DSA3102 Lecture 8 (Reading Section 5.2 & 5.3.1)

Extra Reading: slater.pdf, SVM example, Strong-duality-fails.

Optimization Problem:

$$\min_{x \in \mathbb{R}^n} f_0(x) \quad \text{s.t.} \quad \begin{array}{ll} f_i(x) \leq 0 \quad i \in [m] & \text{Inequality} \\ h_i(x) = 0 \quad i \in [p] & \text{Equality.} \end{array}$$

$$D = \left(\bigcap_{i=0}^{m} \text{dom } f_i \right) \cap \left(\bigcap_{i=1}^{m} \text{dom } h_i \right)$$

<u>Lagrangian</u> $\quad L: \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$

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

<u>Lagrange dual function</u> $\quad g: \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}.$

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

<u>Lagrange dual problem</u>

concave function

$\lambda \geq 0, \ \lambda_i \geq 0 \ \forall i = 1, \ldots, m.$

$$d^* = \max_{\lambda \geq 0, \nu \in \mathbb{R}^p} g(\lambda, \nu)$$

Rmk: This is a convex optimization problem.

Weak Duality

Primal optimal value $p^* = \inf_x \{ f_0(x) : x \text{ feasible} \}$.

$$\Downarrow$$

$$f_i(x) \le 0 \quad \forall i = 1, \dots, m$$
$$h_i(x) = 0 \quad \forall i = 1, \dots, p$$

Thm: Weak duality

i) $\forall \lambda \ge 0_m,\ \nu \in \mathbb{R}^p \quad : g(\lambda, \nu) \le p^*$

ii) $d^* = \max_{\lambda \ge 0, \nu \in \mathbb{R}^p} g(\lambda, \nu) \le p^*$

dual optimal value $\qquad$ Primal optimal value

Strong duality & Slater's condition

If $d^* = p^*$, we say that strong duality holds.

Folklore: Convex Opt Problem + Slater's condition $\Rightarrow$ Strong duality holds.

Convex Opt Problem: $\min_{x \in \mathbb{R}^n} f_0(x)$ s.t. $f_i(x) \le 0 \ \forall i = 1, \dots, m$

$Ax = b$

$A \in \mathbb{R}^{p \times n},\ b \in \mathbb{R}^p$

f_i: convex, $i = 0, 1, \dots, m$

Thm: [Slater's Constraint Qualification].

If the opt. problem is convex & $\exists\ \bar{x} \in D$ (Slater point) s.t.

$$f_i(\bar{x}) < 0 \quad \forall i = 1, \dots, m \quad , \quad A\bar{x} = b$$

then strong duality holds.

Sufficient condition for $d^* = p^*$.

$\bar{x}$: strictly feasible.

Weaker form of Slater's condition

$f_i(x) \leq 0$
$\forall i = 1, \dots, m$

Thm: If $f_1, \dots, f_k$ $(k \leq m)$ are affine, strong duality holds if $\exists\ \bar{x} \in D$ s.t.

$$f_i(\bar{x}) \leq 0 \ , \ \forall i = 1, \dots, k \ , \quad f_i(\bar{x}) < 0 \quad \forall i = k+1, \dots, m$$

$$A\bar{x} = b$$

Affine inequalities do not need to be satisfied strictly

Example: Least Squares Solution to Linear Equation

$$\min_{x \in \mathbb{R}^n} \ x^T x \quad \text{s.t.} \quad Ax = b \quad \text{(Least norm problem)}$$

$$A \in \mathbb{R}^{p \times n}$$

Strong duality holds if $b \in R(A)$.

$p^* = \inf\{x^T x : Ax = b\}$.

$x^* = A^T(AA^T)^{-1} b$

$A[A^T(AA^T)^{-1}]b$

$= \cancel{(AA^T)}(\cancel{AA^T})^{-1} b$

$p^* = (A^T(AA^T)^{-1}b)^T A^T(AA^T)^{-1}b$

$= b^T(AA^T)^{-1}\cancel{AA^T}(\cancel{AA^T})^{-1}b$.

