3-1. Lec 3

# DUALITY

$\tilde{x}$ SATISFYING $f(\tilde{x}) \le 0, h(\tilde{x}) =$ IS PRIMAL FEASIBLE

BV 5.1 The Lagrange Dual Function

We are interested in the optimization problem

"PRIMAL PROBLEM"

$$P: \inf f_0(x) \quad (x \in \mathbb{R}^n)$$

$$\text{(OR min)} \quad \text{S.T.} \quad \begin{array}{ll} f_i(x) \le 0 & i = 1, \ldots, m \\ h_i(x) = 0 & i = 1, \ldots, p \end{array}$$

origin of the name: Dantzig's father.

Assume domain

$$\mathcal{D} = \left(\bigcap_{i=0}^{m} \operatorname{dom} f_i\right) \cap \left(\bigcap_{i=1}^{p} \operatorname{dom} h_i\right) \neq \phi.$$

If $f_0, f_1, \ldots, f_m$ are convex functions AND $h_1, \ldots, h_p$ are affine functions we say P is a convex problem.

But we DO NOT assume this here.

The Lagrangian for P is

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

$$= f_0(x) + \lambda^T \underbrace{f(x)}_{f_1, \ldots, f_m} + \nu^T h(x)$$

$\lambda_i, \nu_i$ are called Lagrange multipliers.

The Lagrange dual function $g: \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R} \cup \{-\infty\}$

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \geq -\infty$$

pointwise inf of family of affine functions, so it is CONCAVE ($-g$ is convex).

3-2.

Let $p^*$ be the minimal value of P.

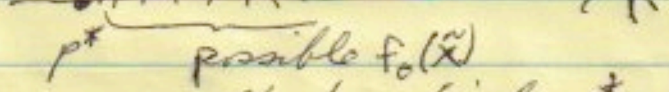
So, for all $\lambda \geq 0$ and all $\nu$ we have,

(primal) for any feasible point $\tilde{x} \in \mathbb{R}^n$ (so $f(\tilde{x}) \leq 0, h(\tilde{x}) = 0$)

$$L(\tilde{x}, \lambda, \nu) = f_0(\tilde{x}) + \underbrace{\lambda^T f(\tilde{x})}_{\leq 0} + \underbrace{\nu^T h(\tilde{x})}_{0}$$

so

$$g(\lambda, \nu) = \inf_{x \in D} L(x, \lambda, \nu) \leq L(\tilde{x}, \lambda, \nu) \leq f_0(\tilde{x})$$




Even if there is no $\tilde{x}$ for which $p^* = f_0(\tilde{x})$, we must have

$$g(\lambda, \nu) \leq p^*.$$

Since $g$ is concave, we write

$$\text{dom}\, g = \{(\lambda, \nu) : g(\lambda, \nu) > -\infty\}$$

and say $(\lambda, \nu) \in \text{dom}\, g$ is dual feasible

BV 5.2 The Lagrange Dual Problem (LDP):

maximize the lower bounds $g(\lambda, \nu)$ $(\lambda \geq 0)$

D: $\sup \; g(\lambda, \nu)$ (a max)

s.t. $\lambda \geq 0$.

This is a convex opt. prob, since $g$ is concave even if P was not.

e.g. LP (linear program) in "standard form"

$$\begin{aligned} &\min\ c^T x \\ &\text{s.t. } Ax = b \\ &\qquad x \geq 0 \end{aligned} \quad \boxed{A} \quad (\mathcal{D} = \mathbb{R}^n)$$

so $f_0(x) = c^T x$, $f(x) = -x$, $h(x) = Ax - b$

The Lagrangian is

$$\begin{aligned} L(x,\lambda,\nu) &= c^T x - \lambda^T x + \nu^T(Ax - b) \\ &= (c - \lambda + A^T \nu)^T x - b^T \nu \end{aligned}$$

The Lagrange dual function is

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

So the Lagrange dual problem is

$$\max_{\lambda \geq 0,\, \nu} g(\lambda,\nu)$$

$$\left.\begin{aligned} \equiv\ &\max\ -b^T\nu \\ &\text{s.t. } c - \lambda + A^T\nu = 0 \\ &\qquad\qquad \lambda \geq 0 \\ &\text{or equivalently} \\ &\min\ b^T\nu \\ &\text{s.t. } A^T\nu + c \geq 0 \end{aligned}\right\} \text{"D"}$$

