S.2.4. EXAMPLES (of DUAL PROGRAMS)

LEAST SQUARES

$$\left\{\begin{array}{ll} \text{minimize:} & x^T x \\ \text{subject to:} & Ax = b \end{array}\right\}$$

THE DUAL IS

$$\text{maximize} \quad \left(-\tfrac{1}{4}\right) v^T A A^T v - b^T v,$$

SLATER'S CONDITION: PRIMAL IS FEASIBLE, IE., $b \in R(A)$

LINEAR PROGRAMS

SLATER'S CONDITION: PRIMAL IS FEASIBLE. BUT:
DUAL OF DUAL IS ORIGINAL PROBLEM. SO IF DUAL IS FEASIBLE, THEN ITS DUAL IS ALSO FEASIBLE $\Rightarrow$

THEOREM: 1) IF EITHER PROBLEM IS FEASIBLE, SO IS THE OTHER, AND OBJECTIVES ARE THE SAME
(WEINBERGER pg 89)
2) IF ONE PROBLEM HAS UNBOUNDED OBJECTIVE, THE OTHER ONE IS INFEASIBLE.
3) BOTH PROBLEMS MAY BE INFEASIBLE

(DIVERSION: WE SHOW THAT DUAL OF DUAL IS ORIGINAL PROBLEM)

WE HAVE SEEN THAT

$$\left\{\begin{array}{ll} \text{minimize:} & c^T x \\ \text{subject to:} & Ax = b \\ & x \geq 0 \end{array}\right\} \quad (1)$$

HAS DUAL

$$\left\{\begin{array}{ll} \text{maximize:} & -b^T v \\ \text{subject to:} & A^T v + c \geq 0 \end{array}\right\} \quad (2)$$

WE WILL CALCULATE THE DUAL OF THE DUAL, (2)

$$(2) \Leftrightarrow \left\{\begin{array}{ll} \text{minimize} & b^T v \\ \text{subject to} & -A^T v - c \leq 0 \end{array}\right\} \Leftrightarrow$$

$$\left\{\begin{array}{ll} \text{minimize} & b^T x \\ \text{subject to} & -A^T x - c \leq 0 \end{array}\right\} \quad (3)$$

LAGRANGIAN IS

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

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

$$g(\lambda) = \inf L(x,\lambda) = \begin{cases} -\infty, & b^T - \lambda^T A^T \neq 0 \\ -\lambda^T c, & b^T - \lambda^T A^T = 0 \end{cases}$$

SO THE DUAL IS

$$\left\{\begin{array}{ll} \text{maximize} & -\lambda^T c \\ \text{subject to} & b^T - \lambda^T A^T = 0 \\ & \lambda \geq 0 \end{array}\right\} \Leftrightarrow$$

$$\left\{\begin{array}{ll} \text{minimize} & \lambda^T c \\ \text{subject to} & b - A\lambda = 0, \\ & \lambda \geq 0 \end{array}\right\} \Leftrightarrow \left\{\begin{array}{ll} \text{minimize} & c^T\lambda \\ \text{subject to} & A\lambda = b \\ & \lambda \geq 0 \end{array}\right\}$$

$$\Leftrightarrow \left\{\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array}\right\}$$

LAGRANGE DUAL OF QCQP

$$\left\{\begin{array}{lll} \text{minimize:} & \frac{1}{2}x^T P_0 x + q_0^T x + r_0 & \\ \text{subject to:} & \frac{1}{2}x^T P_i x + q_i^T x + r_i & i = 1, \ldots, m, \end{array}\right\}$$

$P_0 \in S^n_{++}$, $P_i \in S^n_+$, $i = 1, \ldots, m$

$$L(x,\lambda) = \frac{1}{2}x^T \overbrace{\left[P_0 + \lambda_1 P_1 + \lambda_2 P_2 + \ldots + \lambda_m P_m\right]}^{P(\lambda)} x$$

$$+ \underbrace{\left[q_0^T + \lambda_1 q_1^T + \lambda_2 q_2^T + \ldots + \lambda_m q_m^T\right]}_{q(\lambda)} x + \underbrace{\left[r_0 + \lambda_1 r_1 + \ldots + \lambda_m r_m\right]}_{r(\lambda)}$$

TO FIND $g(\lambda)$ IT IS TOUGH, BUT WHEN $\lambda \geq 0$, WE HAVE A POSITIVE DEFINITE $P(\lambda)$, SO WE CAN MINIMIZE BY TAKING THE GRADIENT TO BE ZERO $\Rightarrow$

$$g(\lambda) = \inf_x L(x,\lambda) = -\left(\frac{1}{2}\right) q^T(\lambda)\, P^{-1}(\lambda)\, q(\lambda) + r(\lambda)$$

$$\left(\nabla \frac{1}{2} x^T P x = Px, \quad \nabla q^T x = q \right)$$

THEREFORE, DUAL IS
$$\begin{cases} \text{maximize} & -\frac{1}{2} q^T(\lambda) P^{-1}(\lambda) q(\lambda) + r(\lambda) \\ \text{subject to} & \lambda \succeq 0 \end{cases}$$

SLATER'S CONDITION: $\exists x$:

$$\left(\frac{1}{2}\right) x^T P_i x + 2 q_i^T x + r_i < 0 \quad \forall i = 1, \dots, m$$

