Lecture 4 -

Summarize duality for LP

(P) $\min c^T x$ s.t. $Ax = b$, $x \geq 0$ ("standard form") $\equiv$ $\min\ f_0(x) \equiv c^T x$ s.t. $f_1(x) = -x \leq 0$, $h(x) = Ax - b = 0$

$$(P)\quad \begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & x \geq 0 \end{array} \quad \equiv \quad \begin{array}{ll} \min & f_0(x) \equiv c^T x \\ \text{s.t.} & f_1(x) = -x \leq 0 \\ & h(x) = Ax - b = 0 \end{array}$$

("standard form")

Lagrangian: $L(x,\lambda,\nu) = c^T x - \lambda^T x + \nu^T(Ax - b)$

$$= x^T(c - \lambda + A^T\nu) - \nu^T b$$

LDF $\quad g(\lambda,\nu) = \inf_x L(x,\lambda,\nu) = \begin{cases} -\nu^T b & \text{if } c - \lambda + A^T\nu = 0 \\ -\infty & \text{otherwise.} \end{cases}$

LDP $\quad \sup_{\lambda \geq 0} g(\lambda,\nu)$

$\boxed{\text{NOTE}}$

$$\equiv \max\ b^T y \quad (\cancel{\text{change}}\ y = -\nu)$$
$$\text{s.t. } A^T y + \lambda = c$$
$$\lambda \geq 0$$

"dual slack variable" in LP parlance

$$(D) \quad \equiv \max\ b^T y \quad \text{s.t. } A^T y \leq c$$

What is the dual of the dual? Next page.

Applay same argument to (D): easier to use the version w/o slack variable λ.

$$(\tilde{P}) \equiv (D) \quad \max\ b^T y \quad \text{s.t. } A^T y \le c \qquad \equiv \quad \min\ -b^T y \quad \text{s.t. } A^T y - c \le 0$$

Lagrangian $\tilde{L}(y, \pi) = -b^T y + \pi^T(A^T y - c)$

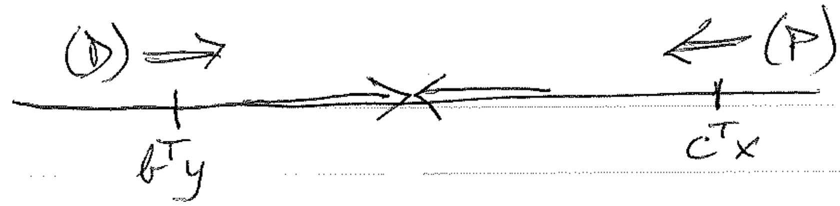
y playing role of x, π playing role of λ (no ν) $\quad = y^T(-b + A\pi) - \pi^T c$

LDF $\quad \tilde{g}(\pi) = \inf_y \tilde{L}(y,\pi) = \begin{cases} -\pi^T c & \text{if } A\pi = b \\ -\infty & \text{otherwise} \end{cases}$

LDP $\quad \sup_{\pi \ge 0} \tilde{g}(\pi) = \min\ c^T\pi \quad \text{s.t. } A\pi = b,\ \pi \ge 0$

NOTE $\rightarrow \pi \ge 0$

Exactly (P)!



For feasible x, y we have

$$b^T y = (Ax)^T y = x^T (A^T y) \le x^T c = c^T x$$

WEAK DUALITY.

<u>Strong duality</u> holds for LP without Slater assumption (relies on SHP thm / Farkas lemma). It says we always have $d^* = p^*$ unless both (P), (D) are infeasible: then $d^* = -\infty$, $p^* = \infty$.

Could have $d^* = p^* = -\infty$ (primal unbounded, dual infeasible)

or $d^* = p^* = \infty$ (dual unbounded, primal infeasible)

Otherwise $\exists\, x^*, y^*$ with zero duality gap: $b^T y^* = c^T x^*$

So, letting $\lambda^* = c - A^T y^*$ "dual slack"

we have $(x^*)^T \lambda^* = 0$

Since $x^* \ge 0$, $\lambda^* \ge 0$ this implies <u>complementarity</u>:

$\forall j \in \{1, \dots, n\}$ either $x_j = 0$ or $\lambda_j = 0$ (or both).