$= b^T(AA^T)^{-1}b$.

$\sum_{i=1}^{p} v_i(a_i^T x - b_i)$

$\parallel$

Lagrangian: $L(x,v) = x^T x + v^T(Ax - b)$

Lagrange dual function

$g(v) = \inf_{x \in \mathbb{R}^n} L(x,v) = \inf_{x \in \mathbb{R}^n}\{x^T x + (A^T v)^T x - v^T b\}$

$\nabla_x [x^T x + (A^T v)^T x - v^T b]$

$= 2x + A^T v = 0$

$\Rightarrow \quad x^* = -\frac{1}{2}(A^T v)$.

$\nabla_x b^T x = b$

$g(v) = (-\frac{1}{2}(A^T v))^T(-\frac{1}{2}(A^T v)) + (A^T v)^T(-\frac{1}{2}A^T v) - v^T b$

$= -\frac{1}{4} v^T A A^T v - v^T b$.

Lagrange dual problem:

$d^* = \max_{v \in \mathbb{R}^p} g(v) = \max_{v \in \mathbb{R}^p} -\frac{1}{4} v^T A A^T v - v^T b$.

$$\nabla_v \left\{ -\frac{1}{4} v^T A A^T v - v^T b \right\}$$
$$= -\frac{1}{2} A A^T v - b = 0, \quad v^* = (AA^T)^{-1}(-2b).$$

$$d^* = -\frac{1}{\cancel{4}} \left[(AA^T)^{-1}(-\cancel{2}b)\right]^T AA^T \left[(AA^T)^{-1}(-\cancel{2}b)\right] - \left[(AA^T)^{-1}(-2b)\right]^T b$$

$$= b^T (AA^T)^{-1} b = p^*$$

Linear Programming

$$\min_{x \in \mathbb{R}^n} c^T x \quad \text{s.t.} \quad \begin{array}{l} Ax = b. \\ x \geq 0 \end{array}$$

Apply weaker form of Slater's condition, we see that strong duality holds if $\exists x \in \mathbb{R}^n$ that is feasible, i.e., $Ax = b$

$$p^* = d^*$$

Entropy Maximization

$$\min_{x \in \mathbb{R}^n_{++}} \sum_{i=1}^n x_i \log x_i \quad \text{s.t.} \quad Ax \leq b, \quad \overset{p=1}{\mathbf{1}^T x = 1}$$

Lagrange dual problem. (derived this last time using conjugates)

$$\max_{\lambda \geq 0, \nu \in \mathbb{R}} \underbrace{-b^T\lambda - \nu - e^{-\nu-1} \sum_{i=1}^{n} e^{-a_i^T\lambda}}_{g(\lambda,\nu)}$$

Qn: Do we have strong duality?
Weaker form of Slater's condition says we have strong duality if $\exists \bar{x} \in \mathbb{R}^n_{++}$ $(\bar{x} > 0)$ s.t. $A\bar{x} \leq b$, $\mathbf{1}^T\bar{x} = 1$.

Simplify the dual problem:

$$\frac{d}{d\nu}\left\{ \cancel{-b^T\lambda} - \nu - e^{-\nu-1} \sum_{i=1}^{n} e^{-a_i^T\lambda} \right\}$$

$$= -1 + e^{-\nu-1} \sum_{i=1}^{n} e^{-a_i^T\lambda} = 0$$

$$\nu^* = \log\left(\sum_{i=1}^{n} e^{-a_i^T\lambda}\right) - 1$$

Substitute this back into the dual problem

$$\max_{\lambda \geq 0} -b^T\lambda - \log\left(\sum_{i=1}^{n} e^{-a_i^T\lambda}\right)$$

$$\equiv \min_{\lambda \geq 0} b^T\lambda + \log\left(\sum_{i=1}^{n} e^{-a_i^T\lambda}\right)$$ geometric program.

