# 262B-Lecture 20

**Date created:** 2021.04.06
**N. of Pages:** 16

- FOC (geometric) : If $x_*$ is a Local min, then

$$\nabla f(x_*)^T \Delta x \geq 0 \qquad \forall \, \Delta x \in T(x_*) \longrightarrow \text{Tangent Cone}$$

(no descent direction in the tangent cone).

- Also, KKT Conditions (need constraint qualification)
(FOC algebraic) :

1 - $\nabla f(x_*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x_*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x_*) = 0$

2 - $\mu \geq 0$

3 - $\mu_j^* = 0$ if $j \notin A(x_*)$

Due to Farkas lemma, $\exists \, (\lambda_*, \mu_*)$ satisfying $\underbrace{1, 2, 3}$

all linear in $(\lambda, \mu)$

$\underbrace{\text{if and only if}}$

$$\nabla f(x_*)^T \Delta x \geq 0 \qquad \forall \, \Delta x \in V(x_*)$$

where :

$$V(x_*) = \left\{ \Delta x \;\middle|\; \begin{array}{ll} \nabla h_i(x_*)^T \Delta x = 0 & i = 1, \dots, m \\ \nabla g_j(x_*)^T \Delta x \leq 0 & j \in A(x_*) \end{array} \right\}$$

Cone of first-order directions

Thm: Consider a local min $x_*$. It satisfies KKT ($\exists\ \lambda_*, \mu_*$ satisfying 1, 2, 3) if and only if

$$\nabla f(x_*)^T \Delta x \geq 0 \qquad \forall\ \Delta x \in V(x_*)$$

$\rightarrow$ Constraint qualifications are to bridge the gap between $V(x_*)$ and $T(x_*)$.

---

Special case: $V(x_*) = T(x_*)$

Def: A point $\underline{y}$ is called quasi-regular if $V(y) = T(y)$. This is the best constraint qualification, but may be hard to check (regularity is easy to check).

---

Thm: $T(y) \subseteq V(y) \qquad \forall y$: feasible

Proof: pick $\Delta x \in T(y)$ and then show $\Delta x \in V(y)$

$$\Rightarrow \exists \underbrace{\{y^{(k)}\}}_{\text{feasible}} \to y \quad \text{s.t.} \quad \frac{\Delta x}{\|\Delta x\|} = \lim_{k\to\infty} \frac{y^{(k)}-y}{\|y^{(k)}-y\|}$$

$$\underset{\text{feasibility}}{0 =} h_i(y^{(k)}) = \underbrace{h_i(y)}_{=0} + \nabla h_i(\underbrace{\tilde{y}^{(k)}}_{\text{mean-value thm}})^T (y^{(k)}-y) \quad (1)$$

$j \in A(y)$:

$$\underset{\text{feasibility}}{0 \geq} g_j(y^{(k)}) = \underbrace{g_j(y)}_{\substack{=0 \\ \text{(active constraint)}}} + \nabla g_j(\bar{y}^{(k)})^T (y^{(k)}-y) \quad (2)$$

(1), (2) as $k \to \infty$: $\nabla h_i(y)^T \Delta x = 0$

$\nabla g_j(y)^T \Delta x \leq 0$

$\Leftarrow$

$\Delta x \in V(y)$

Duality:

Primal problem (P):

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & h_i(x) = 0 \quad i = 1, \dots, m \\ & g_j(x) \leq 0 \quad j = 1, \dots, r \end{aligned}$$

and $x \in X$

Lagrangian:

