4-1 Lecture 4

Let's look at LP again

P: $\min c^T x$ $\equiv$ $\min f_0(x) \equiv c^T x$

$(x\in\mathbb{R}^n)$ s.t. $Ax=b$ s.t. $f_1(x) = -x \le 0$

$x \ge 0$ $h(x) = Ax-b=0$

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

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

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

LDP: $\sup_{\lambda\ge0} g(\lambda,\nu)$

Note

$= \max b^T y$ changing $-\nu$ to y.

s.t. $A^T y + \lambda = c$

$\lambda \ge 0$

"dual slack variable"

$=$

D: $\max b^T y$

s.t. $A^T y \le c$

What is the dual of the dual? Next page.

Apply same argument to D. Easier to use the version without the slack variable λ.

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

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

new "primal" variable — new Lagrange multiplier vector.

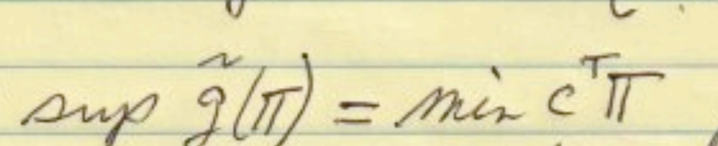
$$= (-b + A\pi)^T y - c^T \pi$$

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

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

Note — EXACTLY P.

So dual of the dual is the primal.



For feasible x, y, we have

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

Weak duality.

Strong duality always holds for LP (refined version of Slater, AKA Farkas lemma) (unless P, D both infeasible) assuming $d^* = p^*$ is finite

Strong duality says that, at optimal x^*, y^*, we have

$$b^T y^* = c^T x^*.$$

Let $\lambda^* = c - A^T y^*$ "dual slack var"

Then $(x^*)^T \lambda^* = 0$ by (†)

Since $x^* \geq 0$ AND $\lambda^* \geq 0$, this implies COMPLEMENTARITY: $x_i^* = 0$ or $\lambda_i^* = 0$ $i = 1, \ldots, n$ (or both).

"ACTIVE PRIMAL CONSTRAINT"

"INACTIVE PRIMAL CONSTRAINT"

Goldman-Tucker for LP:

$\exists x^*, \lambda^*$ with STRICT COMPLEMENTARITY holding: not both $x_i^* = \lambda_i^* = 0$.

4.4 SDP (Semi-Definite Programming)

Does not quite fit in last lecture's framework. Can be derived using BV 5.9 or generalized inequalities, but let's just derive the duality directly here.

$$P: \inf_{X \in S^n} \langle C, X \rangle = \operatorname{tr} CX = \sum c_{ij} x_{ij}$$

p^*

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

$$X \succeq 0 \quad \text{PSD}$$

Define the Lagrangian

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

LDF: $g(\Lambda, \nu) = \inf_{X \in S^n} L(X, \Lambda, \nu)$

$(\Lambda \in S^n, \nu \in \mathbb{R}^p)$

$$= \begin{cases} -b^T \nu & \text{if } C - \Lambda + \sum_{i=1}^{p} \nu_i A_i = 0 \\ -\infty & \text{otherwise.} \end{cases}$$

For any feasible $\tilde{X} \in S^n$, i.e. with $\tilde{X} \succeq 0$ and $\langle A_i, \tilde{X} \rangle = b_i$, $i = 1, \ldots, p$, we have

$$L(\tilde{X}, \Lambda, \nu) \le \langle C, \tilde{X} \rangle \text{ as long as } \Lambda \succeq 0$$

because then $\langle \Lambda, \tilde{X} \rangle \geq 0$ as $\Lambda \succeq 0$ & $\tilde{X} \succeq 0$

Pf Let $\Lambda = G^T G$, $\tilde{X} = H^T H$
(Cholesky or symmetric sq. root)

$$\operatorname{tr} \Lambda \tilde{X} = \operatorname{tr} G^T G H^T H$$
$$= \operatorname{tr} (GH^T)(GH^T)^T$$
$$= \|GH^T\|_F^2 \geq 0$$

ALTERNATIVELY
Follows from $(S_+^n)^* = S_+^n$.

4-5 SDP cont'd

so $g(\Lambda, \nu) = \inf_{X \in S^m} L(X, \Lambda, \nu) \leq p^*$

assuming $\Lambda \succeq 0$.

So again LDP is to maximize the lower bound

$$\sup_{\substack{\Lambda \succeq 0 \\ \nu \in \mathbb{R}^p}} g(\Lambda, \nu)$$

$$\equiv \sup b^T y \quad \text{(changing } -\nu \text{ to } y\text{)}$$

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

$$\Lambda \succeq 0 \quad \text{PSD}$$

Dual D: $\equiv$

$$\sup b^T y \quad (y \in \mathbb{R}^m)$$

$$\text{s.t. } C - \sum_{i=1}^{p} y_i A_i \succeq 0 \quad \text{PSD}$$

Let's now take the dual of the dual:

$$\tilde{L}(y, \Pi) = -b^T y + \left\langle \Pi, \left(\sum_{i=1}^{p} y_i A_i\right) - C \right\rangle$$

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

4-6 SDP cont'd.

$$\underline{\text{LDP}}: \sup_{\Pi \succeq 0} g(\Pi)$$

$$= \inf \langle C, \Pi \rangle$$
$$\text{s.t. } \langle A_i, \Pi \rangle = b_i \quad i = 1, \ldots, p$$
$$\Pi \succeq 0$$

Exactly P!

