07/12/2021 Lecture-20

Minimization

$(\vec{x}, \vec{y}) \in \Omega \subseteq \mathbb{R}^n \times \mathbb{R}^m$
$\hookrightarrow$ convex domain

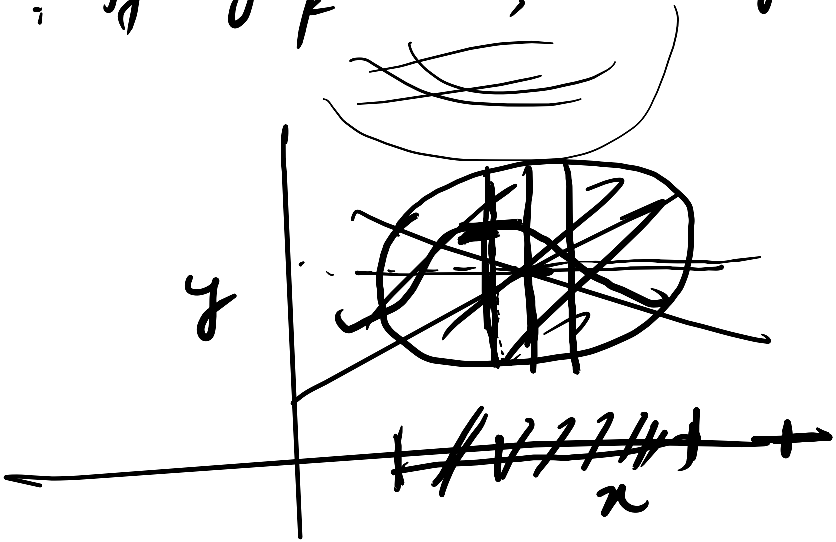
$f : \Omega \to \mathbb{R}$ is convex.

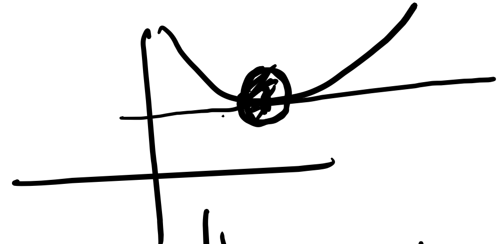
$g(\vec{x}) = \inf_{\vec{y} \in C} f(\vec{x}, \vec{y})$ ($C \to$ some nonempty convex set in $\mathbb{R}^m$).

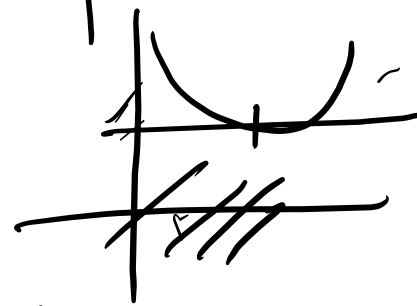
$\text{dom}(g) = \{ \vec{x} \in \mathbb{R}^n \mid (\vec{x}, \vec{y}) \in \text{dom}(f) = \Omega$

If $g(\vec{x}) > -\infty$ for some $\vec{x} \in \text{dom}(g)$, then $g(\vec{x}) > -\infty$ for all $\vec{x} \in \text{dom}(g)$.

Claim: If $g \not\equiv -\infty$, then g is convex.







$\vec{x_1}, \vec{x_2} \in \mathrm{dom}(g)$, $\exists\, \vec{z_1}, \vec{z_2} \in \mathbb{R}^{m}$ s.t.

$(\vec{x_1}, \vec{z_1})$, $(\vec{x_2}, \vec{z_2}) \in \Omega = \mathrm{dom}(f)$.

s.t.:

$$f(\vec{x}_i, \vec{y}_i) \le \underline{\underline{g(\vec{x}_i)}} + \varepsilon, \quad i = 1, 2.$$

$$\varepsilon > 0$$

$$g(\theta \vec{x}_1 + (1-\theta)\vec{x}_2)$$

$$\le \inf_{\vec{y} \in C} f(\underline{\theta \vec{x}_1 + (1-\theta)\vec{x}_2}, \vec{y})$$

$$\le f(\theta \vec{x}_1 + (1-\theta)\vec{x}_2, \theta \vec{y}_1 + (1-\theta)\vec{y}_2)$$

$$\le \theta f(\vec{x}_1, \vec{y}_1) + (1-\theta) f(\vec{x}_2, \vec{y}_2)$$

$$\le \theta g(\vec{x}_1) + (1-\theta) g(\vec{x}_2) + \varepsilon \quad \forall \varepsilon > 0$$

$$\Rightarrow \quad g(\theta \vec{x}_1 + (1-\theta)\vec{x}_2) \leq \theta\, g(\vec{x}_1) + (1-\theta)\, g(\vec{x}_2).$$

