Convex Optimization

Prof. Nati Srebro

Lecture 6:

Constrained Optimization

Lagrangian Duality

Reading: Boyd and Vanndenberghe 4.1-4.4,5.1-5.2,5.5.1

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

$$
s.t. \quad f_i(x) \leq 0 \qquad i = 1 \dots m
$$

$$
h_j(x) = 0 \qquad j = 1 \dots p \qquad \begin{cases} h_j(x) \leq 0 \\ -h_j(x) \leq 0 \end{cases}
$$

$$
f_0, f_1, \dots, f_m : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\} \quad \text{convex}
$$

$$
h_j(x) : \mathbb{R}^n \to \mathbb{R} = \langle a_j, x \rangle - b \text{ linear}
$$

$$
\Downarrow
$$

Feasible set $\left\{x \middle| f_i(x) \leq 0, h_j(x) = 0, f_0(x) < \infty\right\}$ convex

$$
\begin{aligned}
(P) \quad & \min_{x \in \mathbb{R}^n} && f_0(x) \\
& s.t. && f_i(x) \leq 0 \qquad i = 1 \ldots m \\
& && h_j(x) = 0 \qquad j = 1 \ldots p \\
& && Ax = b
\end{aligned}
$$

$f_0, f_1, \ldots, f_m : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ convex

$h_j(x) : \mathbb{R}^n \to \mathbb{R} = \langle a_j, x \rangle - b$ linear

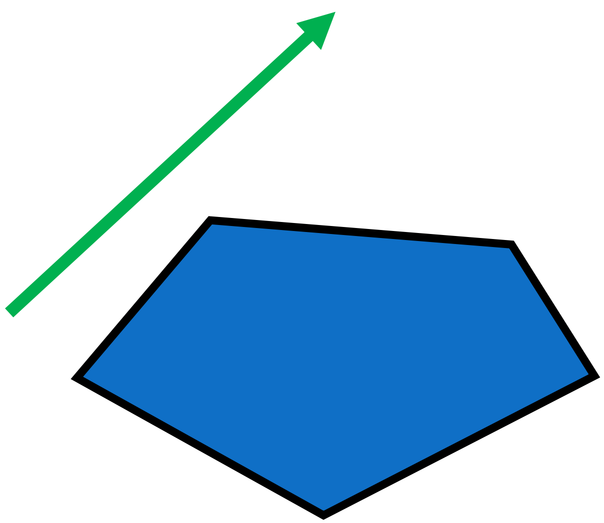
$A \in \mathbb{R}^{p \times n}, b \in \mathbb{R}^p$

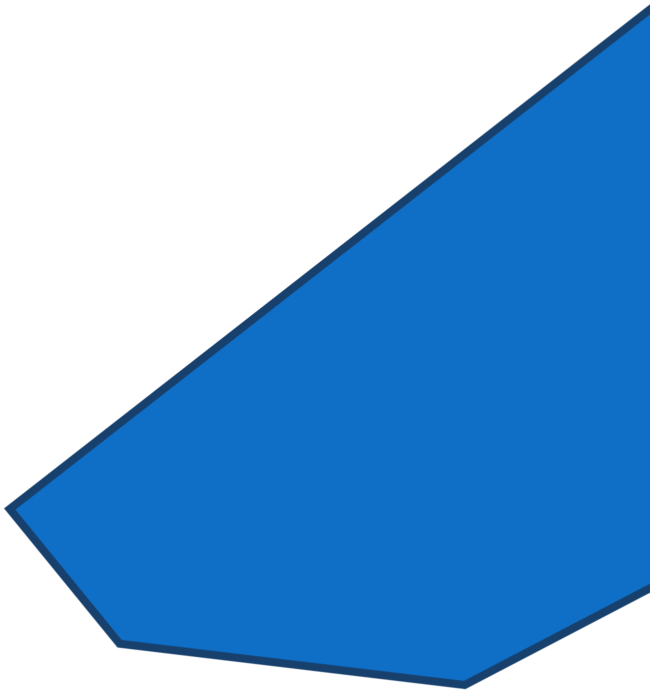
$$\Downarrow$$

Feasible set $\left\{ x \middle| f_i(x) \leq 0, h_j(x) = 0, f_0(x) < \infty \right\}$ convex