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{r} \mu_j g_j(x)$$

Define: $f_* = \inf_{\substack{x \in X \\ h(x) = 0 \\ g(x) \leq 0}} f(x)$

Thm: $\boxed{\inf_{x \in X} L(x, \lambda, \mu) \leq f_*}$ $\forall (\lambda, \mu)$ s.t. $\mu \geq 0$

proof: $\inf_{x \in X} L(x, \lambda, \mu) \leq \inf_{\substack{x \in X \\ h(x) = 0 \\ g(x) \leq 0}} L(x, \lambda, \mu) \leq \inf_{\substack{x \in X \\ h(x) = 0 \\ g(x) \leq 0}} f(x)$

smaller feasible set

$f_*$

(since $\lambda_i h_i(x) = 0$, $\underbrace{\mu_j}_{\geq 0} g_j(x) \leq 0$ over the set)

Dual function: $q(\lambda,\mu) = \inf_{x\in X} l(x, \underbrace{\lambda,\mu})$

So, $q(\lambda,\mu) \leq f_*$ $\forall (\lambda,\mu) : \mu \geq 0$

So, $q(\cdot,\cdot)$ provides a lower bound on the primal problem. If $q(\lambda,\mu) = -\infty$, then the lower bound is useless.

Domain of dual function:

$$D = \{(\lambda,\mu) \mid q(\lambda,\mu) > -\infty\}$$

Thm: Domain $D$ of dual function = convex set

negative dual function over $D$ = convex function

(no assumption of convexity for $f(x)$, $g(x)$, $h(x)$)

Proof: $\forall x, (\lambda,\mu), (\bar{\lambda},\bar{\mu}), \alpha \in [0,1]$

$l(x, \alpha\lambda + (1-\alpha)\bar{\lambda}, \alpha\mu + (1-\alpha)\bar{\mu}) = \alpha\, l(x,\lambda,\mu) + (1-\alpha)\, l(x,\bar{\lambda},\bar{\mu})$ → Linearity of $\underline{l}$ in second & third arguments

$$\Longrightarrow \inf_{x\in X} L(x, \alpha\lambda + (1-\alpha)\bar{\lambda}, \alpha\mu + (1-\alpha)\bar{\mu}) \geq$$

$$\alpha \inf_{x\in X} L(x,\lambda,\mu) + (1-\alpha) \inf_{x\in X} L(x,\bar{\lambda},\bar{\mu})$$

$$\Longrightarrow q(\alpha\lambda + (1-\alpha)\bar{\lambda}, \alpha\mu + (1-\alpha)\bar{\mu}) \geq$$

$$\alpha\, q(\lambda,\mu) + (1-\alpha)\, q(\bar{\lambda},\bar{\mu}) \qquad (*)$$

If $q(\lambda,\mu) > -\infty$, $q(\bar{\lambda},\bar{\mu}) > -\infty \overset{(*)}{\Longrightarrow}$

$$q(\alpha\lambda + (1-\alpha)\bar{\lambda}, \alpha\mu + (1-\alpha)\bar{\mu}) > -\infty$$

$\Longrightarrow$ $D$: convex set $\xrightarrow{(*)}$ $q(\cdot,\cdot)$ = concave on $D$.

---

Best Lower bound on $f_*$:

$$q_* = \sup_{\mu\geq 0} q(\lambda,\mu) \qquad \text{(we can add } (\lambda,\mu)\in D)$$

weak duality: $\quad q_* \leq f_*$

primal
(P)

$$\boxed{\begin{array}{ll} \min & f(x) \\ \text{s.t.} & h(x) = 0 \\ & g(x) \leq 0 \\ & x \in X \end{array}} \longrightarrow \boxed{\begin{array}{ll} \max & q(\lambda, \mu) \\ \text{s.t.} & \mu \geq 0 \end{array}}$$

non-convex

dual
(D)

always convex optimization after $\max q(\cdot,\cdot) \longrightarrow \min -q(\cdot,\cdot)$

If $q_* = f_*$, it is said that strong duality holds or there is a zero duality gap.

