Lecture #7

1

1 Recap: We saw last time, a ~~few weird~~ Lagrangian dual; weak-duality; Strong-duality; minimax.

2. Today, let me quickly mention these optimality conditions again, before ~~we~~ discuss specific examples / extensions

3. Recall our basic result: $0 \in \partial f(x)$ is nec. (suff for cvx) for $x$ to be a local (global) for cvx.

4. In the realm of convexity, let's look at other optimality conditions, all of which are ultimately based around giving more "practical" ways to characterize/verify $0 \in \partial f(x)$.

5. Lemma: $0 \in \partial f(x)$ iff $f$ is finite at $x$ and $f'(x;y) \geq 0 \ \forall y$

Proof: we show that ~~$g \in f'(x;y)$ iff~~ $g \in \partial f(x)$

iff $f'(x;y) \geqslant \langle g, y\rangle \quad \forall y.$

Recall: $f'(x;y) = \lim_{\lambda \downarrow 0} \dfrac{f(x+\lambda y) - f(x)}{\lambda} = \inf_{\lambda > 0} \dfrac{f(x+\lambda y)-f(x)}{\lambda}$

(Why does that inf hold! ~~Since f is cvx the function $\frac{f(x+\lambda y)-f(x)}{\lambda}$ is $\downarrow$.~~

( ~~Consider the function~~ : (a) ~~$h(y) := f(x+y) - f(x)$~~

(b) ~~$\lambda^{-1}h(\lambda y)$ is also cvx (persp. transform)~~

the func $\lambda \mapsto \dfrac{f(x+\lambda y)-f(x)}{\lambda}$ is an incr fn. of $\lambda \geq 0$

[ Because $f$ is cvx, it has increasing slopes

( alternative $h(y) = f(x+y) - f(x)$ is cvx

$\lambda^{-1}h(\lambda y)$ is monotonic

Why does "inf" hold! Since $f$ is cvx the "slopes" (func. $\lambda \mapsto \dfrac{f(x+\lambda y)-f(x)}{\lambda}$ is Discreasing.

(2)

~~Thus we see~~ Now since $f$ is cvx; r

at a point $z = x + \lambda y$ for $\lambda \geq 0$, any $y$

(~~at f(z)~~) $g \in \partial f(x)$ iff

$$f(z) \geqslant f(x) + \langle g, z - x \rangle$$

$$f(x+\lambda y) - f(x) \geqslant \langle g, \lambda y \rangle$$

$\lambda > 0$ $\quad \dfrac{f(x+\lambda y) - f(x)}{\lambda} \geqslant \langle g, y \rangle \quad \forall \lambda > 0, \forall y$

$$\Rightarrow \inf_{\lambda > 0} \text{———} = f'(\,) \geqslant \langle g, y \rangle.$$

(other direction is trivial)

---

Theorem: Let $f$ be a closed, proper ($\not\equiv +\infty$) cvx. fn $f$

Then (a) $\inf f = -f^*(0)$. Thus, $f$ is bdd below iff $0 \in \text{dom}\, f^*$

Proof: $f^*(z) = \sup_x \langle z, x \rangle - f(x)$; plug in $z = 0$.

(b) The minimum set of $f$ is $\partial f^*(0)$. Thus the inf is attained iff $f^*$ is subdiff at 0.

Proof: Recall that $\underline{x}$ is a min of $f$ iff $0 \in \partial f(x)$

But $g \in \partial f(\bar{x})$ $\Leftrightarrow$ $\bar{x} \in \partial f^*(g)$ (for closed cvx fns)

$$\Rightarrow \bar{x}_{\min} \in \partial f^*(0)$$

(satisfied eg if $0 \in ri(\text{dom}\, f^*)$) (b/c then $f^*$ is guaranteed to be subdiff there)

(c) For inf of '$f$' to be finite but unattained it is neces & suff that $f^*(0)$ be finite but $f^{*\prime}(0;y) = -\infty$ for some direction $y$.

c) Use a) + b) (for finite ch.)