<u>ENTROPY MAXIMIZATION</u>:
$$\begin{cases} \text{minimize:} & \sum_{i=1}^{n} x_i \log x_i \\ \text{subject to:} & Ax \preceq b \\ & 1^T x = 1 \end{cases}$$

DUAL IS
$$\begin{cases} \text{maximize:} & -b^T\lambda - \nu - e^{-\nu-1} \sum_{i=1}^{n} e^{-a_i^T \lambda} \\ \text{subject to:} & \lambda \succeq 0 \end{cases}$$

SLATER'S CONDITION: $\underline{\underline{\exists x \succ 0}}$ WITH $Ax \preceq b$, $1^T x = 1$

(NOT $\succeq 0$)

OBSERVE THAT ν IS ONE-DIMENSIONAL AND $\nu \in \mathbb{R}$, THEREFORE TAKING DERIVATIVES:

$$0 = -1 + \left(\sum\right) e^{-\nu-1} \iff$$

$$\left(\sum\right) e^{-\nu-1} = 1 \iff -\nu - 1 = \log\left(\frac{1}{\sum}\right) \iff$$

$$\nu = \log\left[\sum_{i=1}^{n} e^{-a_i^T \lambda}\right] - 1$$

5.2.5 MATRIX GAME

ZERO-SUM MATRIX GAME: 1) PLAYER 1 MAKES A CHOICE $k \in \{1, \dots, n\}$, 2) PLAYER 2 MAKES A CHOICE $l \in \{1, \dots, m\}$

3) PLAYER 1 MAKES A PAYMENT P_{kl} TO PLAYER 2

$P \in \mathbb{R}^{n \times m}$ IS PAYOFF MATRIX.

4) PLAYER 1 MAKES CHOICE i WITH PROBABILITY u_i

5) PLAYER 2 MAKES CHOICE i WITH PROBABILITY v_i

6) CHOICES ARE INDEPENDENT.

$$\Rightarrow \quad \text{EXPECTED PAYOFF} \quad \sum_{k=1}^{n} \sum_{l=1}^{m} u_k v_l P_{kl}$$

FIRST ASSUMPTION: PLAYER 2 KNOWS STRATEGY u OF PLAYER #1

THEN PLAYER 2 WANTS TO MAXIMIZE $u^T P v$, AND EXPECTED PAYOFF IS

$$\sup \{ u^T P v \mid v \succeq 0, \ \mathbf{1}^T v = 1 \} = \max_{i=1,\dots,m} (P^T u)_i$$

THEREFORE, IT MAKES SENSE FOR PLAYER #1 TO SOLVE

$$\begin{pmatrix} \text{OPTIMAL} \\ \text{VALUE} \\ p_1^* \end{pmatrix} \quad \left\{ \begin{array}{ll} \text{minimize} & \max_{i=1,\dots,m} (P^T u)_i \\ \text{subject to:} & u \succeq 0, \ \mathbf{1}^T u = 1 \end{array} \right\} \ 1 \quad \text{(CONVEX)}$$

SECOND ASSUMPTION: PLAYER 1 KNOWS THE STRATEGY v OF PLAYER 2

SO PLAYER 1 SELECTS u TO ACHIEVE

$$\inf \{ u^T P v \mid u \succeq 0, \ \mathbf{1}^T u = 1 \} = \min_{i=1,\dots,n} (Pv)_i$$

SO PLAYER 2 WILL SELECT v ACCORDING TO

$$\begin{pmatrix} \text{OPTIMAL} \\ \text{VALUE} \\ p_2^* \end{pmatrix} \quad \left\{ \begin{array}{ll} \text{maximize} & \min_{i=1,\dots,n} (Pv)_i \\ \text{subject to} & v \succeq 0, \ \mathbf{1}^T v = 1 \end{array} \right\} \ 2 \quad \text{(ALSO CONVEX)}$$

WE EXPECT THAT $p_1^* \geq p_2^*$, BECAUSE KNOWING THE STRATEGY OF THE OPPONENT CANNOT HURT. THIS CAN ALSO BE SHOWN MATHEMATICALLY, INDEED (BOYD EX 5.24),

$$\boxed{\sup_{z \in Z} \inf_{w \in W} f(w,z) \leq \inf_{w \in W} \sup_{z \in Z} f(w,z)} \qquad W \subseteq \mathbb{R}^n, \quad Z \subseteq \mathbb{R}^m$$

SO THE ABOVE EXAMPLE CAN BE GENERALIZED!

<u>Proof</u>: IF BOTH Z, W ARE EMPTY, THEN INEQUALITY BECOMES $-\infty \leq \infty$, BUT THIS IS A TECHNICALITY.

LET W NOT EMPTY, WITH $\tilde{w} \in W$. THEN

$$\inf_{w \in W} f(w,z) \leq f(\tilde{w}, z) \Rightarrow$$

$$\sup_{z \in Z} \inf_{w \in W} f(w,z) \leq \sup_{z \in Z} f(\tilde{w}, z),$$ WHICH HOLDS

FOR ALL $\tilde{w}$, SO IT WILL ALSO HOLD IF WE TAKE THE MEAN.