Example: Convexity alone is not enough.

$$\min_{x\in\mathbb{R},\, y>0} \underbrace{e^{-x}}_{\text{convex}}, \quad \text{s.t.} \quad \underbrace{x^2/y}_{\substack{\| \\ f_1(x,y):\ \text{convex}}} \leq 0.$$

$$\nabla f_1(x,y) = \begin{bmatrix} 2x/y \\ -x^2/y^2 \end{bmatrix} \qquad D = \{(x,y)\in\mathbb{R}^2 : y>0\}.$$

$$\nabla^2 f_1(x,y) = \begin{bmatrix} 2/y & -2x/y^2 \\ -2x/y^2 & 2x^2/y^3 \end{bmatrix} \succeq 0.$$

$$\det(\nabla^2 f_1(x,y)) = \frac{2}{y}\cdot\frac{2x^2}{y^3} - \left(-\frac{2x}{y^2}\right)^2 = 0,$$

$$p^* = e^{-0} = 1.$$

Lagrangian: $L(x,y,\lambda) = e^{-x} + \lambda\dfrac{x^2}{y}$.

Lagrange dual function: $g(\lambda) = \inf\limits_{x\in\mathbb{R},\, y>0}\left\{e^{-x} + \lambda\dfrac{x^2}{y}\right\}$.

$\lambda \geq 0$.

Set $x = t$, $y = t^3$, $t>0$.

$$e^{-t} + \lambda \underbrace{t^2/t^3}_{x^2/y} = e^{-t} + \lambda\cdot\frac{1}{t} \xrightarrow{t\to\infty} 0$$

$\Rightarrow g(\lambda) = 0 \quad \forall \lambda \geq 0.$

Dual Problem: $d^* = \max_{\lambda \geq 0} g(\lambda) = \max_{\lambda \geq 0} 0 = 0.$

Duality Gap $= p^* - d^* = 1 - 0 = 1$

convex

$$\min_{x \in \mathbb{R}, y > 0} e^{-x}, \quad \text{s.t.} \quad x^2/y \leq 0.$$

Does there exist a Slater vector? $(\bar{x}, \bar{y}) \in D$ s.t.

$$\frac{\bar{x}^2}{\bar{y}} < 0$$

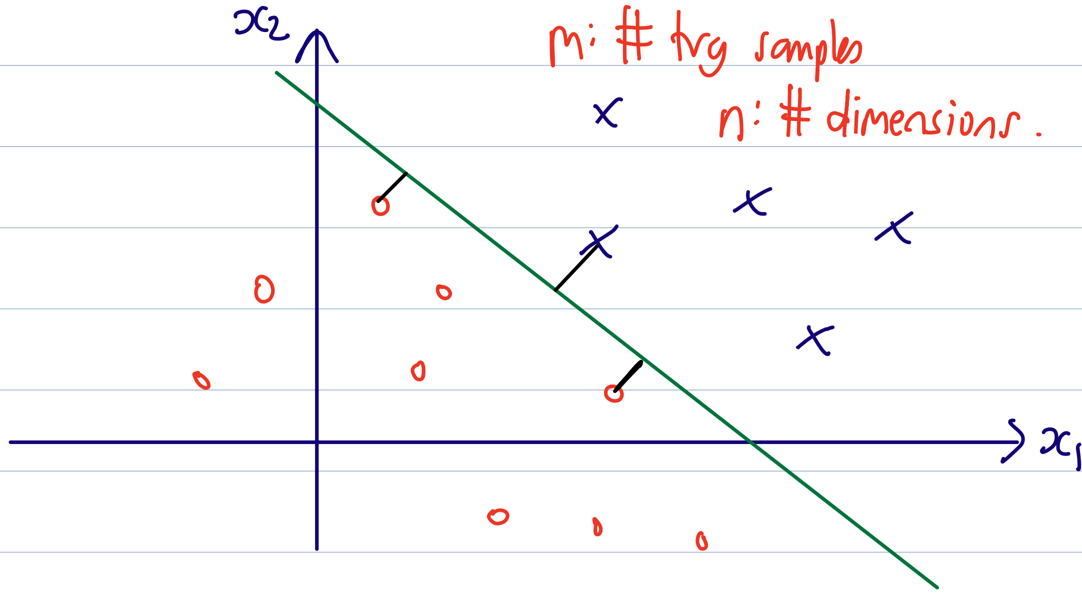
For all feasible $(x, y) \in D$, $x = 0$. There does not exist a Slater vector $(\bar{x}, \bar{y})$.