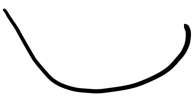
Duality for convex functions (conjugate functions)

$c_1 x_1^{\alpha} + \cdots + c_n x_n^{\alpha}$

$b_1 u_1^{\alpha} + \cdots + b_n u_n^{\alpha}$

$f : \mathbb{R}^n \to \mathbb{R}$

(Ω)

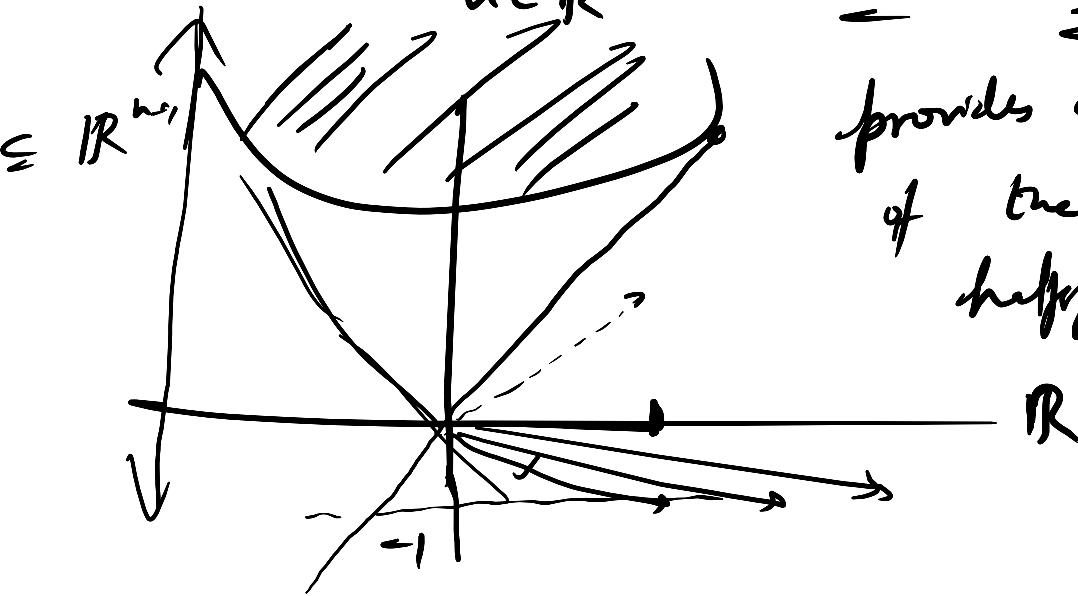


$$f^*(\vec{u}) = \sup_{\vec{x} \in \mathrm{dom}(f)} \left\{ \boxed{\langle \vec{u}, \vec{x} \rangle - f(\vec{x})} \right\}, \quad \vec{u} \in \mathbb{R}^n. \quad \to g(\vec{u})$$

conjugate of f.

$\mathrm{dom}(f^*)$ = wherever supremum exists.

$$f^*(\vec{u}) = \inf_{\alpha\in\mathbb{R}} \{ \underline{\underline{\mathrm{epi}\, f}} : \underline{\underline{H^-_{(\vec{u},-1),\underline{\alpha}}}} \}$$



provides a parametrization of the set of supporting halfspaces of the epigraph of f.

f^* is a convex function

(Why?)

$\rightarrow$ $f^*(\vec{u})$ is the minimal offset value of β such that $\langle \vec{u}, \vec{x} \rangle - \beta$ underestimates f everywhere..

<u>Examples</u>: Conjugates of convex functions on $\mathbb{R}$.

<u>(1) Affine functions</u>: $f(x) = ax + b$.

$\boxed{yx - ax - b}$ is bounded iff $y = a$.

$\mathrm{dom}(f^*) = \{a\}$. $f^*(a) = -b$.

(2) Negative logarithm. $f(x) = -\log x$ on $\mathbb{R}_{>0}$.

$g(x,y) = xy + \log x$ is unbounded above if $y \geq 0$.

$\Rightarrow$ dom $f^*(y) \subseteq \mathbb{R}_{<0}$.

Reaches maximum at $x = -\frac{1}{y}$

$$\frac{\partial g}{\partial x} = y + \frac{1}{x}$$

$$f^*(y) = -1 + \log\left(-\frac{1}{y}\right) = -1 - \log|y|.$$

dom $f^* = \mathbb{R}_{<0}$, $f^*(y) = -1 - \log|y|$.