Example: LP (Linear program)

$$\begin{cases} \min & a_0^T x + b_0 \\ \text{s.t.} & a_i^T x = b_i \quad i=1,\dots,m \longrightarrow \lambda_i \\ & c_j^T x \leq d_j \quad j=1,\dots,r \longrightarrow \mu_j \end{cases}$$

$$L(x,\lambda,\mu) = a_0^T x + b_0 + \sum_{i=1}^{m} \lambda_i (a_i^T x - b_i) + \sum_{j=1}^{r} \mu_j (c_j^T x - d_j)$$

$$\Rightarrow \quad L(x,\lambda,\mu) = \left(a_0^T + \sum_{i=1}^{m} \lambda_i a_i^T + \sum_{j=1}^{r} \mu_j c_j^T\right) x$$

$$+ \left( b_0 + \sum_{i=1}^{m} (-\lambda_i b_i) + \sum_{j=1}^{r} (-\mu_j d_j) \right)$$

① ②

$$\inf_x L(x,\lambda,\mu) = -\infty \quad \text{if} \quad ① \neq 0$$

$$\Rightarrow \quad D = \left\{ (\lambda,\mu) \;\middle|\; a_0^T + \sum_{i=1}^{m} \lambda_i a_i^T + \sum_{j=1}^{r} \mu_j c_i^T = 0 \right\}$$

If $(\lambda,\mu) \in D \Rightarrow q(\lambda,\mu) = ②$

$\Rightarrow$ Dual optimization:

$$\max_{\lambda,\mu} \quad b_0 - \sum_{i=1}^{m} \lambda_i b_i - \sum_{j=1}^{r} \mu_j d_j$$

$$\text{s.t.} \quad \mu \geq 0$$

$$a_0^T + \sum_{i=1}^{m} \lambda_i a_i^T + \sum_{j=1}^{r} \mu_j c_j^T = 0$$

Domain

LP

$\Rightarrow$ Dual of LP = LP

QP (quadratic program):

$$\min \ \frac{1}{2} x^T P_0 x + a_0^T x + b_0$$

s.t. $\quad a_i^T x = b_i \quad i = 1, \dots, m \quad \longrightarrow \quad P_0 \succeq 0$ to ensure convexity

$\quad c_j^T x \le d_j \quad j = 1, \dots, r$

To simplify calculations, assume $P_0 \succ 0$.

$$\Rightarrow L(x, \lambda, \mu) = \frac{1}{2} x^T P_0 x + \left(a_0^T + \sum_{i=1}^{m} \lambda_i a_i^T + \sum_{j=1}^{r} \mu_j c_j^T\right) x + \left(b_0 - \sum_{i=1}^{m} \lambda_i b_i - \sum_{j=1}^{r} \mu_j d_j\right) \quad 2$$

= Convex in $x$ (shown it before: Lagrangian for convex opt is convex)

$$\min_x L(x, \lambda, \mu) \longleftrightarrow \nabla_x L(x, \lambda, \mu) = 0$$

(This is true if solution exists (not $-\infty$), which is the case here since $P_0 \succ 0$)

$$\Rightarrow \underbrace{P_0}_{\text{symmetric}} x + \underbrace{\left(a_0^T + \sum_{i=1}^{m} \lambda_i a_i^T + \sum_{j=1}^{r} \mu_j c_j^T\right)^T}_{w(\lambda,\mu)} = 0$$

$$\Rightarrow q(\lambda,\mu) = \frac{1}{2} w(\lambda,\mu)^T P_0^{-1} P_0 P_0^{-1} w(\lambda,\mu)$$

$$- w(\lambda,\mu)^T P_0^{-1} w(\lambda,\mu) + (\ldots)$$

$\Rightarrow$ Dual optimization:

$$\max_{\lambda,\mu} \; -\frac{1}{2}\left(a_0^T + \sum_{i=1}^{m} \lambda_i a_i^T + \sum_{j=1}^{r} \mu_j c_j\right) P_0^{-1} \left(a_0 + \sum_{i=1}^{m} \lambda_i a_i + \sum_{j=1}^{r} \mu_j c_j\right) + \left(b_0 - \sum_{i=1}^{m} \lambda_i b_i - \sum_{j=1}^{r} \mu_j d_j\right)$$