Support Vector Machines.

Given m data points $x_i \in \mathbb{R}^n$, each associated to a label $y_i \in \{-1, +1\}$.

Find a hyperplane that separates, as much as possible, the two classes of training points.

Let $Z = [y_1 \underline{x}_1, y_2 \underline{x}_2, \cdots, y_m \underline{x}_m] \in \mathbb{R}^{n \times m}$.

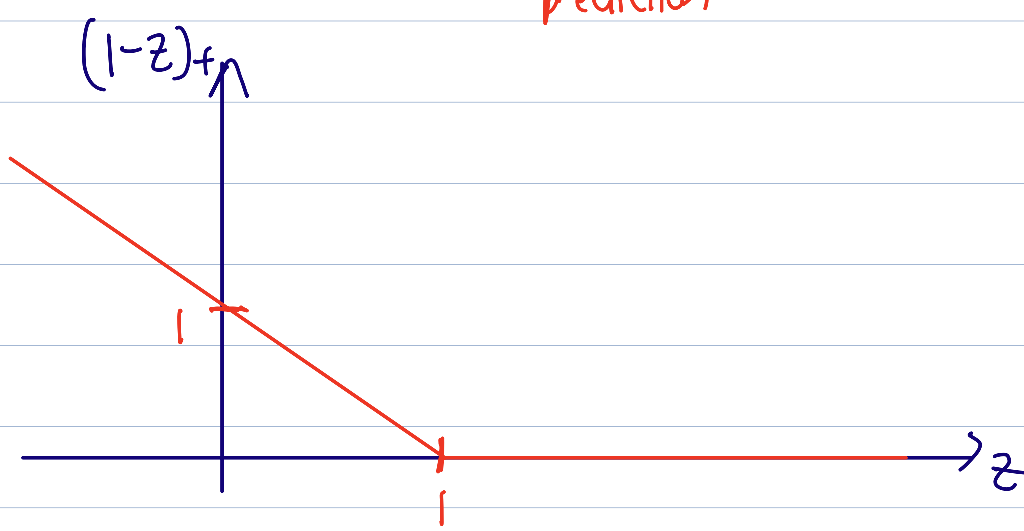


Minimize the hinge loss

$$(z)_+ = \begin{cases} z, & z \geq 0 \\ 0, & z < 0 \end{cases}$$

$$L(\underline{w}, b) = \sum_{i=1}^{m} \underbrace{(1 - y_i(w^T x_i + b))}_{V_i}{}_+$$

Intuition: We want $y_i(\underbrace{w^T x_i + b}_{\text{prediction}})$ positive.



Robustness / Regularization.

$f(w,b) = C\, L(w,b) + \frac{1}{2}\|w\|^2$

$w \in \mathbb{R}^n,\ b \in \mathbb{R}$

$C > 0$ controls the tradeoff between robustness and performance on trg set.

$$p^* = \min_{\substack{w,\ b,\ v \\ \in\ \ \in\ \ \in \\ \mathbb{R}^n\ \mathbb{R}\ \mathbb{R}^m}} \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{m} \underbrace{v_i}_{\text{slack variables}} \qquad \begin{array}{l} v_i \geq 0 \quad \forall i = 1, \dots, m \\ y_i(w^T x_i + b_i) \geq 1 - v_i \\ \quad \forall i = 1, \dots, m. \end{array}$$

$1 - y_i(w^T x_i + b) \leq v_i$

More compactly,

$$\min_{w,b,v} \frac{1}{2}\|w\|^2 + C\, v^T \mathbb{1} \quad \text{s.t.} \quad \begin{array}{l} v \geq 0 \\ v + Z^T w + b y \geq \mathbb{1} \end{array}$$