So again, the dual of the dual is the primal

As with LP

$D \Longrightarrow$ $\Longleftarrow P$

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

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

$$= \langle C, X \rangle - \langle X, \Lambda \rangle$$

So duality gap is $\underbrace{\langle X, \Lambda \rangle}_{\geq 0}$ $\quad X \succeq 0$, $\Lambda \succeq 0$

as we already saw.

Weak duality: $d^* \leq p^*$

Unlike for LP, strong duality does not always hold

But if $\exists X \succ 0$ with $A_i \circ X = b_i$ (strictly primal feasible),

or $\exists \Lambda \succ 0$ with $\Lambda = C - \sum y_i A_i$ and $y \in \mathbb{R}^p$ (strictly dual feasible)

then Slater's condition holds (for primal or dual), so $d^* = p^*$.

SDP cont'd.

If strong duality holds with $p^* = d^*$ finite, then $\exists\, X^*, \Lambda^*$ primal & dual feasible with

$$\langle X^*, \Lambda^* \rangle = \operatorname{tr} X^* \Lambda^* = 0$$

As at the bottom of p. 8, if we let $\Lambda^* = G^T G$, $X^* = H^T H$, then

$$\operatorname{tr} X^* \Lambda^* = \operatorname{tr} G^T G H^T H = \| G H^T \|_F^2$$

so since this is zero, we have

$$G H^T = 0$$

so $X^* \Lambda^* = 0$, i.e. complementarity here means the MATRIX PRODUCT $X^* \Lambda^* = 0$. So $\Lambda^* X^* = 0$ also, i.e. X^*, Λ^* commute. It follows that they share a common orthogonal system of eigenvectors, i.e., $\exists\, Q$ with $Q^T Q = I$

such that

$$Q^T X^* Q = \operatorname{Diag}(\xi_i) = \begin{bmatrix} \xi_1 & & \\ & \ddots & \\ & & \xi_m \end{bmatrix}$$

$$Q^T \Lambda^* Q = \operatorname{Diag}(\lambda_i) = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_m \end{bmatrix}$$

with $\xi_i \lambda_i = 0$, $i = 1, \ldots, m$

EIGENVALUE COMPLEMENTARITY.

– not necessarily strict, i.e. both could be zero, but this is a "nongeneric" case.

4-8 MORE GENERALLY

The KKT equations + Complementary Slackness (aka Complementarity)

Let's return to the convex program

$$\min f_0(x) \text{ S.T. } f(x) \leq 0 \quad \begin{bmatrix} f_1 \\ \vdots \\ f_m \end{bmatrix} \text{ all convex}$$
$$\& \ Ax = b$$

If strong duality holds, so $d^* = p^*$, finite, then $\exists x^*, \lambda^*, \nu^*$ with $\lambda^* \geq 0$, $f(x^*) \leq 0$ and

$$f_0(x^*) = g(\lambda^*, \nu^*)$$
$$= \inf_x \Big(f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(\underbrace{x}_{\text{NOT } x^*}) + \nu^{*}(Ax - b) \Big)$$
$$\leq f_0(x^*) + \sum_{i=1}^m \underbrace{\lambda_i^* f_i(x^*)}_{\leq 0} + \nu^*(\underbrace{Ax^* - b}_{0})$$
$$\leq f_0(x^*).$$

in fact $=$, since $f_0(x^*)$ is on the left side.

$\therefore \lambda_i^* f_i(x^*) = 0, \quad i = 1, \ldots, m$

complementarity.

Nonzero Lagrange multiplier $\Rightarrow$ "active constraint"

"Inactive constraint" $\Rightarrow$ zero Lagrange mult.

But they could both be zero — strict compl. does not have to hold.

Suppose f_0, f_i are all differentiable as well as convex. Then, since x^* minimizes $L(x, \lambda^*, \nu^*)$ over x, we have

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

History: Lagrange for equality constraints
Karush 1939 unpublished M.S. thesis
Fritz John 1948 with Fritz John multiplier for f_0 — then don't need Slater.
Kuhn + Tucker 1951

So KKT, primal feasibility, dual feasibility and complementarity (so $d^* = p^*$) establishes primal + dual optimality.

KKT also extends to <u>nonconvex case</u> too, but in this case they are only NECESSARY conditions for optimality NOT SUFFICIENT. (BV p. 243)

e.g. $\min f_0(x) \equiv x^3$

$\nabla f_0(x) = 0 \not\Rightarrow x$ is optimal.

4-10.

Saddle Point Interpretation

Assume for simplicity there are no equality constraints.

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

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

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

while by def'n

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

Weak duality, $d^* \leq p^*$

does not actually depend on properties of L — we have

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

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!

Pf from Rockafellar 1970

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

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

4-11

For all $w \in W$, we have

$$\left\{ \begin{array}{l} h(w,z) \geq H(z) \quad \forall z \in Z \\ \text{so } \sup_{z \in Z} h(w,z) \geq \sup_{z \in Z} H(z) = \alpha \end{array} \right\}$$

This is true $\forall w \in W$, so

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

(easy proof once you see it!)

Biggest gap:

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

concave in w, convex in z

Then LHS $= -\infty$, $+\infty$.

Strong duality (LHS = RHS) occurs in 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 and von Neumann:

Player 1 chooses w, wants to $\inf$ h.

Player 2 chooses z, wants to $\sup$ h.

Generally, Player 1 wants Player 2 to "go first" to fix z. Player 1 knows that Player 2 will choose z to maximize the best that Player 1 can do; outcome is the the LHS.

But if strong duality holds, order of play makes no difference.