Example: Linear Programing (LP)

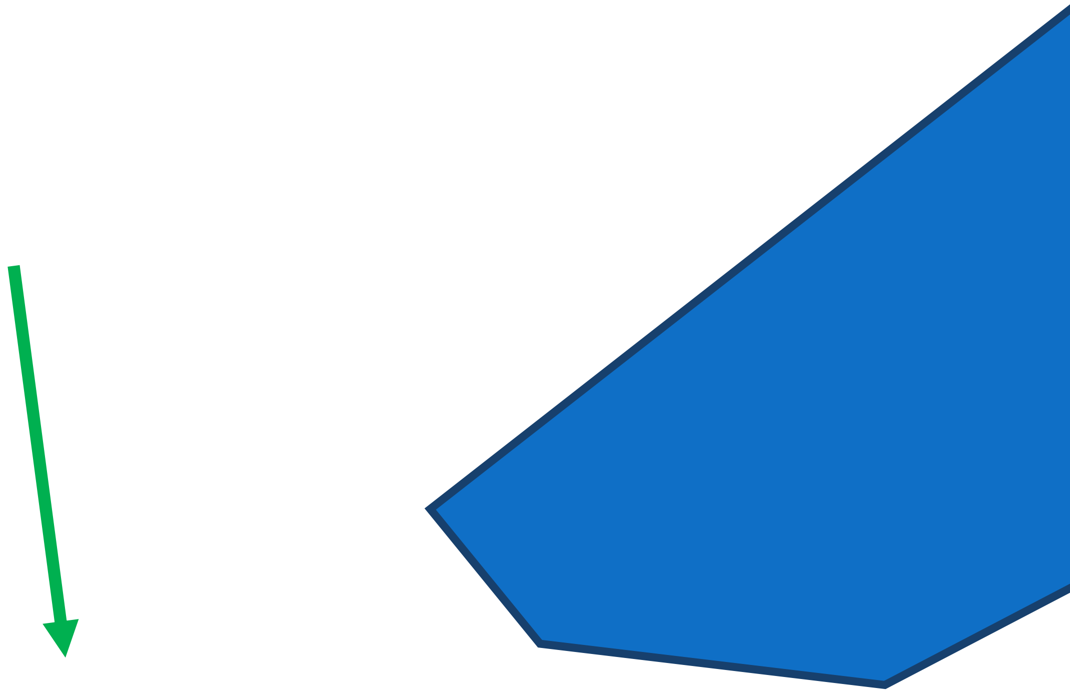
$$
\begin{array}{llll}
(P) & \min\limits_{x \in \mathbb{R}^n} & c^\top x & \\
 & s.t. & Gx \leq h & G \in \mathbb{R}^{m \times n}, h \in \mathbb{R}^m \\
 & & Ax = b & A \in \mathbb{R}^{p \times n}, b \in \mathbb{R}^p
\end{array}
$$





Example: Linear Programing (LP)

$$
\begin{array}{llll}
(P) & \min\limits_{x \in \mathbb{R}^n} & c^\top x & \\
& s.t. & Gx \leq h & G \in \mathbb{R}^{m \times n}, h \in \mathbb{R}^m \\
& & Ax = b & A \in \mathbb{R}^{p \times n}, b \in \mathbb{R}^p
\end{array}
$$



Example of LP: Max Flow

- Vertices: 1…n, source node 1, sink n
- Capacities: C_{ij} on edge $i \to j$

$$
\begin{array}{lll}
\max\limits_{x} & \sum_i x_{1i} & \\
\text{s.t.} & 0 \leq x_{ij} \leq C_{ij} & \forall ij \\
 & \sum_j x_{ji} = \sum_k x_{ik} & \forall i = 2 \ldots n-1
\end{array}
$$