Lagrangian:

$$L(\underbrace{w, b, v}_{\text{primal}}, \lambda, \mu) = \frac{1}{2}\|w\|^2 + C\, v^T \mathbb{1} + \lambda^T(\mathbb{1} - v - Z^T w - b y) + \mu(-v)$$

Lagrange dual function.

$$g(\lambda, \mu) = \inf_{w,b,v} \frac{1}{2}\|w\|^2 + C\, v^T \mathbb{1} + \lambda^T(\mathbb{1} - v - Z^T w - b y)$$

$$+\mu(-v)$$

$$\nabla_w L(w,b,v,\lambda,\mu) = w - Z\lambda = 0 \Rightarrow w = Z\lambda.$$

$$\nabla_b L(w,b,v,\lambda,\mu) = \lambda^T y = 0$$

$$\nabla_v L(w,b,v,\lambda,\mu) = C\mathbb{1} - \lambda - \mu = 0$$

$$C\mathbb{1} = \lambda + \mu.$$

Dual function:

$$g(\lambda,\mu) = \begin{cases} \lambda^T \mathbb{1} - \frac{1}{2}\|Z\lambda\|^2 & \text{s.t. } \lambda^T y = 0 \quad C\mathbb{1} = \lambda + \mu \\ -\infty & \text{else} \end{cases}$$

Dual Problem:

$$d^* = \max_{\lambda \geq 0, \mu \geq 0} \left\{ \lambda^T \mathbb{1} - \frac{1}{2}\lambda^T Z^T Z \lambda \quad \text{s.t. } \lambda^T y = 0,\ C\mathbb{1} = \lambda + \mu \right\}$$

$$= \max_{\lambda \in \mathbb{R}^m} \left\{ \lambda^T \mathbb{1} - \frac{1}{2}\lambda^T Z^T Z \lambda \quad \text{s.t. } \lambda^T y = 0.\ 0 \leq \lambda \leq C\mathbb{1} \right\}$$

Dual Problem:

i) Optimization over m decision variables

ii) Only depends on training samples only through
$\underline{Z^T Z} \in S_+^m$ $\quad \underline{x}_i^T \underline{x}_j$

iii) Kernelization $K(\underline{x}_i, \underline{x}_j) = \exp(-\beta\, \underline{x}_i^T \underline{x}_j)$

Slater's condition holds. $\Rightarrow p^* = d^*$

Slater's condition is not necessary for strong duality.

Problem 5.22 (b)

$$\min_{x \in \mathbb{R}} \; x \qquad \text{s.t. } x^2 \le 0.$$

Problem is convex but $\nexists$ Slater point.

$p^* = 0$

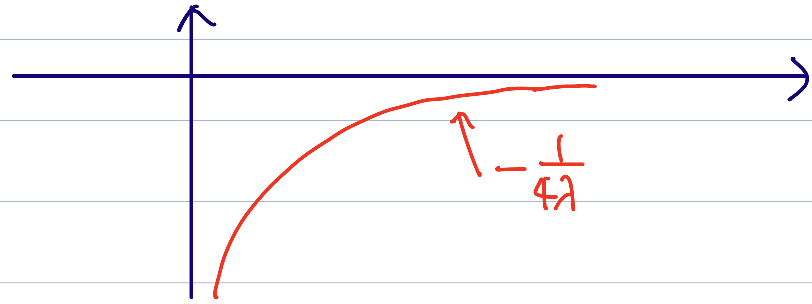
Lagrangian: $L(x, \lambda) = x + \lambda x^2$

Lagrange dual function: $g(\lambda) = \inf_{x \in \mathbb{R}} x + \lambda x^2$

$1 + 2\lambda x = 0 \Rightarrow x^* = -\frac{1}{2\lambda}.$

$g(\lambda) = -\frac{1}{2\lambda} + \lambda\left(-\frac{1}{2\lambda}\right)^2 = -\frac{1}{4\lambda}.$

Dual Optimization: $d^* = \max_{\lambda \geq 0} g(\lambda) = \max_{\lambda \geq 0} -\frac{1}{4\lambda} = 0.$



$p^* = d^* = 0$

Geometric Interpretation

Assume only 1 inequality constraint: $\min_{x \in \mathbb{R}^n} f_0(x)$ s.t. $f_1(x) \leq 0$.

$$G = \{ (f_1(x), f_0(x)) \in \mathbb{R}^2 : x \in D \}$$

ineq. constraint ↓ ↙ obj

set of values of $(f_1(x), f_0(x))$.