Concave & quadratic in $(\lambda,\mu)$

s.t. $\mu \geq 0$

$\Rightarrow$ Dual of QP = QP

Note:

minimization $\leftarrow$ $f_*$

- $+\infty \Rightarrow$ infeasible (P)
- finite $\Rightarrow$ feasible (P)
- $-\infty \Rightarrow$ unbounded from below (P)

maximization $\leftarrow$ $q_*$

- $+\infty \Rightarrow$ unbounded from above (D)
- finite $\Rightarrow$ feasible (D)
- $-\infty \Rightarrow$ infeasible (D)

zero duality gap: $f_* = q_*$

it culd happen that $f_* = q_* = -\infty$ or

$$f_* = q_* = +\infty$$

in these cases either (P) or (D) is infeasible

Note that if (D) is unbounded from above, then (P) is infeasible:

weak duality: $f_* \geq q_* = +\infty$

$$\Rightarrow f_* = +\infty$$

(generalization of Farkas Lemma for non-convex opt)

(P): Two questions:

1- Is $f_*$ finite?

2- If so, does $x_*$ exist?

$$\min \; x \quad \text{s.t.} \; x^2 \leq -1 \;\;\longrightarrow\;\; f_* = +\infty$$

$$\min \; x \;\;\longrightarrow\;\; f_* = -\infty$$

$$\min \; e^x \;\;\longrightarrow\;\; f_* = 0$$

but $x_*$ doesn't exist (since there is no finite $x_*$ corresponding to $f_*$)

To answer above questions & study duality gap, we need the notion of geometric multipliers:

Def: $(\lambda_*, \mu_*)$ is called geometric multipliers if $\mu_* \geq 0$ and $f_* = \inf_{x \in X} l(x, \lambda_*, \mu_*)$

Thm: - If there is no duality gap, then

set of geometric multipliers =

set of dual solutions

- If there is a duality gap ($\underbrace{f_* - q_* > 0}$),

then

set of geometric multipliers = empty.

$\Rightarrow$ existence of geometric multipliers

$\updownarrow$

no duality gap

For simplicity, assume there is no equality constraint.