Example LP: Piecewise Linear Minimization

Unconstrained non-smooth objective:

$$\min_{x \in \mathbb{R}^n} f(x) \qquad\qquad f(x) = \max_{1 \le i \le m} a_i^T x + b_i$$

Equivalent LP:

$$\begin{array}{lll} \min\limits_{x \in \mathbb{R}^n, t \in \mathbb{R}} & t & \\ \text{s.t.} & a_i^T x + b_i \le t & \forall i \end{array}$$

Smooth (even linear!) objective and constraints

Example LP: ℓ_1 Regression

Unconstrained non-smooth objective:

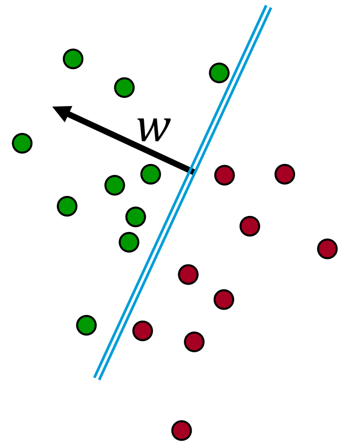
$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^{m} \left| a_i^T x - b_i \right|$$

Equivalent LP:

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n, t \in \mathbb{R}^m} & \sum_{i=1}^{m} t_i \\ \text{s.}t. & t_i \leq a_i^T x - b_i \leq t_i \qquad \forall i = 1 \ldots m \end{array}$$

Example: Linear Separation

- Data: $\mathrm{X}^+ = \{x_1^+, x_2^+, \dots, x_{m^+}^+\}, \mathrm{X}^- = \{x_1^-, x_2^-, \dots, x_{m^-}^-\} \subset \mathbb{R}^n$
- Find hyperplane that separates X^+ and X^-



$$\begin{array}{ll} \text{find} & w \in \mathbb{R}^n, b \in \mathbb{R} \\ \text{s.t.} & w^T x_i^+ \geq b + 1 \quad \forall x_i^+ \in X^+ \\ & w^T x_i^- \leq b - 1 \quad \forall x_i^- \in X^- \end{array}$$

Feasibility problem:

$$\begin{array}{lll} \text{find} & x & \\ \text{s.t.} & f_i(x) \leq 0 & i = 1 \dots m \\ & h_j(x) = 0 & j = 1 \dots p \end{array}$$

Can be thought of as:

$$\begin{array}{lll} \min_x & 0 & \\ \text{s.t.} & f_i(x) \leq 0 & i = 1 \dots m \\ & h_j(x) = 0 & j = 1 \dots p \end{array}$$

Reducing Optimization to Feasibaility

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

Perform binary search over t and solve:

$$\begin{aligned} \text{find} \quad & x \in \mathbb{R}^n & \\ s.t. \quad & f_0(x) \leq t & \\ & f_i(x) \leq 0 & i = 1..m \\ & h_i(x) = 0 & j = 1..p \end{aligned}$$

Example: Quadratic Programming

$S_n^+ = \{\mathbb{R}^{n\times n} \text{ symetric p.s.d.}\}$

$$\begin{array}{lll} \min\limits_{x} & \frac{1}{2}x^T P x + q^T x & P \in S_+^n, q \in \mathbb{R}^n \\ \text{s.t.} & Gx \leq h & G \in \mathbb{R}^{m\times n}, h \in \mathbb{R}^m \\ & Ax = b & A \in \mathbb{R}^{p\times n}, b \in \mathbb{R}^p \end{array}$$