THEN $p_1^* \geq p_2^*$ COMES BY NOTING THAT

$$p_1^* = \inf_u \sup_v \{u^T P v\}, \quad p_2^* = \sup_v \inf_u \{u^T P v\}$$

BUT, IN THIS CASE, WE ACTUALLY HAVE $p_1^* = p_2^*$

<u>Proof</u>: FIRST, WE NOTE THAT (1) IS LINEAR PROGRAM:

$$\left\{ \begin{array}{ll} \text{minimize} & t \\ \text{subject to} & u \succeq 0, \ \mathbb{1}^T u = 1 \\ & P^T u \preceq t \mathbb{1} \end{array} \right\}$$

THE LAGRANGIAN IS

$$t + \lambda^T\left[P^T u - t\mathbb{1}\right] - \mu^T u + \nu\left[1 - \mathbb{1}^T u\right] =$$

$$\nu + (1 - \mathbb{1}^T\lambda)t + (P\lambda - \nu\mathbb{1} - \mu)^T u$$ AND TAKING

THE MINIMUM OVER u, t, WE ARRIVE AT

$$g(\lambda,\mu,\nu) = \begin{cases} \nu, & \mathbb{1}^T\lambda = 1,\ P\lambda - \nu\mathbb{1} - \mu = 0 \\ -\infty, & \text{OTHERWISE.} \end{cases}$$

SO DUAL PROBLEM IS:

$$\left\{\begin{array}{ll} \text{maximize} & \nu \\ \text{subject to:} & \mathbb{1}^T\lambda = 1,\quad P\lambda - \nu\mathbb{1} = \mu,\ \lambda \succeq 0,\ \mu \succeq 0 \end{array}\right\}$$

$$\Leftrightarrow \left\{\begin{array}{ll} \text{maximize} & \nu \\ \text{subject to:} & \lambda \succeq 0,\ \mathbb{1}^T\lambda = 1,\quad P\lambda \succeq \nu\mathbb{1} \end{array}\right\},$$

WHICH IS EQUIVALENT TO (2). BY STRONG DUALITY (1) AND (2) HAVE SAME VALUES FOR THE OPTIMA.

5.3 GEOMETRIC INTERPRETATION

LET THE SET

$$\mathcal{G} = \left\{(f_1(x),\dots,f_m(x), h_1(x),\dots,h_p(x), f_0(x)) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} \mid x \in D\right\}$$

THEN $p^* = \inf\{t \mid (u,v,t) \in \mathcal{G},\ u \preceq 0,\ v = 0\}$

THE LAGRANGIAN IS

$$(\lambda, V, 1)^T(u,v,t) = \sum_{i=1}^m \lambda_i u_i + \sum_{i=1}^p V_i v_i + t$$

ΠΡΟΣΟΧΗ: $\overset{\text{ΝΙ ΚΕΦΑΛΑΙΟ}}{V} \neq \nu \leftarrow$ ΛΑΤΙΝΙΚΟ

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

IF THE INFIMUM IS FINITE, THEN THE INEQUALITY

$(\lambda,\nu,1)^T(u,v,t) \geq g(\lambda,\nu)$ IS A NON-VERTICAL

SUPPORTING PLANE OF $\mathcal{G}$.

IN THIS INTERPRETATION, p^* CORRESPONDS TO A SPECIAL POINT IN $\mathcal{G}$:

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

2) $g(\lambda,\nu)$ DESCRIBES THE SUPPORTING HYPERPLANE CORRESPONDING TO $(\lambda,\nu,1)$.

3) WEAK DUALITY MEANS THAT: ASSUMING $\lambda \succeq 0$

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

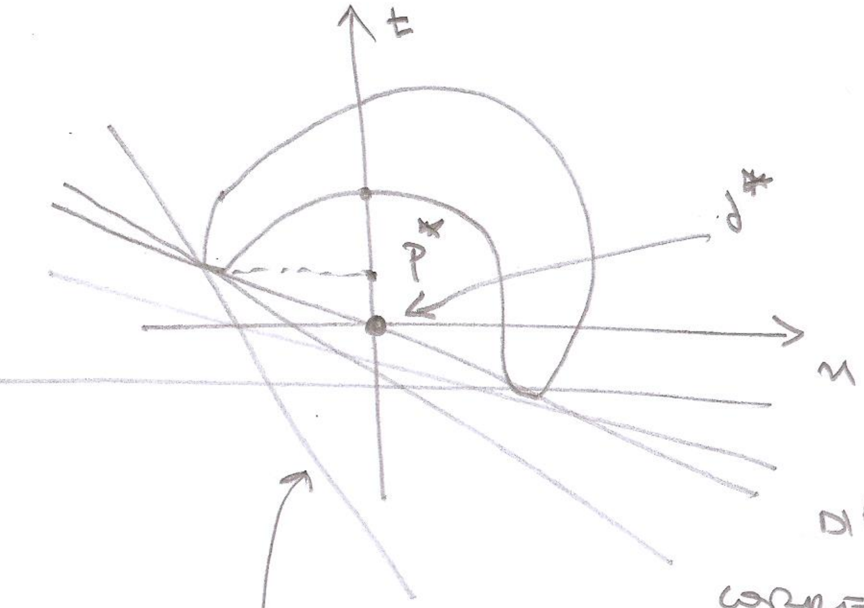
$$\geq \inf\{(\lambda,\nu,1)^T(u,v,t) \mid (u,v,t) \in \mathcal{G},\ u \leq 0,\ v = 0\}$$

$$\left(t \geq \underbrace{\lambda u}_{\leq 0} + \underbrace{\nu v}_{= 0} + t \right)$$

$$\geq \inf\{(\lambda,\nu,1)^T(u,v,t) \mid (u,v,t) \in \mathcal{G}\}$$

$$= g(\lambda,\nu).$$

WE WILL SEE A GRAPHICAL INTERPRETATION WHEN THERE IS A SINGLE INEQUALITY CONSTRAINT:



DIFFERENT HYPERPLANES CORRESPONDING TO DIFFERENT λ

$$(\lambda, 1)^T (u, t) = g(\lambda) \Leftrightarrow \lambda u + t = g(\lambda).$$

SO $u = 0 \Rightarrow t = g(\lambda)$

EPIGRAPH VARIATION

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

$$= \{ (u, v, t) \mid \exists x \in \mathcal{D},\ f_i(x) \le u_i,\ i = 1, \dots, m,\ h_i(x) = v_i,\ i = 1, \dots, p,\ f_0(x) \le t \}$$

IT IS A KIND OF EPIGRAPH OF $\mathcal{G}$, AS IT IS $\mathcal{G}$ PLUS ALL POINTS THAT ARE "WORSE" THAN POINTS IN $\mathcal{G}$.

ALSO: $\mathcal{A}$ IS CONVEX IF ORIGINAL OPTIMIZATION PROBLEM IS CONVEX! INDEED, LET (u_1, v_1, t_1) AND $(u_2, v_2, t_2) \in \mathcal{A}$

THEN $\left\{ \exists x_1 : \begin{array}{l} f_i(x_1) \le u_{1i} \quad i = 1, \dots, m \\ h_i(x_1) = A_i x_1 - b_i = v_{1i}, \ i = 1, \dots, p \\ f_0(x_1) \le t_1 \end{array} \right\}$

LIKEWISE, $\left\{ \exists x_2 : \begin{array}{l} f_i(x_2) \le u_{2i}, \quad i=1,\dots,m \\ h_i(x_2) = A x_2 - b = v_{2i}, \quad i=1,\dots,p \\ f_0(x_2) \le t_2. \end{array} \right\}$

CONSIDER $\quad \theta x_1 + (1-\theta) x_2, \quad \theta \in [0,1]$. THEN

1) $f_i(\theta x_1 + (1-\theta) x_2) \le \theta f_i(x_1) + (1-\theta) f_i(x_2)$

$\le \theta u_{1i} + (1-\theta) u_{2i}$

2) $h_i(\theta x_1 + (1-\theta) x_2) = A_i[\theta x_1 + (1-\theta) x_2] - b_i$

$= \theta [A_i x_1 - b] + (1-\theta) [A_i x_2 - b_i]$

$= \theta h_i(x_1) + (1-\theta) h(x_2) = \theta v_{1i} + (1-\theta) v_{2i}$

3) $f_0(\theta x_1 + (1-\theta) x_2) \le \theta f_0(x_1) + (1-\theta) f_0(x_2)$

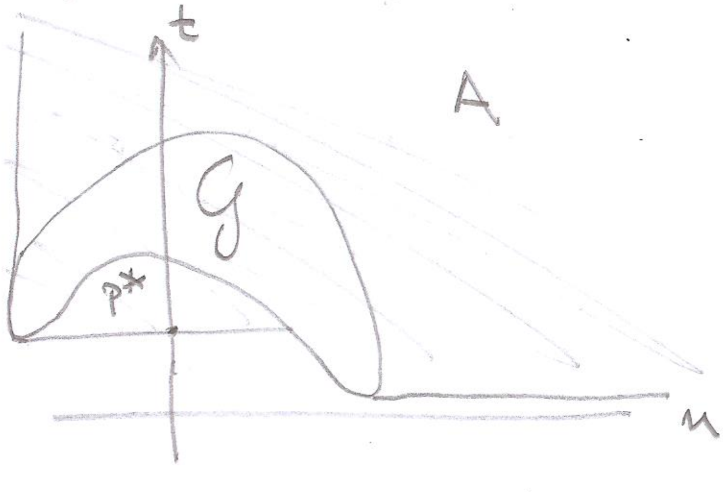
$\le \theta t_1 + (1-\theta) t_2$

$\therefore$ THEREFORE, $\theta x_1 + (1-\theta) x_2$ IS PROOF THAT

$\theta (u_1, v_1, t_1) + (1-\theta)(u_2, v_2, t_2)$ ALSO

BELONGS TO $\mathcal{A}$, SO $\mathcal{A}$ IS CONVEX.