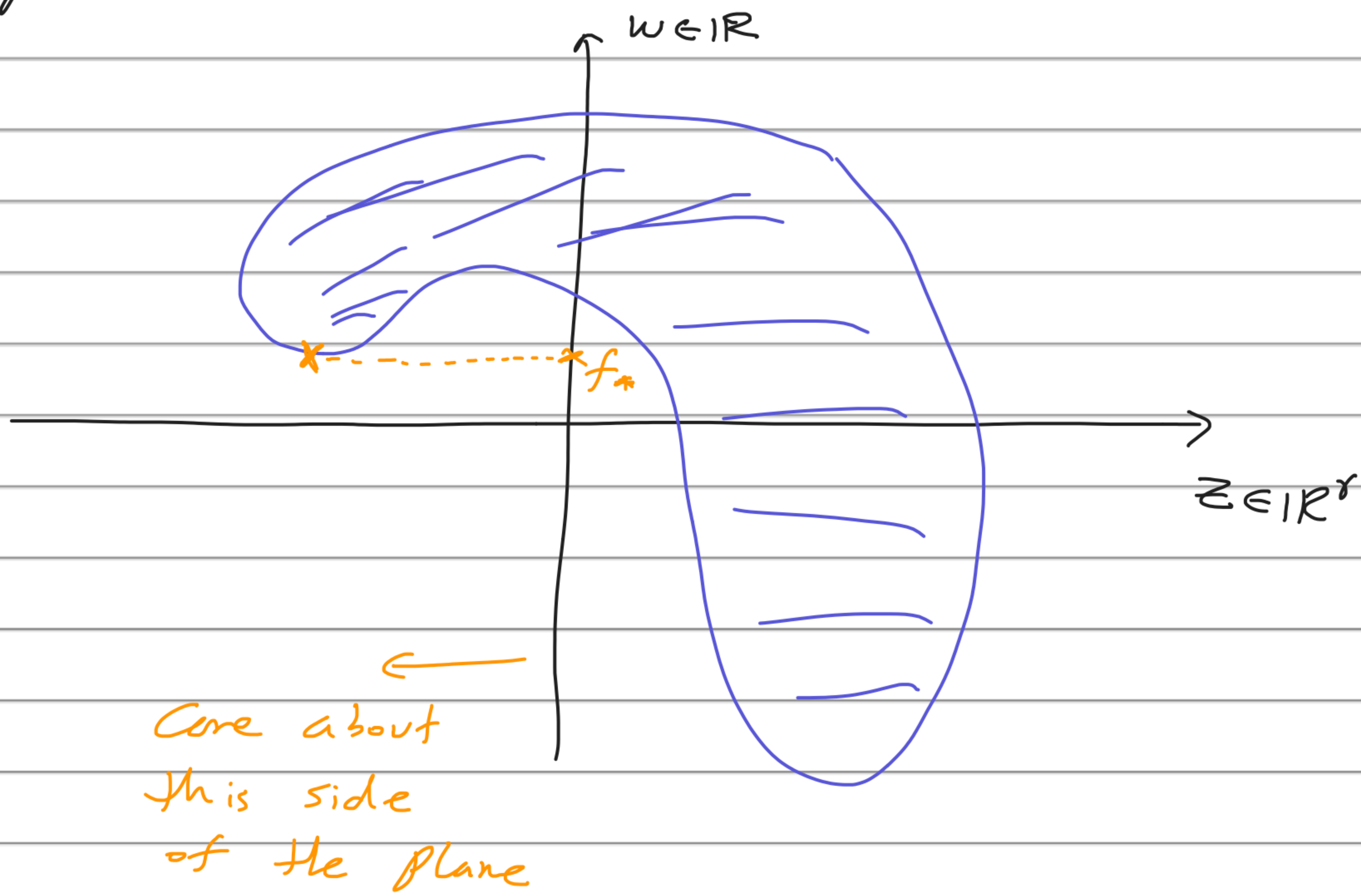
$$\min f(x) \quad \text{s.t.} \quad g(x) \leq 0 \quad , x \in X$$

Define: $S = \{(g(x), f(x)) \mid x \in X\}$

$f_*$ : Find a point $(\underset{\in \mathbb{R}^r}{z}, \overset{\in \mathbb{R}}{w})$ in the set $S$

s.t. $z \leq 0$ and $w$ is as small as

possible.