$P \succcurlyeq 0$ ➔ Problem is convex

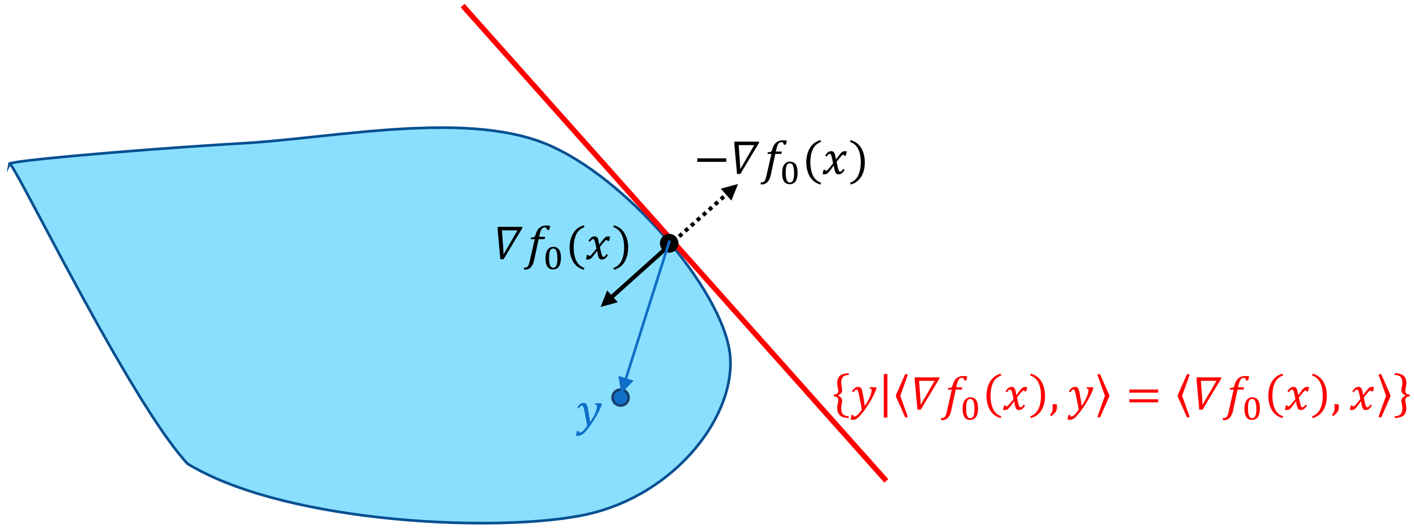
Example: least squares with box constraints:

$$\begin{array}{ll} \min\limits_{x\mathbb{R}^n} & \frac{1}{2}\|Ax - b\|_2^2 = \frac{1}{2}x^\top(A^\top A)x - (b^\top A)x - const \\ \text{s.t.} & l_i \leq x_i \leq u_i \end{array}$$

Optimality Condition

Claim: x is optimal for (P) iff it is feasible and for some $\nabla f_0(x) \in \partial f_0(x)$

$$\forall \text{ feasible } y, \qquad \langle \nabla f_0(x), y - x \rangle \geq 0$$



Either $\nabla f_0(x) = 0$, and x could be in the interior of the feasible set,

Or, if $\nabla f_0(x) \neq 0$, $\{y | \langle \nabla f_0(x), y \rangle = \langle \nabla f_0(x), x \rangle\}$ is a supporting hyperplane of the feasible set.

x is optimal $\Leftrightarrow$ x feasible and $\forall$ feasible y, $\langle \nabla f_0(x), y - x \rangle \geq 0$

- Minimization over non-negative orthant:

$$\boxed{\begin{array}{lll} \min\limits_{x \in \mathbb{R}^n} & f_0(x) & \\ s.t. & x[i] \geq 0 & i = 1..n \end{array}}$$

x optimal iff $x \geq 0$ and $\quad x[i] > 0 \rightarrow \nabla f(x)[i] = 0$

$\quad x[i] = 0 \rightarrow \nabla f(x)[i] \geq 0$

- Unconstrained Optimization:

$$\boxed{\min\limits_{x \in \mathbb{R}^n} \quad f_0(x)}$$

x optimal $\Leftrightarrow \nabla f_0(x) = 0$

x is optimal $\Leftrightarrow$ x feasible and $\forall$ feasible y, $\langle \nabla f_0(x), y - x \rangle \geq 0$