A GRAPHICAL INTERPRETATION OF THE PREVIOUS CASE:



IN THIS INTERPRETATION,

1) $$p^* = \inf\{ t \mid (0,0,t) \in \mathcal{A} \}.$$

2) IF $\lambda \geq 0$,

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

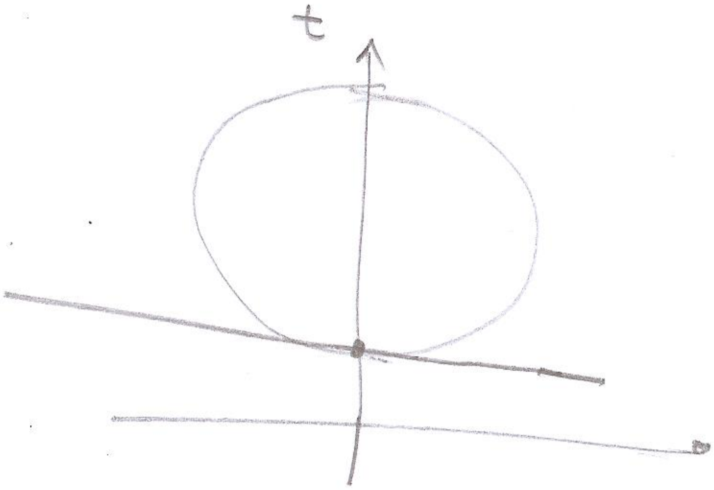
($\mathcal{A}$ IS LARGER THAN $\mathcal{G}$, BUT INF DOES NOT CHANGE)

AND IF IT EXISTS, IT <u>ALSO</u> DEFINES A SUPPORTING HYPERPLANE of $\mathcal{A}$. 3) SINCE $p^* \in \mathrm{bd}\,\mathcal{A}$, IT FOLLOWS THAT FOR p^* AS WELL,

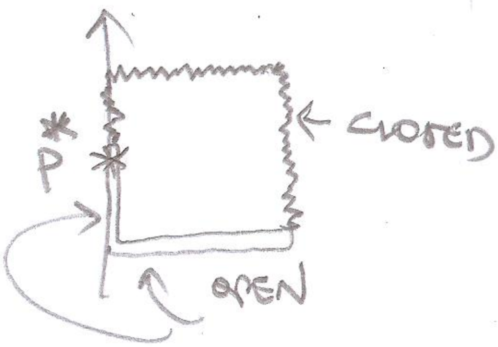
$$p^* = (\lambda,\nu,1)^T (0,0,p^*) \geq g(\lambda,\nu),$$

SO WE HAVE WEAK DUALITY.

STRONG DUALITY MEANS THAT $\mathcal{A}$ LOOKS LIKE THIS:



AND NOT LIKE THIS:



SLATER'S CONDITION ENSURES THAT SITUATION IS LIKE ON THE LEFT, AND NOT THE RIGHT

5.4 SADDLE POINT INTERPRETATION of $L(x^*, \lambda^*, \nu^*)$

NOTE THAT

$$\sup_{\lambda \geq 0} L(x, \lambda, \nu) =$$

$$\sup_{\lambda \geq 0, \nu} \left\{ f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right\}$$

$$= \begin{cases} \infty, & \exists i : h_i(x) \neq 0, \ \exists i : f_i(x) > 0 \\ f_0(x), & \forall i \ h_i(x) = 0, \ f_i(x) \leq 0. \end{cases} \Rightarrow$$

$$\inf_x \sup_{\lambda \geq 0} L(x, \lambda, \nu) = p^*$$

WE ALSO HAVE $\quad d^* = \sup_{\lambda \geq 0} \inf_x L(x, \lambda, \nu)$

THEREFORE:

WEAK DUALITY: $\sup_{\lambda \geq 0} \inf_x L(x, \lambda, \nu) \leq \inf_x \sup_{\lambda \geq 0} L(x, \lambda, \nu)$

STRONG DUALITY: $\sup_{\lambda \geq 0} \inf_x L(x, \lambda, \nu) = \inf_x \sup_{\lambda \geq 0} L(x, \lambda, \nu)$

THESE ARE RELATED TO

MAX-MIN EQUALITY: $\sup_{z \in Z} \inf_{w \in W} f(w, z) \leq \inf_{w \in W} \sup_{z \in Z} f(w, z)$

(ALWAYS HOLDS, WE SHOWED PROOF)

STRONG MAX-MIN-PROPERTY:

$$\sup_{z \in Z} \inf_{w \in W} f(w, z) = \inf_{w \in W} \sup_{z \in Z} f(w, z)$$

(NOT ALWAYS HOLDS)

DEFINITION: A PAIR $\tilde{w} \in W$, $\tilde{z} \in Z$ IS A SADDLE-POINT for f, W, Z, IF $\forall w \in W, \forall z \in Z$,

$$f(\tilde{w}, z) \le f(\tilde{w}, \tilde{z}) \le f(w, \tilde{z}), \quad \text{I.E.}$$

$$1)\ f(\tilde{w}, \tilde{z}) = \inf_{w \in W} f(w, \tilde{z}), \qquad 2)\ f(\tilde{w}, \tilde{z}) = \sup_{z \in Z} f(\tilde{w}, z)$$

PROPERTY: IF THERE IS A SADDLE POINT, THEN THE STRONG MAX-MIN PROPERTY HOLDS FOR f.

PROOF:

$$\sup_{z \in Z} \inf_{w \in W} f(w, z) \ge \inf_{w \in W} f(w, \tilde{z}) \overset{\text{(ASSUMPTION)}}{=} \sup_{z \in Z} f(\tilde{w}, z) \ge \inf_{w \in W} \sup_{z \in Z} f(w, z)$$

AND RESULT FOLLOWS FROM MAX-MIN INEQUALITY.