"active constraint" "inactive constraint" (or degenerate)

Goldman Tucker theorem for LP:

$\exists\, x^*, \lambda^*$ with STRICT COMPLEMENTARITY:

not both $x_j^* = \lambda_j^* = 0$

GENERALIZED INEQUALITIES

Somewhat different notation from book

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

$$\text{s.t. } F_i(x) \preceq_K 0 \quad i=1,\dots,m$$

$$\& \ h(x) = 0$$

means: $-F_i(x) \in K$

e.g. PSD cone.

Lagrangian

$$L(x,\Lambda,\nu) = f_0(x) + \sum_{i=1}^m \langle \Lambda_i, F_i(x) \rangle + \nu^T h(x)$$

$\Lambda_1,\dots,\Lambda_m$

$\mathrm{tr}\,\Lambda F(x)$ for PSD cone.

LDF $\quad g(\Lambda,\nu) = \inf_x L(x,\Lambda,\nu)$

We have $g(\Lambda,\nu) \le p^*$ for $\Lambda_i \succeq_{K^*} 0 \quad i=1,\dots,m$

means $\Lambda_i \in K^*$ ← dual cone

because, for feasible $\tilde{x}$,

$$\sum_{i=1}^m \langle \Lambda_i, F_i(\tilde{x}) \rangle \le 0 \quad (\text{def}^n \text{ of } K^*)$$

$$\nu^T h(\tilde{x}) = 0$$

(regardless of ν)

so $\inf_x L(x,\Lambda,\nu) \le L(\tilde{x},\Lambda,\nu) \le f_0(\tilde{x})$.

LDP $\quad d^* = \sup_{\Lambda_i \succeq_{K^*} 0} g(\Lambda, \nu) \le p^*$

(weak duality)

If f_0 is convex, F_i are K-convex, and $h(x) = Ax - b$,

$$F_i(\theta x + (1-\theta)y) \preceq_K \theta F_i(x) + (1-\theta) F_i(y)$$

+ Slater cond holds, i.e. $\exists \tilde{x}$ ($\in \operatorname{relint} \mathcal{D}$) with

$$F_i(\tilde{x}) \prec_{K^*} 0, \quad A\tilde{x} = b$$

then strong duality holds: $d^* = p^*$.

Example SDP in primal standard form.

$$(P) \quad \inf \langle C, X \rangle \quad = \operatorname{tr} CX = \sum c_{ij} x_{ij}$$

$$\text{s.t. } \langle A_i, X \rangle = b_i \quad i = 1, \dots, p.$$

$X \in S^n \qquad X \succeq 0$ $\boxed{\text{JUST ONE INEQ}}$ $\quad v^T X v \ge 0 \;\; \forall v.$

$\uparrow$ PSD.

$$L(X, \Lambda, \nu) = \langle C, X \rangle - \langle \Lambda, X \rangle + \sum_{i=1}^{p} \nu_i \left(\langle A_i, X \rangle - b_i \right)$$

$$g(\Lambda, \nu) = \inf_X \langle C - \Lambda, X \rangle + \sum_{i=1}^{p} \nu_i \langle A_i, X \rangle - \nu^T b$$

$$= \begin{cases} -\nu^T b & \text{if } \left(\sum_{i=1}^{p} \nu_i A_i = \Lambda - C \right) \\ -\infty & \text{otherwise.} \end{cases}$$

LDP is $d^* = \sup_{\Lambda \succeq 0} g(\Lambda, \nu)$

(PSD, as $K = K^*$ for semidefinite cone)

$y = -\nu$

(D) $\equiv \sup_{\Lambda, y} \; b^T y$

s.t. $\left(\sum_{i=1}^{p} y_i A_i\right) + \Lambda = C$ ← "SLACK MATRIX".

$\Lambda \succeq 0$

↑ PSD

Just as in LP, C in primal obj ⟷ dual constr.
b " " cons. ⟷ dual obj.

Version without slack matrix:

$$\sup_y b^T y$$

s.t. $\sum_{i=1}^{p} y_i A_i \preceq C$

(means $C - \sum y_i A_i \succeq 0$).

As in LP, let's take dual of dual.

