \in \Gamma_0(\mathbb{R}^n)$ strictly cvx $\Rightarrow$ at most one minimizer
③ $f \in \Gamma_0(\mathbb{R}^n)$ coercive $\Rightarrow$ at least one minimizer

$f: \mathbb{R}^n \to [-\infty, \infty]$ is coercive if $\lim_{\|x\| \to \infty} f(x) = \infty$ (all sub-level sets bdd)

(if $f \in \Gamma_0(\mathbb{R}^n)$, it is coercive iff $\exists\, \alpha$ st. $\{x : f(x) \le \alpha\} \ne \phi$ and is bdd.)

▷ LAGRANGIAN DUALITY

$$(P)\quad \begin{aligned} &\min f_0(x) \\ &f_i(x) \le 0,\ i=1,\dots,m \\ &Ax = b \end{aligned}$$

Assuming convexity, so f_i, $i=0,1,\dots,m$ convex.

(spend time on this: f cvx $\Rightarrow$ sub-level sets cvx)

Lagrangian* $\mathcal{L}(x, \underbrace{\lambda, \nu}_{\text{Lagrange multiplies or Dual Variables}}) = f_0(x) + \sum_{i=1}^m \lambda_i \cdot f_i(x) + \langle \nu, Ax - b \rangle$

*can check to dualize, to keep explicit

and Dual Function is $g(\lambda, \nu) = \inf_x \mathcal{L}(x, \lambda, \nu)$, which is always concave (even if f_i weren't convex!)

let $p^* = \min_{f_i(x) \le 0,\ Ax=b.} f_0(x)$ Suppose $\lambda \ge 0$, ν arbitrary, $\tilde{x}$ primal feasible

$$\begin{aligned} \text{Then } g(\lambda,\nu) &= \inf_x \mathcal{L}(x,\lambda,\nu) \\ &\le \mathcal{L}(\tilde{x}, \lambda, \nu) = f_0(\tilde{x}) + \sum \underbrace{\lambda_i}_{\ge 0} \underbrace{f_i(\tilde{x})}_{\le 0} + \langle \nu, \underbrace{A\tilde{x} - b}_{=0} \rangle \\ &\le f_0(\tilde{x}) \end{aligned}$$

So

$$(D)\quad d^* = \max_{\substack{\lambda \ge 0 \\ \nu}} g(\lambda, \nu) \quad \text{"Dual Problem"}$$
} ALWAYS "CONVEX"

then "Weak Duality": $d^* \le p^*$.

Why solve dual?
- if P not convex, D is, so we get a lower bound. Upper bnd via feasible pt.
- exploit sparsity, structure: tradeoff smoothness for strong convexity, ...
- linear operators $\to$ bad for prox, good for gradient

▷ DUALITY EX: LINEAR PROGRAMMING

LP is $\min_{x\in\mathbb{R}^n} \langle c, x\rangle$ $\quad \mathcal{L}(x,\lambda,\nu) = \langle c,x\rangle - \sum_i \lambda_i x_i + \langle \nu, Ax-b\rangle$

$Ax = b$

$x \geq 0$, $\quad (f_i(x) = -x_i)$

$$g(\lambda,\nu) = \inf_x \; -\langle \nu, b\rangle + \langle c,x\rangle - \langle \lambda, x\rangle + \langle A^T\nu, x\rangle$$

$$= -\langle \nu, b\rangle + \inf_x \langle c - \lambda + A^T\nu, x\rangle$$

$$= \begin{cases} -\infty & \text{if } c-\lambda+A^T\nu \neq 0, \\ -\langle \nu, b\rangle & \text{else} \end{cases}$$

So (D) $\max_{\nu,\lambda} -\langle \nu, b\rangle$

$\left.\begin{array}{l}\lambda \geq 0 \\ c + A^T\nu = \lambda\end{array}\right\}$ i.e., $c + A^T\nu \geq 0$ $\Big\}$ also a LP! ($\mathbb{R}^n_+$ is a self-dual cone)

▷ SEMI-DEFINITE PROGRAMMING (SDP)

S^n = symm. $n\times n$ matrices

skip QP, QCQP, SOCP, S-lemma

(SDP) $\min_{X\in S^n} \langle C, X\rangle := \text{trace}(C^T X) = \langle \text{vec}(C), \text{vec}(X)\rangle$