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

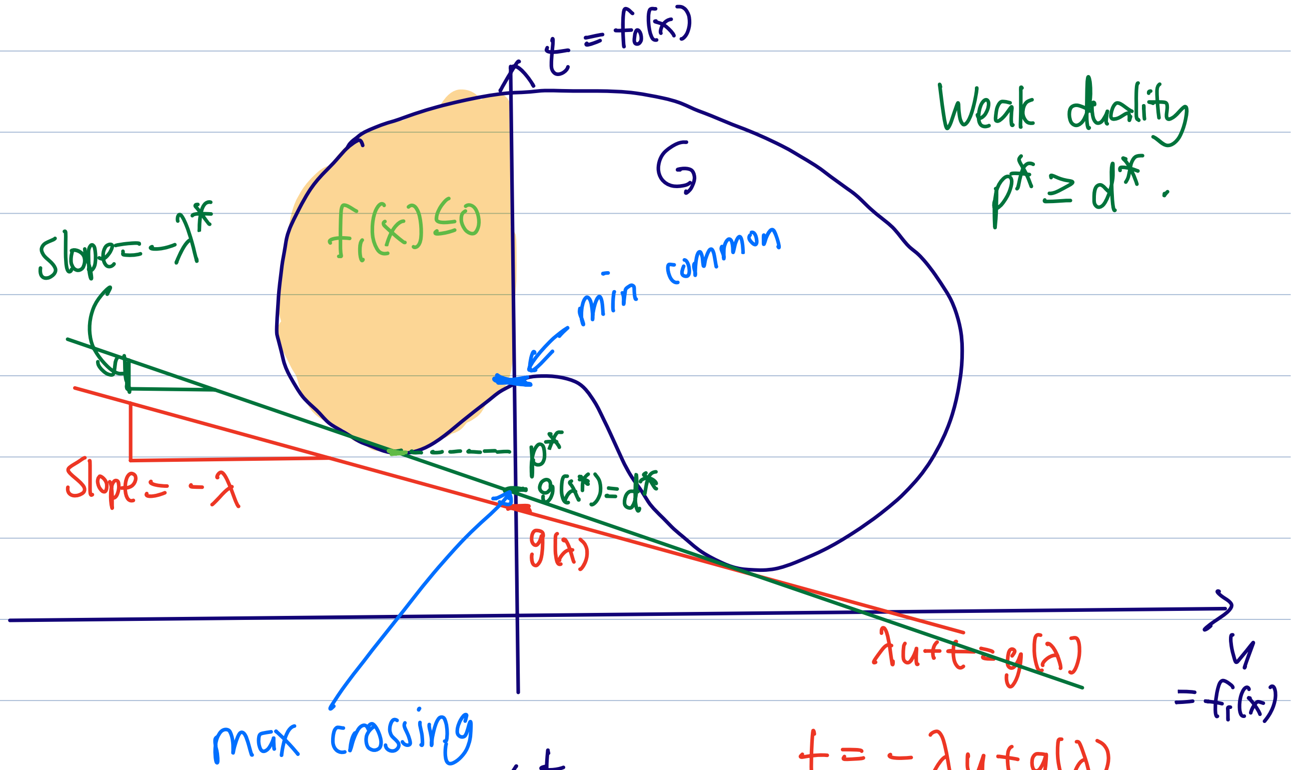
$g(\lambda) = \inf_x \underbrace{f_0(x)}_{t} + \lambda \underbrace{f_1(x)}_{u}$