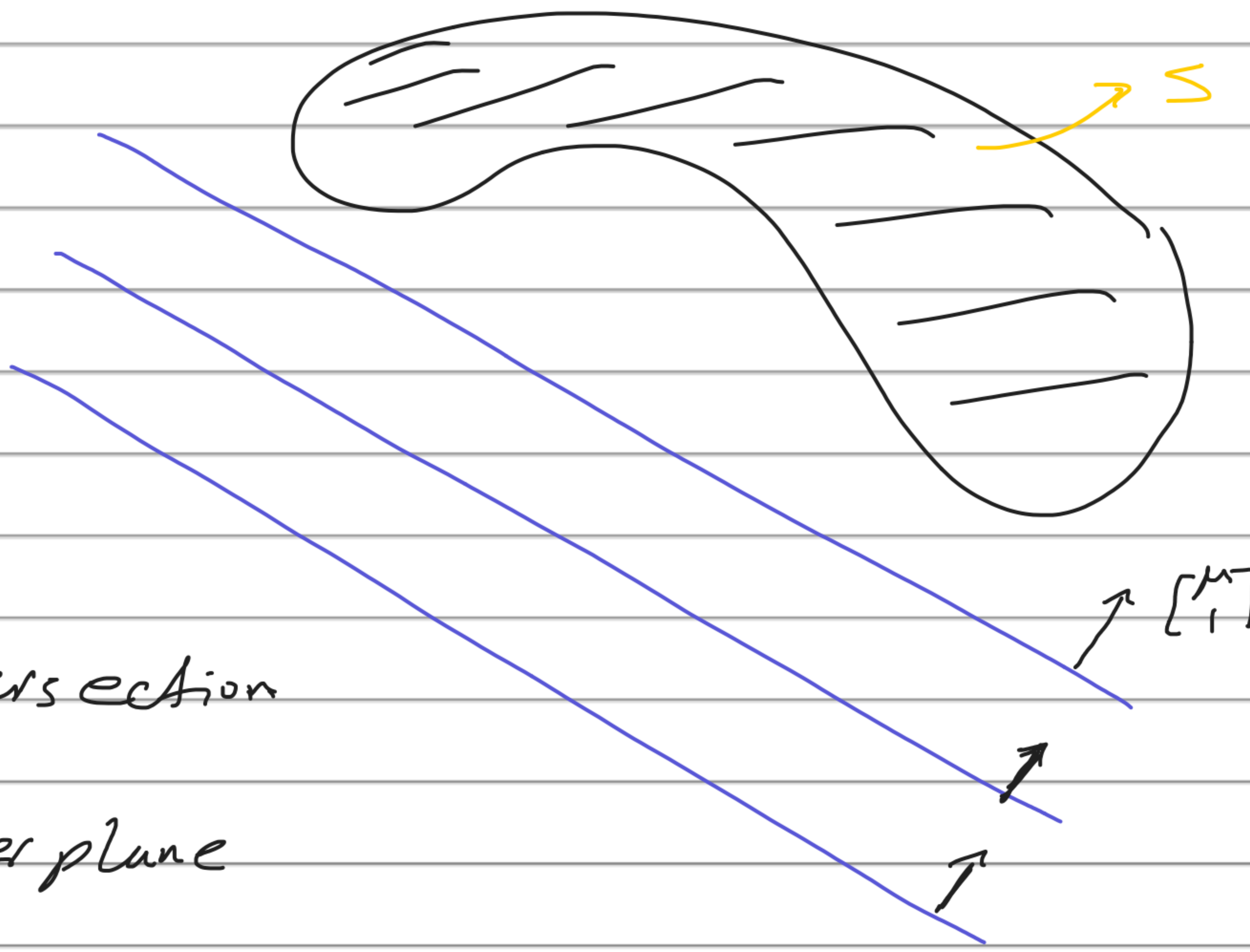



Also, $L(x, \mu) = f(x) + \mu^T g(x)$

$$= [\mu^T \quad 1] \begin{bmatrix} g(x) \\ f(x) \end{bmatrix} \begin{matrix} \to z \\ \to w \end{matrix}$$

$$\Rightarrow \inf_{x \in X} L(x, \mu) = \inf_{(z,w) \in S} [\mu^T \quad 1] \begin{bmatrix} z \\ w \end{bmatrix}$$