- Equality constrained minimization:

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

x optimal iff $Ax = b$ and $\nabla f(x)^\top (y - x) \geq 0 \;\; \forall\, Ay = b = Ax$

$\Updownarrow$

$Ax = b$ and $\nabla f(x)^\top v = 0 \;\forall\, Av = 0$

$\Updownarrow$

$Ax = b$ and $\nabla f(x) \perp null(A)$

$\Updownarrow$

$Ax = b$ and $\nabla f(x) \in image(A^\top)$

$\Updownarrow$

$\boldsymbol{Ax = b}$ **and** $\boldsymbol{\nabla f(x) = A^\top u}$ **for some** $\boldsymbol{u} \in \mathbb{R}^{\boldsymbol{m}}$

$A(\underbrace{y - x}_{\boldsymbol{v}}) = 0$

And so also

$A(-v) = 0$ and

$\nabla f(x)^T v \leq 0$

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

Lagrange multipliers

$$p^* = \inf_{x\in\mathbb{R}^n} \sup_{\substack{\lambda\in\mathbb{R}^m \\ \nu\in\mathbb{R}^p \\ \lambda_i\ge 0}} \underbrace{f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \nu_j h_j(x)}_{L(x,(\lambda,\nu))\ \text{Lagrangian}}$$

- $h_j(x) = 0$
- $f_i(x) \le 0 \Rightarrow \lambda_i f_i(x) \le 0$

$$\Rightarrow \sup_{\lambda_i\ge 0} \lambda_i f_i(x) = 0$$

$$\sup_{\lambda\ge 0,\nu} L(x,(\lambda,\nu)) = \begin{cases} f_0(x) & x \text{ is feasible} \\ \infty & \text{otherwise} \end{cases}$$

- if $f_i(x) > 0 \quad \Rightarrow \quad \lambda_i \to \infty$
- if $h_i(x) > 0 \quad \Rightarrow \quad \nu_i \to \infty$
- if $h_i(x) < 0 \quad \Rightarrow \quad \nu_i \to -\infty$

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

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

<u>Theorem</u>:

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

<u>Proof</u> (if $x^* = \min_x \sup_{\lambda>0,\nu} L\big(x,(\lambda,\nu)\big)$ exists):

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

<u>Proof</u> (general case):

For any $\tilde{x}$: $\sup_{\lambda\ge 0,\nu} \inf_x L(x,(\lambda,\nu) \le \sup_{\lambda\ge 0,\nu} L\big(\tilde{x},(\lambda,\nu)\big)$. Since this holds for any $\tilde{x}$, we can take an infimum over it and the inequality holds.

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

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

<u>Theorem</u> (Weak Duality):

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

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

Dual Problem:

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

Denote its optimal value d^* (or $d^* = -\infty$ if infeasible),

And its optimum (if exists) λ^*, ν^*

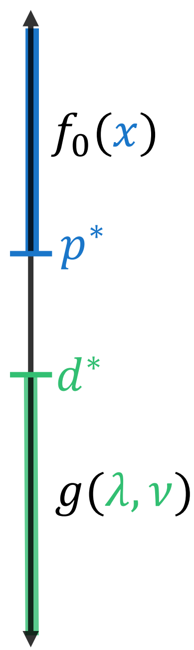
$$
\begin{aligned}
\min_{x \in \mathbb{R}^n} \quad & f_0(x) & \\
s.t. \quad & f_i(x) \le 0 & i = 1..m \\
& h_i(x) = 0 & j = 1..p
\end{aligned}
$$

$$
\begin{aligned}
\max_{\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p} \quad & g(\lambda, \nu) \\
s.t. \quad & \lambda_i \ge 0
\end{aligned}
$$

