Convex Optimization

Prof. Nati Srebro

Lecture 7:

Geometric Interpretation of Duality

Recommended reading with more precise and general treatment: Boyd and Vandenberghe 5.3

KKT Optimality Conditions

Boyd and Vandenberghe 5.5

Additional optional reading: Sections 5.6,5.8

$$\begin{array}{lll} \min\limits_{x \in \mathbb{R}^n} & f_0(x) & \\ s.t. & f_i(x) \leq 0 & i = 1..m \\ & h_i(x) = 0 & j = 1..p \end{array}$$

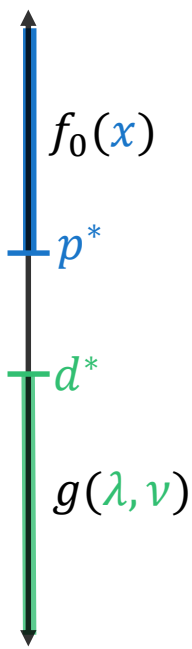
$$\begin{array}{ll} \max\limits_{\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p} & g(\lambda, \nu) \\ s.t. & \lambda_i \geq 0 \end{array}$$

$$L\big(x, (\lambda, \nu)\big) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \nu_j h_j(x)$$

$$g(\lambda, \nu) = \inf_x L\big(x, (\lambda, \nu)\big)$$

Weak Duality (always holds):

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



Example: Correlation Clustering

Data points $i = 1..n$, W_{ij} =similarity between points i and j

Find cluster assignments $x_i \in \pm 1$ maximizing:

$$x^\top W x = \sum_{x_i = x_j} W_{ij} - \sum_{x_i \neq x_j} W_{ij}$$

$$\boxed{\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & -x^\top W x \\ s.t. & x_i^2 = 1 \end{array}}$$

$$\boxed{\begin{array}{ll} \max\limits_{\nu \in \mathbb{R}^n} & -\sum \nu_i \\ s.t. & diag(\nu) \succcurlyeq W \end{array}}$$

$$L(x, \nu) = -x^\top W x + \sum_i \nu_i \left(x_i^2 - 1\right) = -x^\top W x + x^\top diag(\nu) x - \sum \nu_i$$
$$= x^\top (diag(\nu) - W) x - \sum \nu_i$$

$$g(\nu) = \inf_x L(x, \nu) = \begin{cases} -\infty & \text{If } \lambda_{min}(diag(\nu) - W) < 0 \\ \sum \nu_i & \text{If } (diag(\nu) - W) \succcurlyeq 0 \end{cases}$$

(P) not convex, but weak duality still holds: $d^* \leq p^*$

$$\begin{array}{lll} \min\limits_{x\in\mathbb{R}^n} & f_0(x) & \\ s.t. & f_i(x) \le 0 & i = 1..m \\ & h_i(x) = 0 & j = 1..p \end{array}$$

$$\begin{array}{ll} \max\limits_{\lambda\in\mathbb{R}^m,\nu\in\mathbb{R}^p} & g(\lambda,\nu) \\ s.t. & \lambda_i \ge 0 \end{array}$$

<u>Slater's Condition</u>:

$\exists x$ s.t.: $x \in \text{rel}-\text{interior}\big(dom(f_0)\big)$

(interior relative to the affine subspace spanned by it)

$f_i(x) < 0$ for non-linear f_i

$f_i(x) \le 0$ for linear f_i

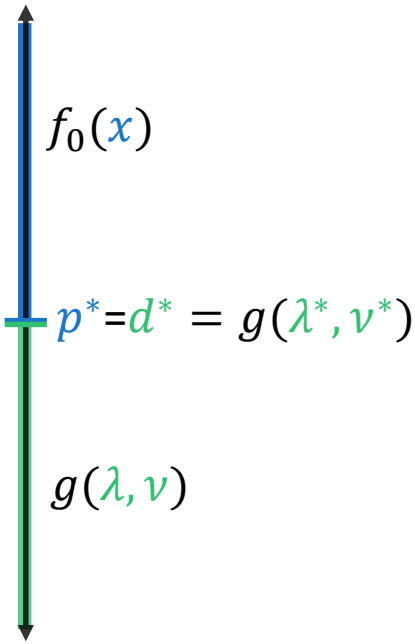
$h_j(x) = 0$

<u>Theorem</u>: If (P) is a convex problem and Slater's Condition holds, then we have strong duality, i.e.

$$p^* = d^*$$

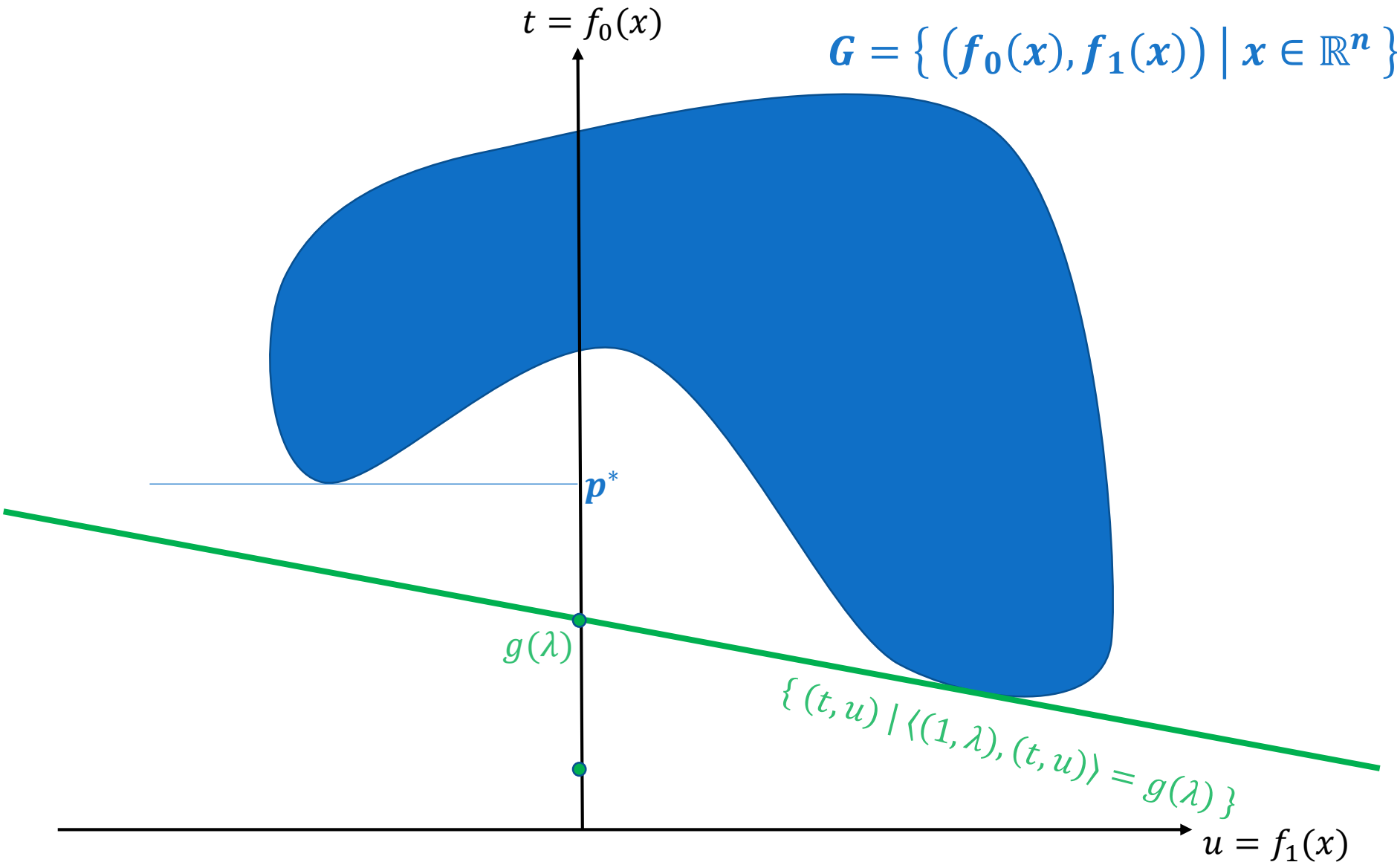
Furthermore, if $p^* = d^* > -\infty$, then the dual optimum is attainted:

$$\exists \lambda^*, \nu^* \;\; s.t. \;\; g(\lambda^*, \nu^*) = d^* = p^*$$



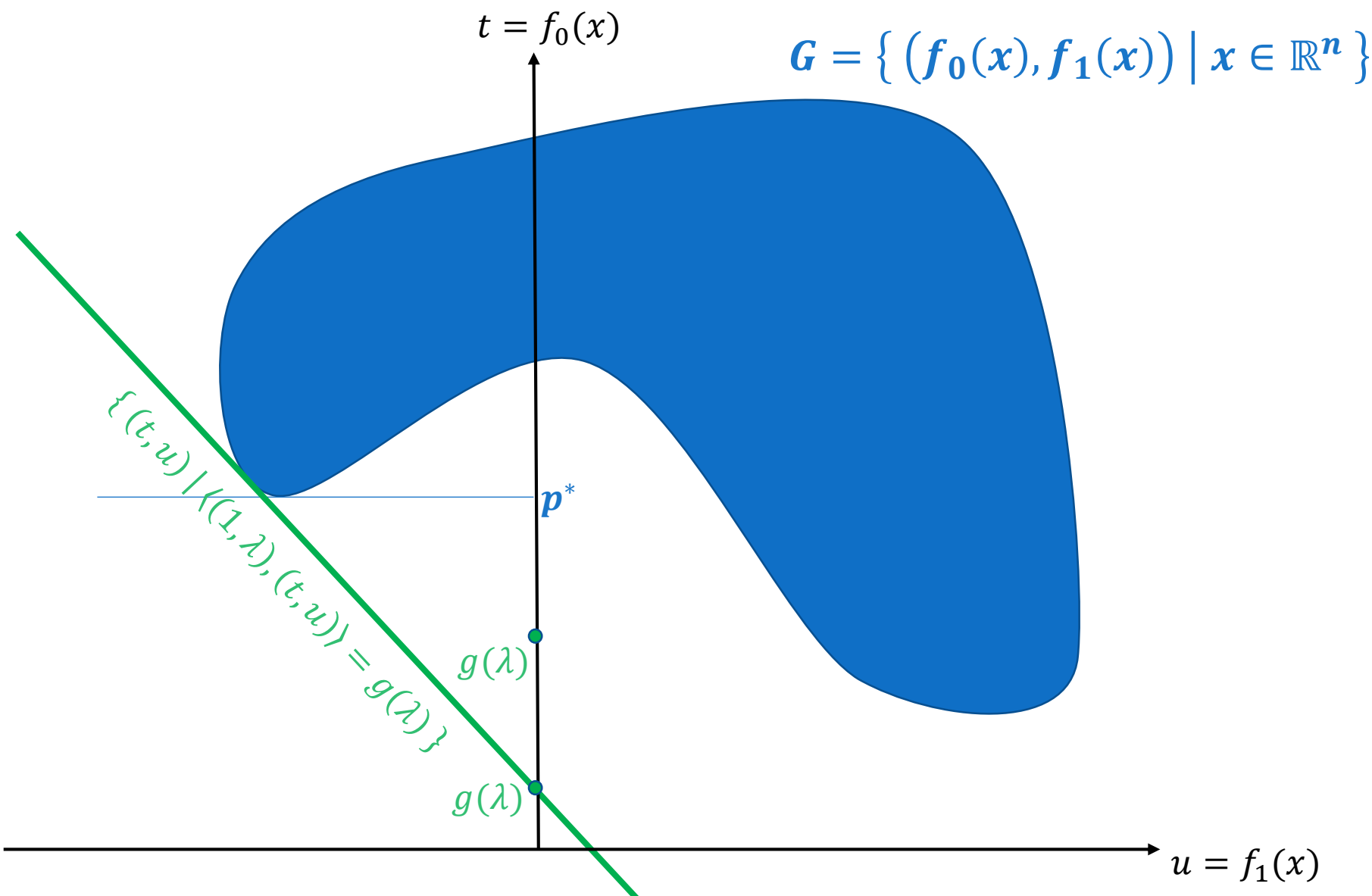
$$\boxed{\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & f_0(x) \\ s.t. & f_1(x) \leq 0 \end{array}}$$

$$g(\lambda) = \inf_x f_0(x) + \lambda f_1(x)$$



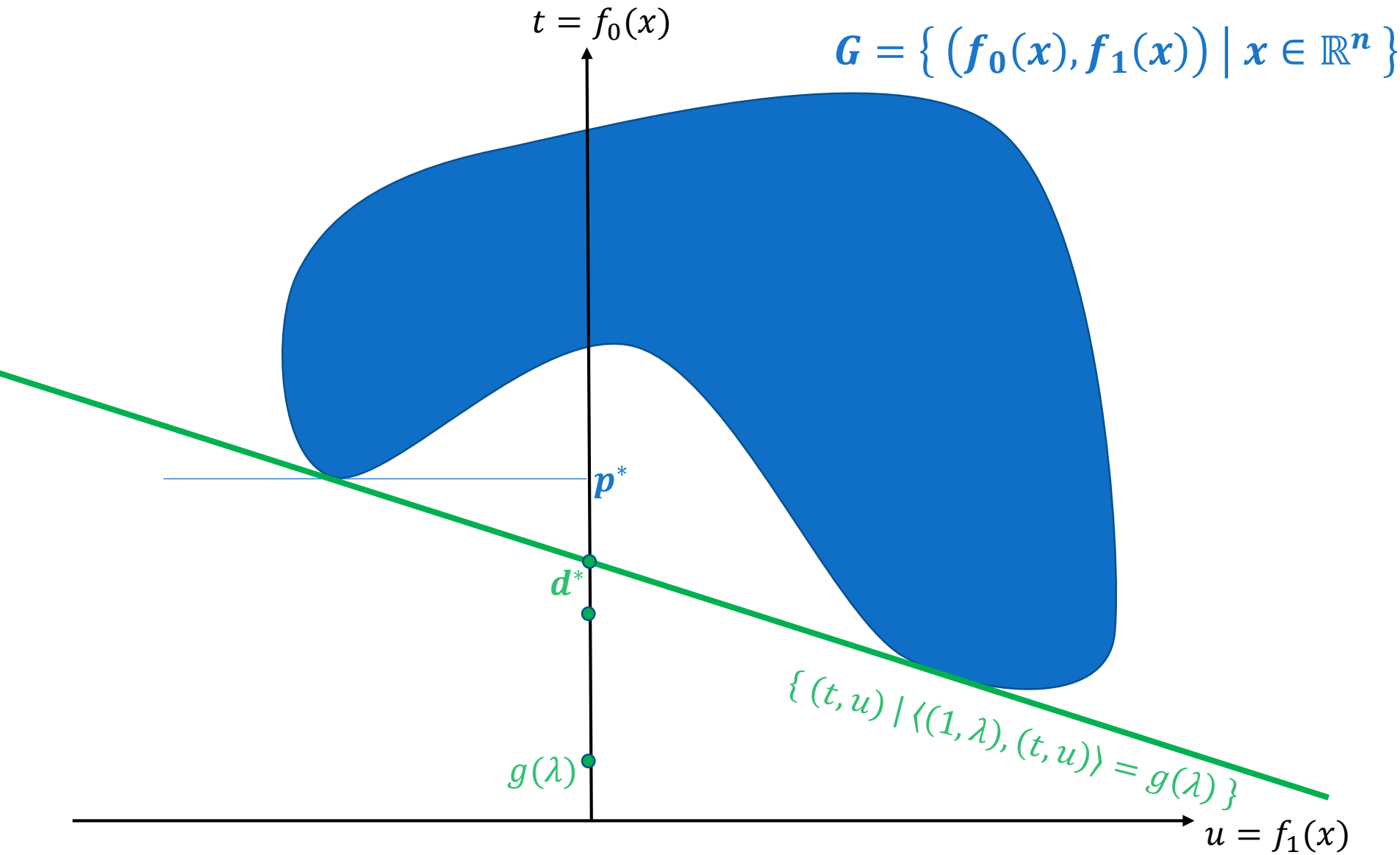
$$\boxed{\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & f_0(x) \\ s.t. & f_1(x) \leq 0 \end{array}}$$

$$g(\lambda) = \inf_x f_0(x) + \lambda f_1(x)$$



$$\boxed{\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & f_0(x) \\ s.t. & f_1(x) \leq 0 \end{array}}$$