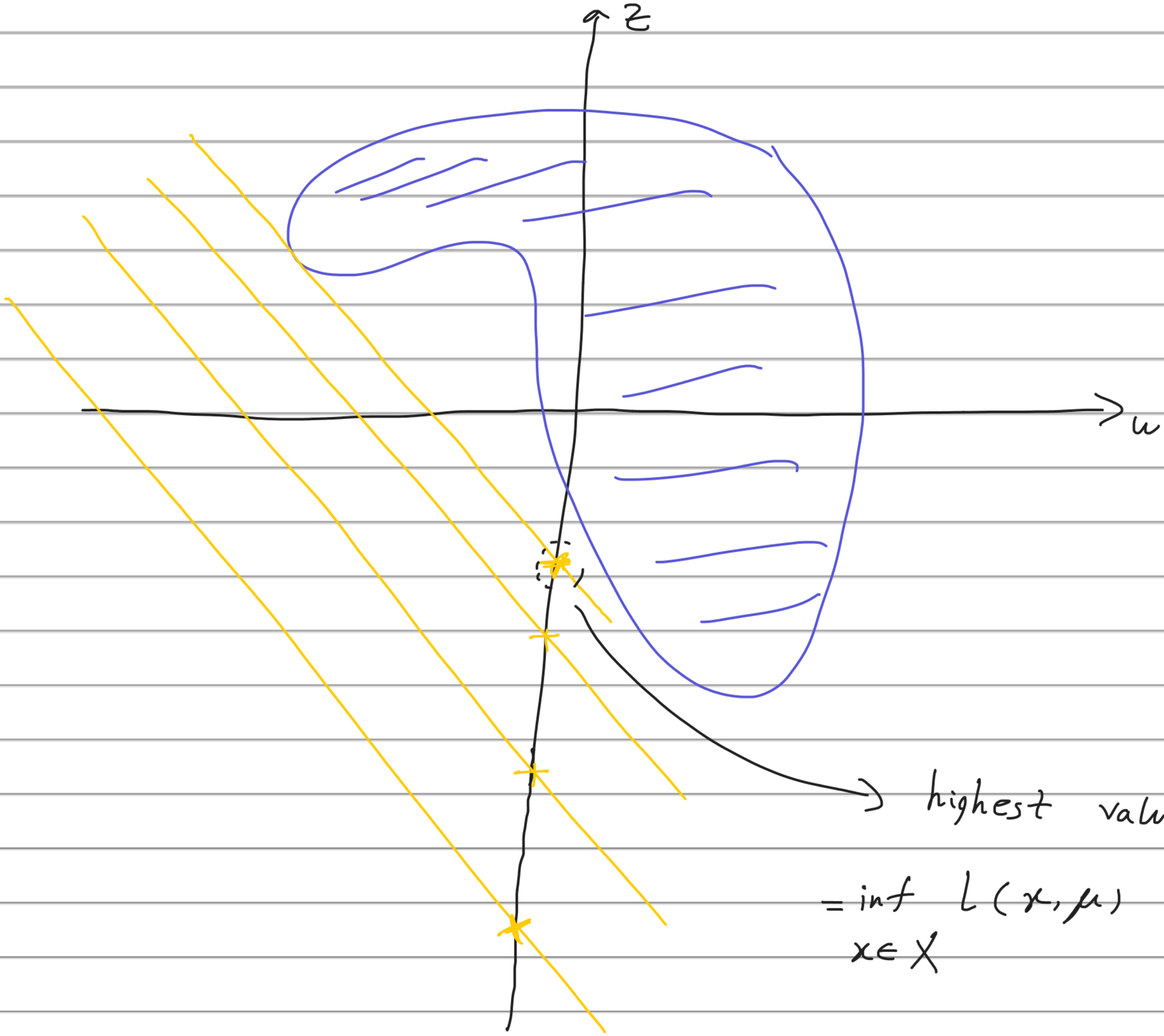
- Consider all hyperplanes with the normal vector $\begin{bmatrix} \mu \\ 1 \end{bmatrix}$ that include $S$ on their positive half space.




- Find the intersection of each hyperplane with the w-axis

- pick the hyperplane giving the highest level of interception with the w-axis.

- The value on the w-axis is $\inf_{x \in X} L(x, \mu)$




highest value

$$= \inf_{x \in X} L(x, \mu)$$


$$\inf_{x \in X} L(x, \mu) = [\mu^T \quad 1] \begin{bmatrix} A \\ B \end{bmatrix}$$

$$= [\mu^T \quad 1] \begin{bmatrix} 0 \\ B^- \end{bmatrix}$$

$$= B^-$$