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

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

Weak Duality (always holds): $d^* \le p^*$



<u>Certificate of sub-optimality:</u>

For *any* feasible x, and *any* feasible (λ, ν):
$f_0(x) - p^* \le f_0(x) - g(\lambda, \nu)$

<u>Certificate of infeasibility:</u>

If we can show a feasible sequence (λ_t, ν_t) s.t.
$g(\lambda_t, \nu_t) \to +\infty$, then (P) is infeasible ($p^* = +\infty$)

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

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

Slater's Condition: $\exists x$ s.t.: $x \in \mathrm{rel{-}interior}\big(dom(f_0)\big)$

(interior relative to the affine subspace spanned by $dom(f_0)$)

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

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

$h_j(x) = 0$

Theorem: 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 dual optimum attainted

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

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

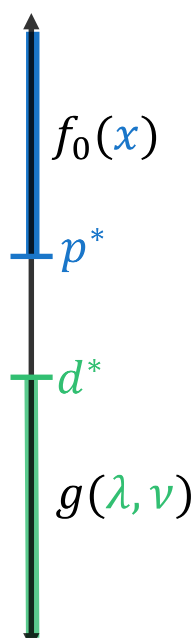
$$\begin{aligned} \max_{\lambda\in\mathbb{R}^m, \nu\in\mathbb{R}^p} \quad & g(\lambda, \nu) \\ s.t. \quad & \lambda_i \ge 0 \end{aligned}$$

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

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

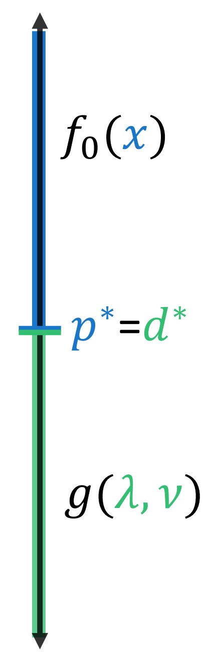
Weak Duality (always holds):

$$d^* \le p^*$$



Convex + Slater ➔ Strong Duality

$$d^* = p^*$$



Example: Linear Programming

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & c^\top x \\ s.t. \quad & Gx \le h \quad G \in \mathbb{R}^{m\times n}, h \in \mathbb{R}^m \\ & Ax = b \quad A \in \mathbb{R}^{p\times n}, b \in \mathbb{R}^p \end{aligned}$$

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

$$L(x, (\lambda, \nu)) = c^\top x + \sum_{i=1}^{m} \lambda_i (g_i^\top x - h_i) + \sum_{j=1}^{p} \nu_j (a_i^\top x - b_i)$$

$$= c^\top x + \lambda^\top (Gx - h) + \nu^\top (Ax - b) = (c^\top + \lambda^\top G + \nu^\top A)x - \lambda^\top h - \nu^\top b$$

$$g(\lambda, \nu) = \inf_x L(x, (\lambda, \nu)) = \begin{cases} -\lambda^\top h - \nu^\top b & \text{If } c^\top + \lambda^\top G + \nu^\top A = 0 \\ -\infty & \text{otherwise} \end{cases}$$

Example: Linear Programming

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & c^\top x \\ s.t. \quad & Gx \le h \quad G \in \mathbb{R}^{m\times n}, h \in \mathbb{R}^m \\ & Ax = b \quad A \in \mathbb{R}^{p\times n}, b \in \mathbb{R}^p \end{aligned}$$

$$\begin{aligned} \max_{\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p} \quad & -h^\top \lambda - b^\top \nu \\ s.t. \quad & G^\top \lambda + A^\top \nu + c = 0 \\ & \lambda \ge 0 \end{aligned}$$

$$L\big(x, (\lambda, \nu)\big) = c^\top x + \sum_{i=1}^{m} \lambda_i (g_i^\top x - h_i) + \sum_{j=1}^{p} \nu_j \big(a_i^\top x - b_i\big)$$

$$= c^\top x + \lambda^\top (Gx - h) + \nu^\top (Ax - b) = (c^\top + \lambda^\top G + \nu^\top A)x - \lambda^\top h - \nu^\top b$$

$$g(\lambda, \nu) = \inf_x L\big(x, (\lambda, \nu)\big) = \begin{cases} -\lambda^\top h - \nu^\top b & \text{If } c^\top + \lambda^\top G + \nu^\top A = 0 \\ -\infty & \text{otherwise} \end{cases}$$