ANOTHER LP
(could be converted to "standard form")

## Weak Duality

$p^*$ is optimal value of primal

Let $d^*$ be optimal value of (Lagrange) dual

Since $g(\lambda, \nu) \leq p^*$ $\forall \lambda \geq 0, \nu$, we have immediately that

$$d^* \leq p^*.$$

The (optimal) duality gap is $p^* - d^*$ $(\geq 0)$.

$$\xrightarrow{D} \; d^* \; \text{———} \; p^* \; \xleftarrow{P} \quad \mathbb{R}$$

## Strong Duality

Means $d^* = p^*$.

Does not always hold.

**Strong Duality Theorem** If P is convex (or $f_0, f_1, \ldots, f_m$ are convex and $h_1, \ldots, h_p$ are affine) and Slater's condition holds, i.e. $\boxed{h(x) = Ax - b}$

also called a "constraint qualification" → $\exists \; \tilde{x} \in \operatorname{relint} \mathcal{D}$ such that

$$f_i(\tilde{x}) < 0, \; i = 1 \ldots m, \text{ and } A\tilde{x} = b,$$

STRICT

then $d^* = p^*$.

REFINEMENT (WEAKER CONDITION)

Can allow $f_i(x) = 0$ if $f_i$ is affine inequality.

So, for LP, Slater's condition reduces to → $\mathcal{D} = \mathbb{R}^n$.

<u>primal feasibility.</u> [If P is unbounded below, then $p^* = -\infty$, so $d^* = -\infty$ ∴ D is infeasible. If D is unbounded above, then $d^* = \infty$, so $p^* = \infty$ ∴ P is infeasible. But they could both be infeasible, in which case $d^* = -\infty < p^* = \infty$.

To prove the theorem we introduce some geometry.

IN GENERAL (without assuming convexity or Slater)

Let $\mathcal{G} = \{ (f_1(x), \ldots, f_m(x), h_1(x), \ldots, h_p(x), f_0(x)) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} : x \in \mathcal{D} \}$

Clearly, $p^* = \inf \{ t : (u, v, t) \in \mathcal{G}, u \leq 0, v = 0 \}$

and $g(\lambda, \nu) = \inf \{ (\lambda, \nu, 1)^T (u, v, t) : (u, v, t) \in \mathcal{G} \}$