PROPERTY: LET $x^*, (\lambda^*, \nu^*)$ BE PRIMAL/DUAL OPTIMAL AND STRONG DUALITY HOLDS. THEN THEY FORM A SADDLE POINT FOR THE LAGRANGIAN, AND $L(x^*, \lambda^*, \nu^*) = p^* = d^*$ for $\lambda \ge 0$

$$L(x^*, \lambda, \nu) \overset{1}{\le} p^* \overset{2}{=} L(x^*, \lambda^*, \nu^*) \overset{3}{\le} L(x, \lambda^*, \nu^*)$$

PROOFS OF 2, 3, WE SEE LATER.

PROOF OF 1 BY CONTRADICTION. LET

$$L(x^*, \lambda, \nu) > f_0(x^*) \Rightarrow$$

$$\cancel{f_0(x^*)} + \underbrace{\sum_{i=1}^{m} \lambda_i f_i(x^*)}_{\le 0} + \sum_{i=1}^{p} \nu_i \cancel{h_i(x^*)} > f_0(x^*) \Rightarrow$$

CONTRADICTION.

QED

PROPERTY: THE INVERSE HOLDS AS WELL: (LAGRANGE 1797) 27

IF $x^*, (\lambda^*, \nu^*)$ ARE A SADDLE POINT OF THE LAGRANGIAN OVER THE SETS $\lambda \geq 0$, $x \in \mathbb{R}^n$, THEN:

1) WE HAVE STRONG DUALITY WITH $p^* = d^* = L(x^*, \lambda^*, \nu^*)$

2) $f_0(x^*) = p^*$

3) $g(\lambda^*, \nu^*) = d^*$

PROOF OF 1)

THE ASSUMPTIONS SAYS: $\forall \lambda \geq 0, \nu, x,$

$$\boxed{L(x^*, \lambda, \nu) \underset{A}{\leq} L(x^*, \lambda^*, \nu^*) \underset{B}{\leq} L(x, \lambda^*, \nu^*)}$$

WORKING AS ABOVE,

$$\inf_x L(x, \lambda^*, \nu^*) = \sup_{\lambda \geq 0, \nu} L(x^*, \lambda, \nu) = L(x^*, \lambda^*, \nu^*)$$

$$\Rightarrow \sup_{\lambda \geq 0, \nu} \inf_x L(x, \lambda, \nu) \geq \inf_x \sup_{\lambda \geq 0, \nu} L(x, \lambda, \nu)$$

BUT WE KNOW INVERSE INEQUALITY, SO WE HAVE STRONG DUALITY.

PROOF OF 2)

FROM A $\quad L(x^*, \lambda, \nu) \leq L(x^*, \lambda^*, \nu^*)$

$$\Rightarrow f_0(x^*) + \sum_{i=1}^m \lambda_i f_i(x^*) + \sum_{i=1}^p \nu_i h_i(x^*)$$

$$\leq f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \sum_{i=1}^p \nu_i^* h_i(x^*) \quad \forall \lambda \geq 0, \forall \nu \in \mathbb{R}^p$$

A) FOR THE ABOVE TO HOLD $\forall \nu_i$, WE HAVE

$h_i(x^*) = 0$.

B) FOR THE ABOVE TO HOLD FOR ARBITRARILY LARGE λ_i, WE HAVE $f_i(x^*) \leq 0$.

C) FOR THE ABOVE TO HOLD FOR ARBITRARILY SMALL $\lambda_i \geq 0$, WE HAVE $\sum_{i=1}^{m} \lambda_i^* f_i(x^*) = 0$ (COMPLEMENTARY SLACKNESS!)

FROM (B):

$$L(x^*, \lambda^*, \nu^*) \leq L(x, \lambda^*, \nu^*)$$

$$\Rightarrow f_0(x^*) + \cancel{\sum_{i=1}^{m} \lambda_i^* f_i(x_i^*) + \sum_{i=1}^{p} \nu_i h_i(x^*)}$$

$$\leq f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

SO IF x ALSO FEASIBLE, $f_0(x^*) \leq f_0(x)$.

PROOF OF 3)

$$g(\lambda^*, \nu^*) = \inf_x L(x, \lambda^*, \nu^*) = L(x^*, \lambda^*, \nu^*) = d^*$$

5.4.4 PRICE/TAX INTERPRETATION

ORIGINAL PROBLEM: minimize $f_0(x)$

subject to: $f_i(x) \leq 0 \quad i=1,\dots,m$

LET US PAY FOR VIOLATIONS WITH PRICE λ_i PER UNIT OF VIOLATION:

NEW OBJECTIVE: $f_0(x) + \sum \lambda_i f_i(x)$

STRONG DUALITY: THERE IS A SET OF SHADOW PRICES FOR WHICH VIOLATING THE CONSTRAINT AND PAYING OFFERS NO ADVANTAGE

5.5 OPTIMALITY CONDITIONS

OBSERVE THAT ANY DUAL FEASIBLE PAIR (λ, ν) IS A CERTIFICATE ABOUT THE LOWER BOUND $p^* \geq g(\lambda, \nu)$.