Example: Linear Programming

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & c^\top x \\ s.t. \quad & Gx \le h \quad G \in \mathbb{R}^{m\times n}, h \in \mathbb{R}^m \\ & Ax = b \quad A \in \mathbb{R}^{p\times n}, b \in \mathbb{R}^p \end{aligned}$$

$$\begin{aligned} \max_{\lambda\in\mathbb{R}^m,\nu\in\mathbb{R}^p} \quad & -h^\top\lambda - b^\top\nu \\ s.t. \quad & G^\top\lambda + A^\top\nu + c = 0 \\ & \lambda \ge 0 \end{aligned}$$

$$\begin{aligned} \max_{z\in\mathbb{R}^n,w\in\mathbb{R}^m} \quad & c^\top z \\ s.t. \quad & \begin{bmatrix} -I \\ 0 \end{bmatrix} w + \begin{bmatrix} G \\ A \end{bmatrix} z + \begin{bmatrix} h \\ b \end{bmatrix} = 0 \\ & w \ge 0 \end{aligned}$$

$Gz + h = w \ge 0$

$$\begin{aligned} \min_{\lambda\in\mathbb{R}^m,\nu\in\mathbb{R}^p} \quad & \begin{bmatrix} h^\top & b^\top \end{bmatrix} \begin{bmatrix} \lambda \\ \nu \end{bmatrix} \\ s.t. \quad & \begin{bmatrix} -I & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ \nu \end{bmatrix} \le 0 \quad w \\ & \begin{bmatrix} G^\top & A^\top \end{bmatrix} \begin{bmatrix} \lambda \\ \nu \end{bmatrix} = -c \quad z \end{aligned}$$

$x = -z$

$$\begin{aligned} \max_{x\in\mathbb{R}^n} \quad & -c^\top x \\ s.t. \quad & Gx \le h \quad G \in \mathbb{R}^{m\times n}, h \in \mathbb{R}^m \\ & Ax = b \quad A \in \mathbb{R}^{p\times n}, b \in \mathbb{R}^p \end{aligned}$$

Example: Linear Programming

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & c^\top x \\ s.t. \quad & Gx \le h \quad G \in \mathbb{R}^{m \times n}, h \in \mathbb{R}^m \\ & Ax = b \quad A \in \mathbb{R}^{p \times n}, b \in \mathbb{R}^p \end{aligned}$$

$$\begin{aligned} \max_{\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p} \quad & -h^\top \lambda - b^\top \nu \\ s.t. \quad & G^\top \lambda + A^\top \nu + c = 0 \\ & \lambda \ge 0 \end{aligned}$$

Strong Duality?

- (P) is feasible ➔ Slater ➔ Strong Duality

  i.e. $p^* < +\infty$ ➔ $p^* = d^*$
- $d^* > -\infty$ ➔ (D) is feasible ➔ Slater ➔ $p^* = d^*$
- But: it is possible to have both infeasible and $-\infty = d^* < p^* = +\infty$

$$\begin{aligned} \min_{x \in \mathbb{R}} \quad & x \\ s.t. \quad & \begin{bmatrix} 0 \\ 1 \end{bmatrix} x \le \begin{bmatrix} -1 \\ 1 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} \max_{\lambda \in \mathbb{R}^2} \quad & \begin{bmatrix} 1 \\ -1 \end{bmatrix} \lambda \\ s.t. \quad & [0 \quad 1]\, \lambda + 1 = 0 \\ & \lambda \ge 0 \end{aligned}$$