Unattained iff $\partial f^*(0) = \phi$, but this happens, when $\exists y$ st $f^{*\prime}(0;y) = -\infty$ happens

[(~~Intuition~~: Reason: $\partial f(x) = \{ g \in \mathbb{R}^n \mid f'(x;y) \geq \langle g, y \rangle \ \forall y \}$ ]

d) The minimum set of '$f$' is nonempty and bounded if and only if $0 \in \text{int}(\text{dom } f^*)$

e) The min set is a singleton $\{x\}$ iff $f^*$ is differentiable at $0$ and $x = \nabla f^*(0)$

(notice if $\partial f^*(0)$ is a singleton $\Rightarrow$ $f^*$ is diff at $0$ )

---

~~Now recall Lagrangian from last time;~~

~~$\min f(x)$~~

~~st $h_i(x) \leq 0 \quad 1 \leq i \leq m$~~

~~$L(x, \lambda) := f(x) + \lambda_1 h_1 + \dots + \lambda_m h_m$~~

Let (P) be the primal problem:

$$\left.\begin{array}{ll} \min\limits_{x \in \Omega} f(x) & \\ h_i(x) \leq 0 & 1 \leq i \leq r \\ h_i(x) = 0 & i = r+1, \dots m \end{array}\right] (P)$$

The Lagrangian of (P) is the function $\mathbb{R}^n \times \mathbb{R}^n$ defined by:

$$\mathcal{L}(x, \lambda) := \begin{cases} f(x) + \lambda_1 h_1(x) + \dots + \lambda_m h_m(x), & \underrightarrow{\lambda_1, \dots, \lambda_r \geq 0}_{x \in C} \\ -\infty \text{ if infeas. } \lambda \\ +\infty \text{ if } x \notin C. \end{cases}$$

we look at $\inf_{x \in C} \sup_{\lambda \in E} L(x,\lambda)$

and worry about $\sup_{\lambda \in E} \inf_{x \in C} L(x,\lambda)$ } the minimax setup

☆ $L$ as specified above contains complete info about (P)

☆ $L$ is finite on $E \times C$; else infn.

☆ $f(x) = L(x,0)$

$h_i(x) = L(x,e_i) - L(x,0)$ $\forall\, i=1,\dots,m,\ x \in C$

Now recall a pair $(\bar{x},\bar{\lambda})$ called a saddle-point of $L$

if $L(\bar{x},\lambda) \le L(\bar{x},\bar{\lambda}) \le L(x,\bar{\lambda})$ $\forall x,\lambda$

(and for this $\inf_x \sup_\lambda L = \sup_\lambda \inf_x L \ni L(\bar{x},\bar{\lambda})$)

Main theorem: Let (P) be a convex problem as stated above. Let $\bar{x}, \bar{\lambda}$ be vectors in $\mathbb{R}^n$ & $\mathbb{R}^m$. In order that $\bar{\lambda}$ be a KKT vector for (P) and $\bar{x}$ an optimal solution to (P) it is necessary and sufficient that $(\bar{x},\bar{\lambda})$ is a saddle-point of the Lagrangian.

Moreover, this condition holds iff $\bar{x}$ and $\bar{\lambda}$ satisfy

(a) $\lambda_i \geq 0$ for $1 \le i \le r$

(b) $h_i(\bar{x}) \le 0$ " $1 \le i \le r$

(c) $\lambda_i h_i(\bar{x}) = 0$ $1 \le i \le r$

(d) $h_i(\bar{x}) = 0$ $i = r+1,\dots,m$

(e) $0 \in (\partial f(\bar{x}) + \lambda_1 \partial h_1(\bar{x}) + \dots + \lambda_m \partial h_m(\bar{x}))$

SP $\Rightarrow$ $\bar{x} \in \operatorname{argmin} L(x,\bar{\lambda})$ + this!

Alternative derivation of KKT condition

Assume for simplicity that we have

$$(P)\begin{cases} \min & f(x) \\ \text{st} & h_i(x) \leq 0 \quad 1 \leq i \leq m \\ & x \in \mathbb{R}^n \end{cases}$$

(let $h_i$, $f$ finite valued ~~and continuous~~) on $\mathbb{R}^n$; hence continuous)

(2) Also assume that $\exists\, x \in \mathbb{R}^n$ such that

$$h_1(x) < 0, \ldots, h_m(x) < 0$$

Write: $\quad C_i = \{x \mid h_i(x) \leq 0\} \quad 1 \leq i \leq m$

$$(P) \equiv \min_x \phi(x) := f(x) + \delta_{C_1}(x) + \cdots + \delta_{C_m}(x).$$

The optimal solution to (P) are vectors $\bar{x}$ s.t.

$$0 \in \partial\phi(\bar{x})$$

The assumption $\exists\, x$ st $h_i(x) < 0 \Longrightarrow$ h.