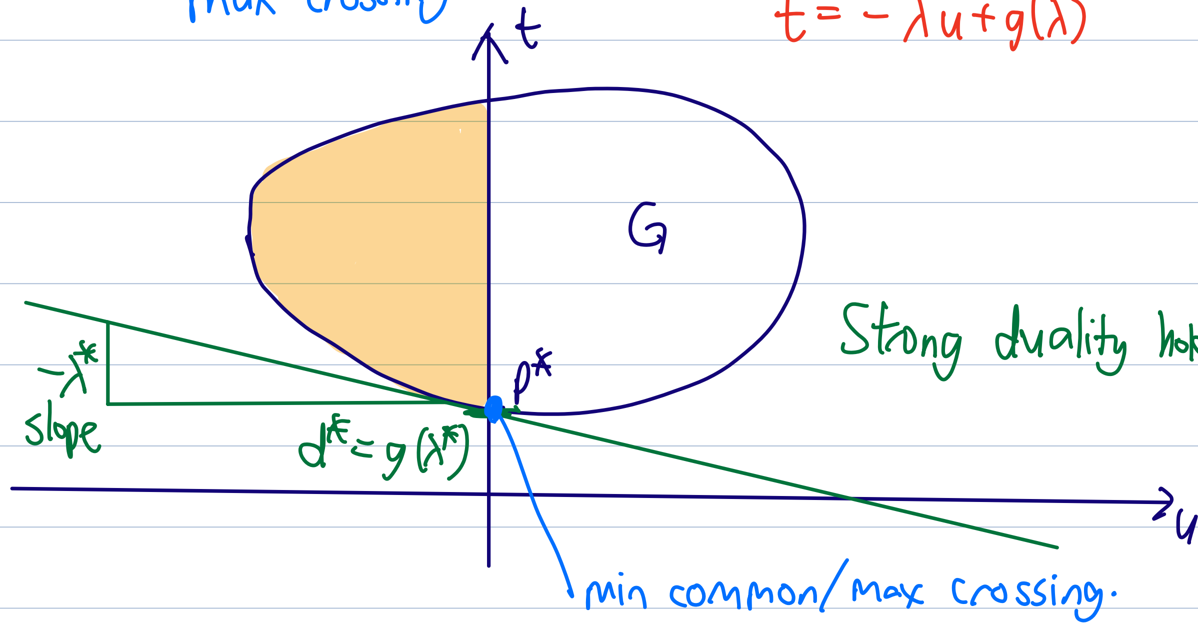
To evaluate the dual function at λ, minimize the affine function

$\underline{\lambda u + t} = \begin{pmatrix} \lambda \\ 1 \end{pmatrix}^T \begin{pmatrix} u \\ t \end{pmatrix}$ over all $(u,t) \in G$.

$$g(\lambda) = \inf\left\{ \begin{pmatrix}\lambda\\1\end{pmatrix}^T \begin{pmatrix}u\\t\end{pmatrix} : (u,t) \in G \right\}.$$

If the infimum is finite, then $\begin{pmatrix}\lambda\\1\end{pmatrix}^T \begin{pmatrix}u\\t\end{pmatrix} \geq g(\lambda) \quad \forall (u,t) \in G$.
defines a supporting hyperplane to G.





Proof of Slater's Thm (not tested)

Convex primal problem: $\min_{x \in \mathbb{R}^n} f_0(x)$ s.t. $f_1(x) \leq 0$.

No equality constraints

f_0, f_1: convex.

$p^* = \inf\{f_0(x) : f_1(x) \leq 0\}$ finite.

Slater's condition: $\exists \bar{x} \in \underbrace{\mathrm{dom} f_0 \cap \mathrm{dom} f_1}_{D}$ s.t. $f_1(\bar{x}) < 0$.

Lagrangian: $L(x, \lambda) = f_0(x) + \lambda f_1(x)$

Lagrange dual function:

$$g(\lambda) = \inf_{x \in D} L(x, \lambda).$$

Lagrange dual problem: $d^* = \max_{\lambda \geq 0} g(\lambda)$

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

Suffices for us to prove: $p^* \leq d^*$.