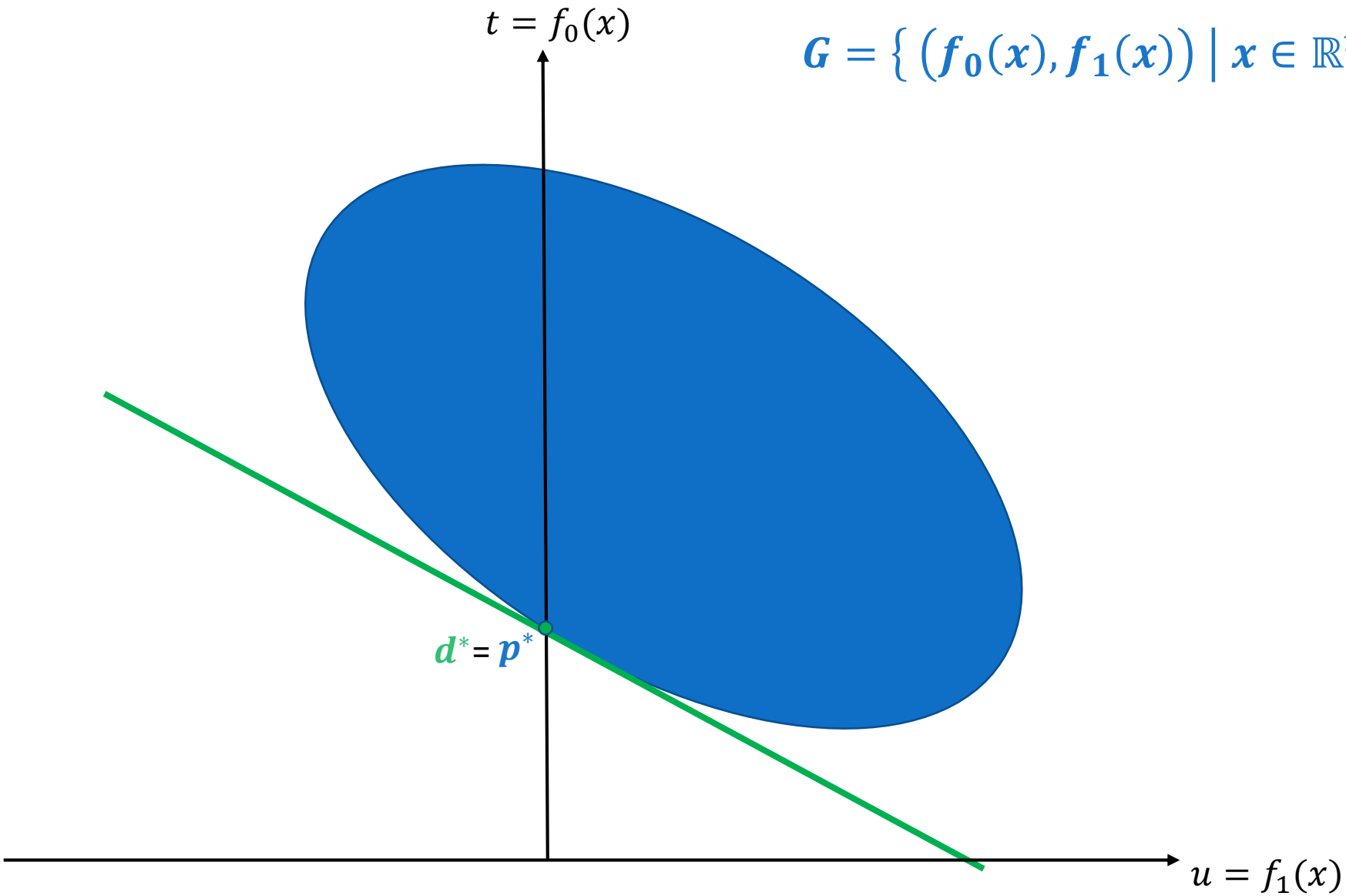
$$g(\lambda) = \inf_x f_0(x) + \lambda f_1(x)$$



$$\boxed{\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & f_0(x) \\ s.t. & f_1(x) \leq 0 \end{array}}$$

$$g(\lambda) = \inf_{x} f_0(x) + \lambda f_1(x) = \inf_{(t,u) \in G} \langle (1, \lambda), (t, u) \rangle$$

$$\boldsymbol{G = \{ \left(f_0(x), f_1(x) \right) \mid x \in \mathbb{R}^n \}}$$

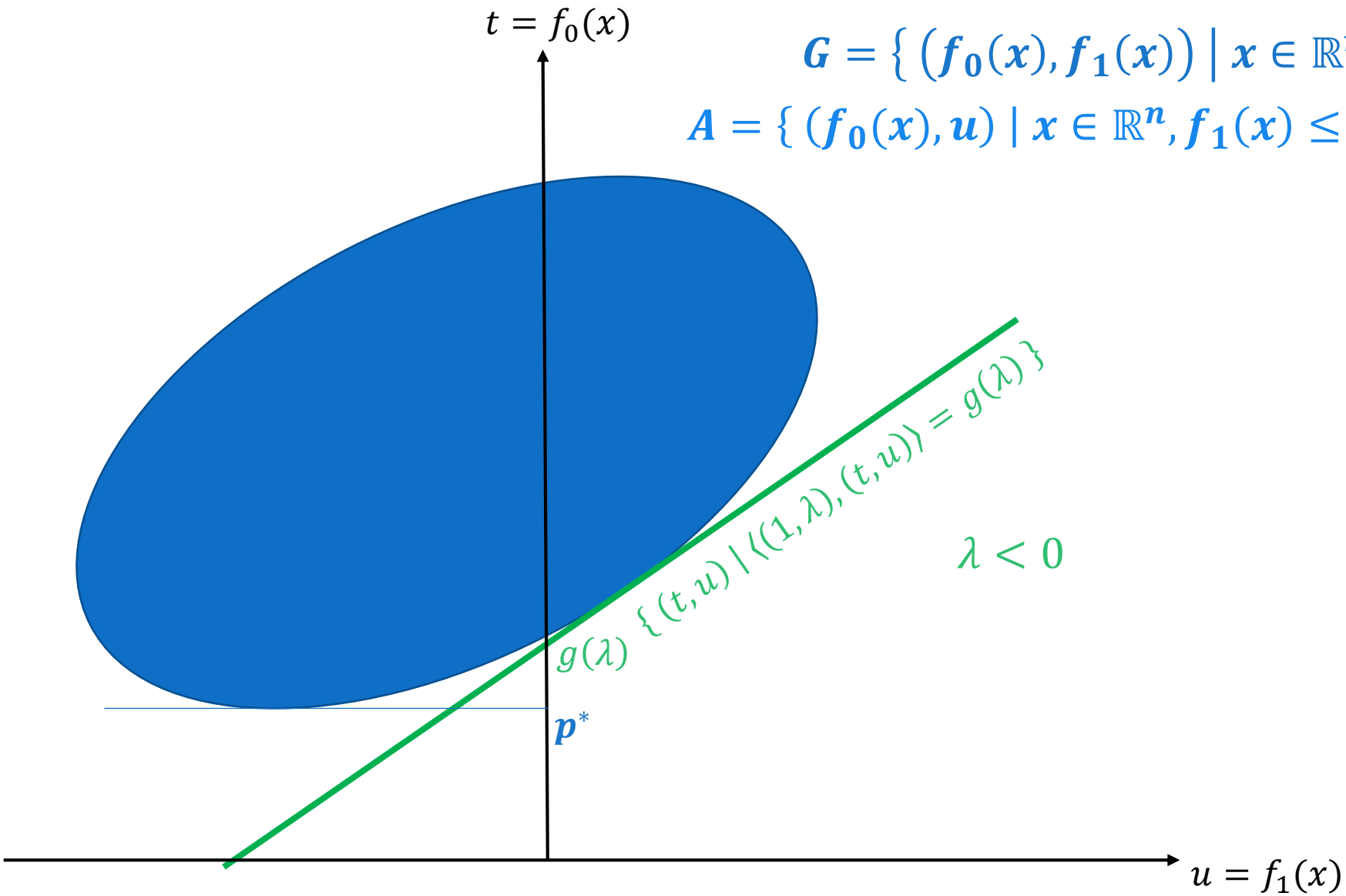


$$\boxed{\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & f_0(x) \\ s.t. & f_1(x) \leq 0 \end{array}}$$

$$g(\lambda) = \inf_x f_0(x) + \lambda f_1(x) = \inf_{(t,u) \in G} \langle (1, \lambda), (t, u) \rangle$$

$$\boldsymbol{G = \{ \left(f_0(x), f_1(x) \right) \mid x \in \mathbb{R}^n \}}$$

$$\boldsymbol{A = \{ \left(f_0(x), u \right) \mid x \in \mathbb{R}^n, f_1(x) \leq u \}}$$