IT ALSO PROVIDES AN UPPER BOUND ON THE DISTANCE BETWEEN THE OPTIMAL VALUE AND A VALUE WE HAVE:

$$f_0(x) - p^* \leq \underbrace{f_0(x) - g(\lambda, \nu)}_{\epsilon} \Rightarrow x \text{ IS } \epsilon\text{-SUBOPTIMAL}$$

THIS ALSO WORKS INVERSELY: IF WE HAVE PRIMAL FEASIBLE x AND DUAL FEASIBLE (λ, ν), WE KNOW THAT

$$p^* \in [g(\lambda, \nu), f_0(x)], \quad d^* \in [g(\lambda, \nu), f_0(x)]$$

THIS ALLOWS US TO TERMINATE ITERATIVE ALGORITHMS WHICH PRODUCE $x^{(k)}, \lambda^{(k)}, \nu^{(k)}, k = 1, \ldots,$

5.5.2 COMPLEMENTARY SLACKNESS

LET x^*, λ^*, ν^* PRIMAL AND DUAL OPTIMAL AND STRONG DUALITY HOLDS

$$
\begin{aligned}
f_0(x^*) &= g(\lambda^*, \nu^*) \\
&= \inf_x \left\{ f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x) \right\} \\
&\leq f_0(x^*) + \sum_{i=1}^{m} \underbrace{\lambda_i^*}_{\geq 0} \underbrace{f_i(x^*)}_{\leq 0} + \sum_{i=1}^{p} \nu_i^* \underbrace{h_i(x^*)}_{= 0} \\
&\leq f_0(x^*)
\end{aligned}
$$

$\Rightarrow$ INEQUALITIES ARE ACTUALLY EQUALITIES $\Rightarrow$

TWO IMPORTANT CONCLUSIONS

1 $$\sum_{i=1}^{m} \lambda_i f_i(x^*) = 0 \Rightarrow$$

$$\boxed{\lambda_i^* f_i(x^*) = 0, \quad i=1,\dots,m}$$ (COMPLEMENTARY SLACKNESS)

(PUT DIFFERENTLY: $\lambda_i^* > 0 \Rightarrow f_i(x^*) = 0$

$f_i(x^*) < 0 \Rightarrow \lambda_i^* = 0$.)

2 x^* minimizes $L(x, \lambda^*, \nu^*)$ over x

3 $p^* = d^* = L(x^*, \lambda^*, \nu^*)$

5.5.3 KKT OPTIMALITY CONDITIONS

LET $f_0, \dots, f_m, \; h_1, \dots, h_p$ BE DIFFERENTIABLE ($\Rightarrow$ THEY HAVE OPEN DOMAINS) BUT NOT NECESSARILY CONVEX.

LET x^*, λ^*, ν^* PRIMAL/DUAL OPTIMA. BY 2 ABOVE:

$$\nabla_x L(x^*, \lambda^*, \nu^*) = 0 \Leftrightarrow$$

$$\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^{p} \nu_i^* \nabla h_i(x^*) = 0$$

$\uparrow$ DISCUSS GEOMETRIC INTUITION!

WRITING EVERYTHING WE KNOW TOGETHER,

$$\left\{\begin{array}{lll} f_i(x^*) \le 0, & i=1,\dots,m & (1) \\ h_i(x^*) = 0, & i=1,\dots,p & (2) \\ \lambda_i^* \ge 0, & i=1,\dots,m & (3) \\ \lambda_i^* f_i(x^*) = 0, & i=1,\dots,m & (4) \\ \multicolumn{2}{l}{\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^{p} \nu_i^* \nabla h_i(x^*) = 0} & (5) \end{array}\right\} \text{KKT CONDITIONS}$$

SO WE HAVE PROVED THE FOLLOWING:

THEOREM: FOR ANY OPTIMIZATION PROBLEM WITH DIFFERENTIABLE OBJECTIVES AND CONSTRAINT FUNCTIONS WHERE WE HAVE STRONG DUALITY, ANY SET x^*, λ^*, ν^* SATISFIES THE KKT CONDITIONS.

NOW ASSUME THAT THE PROBLEM IS CONVEX. THE KKT CONDITIONS BECOME SUFFICIENT, AS WE NOW SHOW:
LET x^*, λ^*, ν^* SATISFY THE KKT CONDITIONS
①, ③ $\Rightarrow$ x^* IS FEASIBLE. ③ $\Rightarrow$ $L(x, \lambda^*, \nu^*)$ IS CONVEX. ⑤ $\Rightarrow$ THE GRADIENT OF $L(x, \lambda^*, \nu^*)$ IS ZERO AT x^* $\Rightarrow$ x^* MINIMIZES $L(x, \lambda^*, \nu^*)$.
THEREFORE

$$g(\lambda^*, \nu^*) = L(x^*, \lambda^*, \nu^*)$$
$$= f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) + \sum_{i=1}^{p} \nu_i^* h_i(x^*) \overset{②,④}{=}$$
$$= f_0(x^*) \Rightarrow \lambda^*, \nu^*, x^* \text{ OPTIMAL.}$$

SO WE HAVE PROVED ANOTHER THEOREM:

THEOREM: PROBLEM IS CONVEX WITH DIFFERENTIABLE f_i, KKT SATISFIED FOR x^*, λ^*, ν^*
$\Rightarrow$ x^*, λ^*, ν^* ARE PRIMAL/DUAL OPTIMAL; DUALITY GAP IS ZERO.