Lagrangian $\tilde{L}(y, \Pi) = -b^T y + \left\langle \Pi, \sum_{i=1}^{p} y_i A_i - C \right\rangle$
(no ν)

LDF $\tilde{g}(\pi) = \inf_y \tilde{L}(y, \Pi) = \begin{cases} -\langle C, \Pi \rangle & \text{if } \langle A_i, \Pi \rangle = b_i, \; i = 1, \ldots, p \\ -\infty & \text{otherwise.} \end{cases}$

LDP $\sup_{\Pi \succeq 0} g(\Pi) \equiv \inf_{\Pi} \langle C, \Pi \rangle$

s.t. $\langle A_i, \Pi \rangle = b_i \quad i = 1, \ldots, p$

$\Pi \succeq 0$

EXACTLY (P)!!

As with LP $\Rightarrow$ (D) $\qquad\qquad$ $\Leftarrow$ (P)

$$b^T y \qquad\qquad\qquad \langle C, X\rangle$$

$$b^T y = \sum_{i=1}^{\ell} \langle A_i, X\rangle y_i = \langle \underbrace{\sum y_i A_i}_{C-\Lambda}, X\rangle = \langle C, X\rangle - \langle X, \Lambda\rangle$$

What is sign of $\langle X, \Lambda\rangle = \operatorname{tr} X\Lambda$

We know $X \succeq 0$, $\Lambda \succeq 0$ $\qquad$ ($Q^TQ = I$) (D diagonal)

Let $X = M^2$ (sym sq. root: if $X = QDQ^T$ ($D \geq 0$) then $M = QD^{1/2}Q^T$)

$$\langle X, \Lambda\rangle = \operatorname{tr} X\Lambda = \operatorname{tr} MM\Lambda$$
$$= \operatorname{tr} \underbrace{M\Lambda M}_{\succeq 0} \geq 0$$

so $b^T y \leq \langle C, X\rangle$ when X, y feasible

weak duality.

DOES NOT ALWAYS HOLD $\rightarrow$ Strong duality says $d^* = p^*$, so assuming these are both attained, $\exists X^*, y^*, \Lambda^*$ S.T. $b^T y^* = \langle C, X^*\rangle$

i.e. $\langle X^*, \Lambda^*\rangle = 0$.

Let $X^* = M^2$, $\Lambda^* = P^2$ $\quad M = M^T \quad P = P^T$

$$\langle X^*, \Lambda^*\rangle = \langle M^2, P^2\rangle = \operatorname{tr} MMPP = \operatorname{tr}(MP)(PM)$$
$$= \|PM\|_F^2 = 0$$

so $PM = 0 = MP$

so $X^*\Lambda^* = MMPP = 0 = \Lambda^* X^*$

so X^*, Λ^* share common system of eig-vectors

$\exists Q$ s.t. $Q^TQ = I$

with $Q^T X^* Q = D = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix}$, $Q^T \Lambda^* Q = \begin{bmatrix} e_1 & & \\ & \ddots & \\ & & e_n \end{bmatrix}$

$\forall j \in [1, \dots, n]$: $d_j = 0$ or $e_j = 0$ or both

(eigenvalue complementarity).

MORE GENERALLY

The KKT Equations + Complementary Slackness

Suppose that P & D optimal values are attained + equal (so strong duality holds.),

(BV p. 242).

for the convex program

$$\min_{x \in D} f_0(x) \quad \text{S.T.} \quad f(x) \leq 0 \quad f = \begin{bmatrix} f_1 \\ \vdots \\ f_m \end{bmatrix} \text{ all convex}$$

$$Ax = b \quad \text{affine equality constr.}$$

Then $\exists\, x^*, \lambda^*, \nu^*$ S.T. $\lambda^* \geq 0$, $f(x^*) \leq 0$ &

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

$$= \inf_x \Big(f_0(x) + \sum_{i=1}^{m} \underbrace{\lambda_i^*}_{\geq 0} \underbrace{f_i(x)}_{\leq 0 \text{ NO}^*} + (\nu^*)^T (Ax - b) \Big)$$

by def'n of the dual problem

$$\leq f_0(x^*) + \sum_{i=1}^{m} \underbrace{\lambda_i^*}_{\geq 0} \underbrace{f_i(x^*)}_{\leq 0} + (\nu^*)^T \underbrace{(Ax^* - b)}_{0}$$

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

i.e "=", that is $\lambda_i^* f_i(x^*) = 0 \quad i = 1, \dots, m$

COMPLEMENTARITY.

Nonzero Lagrange multipliers must correspond to "ACTIVE" constraints.
"Inactive" constraints must corresp to zero Lagrange multipliers.