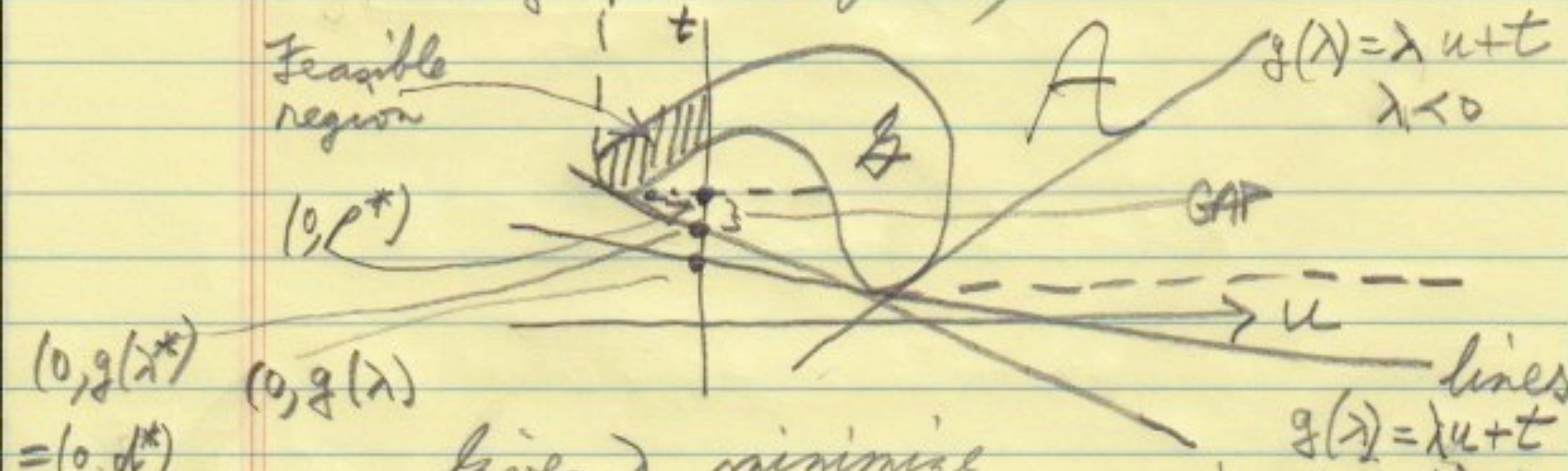
Now let

$$A = \mathcal{G} + (\mathbb{R}^m_+ \times \{0\} \times \mathbb{R}_+) \subseteq \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}$$

$$\doteq \left\{ (u, v, t) : \exists x \in \mathcal{D} \text{ s.t. } \begin{array}{l} f_i(x) \leq u_i, \ i = 1, \ldots, m \\ h_i(x) = v_i, \ i = 1, \ldots, p \\ f_0(x) \leq t \end{array} \right\}$$

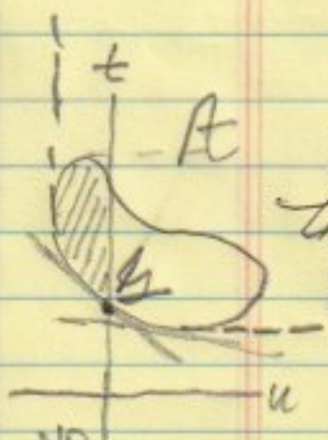
We can replace $\mathcal{G}$ by $A$ in the def'n of $g(\lambda, \nu)$ <u>if</u> $\lambda \geq 0$.

A nonconvex example
with just one inequality constraint.




Given $\lambda$, minimize
$(\lambda,1)^T(u,t)$ over
$\mathcal{A}=\{(u,t): \exists\, x\in\mathcal{D}$ with $f_0(x)\le t$,
$f_1(x)\le u\}$
to give $g(\lambda)$.

Lines $g(\lambda)=\lambda u+t$ i.e. $t=g(\lambda)-\lambda u$
are supporting hyperplanes for $\mathcal{G}$ $\forall\lambda$
and " " " $\mathcal{A}$ $\forall\lambda\geq 0$




If P is a convex problem ($f_i$ all convex, $\underline{h_i \text{ affine}}$ $Ax=b$)
then $\mathcal{A}$ is a convex set. Pf: HW
But $\mathcal{G}$ might not be. $p\,\boxed{\phantom{xx}}_{\,n}$

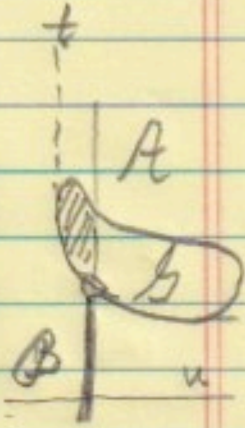
Proof of Strong Duality Thm
with slightly stronger hypotheses:
instead of assuming $\exists\,\tilde{x}\in \operatorname{relint}\mathcal{D}$
with $f_i(\tilde{x})<0$ and $A\tilde{x}=b$
assume $\tilde{x}\in \operatorname{int}\mathcal{D}$. (e.g., holds if $\mathcal{D}=\mathbb{R}^n$)
<u>and</u> assume $A$ has full rank:
rank $A=p$ $\quad\overset{m}{\boxed{\phantom{xxxx}}}\,p$

3-7

Since P is a convex problem, $\mathcal{A}$ is a convex set.

Let $\mathcal{B} = \{(0,0,s) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} : s < p^*\}$

We claim $\mathcal{A} \cap \mathcal{B} = \emptyset$. Suppose not, then $\exists (u,v,t) \in \mathcal{A} \cap \mathcal{B}$. Since $(u,v,t) \in \mathcal{A}$, $\exists x$ with $f_i(x) \le 0$, $i=1,\dots,m$, $Ax-b=0$, and $f_0(x) \le t < p^*$ (since $(u,v,t) \in \mathcal{B}$ as well). Contradiction to def. of $p^*$.