WE ALSO HAVE ANOTHER THEOREM:

THEOREM: LET CONVEX OPTIMIZATION PROBLEM WITH DIFFERENTIABLE f_i THAT SATISFIES SLATER'S CONDITION. THEN

x IS OPTIMAL $\Leftrightarrow$ $\exists (\lambda, \nu)$: $x, (\lambda, \nu)$ SATISFY THE KKT CONDITIONS (EASY PROOF)

NOTE: SLATER'S CONDITION $\Rightarrow$ ① GAP = 0; ② DUAL IS ATTAINED FOR SOME λ^*, ν^*

WHY ARE KKT CONDITIONS IMPORTANT?

1) SOMETIMES WE CAN SOLVE THEM (BUT NOT OFTEN)

2) SOME ALGORITHMS TRY TO ITERATIVELY SOLVE KKT CONDITIONS

EXAMPLE #1 minimize: $\left(\frac{1}{2}\right)x^T P x + q^T x + r$

subject to: $Ax = b$, , $P \in S_+^n$, $A_{m \times n}$

KKT CONDITIONS: $\{Ax^* = b, \quad Px^* + q + A^T v^* = 0\}$

$$\Leftrightarrow \begin{array}{c} \\ n \{ \\ m \{ \end{array} \begin{array}{cc} \overbrace{\ }^{n} & \overbrace{\ }^{m} \\ \multicolumn{2}{c}{\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix}} \end{array} \begin{bmatrix} x^* \\ v^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}$$

SO WE NEED TO SOLVE A LINEAR SYSTEM, AND WE ARE DONE

EXAMPLE #2 WATER-FILLING

$$\left\{ \begin{array}{l} \text{minimize: } -\sum_{i=1}^{n} \log(\alpha_i + x_i) \\ \text{subject to: } x \succeq 0, \quad 1^T x = 1 \end{array} \right\}$$

PROBLEM ARISES IN TELECOMMUNICATIONS, AND IS TYPICAL OF A LARGER CLASS OF OPTIMIZATION PROBLEMS

KKT CONDITIONS: $x^* \succeq 0, \quad 1^T x^* = 1, \quad \lambda^* \succeq 0,$

$$\left\{ \begin{array}{l} \lambda^* x^* = 0, \quad i = 1, \ldots, n \\ -\dfrac{1}{\alpha_i + x_i^*} - \lambda_i^* + v^* = 0, \quad i = 1, 2, \ldots, n \end{array} \right\}$$

(REMOVING λ_i^*)

$$\Leftrightarrow \left\{ \begin{array}{l} x^* \succeq 0, \quad 1^T x^* = 1, \quad x_i^*\left(v^* - \dfrac{1}{\alpha_i + x_i^*}\right) = 0 \\ v^* \geq \dfrac{1}{\alpha_i + x_i^*}, \quad i = 1, \ldots, n \end{array} \right\}$$

FOR ALL POSITIVE x_i, THE DERIVATIVE MUST BE THE SAME

1) IF $\nu^* < \frac{1}{\alpha_i}$, BY LAST CONDITION $x_i^* > 0 \Rightarrow$

$$\nu^* = \frac{1}{\alpha_i + x_i^*} \Rightarrow x_i^* = \frac{1}{\nu^*} - \alpha_i$$

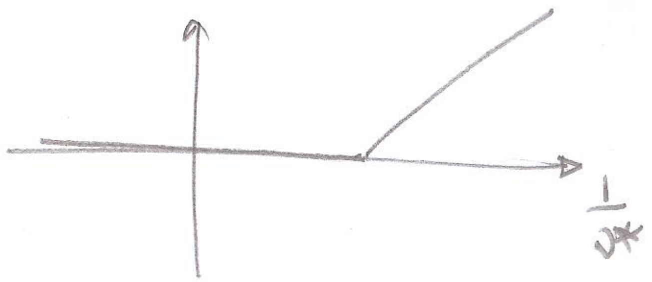
2) IF $\nu^* \geq \frac{1}{\alpha_i}$, THEN $x_i^* > 0$ IS NOT POSSIBLE, BECAUSE

THEN $\nu^* = \frac{1}{\alpha_i + x_i}$, WHICH CONTRADICTS $\nu^* \geq \frac{1}{\alpha_i}$

$$\Rightarrow \quad x_i^* = \begin{cases} \frac{1}{\nu^*} - \alpha_i, & \nu^* < \frac{1}{\alpha_i} \\ 0, & \nu^* > \frac{1}{\alpha_i} \end{cases}$$

$$\mathbb{1}^T x^* = 1 \Rightarrow \sum_{i=1}^{m} \max\left\{0, \frac{1}{\nu^*} - \alpha_i\right\} = 1$$

THIS IS A SUM OF FUNCTIONS OF THE FORM



SO IT IS INCREASING WITH RESPECT TO $\frac{1}{\nu^*}$ AND HITS 1 WITH POSITIVE RATE, SO IT HAS UNIQUE SOLUTION

WATER-FILLING INTERPRETATION