But STRICT COMP does not necessarily hold.
IT DOES FOR LP.

Suppose f, f_i are all differentiable as well as convex. Then since x^* minimizes $L(x, \lambda^*, \nu^*)$ over x, we have:

$$(\mathrm{KKT}): \quad 0 = \nabla f_0(x^*) + \left(\sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*)\right) + A^T \nu^*$$

History: Lagrange for equality constraints only
Karush 1939: unpublished Master's thesis!
Fritz John 1948 (with the "Fritz John" multiplier for f_0)
Kuhn + Tucker 1951.

We see that (KKT) along with feasibility
$(f_i(x^*) \le 0_{\,i=1,\dots,m},\ Ax^* = b,\ \lambda^* \ge 0)$
and complementarity: $\lambda_i^* f_i(x^*) = 0$ (so $d^* = p^*$)
shows primal + dual optimality.

KKT also extends to nonconvex case too (BV, p. 243) but in this case they are only NECESSARY conditions for optimality, not sufficient
(e.g., unconstrained case: $\min f_0(x) = x^3$
$\nabla f_0(x) = 0 \not\Rightarrow x$ is optimal).

<u>Saddle Point Interpretation</u>. Assume for simplicity there are no equality constraints.

Note $\sup_{\lambda \succeq 0} L(x,\lambda) = \sup_{\lambda \succeq 0} f_0(x) + \lambda^T f(x)$

$$= \begin{cases} f_0(x) & \text{if } f(x) \preceq 0 \quad (f_i(x) \le 0,\ i=1,\dots,m) \\ \infty & \text{otherwise.} \end{cases}$$

Thus $p^* = \inf_{x\ (\in \mathcal{D})} \sup_{\lambda \succeq 0} L(x,\lambda)$.

while by defn

$$d^* = \sup_{\lambda \succeq 0} \inf_{x\ (\in \mathcal{D})} L(x,\lambda).$$

Weak Duality: $d^* \le p^*$

does not actually depend on properties of L, we have

$$(*) \qquad \sup_{z\in Z} \inf_{w \in W} h(w,z) \le \inf_{w\in W} \sup_{z \in Z} h(w,z)$$

for any function $h: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, and any $W \subseteq \mathbb{R}^n$, $Z \subseteq \mathbb{R}^m$ (as long as not both W and Z are empty!)

<u>Pf</u> (from RTR '70) (also see ex. 5.24)

Let $H(z) = \inf_{w\in W} h(w,z)$

and $\alpha = \sup_{z \in Z} H(z)$: (LHS of (*))

For all $w \in W$, we have

$$\{ \ h(w,z) \ge H(z) \quad \forall z \in Z \ \}$$

so $\{ \sup_{z\in Z} h(w,z) \ge \sup_{z \in Z} H(z) = \alpha. \}$

continued on next page

(Pf that sup inf $\le$ inf sup, cont'd)

This is true for all $w \in W$, so

$$\underset{\text{of }(*)}{\text{RHS}} = \inf_{w\in W}\ \sup_{z\in Z}\ h(w,z) \ge \alpha .$$

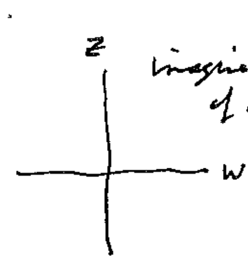
(easy proof once you see it!).

The biggest gap between LHS & RHS:

let $W = Z = \mathbb{R}$, $h(w,z) = -w^2 + z^2$.

concave in w, convex in z

then LHS $= -\infty$, RHS $= \infty$.



imagine graph of h here.

Strong duality (LHS = RHS) occurs in the opposite case: when h is CONVEX in w and CONCAVE in z.

e.g. $h(w,z) = w^2 - z^2$

then LHS = RHS = 0.

GAME INTERPRETATION (goes back to Morgenstern + Von Neumann)

choice of W = Player 1 (wants to INF h), choice of Z = Player 2 (wants to SUP h).

In general, Player 1 wants Player 2 to "go first" (FIX Z) ~~(sup, then inf)~~ so he can take account of ~~his~~ her choice. ~~His choice~~ He knows that Player 2 will choose the z that maximizes the best that Player 1 can do (LHS).

* But if strong duality holds, it makes no difference