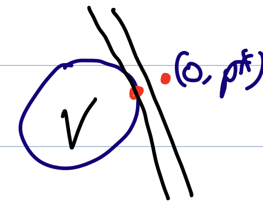
Pf: Consider

$V = \{(u, w) \in \mathbb{R} \times \mathbb{R} : f_0(x) \leq w, f_1(x) \leq u, \forall x \in D\}$.

i) V is convex

ii) V is "upward closed"; $(u, w) \in V$, $(u', w') \in V$

for any $(u', w') \geq (u, w)$.



Claim: $(0, p^*) \notin \text{int}(V)$

Suppose $(0, p^*) \in \text{int}(V)$. $\exists \varepsilon > 0$ s.t. $(0, p^* - \varepsilon) \in \text{int}(V)$.
this contradicts the optimality of p^*.

$(0, p^*) \in \text{bd}(V)$ or $(0, p^*) \notin V$.

By the supporting hyperplane theorem, $\exists$ a hyperplane passing through $(0, p^*)$ and supporting the set V.

$\exists (\lambda, \lambda_0) \in \mathbb{R}^2$ s.t. $(\lambda, \lambda_0) \neq (0, 0)$ s.t.

$$\begin{pmatrix} \lambda \\ \lambda_0 \end{pmatrix}^T \begin{pmatrix} u \\ w \end{pmatrix} = \lambda u + \lambda_0 w \geq \begin{pmatrix} \lambda \\ \lambda_0 \end{pmatrix}^T \begin{pmatrix} 0 \\ p^* \end{pmatrix} = \lambda_0 p^* \quad — (*).$$

for all $(u, w) \in V$

This means by the "upward closed" property that $\lambda \geq 0$ and $\lambda_0 \geq 0$.
Suppose, to the contrary, $\lambda_0 < 0$, we can make w larger contradicting $(*)$

i) $\lambda_0 = 0$: This means that $\lambda > 0$.

$$\inf_{(u,w)\in V} \lambda u \geq 0.$$

On the other hand, by the def^{n} of V and since $\lambda > 0$, we have

$$\inf_{(u,w)\in V} \lambda u = \inf_{x} \lambda f_1(x) = \lambda \inf_x f_1(x) \leq \lambda f_1(\bar{x}) < 0.$$

$\Rightarrow\Leftarrow$ $\lambda_0 = 0$ not possible.

ii) $\lambda_0 > 0$ $\quad \begin{pmatrix}\lambda \\ \lambda_0\end{pmatrix}^T \begin{pmatrix}u \\ w\end{pmatrix} = \lambda u + \lambda_0 w \geq \begin{pmatrix}\lambda \\ \lambda_0\end{pmatrix}^T \begin{pmatrix}0 \\ p^*\end{pmatrix} = \lambda_0 p^*$

$$\lambda u + \lambda_0 w \geq \lambda_0 p^* \qquad\qquad d^* \geq p^*$$

$$\Rightarrow \frac{\lambda}{\lambda_0} u + w \geq p^* \qquad\qquad \tilde{\lambda} = \frac{\lambda}{\lambda_0}$$

$$\Rightarrow \tilde{\lambda} u + w \geq p^*$$

$$\Rightarrow \inf_{(u,w)\in V} \{ \tilde{\lambda} \overset{f_1}{\overset{\shortparallel}{u}} + \overset{f_0}{\overset{\shortparallel}{w}} \} \geq p^*$$

$$\Rightarrow g(\tilde{\lambda}) = \inf_x \{ f_0(x) + \tilde{\lambda} f_1(x) \} \geq p^*$$

Maximize the LHS over all $\tilde{\lambda} \geq 0$

$$\Rightarrow \max_{\tilde{\lambda}\geq 0} g(\tilde{\lambda}) \geq p^*$$

$\Rightarrow \quad d^* \geq p^*$ $\nearrow$ combine with weak duality $d^* \leq p^*$ to get $d^* = p^*$.

(qed).