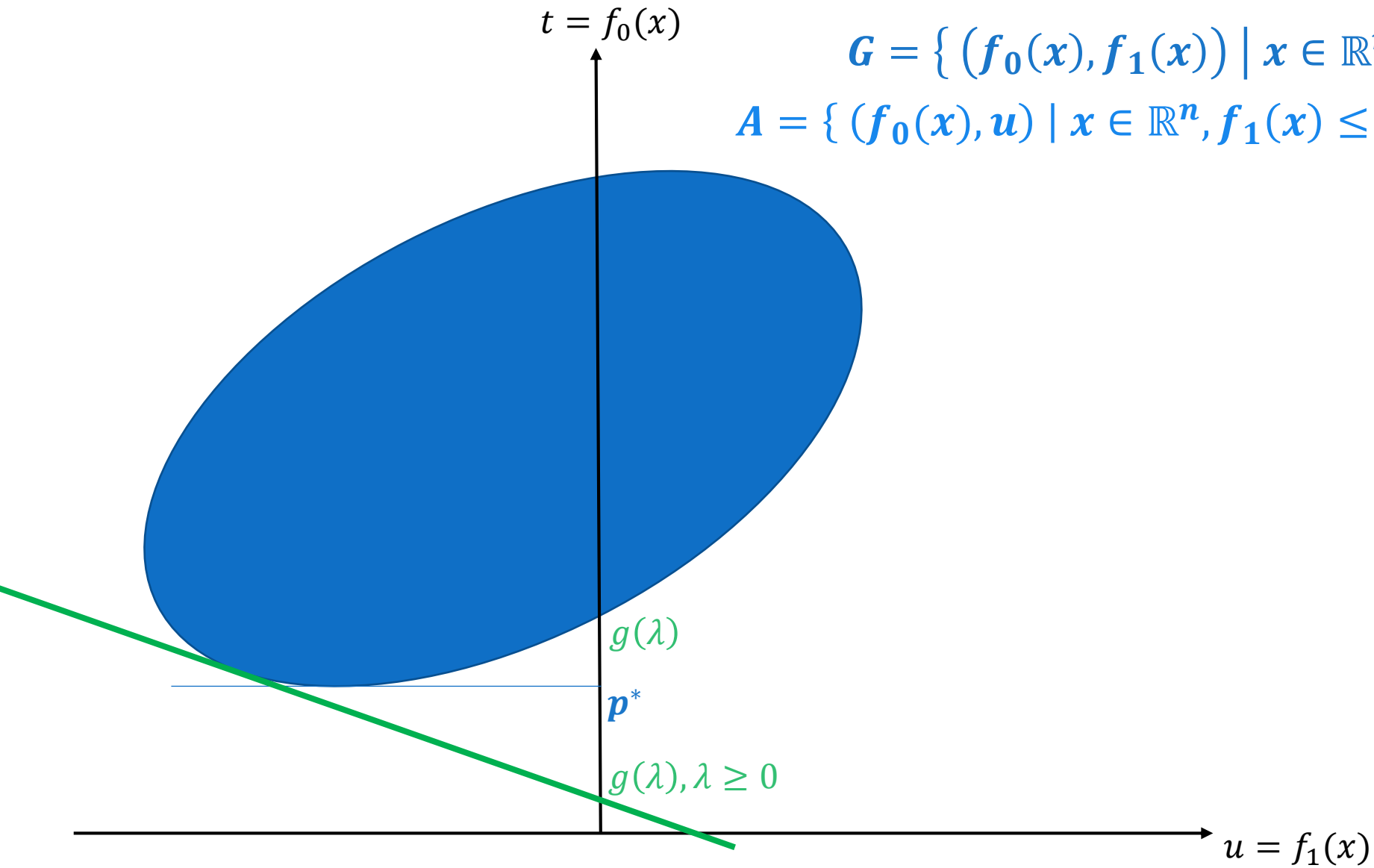


$$\boxed{\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & f_0(x) \\ s.t. & f_1(x) \leq 0 \end{array}}$$

$$g(\lambda) = \inf_x f_0(x) + \lambda f_1(x) = \inf_{(t,u) \in G} \langle (1, \lambda), (t, u) \rangle$$

$$\boldsymbol{G} = \{ (\boldsymbol{f_0}(\boldsymbol{x}), \boldsymbol{f_1}(\boldsymbol{x})) \mid \boldsymbol{x} \in \mathbb{R}^{\boldsymbol{n}} \}$$

$$\boldsymbol{A} = \{ (\boldsymbol{f_0}(\boldsymbol{x}), \boldsymbol{u}) \mid \boldsymbol{x} \in \mathbb{R}^{\boldsymbol{n}}, \boldsymbol{f_1}(\boldsymbol{x}) \leq \boldsymbol{u} \}$$

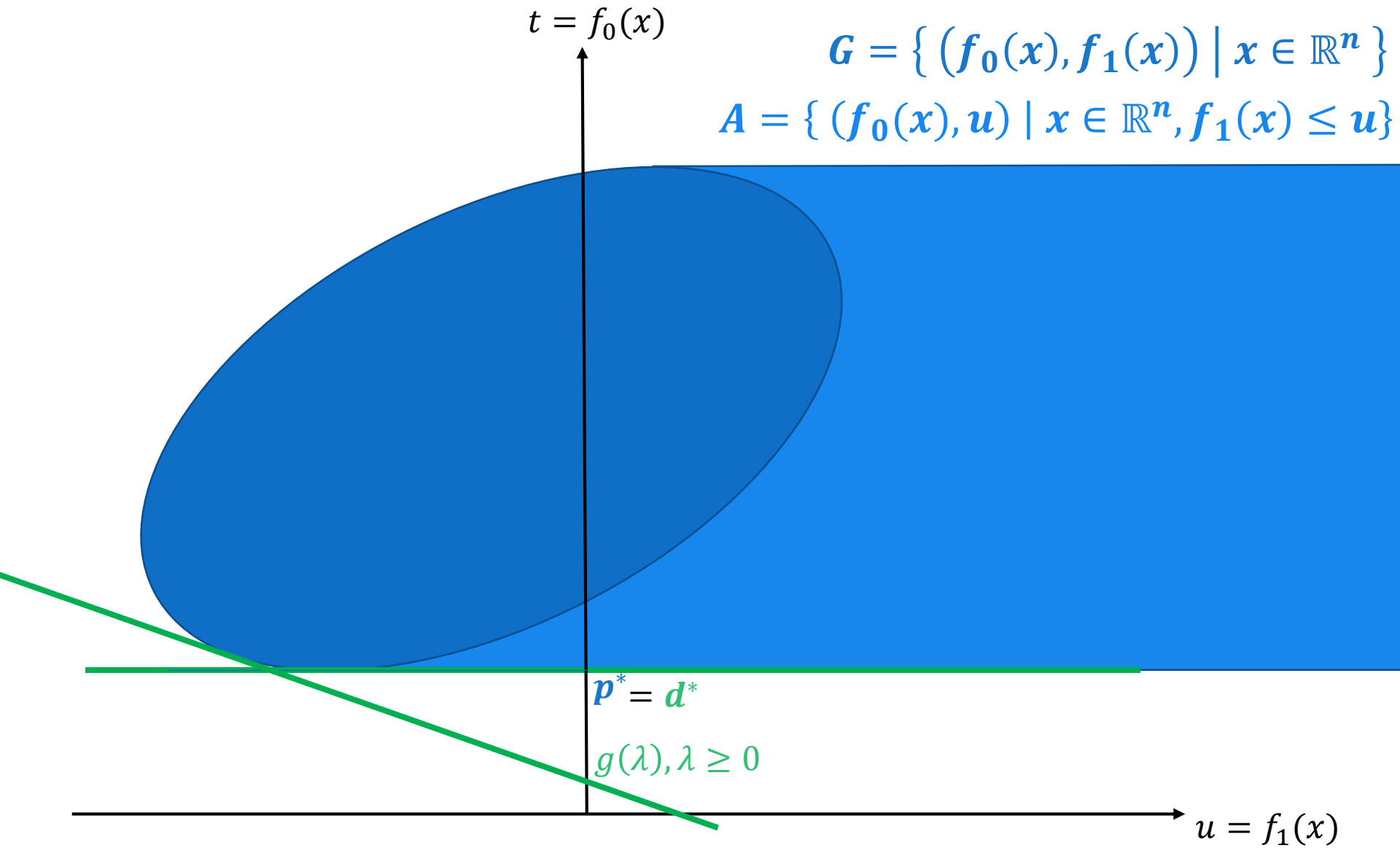


$$\boxed{\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & f_0(x) \\ s.t. & f_1(x) \leq 0 \end{array}}$$