$$\text{int}\, C_1 \cap \cdots \cap \text{int}\, C_m \neq \emptyset$$

(Since the $h_i$ are finite, and cont. on their interior)

$h_i(x)$ is finite at $x$ and so it's cont. then etc. it has to have a neighborhood)

From Rockafellar's theorem on sub diff. we follow that

$$\partial\phi(x) = \partial f(x) + \cdots + \partial\delta_{C_m}(x)$$

$\partial\delta_{C_i} \equiv \mathcal{N}_{C_i}$ (normal cone) we derived it

and an exercise shows that (see sub diff calculus rules from Lect. 3)

$$\partial\delta_{C_i} = \begin{cases} \bigcup\{\lambda_i \partial h_i(x) \mid \lambda_i \geq 0\} & \text{if } h_i(x) = 0 \\ \{0\} & \text{if } h_i(x) < 0 \\ \emptyset & \text{if } h_i(x) > 0 \end{cases}$$

Thus, $\partial\phi(x) \neq$ empty iff

$x$ satisfies $h_i(x) \leq 0 \quad \forall\ 1 \leq i \leq m$.

( Since, contrary to what I said in last 3, $\phi + \phi = \phi$ )

Thus $\partial\phi(x) = \bigcup \{ \partial f(x) + \lambda_1 \partial h_1(x) + \dots + \lambda_m \partial h_m(x) \}$

over all choices of $\lambda_i \geq 0$

Such that

$$\lambda_i h_i(x) = 0$$

( Since if $h_i(x) < 0$, $\partial \delta_{C_i} = \{0\}$ : here

while for $h_i(x) = 0$, $\partial \delta_{C_i} = \{ \lambda_i \partial h_i(x) \mid \lambda_i \geq 0 \}$ )

So either $\lambda_i \geq 0$, or $h_i(x) = 0$ & we cannot jointly have

$\lambda_i > 0$, $h_i > 0$ : that would violate.

Thus, $0 \in \partial\phi(\bar{x})$ iff $\exists\ \lambda_1, \dots, \lambda_m$

which satisfy the KKT conditions. $\square$

Example: $\min \frac{1}{2}\|Ax-b\|^2$

st $\|x\|_2 \leq 1$ $\qquad C = \{x \mid \|x\|_2 \leq 1\}$

unc. $\phi(x) = \frac{1}{2}\|Ax-b\|^2 + \delta_C(x)$

$\partial\phi(x) = A^T(Ax-b) + \partial\delta_C(x)$

$\exists\, x$ st $\|x\| < 1$ obviously; $\Leftrightarrow$ KKT $\quad 0 \in A^T(Ax-b) + \lambda\, \partial\|x\|_2$

$\lambda \geq 0$; $\lambda(\|x\|_2 - 1) = 0$

6.5

Recall that $\partial \delta_C(x) = N_C(x)$

$$= \{ g \mid 0 \geq \langle g, y-x \rangle \ \forall y \in C \}$$

Since this is a cone if $g \in N_C(x)$

$$\Rightarrow \lambda g \in N_C(x) \ \forall \ \lambda \geq 0 .$$

if $x \neq 0$, $\partial \|x\|_2 = \begin{cases} \frac{x}{\|x\|}, & x \neq 0 \\ B_0(1), & x = 0. \end{cases}$ 7

~~Ex: For diff fn~~

Sometimes ! ~~very rarely~~ we can directly solve the KKT conditions (bunch of non-linear equations)

Usually, that is too hard, and we proceed to:

(a) iteratively solve them (Newton method etc)
(b) Some generate a seq $(x^k)$ that satisfies KKT in the limit
(c) several other variants

Let's try to now briefly extend above duality-optimality theory to the conic world.

Facts about LP-duality: Slater

Consider LP: $\min\ c^T x$ st $Ax \leq b$

Dual: $\max\ b^T \lambda$ st $A^T\lambda + c = 0$, $\lambda \geq 0$

A • If either $p^*$ or $d^*$ finite, then $p^* = d^*$ and both primal / dual solution attained.

(recall $v(u) = \inf \{ c^T x \mid Ax - b \leq u \}$

~~$v(0) = p^*$ if $p^*$ is finite~~

• if $p^* = -\infty$ then $d^* = -\infty$ (weak dual.)
• if $p^* = +\infty$ " $d^* = +\infty$ (weak dual.)

Known results about LP-duality

Just skip maybe to save time

1. $p^* = -\infty \Rightarrow d^* = -\infty$
2. $d^* = +\infty \Rightarrow p^* = +\infty$
3. if $p^*$ or $d^*$ finite, then $p^* = d^*$ and both (P) + (D) have optimal solutions.