$\langle A_i, X\rangle = b_i, \; i=1,\dots,p \quad (A_i \in S^n)$

$X \succeq 0$, i.e., $X \in S^n_+$

can rewrite as $\mathcal{A}: \mathbb{R}^{n^2} \to \mathbb{R}^p$, $\mathcal{A}(X) = \tilde{A}\cdot\text{vec}(X)$, rows of $\tilde{A}$ are $\text{vec}(A_i)^T$

NOT as $\tilde{A}\cdot X$

If X constrained to be diagonal, this is an LP. (or SOCP if Schur complement)

$\underbrace{\text{diagonal}}_{\text{Don't do this in software}}$

Dual is $\max_{\nu\in\mathbb{R}^p} -\langle \nu, b\rangle$

s.t. $\underbrace{\mathcal{A}^*(\nu) + C}\succeq 0$

i.e., $\sum_i \nu_i A_i + C \succeq 0$

$\Big\}$ Linear Matrix Inequality (LMI)

▷ STRONG DUALITY + SLATER'S CONDITION

Weak duality is $d^* \le p^*$, always true (regardless of convexity)

recall $d^* = \sup_{\lambda \ge 0, \nu} g(\lambda,\nu) = \sup_{\lambda \ge 0, \nu} \inf_x \mathcal{L}(x,\lambda,\nu)$

turns out,

$$p^* = \inf_{\substack{f_i(x)\le 0 \\ Ax=b}} f_0(x) = \inf_x \sup_{\lambda \ge 0, \nu} \mathcal{L}(x,\lambda,\nu)$$

why? if x is feasible, $\langle \nu, Ax-b\rangle = 0$

$\lambda_i f_i(x) \le 0$, so for sup, $\lambda_i = 0$.

So $d^* \le p^*$ is due to

max-min ineq: $\sup_{y\in Y} \inf_{x \in X} f(x,y) \le \inf_{x\in X} \sup_{y \in Y} f(x,y)$.

When we have "=", $d^* = p^*$, call it STRONG DUALITY

i.e. strong duality doesn't imply values are finite, i.e. (D) or (P) feasible ...

(slightly stronger: $\exists$ saddle-pts x^*, y^* that achieve this fcn value

i.e., $f(x^*,y^*) = \inf_{x\in X} f(x,y^*) = \sup_{y \in Y} f(x^*,y)$ $\Rightarrow$ strong duality)

Strong Duality is great, but

1) rare, if (P) not convex

2) common, if (P) convex and "generic" * if convex and doesn't hold, "degenerate"

cf. Dmitriy "Dima" Drusvyatskiy

+ Henry Wolkowicz Survey in 2017 Found. Trends Opt.

to get SD, need a CQ, like SLATER:

Def For (P) $\min f_0(x)$, $f_i(x) \le 0$, $Ax=b$, we say Slater's Cond hold if $\exists\, x$ s.t. $x \in \text{relint}(\text{dom}(f_0))$ and $f_i(x) < 0$ $\forall i$, and $Ax=b$.

i.e. Strictly Feasible

that is, only needed if f_i not affine.

Thm If (P) is convex and Slater's holds, $d^* = p^*$. [Proof via Farkas Lemma, Separating hyperplanes (to epigraphs)]

If $p^* < \infty$, then dual admits an optimal soln. (+ vice-versa)

Note: Slater for (P) $\not\Leftrightarrow$ Slater for (D)

how could you not? see side

Corollary: LP's are nice, i.e., Slater's Conditions always hold (all f_i are affine), if feasible.

So either $d^* = -\infty$ and $p^* = +\infty$ (both infeasible)

or $d^* = p^*$ (both $-\infty$, both finite, or both $+\infty$)

Corollary SDP's are not nice.

Often preprocessing can help (easy for LP, not for SDP)

To certify, find feasible (λ^*, ν^*) and x^* such that $g(\lambda^*, \nu^*) = f_0(x^*)$

$$g(\lambda^*,\nu^*) = \inf_x \mathcal{L}(x, \lambda^*, \nu^*) \leq \mathcal{L}(x^*, \lambda^*, \nu^*) = f_0(x^*) + \underbrace{\sum \lambda_i^* f_i(x^*)}_{\leq 0} + \underbrace{\langle \nu^*, Ax^* - b\rangle}_{=0}$$

equality.

$$\leq f_0(x^*)$$

$$= g(\lambda^*, \nu^*) \quad \text{our claim. Then } = 0$$

▷ OPTIMALITY AND KKT CONDITIONS

(P) $\min_x f_0(x)$

$f_i(x) \leq 0, \ i = 1, \dots, p$

$Ax = b$

If we think x^* is optimal, and want to certify it, we want (a bit more than) strong duality (since want feasibility)