$$g(\lambda) = \inf_x f_0(x) + \lambda f_1(x) = \inf_{(t,u) \in G} \langle (1, \lambda), (t, u) \rangle$$

$$\boldsymbol{G = \{ (f_0(x), f_1(x)) \mid x \in \mathbb{R}^n \}}$$

$$\boldsymbol{A = \{ (f_0(x), u) \mid x \in \mathbb{R}^n, f_1(x) \leq u \}}$$



Slater Not Satisfied

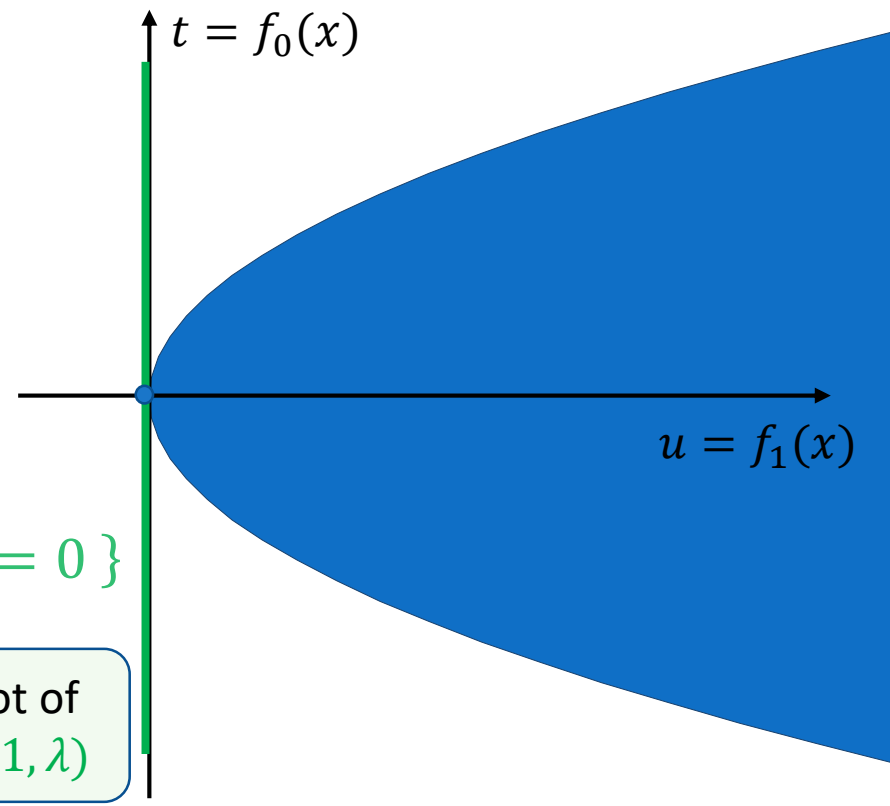
$$\min_{x,y\in\mathbb{R}} \quad x$$
$$s.t. \quad x^2 + y^2 \leq 0$$

$p^* = 0$

$$\max_{\lambda\in\mathbb{R}} \quad -1/4\lambda$$
$$s.t. \quad \lambda \geq 0$$

$d^* = 0$

but not attained



$\{\ (t,u) \mid \langle(\mathbf{0},\lambda),(t,u)\rangle = 0\ \}$

normal not of the form $(1,\lambda)$

Slater Not Satisfied

$$\min_{x,y\in\mathbb{R}} \quad x \qquad s.t. \quad x^2 + y^2 \leq 0$$

$p^* = 0$

$$\max_{\lambda\in\mathbb{R}} \quad -1/4\lambda \qquad s.t. \quad \lambda \geq 0$$

$d^* = 0$

but not attained

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

$p^* = 1$

$$\max_{\lambda\in\mathbb{R}} \quad 0 \qquad s.t. \quad \lambda \geq 0$$

$d^* = 0$

$$\begin{array}{lll} \min\limits_{x \in \mathbb{R}^n} & f_0(x) & \\ s.t. & f_i(x) \leq 0 & i = 1..m \\ & h_i(x) = 0 & j = 1..p \end{array}$$

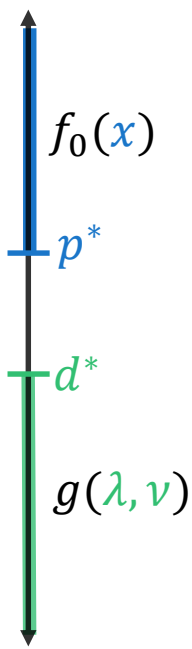
$$\begin{array}{ll} \max\limits_{\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p} & g(\lambda, \nu) \\ s.t. & \lambda_i \geq 0 \end{array}$$

$$L\big(x, (\lambda, \nu)\big) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \nu_j h_j(x)$$

$$g(\lambda, \nu) = \inf_x L\big(x, (\lambda, \nu)\big)$$

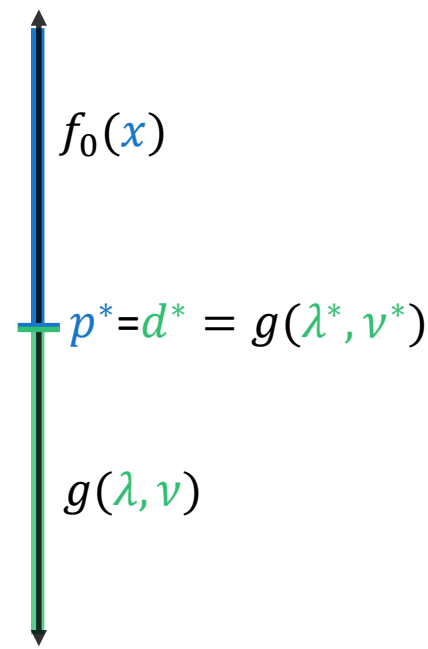
Weak Duality (always holds):

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



Convex + Slater ➔ Strong Duality

$$d^* = p^*$$



Example: Min Norm Solution

$$\boxed{\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & \frac{1}{2}\|x\|_2^2 \\ s.t. & Ax = b \end{array}}$$

$\nu \in \mathbb{R}^m$

$A \in \mathbb{R}^{m \times n}, m < n$

$$\boxed{\max_{\nu \in \mathbb{R}^m} \quad -\tfrac{1}{2}\nu^\top (AA^\top)\nu - \nu^\top b}$$

$$\nu^* = -(AA^\top)^{-1} b$$

$$L(x, \nu) = \tfrac{1}{2}x^\top x + \nu^\top (Ax - b) = \tfrac{1}{2}x^\top x + \nu^\top Ax - \nu^\top b$$

$$g(\nu) = \inf_x L(x, \nu) = \tfrac{1}{2}(-\nu^\top A)(-A\nu) + \nu^\top A(-A\nu) - \nu^\top b$$

$$= -\tfrac{1}{2}\nu^\top (AA^\top)\nu - \nu^\top b$$

$$0 = \nabla_x L(x, \nu) = x^\top + \nu^\top A \qquad \Rightarrow x = -A^\top \nu$$

$$p^* = d^* = \tfrac{1}{2}b^\top (AA^\top)^{-1} b$$

$$\begin{array}{lll} \min\limits_{x\in\mathbb{R}^n} & f_0(x) & \\ s.t. & f_i(x) \le 0 & i = 1..m \\ & h_i(x) = 0 & j = 1..p \end{array}$$

$$\begin{array}{ll} \max\limits_{\lambda\in\mathbb{R}^m, \nu\in\mathbb{R}^p} & g(\lambda,\nu) \\ s.t. & \lambda_i \ge 0 \end{array}$$

$$L\big(x,(\lambda,\nu)\big) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \nu_j h_j(x)$$

$$g(\lambda,\nu) = \inf_x L\big(x,(\lambda,\nu)\big)$$

Strong Duality ($p^* = d^*$) and primal x^* and dual (λ^*,ν^*) opt attained ➔

$$f_0(x^*) = g(\lambda^*,\nu^*) = \inf_x \; f_0(x) + \sum \lambda_i^* f_i(x) + \sum \nu_j^* h_j(x)$$

$$\le \quad f_0(x^*) + \underbrace{\sum \lambda_i^* f_i(x^*)}_{\le 0} + \cancel{\sum \nu_j^* h_j(x^*)} \le f_0(x^*)$$