Thus, if LP is feasible (either (P) or in (D)) Strong duality

if $p^* = +\infty$ (infeas primal, then SD fails)

---

SDP - duality:

(Primal form) $\quad p^* = \inf \{ \langle C, X \rangle \mid \langle A_i, X \rangle = b_i, \ X \succeq 0 \}$

How to handle the $X \succeq 0$ constraint?

# Introduce the "Conic Lagrangian"

$$L(X, \nu, Y) := \langle C, X \rangle - \sum_i \nu_i ( \langle A_i, X \rangle - b_i ) - \operatorname{tr}(YX)$$

where $\nu \in \mathbb{R}^m$, and $Y \succeq 0$ : matrix-dual variable.

# Impt fact: If $X, Y \succeq 0 \Rightarrow \operatorname{Tr}(XY) \geq 0$

$\Rightarrow \quad p^* \geq L(X, \nu, Y) \quad \forall$ feasible $X$, $\forall \nu, Y$

$\Rightarrow \quad p^* \geq \sup_{\nu, Y \succeq 0} L(X, \nu, Y)$

in fact as before $\boxed{p^* \geq d^* = \sup_{\nu, Y} \inf_{x} L(X, \nu, Y)}$

Simplifying $\inf_X L(X, \nu, Y)$ we obtain

$$g(\nu, Y) = \begin{cases} +b^T\nu & \text{if } C - \sum_i \nu_i A_i - Y = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

So dual problem:

$$\max_{\nu,\, Y \succeq 0} \; b^T\nu \quad \text{st} \quad C - \sum_i \nu_i A_i = Y \succeq 0$$

$$\Rightarrow \boxed{\max_{\nu} \; b^T\nu \quad \sum_i \nu_i A_i \preceq C} = \boxed{\min \; b^T x \quad \text{st } x_1 A_1 + \dots + x_m A_m - C \succeq 0}$$

This so called - conic form of SDP

Some points: Strong duality holds if either primal or dual is strictly feasible
(In LPs, mere feasibility suffices)
(Explicit example in slides linked to.)

Ex. 5.44 + last 2 pages of Ch. 5 of BV

---

Lemma: Let $A_1, \dots, A_m, C$ be symm. matrices — Skip the following

The system: $x_1 A_1 + \dots + x_m A_m - C \succ 0$ has NO solution ($x \in \mathbb{R}^m$)

iff $\exists$ a symm matrix $Y$ st.

$$\mathrm{tr}(A_i Y) = 0, \quad 1 \le i \le m$$
$$\mathrm{tr}(CY) \ge 0$$
$$Y \succeq 0.$$

Ex. 5.44 in BV

Proof: If $L = x_1 A_1 + \dots + x_m A_m \succ C$ has no soln $\Rightarrow$ Lin subspace $L$ of cone matrix is disjoint from the interior of $P^n$. $\Rightarrow \exists$ a separating hyperplane s.t. $\{X \mid \mathrm{tr}(YX) = 0\}$ we may assume that $\mathrm{tr}(YX) \ge 0 \;\; \forall X \in P^n \Rightarrow Y \succeq 0$

and $\langle A_i, Y \rangle = 0$ $\quad (A_i - \frac{1}{m} C) \quad \langle A_i - \frac{1}{m} C, Y \rangle = 0 \quad \mathrm{tr}(CY) \ge 0$

That lin variety

$\Rightarrow$ we'll have $H = \{ X \mid \mathrm{tr}(\cancel{(Y-C)X}) = 0 \}$

we can chose st $C = \mathrm{tr}(Y X_0)$

$\mathrm{tr}((Y-C)X) = 0 \quad \{A \mid \langle Y, X \rangle = \langle C, X \rangle \}$

are like $< 0$ or $\geq 0$

---

Most weight cvx fn on $\mathbb{P}^n$: $= \log\det(X)$.

The Set

$$A(x) := \{ A_0 + x_1 A_1 + \cdots + x_m A_m \succeq 0 \}$$

called a "Spectrahedron"

coincides with solution set of an SDP

---

Rank of SDP

Barvinok: If (P) SDP is solvable, then $\exists$ a solution whose rank $r$ satisfies $\frac{r(r+1)}{2} \leq m$

($m$ is # constraints)

Reading assignment: Burer & Monteiro: A nonlinear prog algorithm via Low-rank factorization.