Want saddle-pts of min-max system

i.e., ~~$f_0 \leq x$~~

$$\mathcal{L}(x^*, \lambda^*, \nu^*) = \inf_x \mathcal{L}(x, \lambda^*, \nu^*),$$

i.e., $0 \in \partial_x \mathcal{L}(x^*, \lambda^*, \nu^*)$.

we also have $d^* = p^*$, so

$$f_0(x^*) = p^* = d^* = g(\lambda^*, \nu^*)$$

$$= \inf_x f_0(x) + \sum_{i=1}^{p} \lambda_i^* f_i(x) + \langle \nu^*, Ax - b\rangle$$

$$\leq f_0(x^*) + \underbrace{\sum \lambda_i^* f_i(x^*)}_{=0.} + \underbrace{\langle \nu^*, Ax^* - b\rangle}_{=0}$$

$$\leq p^* = f_0(x^*)$$

wk-duality

So $\sum \lambda_i^* f_i(x^*) \leq 0$ (feasibility)

but $\sum \lambda_i^* f_i(x^*) = 0$ (optimality)

since each term is ≤ 0, no cancellation, so equiv. to

$$\lambda_i^* f_i(x^*) = 0 \quad \forall i = 1, \dots, p \quad \text{"COMPLEMENTARY SLACKNESS"}$$

hence

KKT Conditions : "Karush-Kuhn-Tucker"

1. "stationarity" $0 \in \partial_x \mathcal{L}(x, \lambda, \nu)$
2. "primal feas." $f_i(x) \leq 0$, $Ax = b$
3. "dual feas." $\lambda_i \geq 0$
4. "comp. slack" $\lambda_i f_i(x) = 0$.

Thm Let f_i be differentiable, but (P) possibly non-convex. If (x), (λ, ν) are primal/dual optimal, w/ no duality gap, then the KKT conditions ~~are~~ must hold. i.e., KKT are necessary.

(✱) Thm If (P) is convex and (x, λ, ν) satisfy the KKT conditions, then these are primal/dual optimal and there's no duality gap.

i.e., convexity + KKT is sufficient

Used for special cases... eg. prox. (see ex. to left) skipped)

▷ (OPTIONAL: FENCHEL-ROCKAFELLAR DUALITY)

(P) $\min_x f(x) + g(Lx)$, $f, g \in \Gamma_0$, $L: \mathbb{R}^n \to \mathbb{R}^m$ linear   cf. PLC book

(D) $\min_v f^*(L^* v) + g^*(-v)$

<u>Thm</u> (15.23, PLC 1st ed.) If (CQ holds)*, then $p^* = d^*$ and dual soln is achieved.

precisely, $\inf(f + g \circ L) = -\min(f^* \circ L^* + g^* \circ \underset{\text{flip.}}{\vee})$

CQ in finite dim is either

a) $\operatorname{relint}(\operatorname{dom} g) \cap L(\operatorname{relint}(\operatorname{dom} f)) \neq \emptyset$

or

b) if $\operatorname{epi}(f)$ is polyhedral, $\operatorname{dom}(g) \cap L(\operatorname{dom} f) \neq \emptyset$

<u>Why solve dual?</u> $\operatorname{prox}_{g \circ L}$ isn't easy, but $\operatorname{prox}_{g^*}$ is.

<u>Connection w/ standard Lagrangian Duality</u>

(P) $\min_x f(x) + \underbrace{g(Lx)}_{= \sup_v \langle v, Lx\rangle - g^*(v)} = \boxed{\min_x \sup_v f(x) + \langle Lx, v\rangle - g^*(v)}$

if we can flip min/max

– Saddle-pt

$$\cdots = \sup_v -g^*(v) + \min_x f(x) + \langle x, L^* v\rangle$$

$$= \sup_v -g^*(v) - f^*(-L^* v)$$

$$= -\left(\inf_v g^*(-v) + f^*(L^* v)\right).$$

– or, $\min f(x) + g(z)$ s.t. $Lx = z$,

apply Lagr. duality.

<u>Thm</u> let $f \in \Gamma_0(\mathbb{R}^n)$

f strongly cvx iff f^* has L-Lipschitz gradient (+ vice-versa).

(like Fourier duality: smoothness ~~tradeoff~~ if $\hat{f}$ ~~to~~ decay

here, smoothness if f^* "more convex")