➔ $x^* = \arg\min L\big(x,(\lambda^*,\nu^*)\big)$

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

Convex problem, $\lambda^* \ge 0$

➔ $L\big(x,(\lambda^*,\nu^*)\big)$ convex in x

Example: Min Norm Solution

$$\boxed{\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & \frac{1}{2}\|x\|_2^2 \\ s.t. \quad & Ax = b \end{aligned}}$$

$A \in \mathbb{R}^{m \times n}, m < n$

$$\boxed{\max_{\nu \in \mathbb{R}^m} \quad -\frac{1}{2}\nu^\top (AA^\top)\nu - \nu^\top b}$$

$$\nu^* = -(AA^\top)^{-1} b$$

$$L(x, \nu) = \frac{1}{2}x^\top x + \nu^\top (Ax - b) = \frac{1}{2}x^\top x + \nu^\top Ax - \nu^\top b$$

$$0 = \nabla_x L(x^*, \nu^*) = \nabla_x \left(\frac{1}{2}{x^*}^\top x^* + (-(AA^\top)^{-1} b)^\top (Ax^* - b) \right)$$

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

$$\Rightarrow x^* = A^\top (AA^\top)^{-1} b$$

$$\begin{array}{lll} \min\limits_{x\in\mathbb{R}^n} & f_0(x) & \\ s.t. & f_i(x) \le 0 & i = 1..m \\ & h_i(x) = 0 & j = 1..p \end{array}$$

$$\begin{array}{ll} \max\limits_{\lambda\in\mathbb{R}^m, \nu\in\mathbb{R}^p} & g(\lambda,\nu) \\ s.t. & \lambda_i \ge 0 \end{array}$$

$$L\big(x,(\lambda,\nu)\big) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \nu_j h_j(x)$$

$$g(\lambda,\nu) = \inf_x L\big(x,(\lambda,\nu)\big)$$

Strong Duality ($p^* = d^*$) and primal x^* and dual (λ^*, ν^*) opt attained ➔

$$f_0(x^*) = g(\lambda^*,\nu^*) = \inf_x \; f_0(x) + \sum \lambda_i^* f_i(x) + \sum \nu_j^* h_j(x)$$

$$\le \quad f_0(x^*) + \sum \underbrace{\lambda_i^* f_i(x^*)}_{\le 0} + \cancel{\sum \nu_j^* h_j(x^*)} \le f_0(x^*)$$

➔for all i, $\lambda_i^* f_i(x^*) = 0$

$$\left.\begin{array}{l} \lambda_i^* > 0 \;\Rightarrow\; f_i(x^*) = 0 \\ f_i(x^*) < 0 \;\Rightarrow\; \lambda_i^* = 0 \end{array}\right\} \text{Complementary Slackness}$$

For all optimal x^* and dual optimal (λ^*, ν^*)

Example: Max Flow

$$\begin{array}{lll} \max\limits_{x\in\mathbb{R}^{n\times n}, x\geq 0} & x_{n1} & \\ s.t. & x_{ij} \leq C_{ij} \quad \forall_{ij\neq n1} & \lambda_{ij} \\ & \sum_j x_{ji} = \sum_k x_{ik} \quad \forall_i & \nu_i \end{array}$$

$$\begin{array}{ll} \min\limits_{\lambda,\nu} & \sum_{ij\neq n1} \lambda_{ij} C_{ij} \\ 0 \leq \lambda_{ij} & \nu_i - \nu_j \leq \lambda_{ij} \quad \forall_{ij\neq n1} \\ & \nu_n \leq \nu_1 - 1 \end{array}$$

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

$$L(x,(\lambda,\nu)) = x_{n1} + \sum_{ij\neq n1} \lambda_{ij}(C_{ij} - x_{ij}) + \sum_i \nu_i \left(\sum_j x_{ji} - \sum_k x_{ik}\right)$$

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

$$= \sup_{x\geq 0} \sum_{ij\neq n1} (-\lambda_{ij} + \nu_i - \nu_j) x_{ij} + (1 + \nu_n - \nu_1) x_{n1} + \sum_{ij\neq n1} \lambda_{ij} C_{ij}$$

$$= \begin{cases} \sum \lambda_{ij} C_{ij} & \text{If } (1 + \nu_n - \nu_1) \leq 0 \text{ and } \forall_{ij\neq n1}(-\lambda_{ij} + \nu_i - \nu_j) \leq 0 \\ +\infty & \text{otherwise} \end{cases}$$

Either $x_{ij} = C_{ij}$ or $\lambda_{ij} = 0$

Example: Max Flow

$$\max_{x \in \mathbb{R}^{n \times n}, x \geq 0} \quad x_{n1}$$
$$s.t. \quad x_{ij} \leq C_{ij} \quad \forall_{ij \neq n1}$$
$$\sum_j x_{ji} = \sum_k x_{ik} \quad \forall_i$$

$$\min_{\lambda, \nu} \quad \sum_{ij \neq n1} \lambda_{ij} C_{ij}$$
$$0 \leq \lambda_{ij} \quad \nu_i - \nu_j \leq \lambda_{ij} \quad \forall_{ij \neq n1}$$
$$\nu_n \leq \nu_1 - 1$$

$\Downarrow$

At the optimum:

$$\lambda_i = \max(\nu_i - \nu_j, 0) = [\nu_i - \nu_j]_+$$

$\Downarrow$

$$\min_{\nu} \quad \sum_{ij \neq n1} [\nu_i - \nu_j]_+ C_{ij}$$
$$\nu_n \leq \nu_1 - 1$$

Complimentary slackness:

$\lambda_{ij} = \nu_i - \nu_j > 0$ (edge in cut) ➔ $x_{ij} = C_{ij}$ (edge saturated)

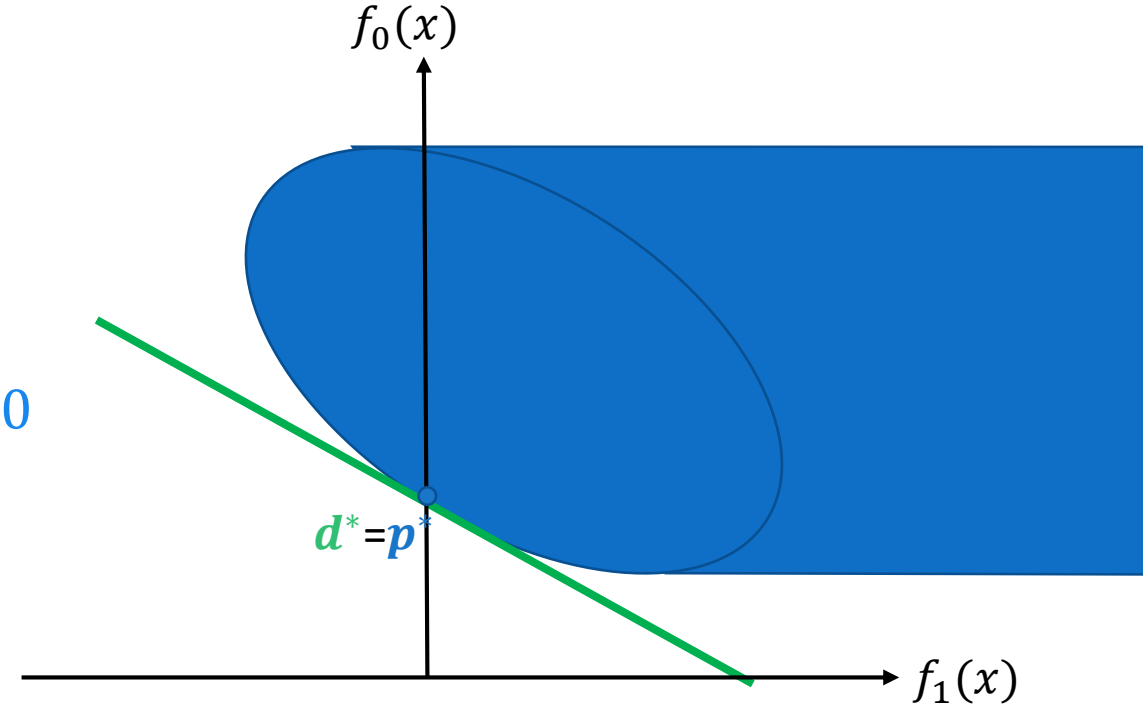
$x_{ij} < C_{ij}$ (not saturated) ➔ $\nu_i = \nu_j$ (not in cut)