Now apply the separating hyperplane thm:

$\exists (\tilde{\lambda}, \tilde{\nu}, \tilde{\mu}) \neq 0$ and $\alpha$ s.t.

(BV write $\mu$, not $\tilde{\mu}$)

(1) $(u,v,t) \in \mathcal{A} \Longrightarrow \tilde{\lambda}^T u + \tilde{\nu}^T v + \tilde{\mu} t \ge \alpha$

(2) $(u,v,t) \in \mathcal{B} \Longrightarrow \tilde{\lambda}^T u + \tilde{\nu}^T v + \tilde{\mu} t \le \alpha$

From (1), we see $\tilde{\lambda} \ge 0$ and $\tilde{\mu} \ge 0$. [Otherwise, $\tilde{\lambda}^T u + \tilde{\mu} t$ is unbounded below over $\mathcal{A}$: just let $u_i$ or $t \to \infty$ corresp to $\tilde{\lambda}_i < 0$ or $\tilde{\mu} < 0$, which violates (1).]

(2) says $\tilde{\mu} t \le \alpha$ $\forall t < p^*$. So (1)+(2) gives,

$\forall x \in \mathcal{D}$, (3) $\tilde{\lambda}^T f(x) + \tilde{\nu}^T (Ax-b) + \tilde{\mu} f_0(x) \ge \alpha \ge \tilde{\mu} p^*$

(taking $u = f(x)$, $v = Ax-b$, $t = f_0(x)$, so $(u,v,t) \in \mathcal{A}$).

Case 1 $\tilde{\mu} > 0$. Divide (3) by $\tilde{\mu}$ to get

$$L\left(x, \frac{\tilde{\lambda}}{\tilde{\mu}}, \frac{\tilde{\nu}}{\tilde{\mu}}\right) \ge p^*, \quad \forall x \in \mathcal{D}.$$

"FRITZ JOHN MULTIPLIER"

Minimize over $x$ to get

$$g(\hat{\lambda}, \hat{\nu}) \ge p^* \quad \text{with } \hat{\lambda} = \frac{\tilde{\lambda}}{\tilde{\mu}}, \hat{\nu} = \frac{\tilde{\nu}}{\tilde{\mu}}$$

By weak duality, $g(\hat{\lambda}, \hat{\nu}) \le p^*$, so

$$d^* = \sup_{\lambda \ge 0} g(\lambda, \nu) = g(\hat{\lambda}, \hat{\nu}) = p^*$$

i.e. strong duality holds. contd.

3-8.

<u>Case 2</u> $\tilde{\mu} = 0$. From (3), we have

$$\forall x \in \mathcal{D}, \quad \tilde{\lambda}^T f(x) + \tilde{\nu}^T (Ax - b) \geq 0 \quad (4)$$

Apply this to the Slater point $\tilde{x}$
(for which $f(\tilde{x}) < 0$ and $A\tilde{x} = b$), so

$$\tilde{\lambda}^T f(\tilde{x}) \geq 0$$

hence $\tilde{\lambda} = 0$ (as $\tilde{\lambda} \geq 0$, $f(\tilde{x}) < 0$).

Since $(\tilde{\lambda}, \tilde{\nu}, \tilde{\mu}) \neq 0$, we must have $\tilde{\nu} \neq 0$.

Then from (4), $\forall x \in \mathcal{D}$

$$\tilde{\nu}^T (Ax - b) \geq 0$$

i.e. $x^T (A^T \tilde{\nu}) - b^T \tilde{\nu} \geq 0 \quad \forall x \in \mathcal{D}$.

But $\tilde{x}$ satisfies $\tilde{x}^T (A^T \tilde{\nu}) - b^T \tilde{\nu} = 0$

and since $\tilde{x} \in \operatorname{int} \mathcal{D}$, can perturb $\tilde{x} \to \hat{x}$

so $\hat{x}^T (A^T \tilde{\nu}) - b^T \tilde{\nu} < 0$

unless $A^T \tilde{\nu} = 0$ — but that contradicts our assumption that the rows of $A$ (columns of $A^T$) are linearly independent.

QED.