Either $x_{ij} = C_{ij}$ or $\lambda_{ij} = 0$

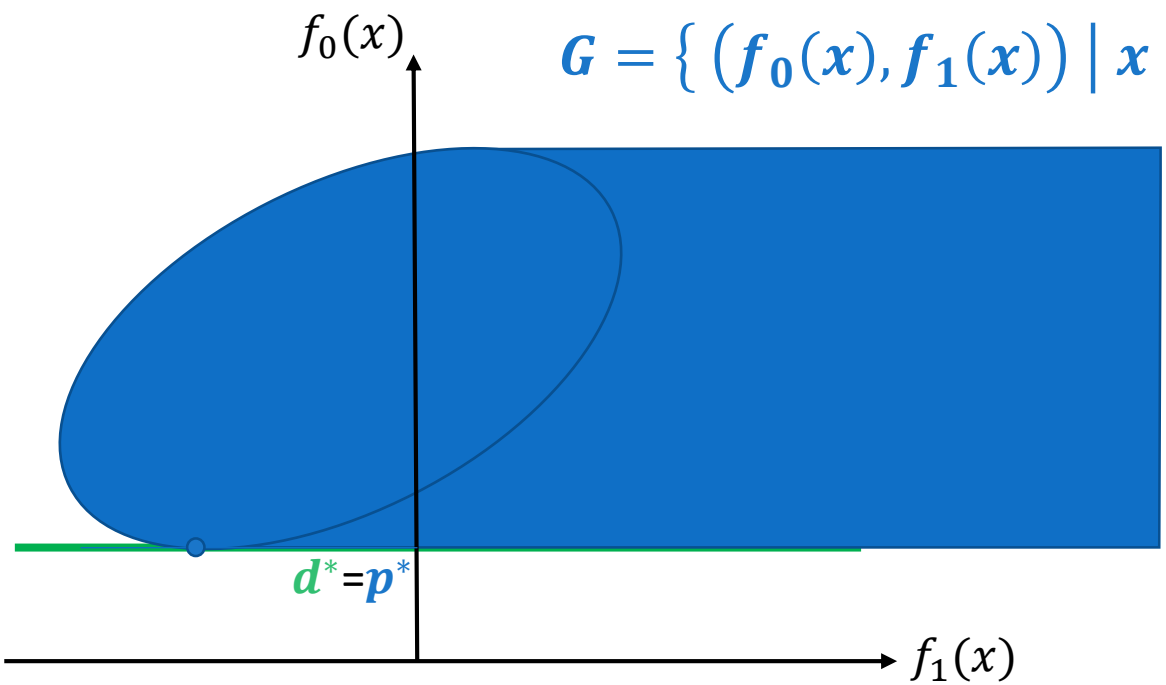
Complementary Slackness

Constraint tight: $f_1(x^*) = 0$

Dual lax: $\lambda^* > 0$



$$\boldsymbol{G} = \{ (\boldsymbol{f_0}(\boldsymbol{x}), \boldsymbol{f_1}(\boldsymbol{x})) \mid \boldsymbol{x} \in \mathbb{R}^{\boldsymbol{n}} \}$$



Constraint lax: $f_1(x^*) < 0$

Dual tight: $\lambda^* = 0$

What do we know about the opt? (assuming Strong Duality and Convexity)

x^* optimal for (P) $\Rightarrow$

λ^*, ν^* optimal for (D) $\Leftarrow$

KKT Conditions

$f_i(x^*) \leq 0 \; \forall_{i=1\ldots m}$

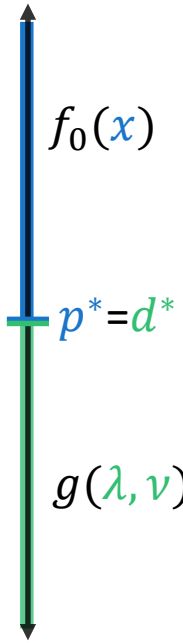
$h_j(x^*) = 0 \; \forall_{j=1\ldots p}$

$\lambda_i^* \geq 0 \; \forall_{i=1..m}$

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

or $0 \in \partial_x L\big(x^*, (\lambda^*, \nu^*)\big)$

$\lambda_i^* f_i(x^*) = 0 \; \forall_{i=1..m}$



If conditions hold:

$\lambda^* \geq 0$ ➔ $L\big(x, (\lambda^*, \nu^*)\big)$ is convex in x ➔ $x^* = \arg\min L\big(x, (\lambda^*, \nu^*)\big)$

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

$$= f_0(x^*) + \cancel{\textstyle\sum_i \lambda_i^* f_i(x^*)} + \cancel{\textstyle\sum_i \nu_i^* h_i(x^*)} = f_0(x^*)$$

➔ x^* optimal for (P) and λ^*, ν^* optimal for (D)

What do we know about the opt? (assuming Strong Duality and Convexity)

x^* optimal for (P)

λ^*, ν^* optimal for (D)

$\Rightarrow$

$\Leftarrow$

<u>KKT Conditions</u>

$$f_i(x^*) \leq 0 \; \forall_{i=1\ldots m}$$

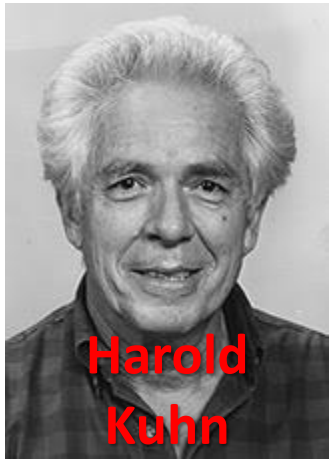
$$h_j(x^*) = 0 \; \forall_{j=1\ldots p}$$

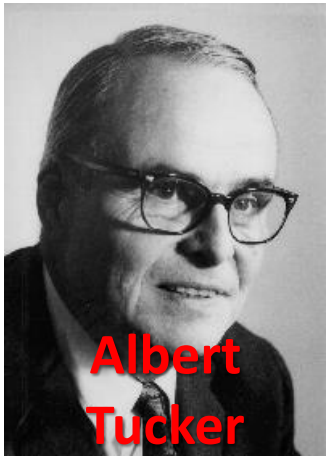
$$\lambda_i^* \geq 0 \; \forall_{i=1..m}$$

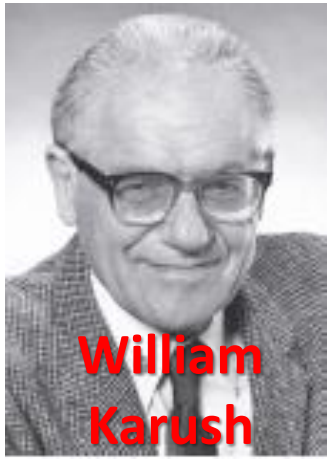
$$\nabla_x L\big(x^*, (\lambda^*, \nu^*)\big) = 0$$

$$\text{or } 0 \in \partial_x L\big(x^*, (\lambda^*, \nu^*)\big)$$

$$\lambda_i^* f_i(x^*) = 0 \; \forall_{i=1..m}$$







What do we know about the opt? (assuming Strong Duality and Convexity)

x^* optimal for (P) $\Rightarrow$

λ^*, ν^* optimal for (D) $\Leftarrow$

<u>KKT Conditions</u>

$f_i(x^*) \leq 0 \; \forall_{i=1\ldots m}$

$h_j(x^*) = 0 \; \forall_{j=1\ldots p}$

$\lambda_i^* \geq 0 \; \forall_{i=1..m}$

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

or $0 \in \partial_x L\big(x^*, (\lambda^*, \nu^*)\big)$

$\lambda_i^* f_i(x^*) = 0 \; \forall_{i=1..m}$

x^* optimal for (P) $\Leftrightarrow \exists_{\lambda^*, \mu^*}$ s.t. x^*, λ^*, μ^* satisfy KKT

Optimality Condition

- For general convex constrained problems (+Slater):

$$x^* \text{ optimal for (P)} \Leftrightarrow \exists_{\lambda^*,\mu^*} \text{ s.t. } x^*, \lambda^*, \mu^* \text{ satisfy KKT}$$

- For unconstrained:

  Slater always holds if feasible, and KKT reduces to:

$$\nabla f_0(x^*) = 0$$

- With only equality constraints:

$$\min_x f_0(x) \text{ s.t. } Ax = b$$

  KKT reduces to:

$$Ax = b \text{ and } \nabla f_0(x^*) + A^\top \nu = 0$$

$\sum_i \nu_i a_i$

  i.e.:

$$Ax = b \text{ and } \nabla f_0(x) \in image(A^\top) = rowspan(A)$$

Example: Allocation with Decaying Benefits

$$\boxed{\begin{array}{ll} \min & -\sum_{i=1}^{n} \log(\alpha_i + x_i) \\ \text{s.t.} & x_i \geq 0 \\ & \sum_{i=1}^{n} x_i = 1 \end{array}} \quad \begin{array}{l} \\ \lambda_i \; i = 1..n \\ \nu \end{array}$$

$$L(x, (\lambda, \nu)) = -\sum \log(\alpha_i + x_i) - \sum \lambda_i x_i + \nu(\sum x_i - 1)$$

$$0 = \frac{\partial}{\partial x_i} L(x^*, (\lambda^*, \nu^*)) = -\frac{1}{\alpha_i + x_i^*} - \lambda_i^* + \nu^* \Rightarrow \lambda_i^* = \nu^* - \frac{1}{\alpha_i + x_i^*}$$

KKT: $x_i^* \geq 0 \quad \sum x_i^* = 1 \quad \lambda_i^* \geq 0 \quad \lambda_i^* x_i^* = 0 \quad \lambda_i^* = \nu^* - \frac{1}{\alpha_i + x_i^*}$

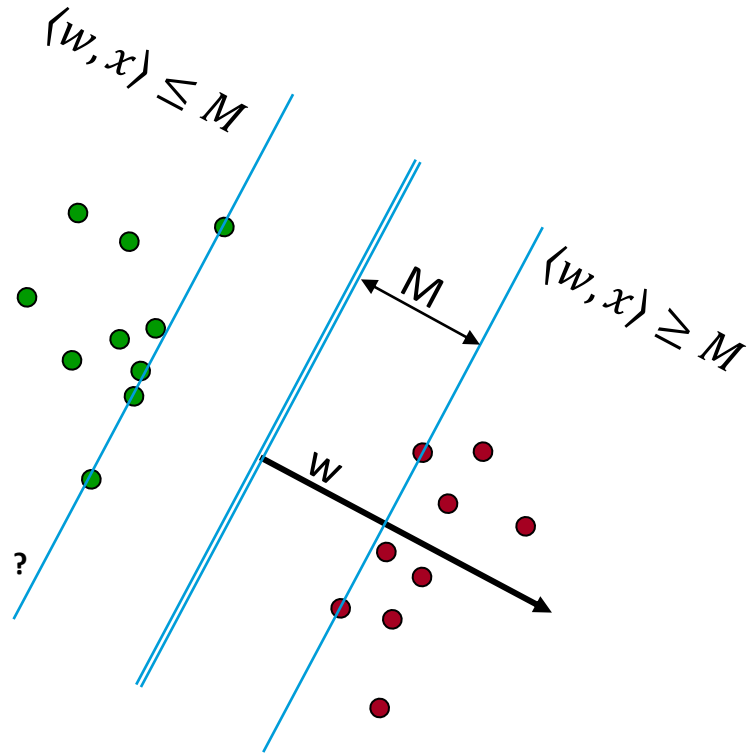
$x_i^* \geq 0 \quad \sum x_i^* = 1 \quad \nu^* \geq \frac{1}{\alpha_i + x_i^*} \quad x_i^* \left(\nu^* - \frac{1}{\alpha_i + x_i^*}\right) = 0$

$$\left.\begin{array}{l} \text{If } \nu^* < \frac{1}{\alpha_i} \Rightarrow x_i^* > 0 \Rightarrow x_i^* = \frac{1}{\nu^*} - \alpha_i \\ \text{If } \nu^* \geq \frac{1}{\alpha_i} \Rightarrow x_i^* = 0 \end{array}\right\} \Rightarrow x_i^* = \left[\frac{1}{\nu^*} - \alpha_i\right]_+$$

ν^* is unique threshold s.t. $\sum_i \left[\frac{1}{\nu^*} - \alpha_i\right]_+ = 1$

Example: Large Margin Discrimination

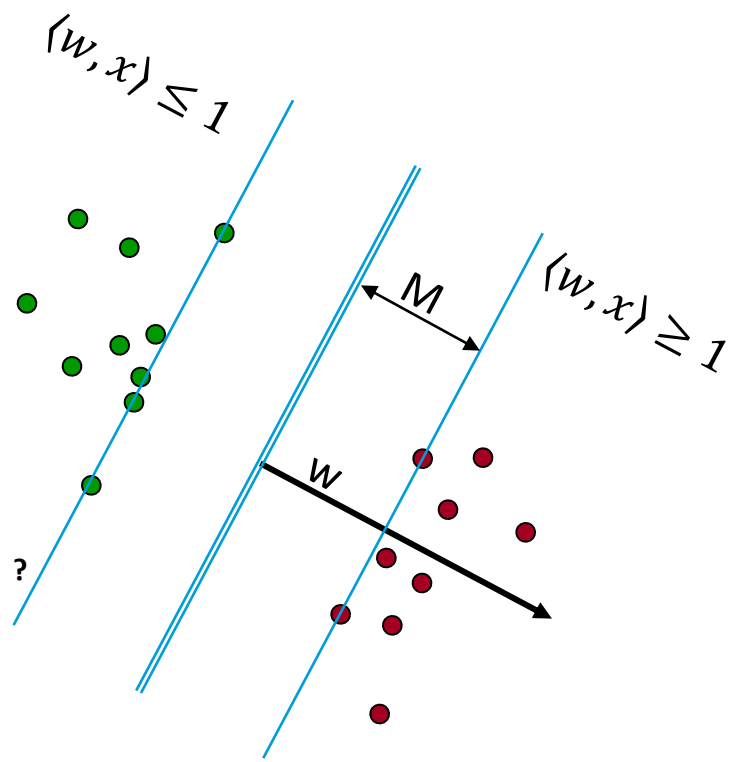
$$\begin{array}{ll} \min\limits_{w \in \mathbb{R}^n} & \frac{1}{2}\|w\|^2 \\ \text{s.t.} & y_i \langle w, x_i \rangle \geq 1 \quad i = 1..m \end{array}$$



|w|=1

Example: Large Margin Discrimination

$$\begin{array}{ll} \min\limits_{w \in \mathbb{R}^n} & \frac{1}{2}\|w\|^2 \\ \text{s.t.} & y_i \langle w, x_i \rangle \geq 1 \quad i = 1..m \end{array}$$



Margin: M = 1/|w|

Example: Large Margin Discrimination

$$\boxed{\begin{array}{ll} \min\limits_{w \in \mathbb{R}^n} & \frac{1}{2}\|w\|^2 \\ \text{s.t.} & y_i\langle w, x_i\rangle \geq 1 \quad i = 1..m \end{array}} \quad \lambda_i$$

$$L(w, \lambda) = \frac{1}{2}\|w\|^2 + \sum_{i=1}^{m} \lambda_i (1 - y_i\langle w, x_i\rangle)$$

KKT:

$$y_i\langle w^*, x_i\rangle \geq 1$$

$$\lambda_i^* \geq 0$$

$$\lambda_i^*(1 - y_i\langle w^*, x_i\rangle) = 0$$

$$w^* = \sum_{i=1}^{m} \lambda_i^* y_i x_i$$

$$0 = \nabla_w L(w^*, \lambda^*) = w^* - \sum_i \lambda_i^* y_i x_i$$

Optimality Condition (assuming Strong Duality)

x^* optimal for (P) $\Rightarrow$

λ^*, ν^* optimal for (D) $\Leftarrow$

<u>KKT Conditions</u>

$f_i(x^*) \leq 0 \; \forall_{i=1\ldots m}$

$h_j(x^*) = 0 \; \forall_{j=1\ldots p}$

$\lambda_i^* \geq 0 \; \forall_{i=1..m}$

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

or $0 \in \partial_x L\big(x^*, (\lambda^*, \nu^*)\big)$

$\lambda_i^* f_i(x^*) = 0 \; \forall_{i=1..m}$

x^* optimal for (P) $\Leftrightarrow \exists_{\lambda^*, \mu^*}$ s.t. x^*, λ^*, μ^* satisfy KKT