(3) <u>Exponential</u>; $f(x) = e^x$;

$xy - e^x$ is unbounded if $y < 0$

For $y > 0$, $xy - e^x$ reaches its maximum

at $x = \log y$.

$$f^*(y) = y\log y - y$$

$$\mathrm{dom}(f^*) = \mathbb{R}_{>0}$$

Example. Conjugates of convex functions on $\mathbb{R}^n$:

(1) (Strictly convex quadratic functions)

$$f(\vec{x}) = \frac{1}{2}\langle \vec{x}, Q\vec{x}\rangle \quad Q \text{ — positive-definite}$$

$$= \langle \vec{y}, \vec{x}\rangle - \frac{1}{2}\langle \vec{x}, Q\vec{x}\rangle \text{ is bounded above}$$

$\vec{\xi}$

(as a function of $\vec{x}$) for all $\vec{\xi}$.

A scalar fn attains its maximum at $\vec{x} = Q^{-1}\vec{\xi}$.

$$\left(f^*(\vec{\xi}) = \frac{1}{2}\langle \vec{\xi}, Q^{-1}\vec{\xi}\rangle \right.$$

$$\left. \operatorname{dom}(f^*) = \mathbb{R}^n \right)$$

$$-\frac{1}{2}\langle \vec{x}-\vec{\alpha}, Q(\vec{x}-\vec{\alpha})\rangle$$

$$= \langle \vec{y}, \vec{x}\rangle - \frac{1}{2}\langle \vec{x}, Q\vec{x}\rangle$$

+ const.

$$-\frac{1}{2}\langle \vec{x}, Q\vec{x}\rangle + \langle \vec{\alpha}, Q\vec{x}\rangle - \frac{1}{2}\langle \vec{\alpha}, Q\vec{\alpha}\rangle$$

$$\langle Q^T\vec{\alpha}, \vec{x}\rangle = \langle \vec{y}, \vec{x}\rangle$$

$$Q\alpha = \vec{y}, \quad \alpha = Q^{-1}\vec{y}$$

min. $\langle \vec{z}, Q\vec{z}\rangle \geq 0$

2 (1) f^* is convex

(2) $(f^*)^* = f$

(3) (Fenchel's inequality)

$$\left[\langle \vec{y}, \vec{x} \rangle \le f(\vec{x}) + f^*(\vec{y})\right.$$

$$f^*(\vec{y}) = \sup_{\vec{x}} \langle \vec{y}, \vec{x} \rangle - f(\vec{x})$$

$$\ge \langle \vec{y}, \vec{x} \rangle - f(\vec{x})$$

$$\langle \vec{y}, \vec{x} \rangle \le \frac{1}{2} \boxed{\langle \vec{x}, Q\vec{x} \rangle} + \frac{1}{2} \langle \vec{y}, Q^{-1}\vec{y} \rangle.$$

Economic interpretation of the conjugate

$\vec{r} = (r_1, \ldots, r_n)$ resources consumed.

$\underline{S(\vec{r})} \rightarrow$ sales revenue derived from the product (as a function of resource consumed).

Total amount paid for resource $\langle \vec{p}, \vec{r} \rangle$

Profit: $\underline{S(\vec{r})} - \langle \vec{p}, \vec{r} \rangle$.

Maximum possible profit: $M(\vec{p}) = \sup_{\vec{r}} \left(S(\vec{r}) - \langle \vec{p}, \vec{r} \rangle \right)$

as a function of resource prices

$$M(p) = (-S)^*(-p).$$

Maximum profit is related to the conjugate of gross sales (as a function of resources consumed).

$$\|(x_1, \ldots, x_n)\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}.$$

Unconstrained optimization

$$\mathcal{D} = \boxed{\mathbb{R}^n}$$

decision space

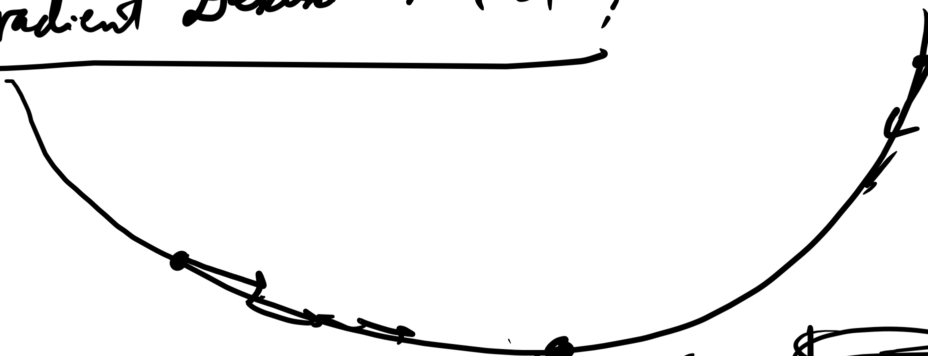
$$f_0 : \mathbb{R}^n \to \mathbb{R}.$$

$$\boxed{\underset{\vec{x} \in \mathbb{R}^n}{\text{minimize}} \; f_0(\vec{x})}$$

If f_0 is convex, it has a unique minimum value (if one exists)

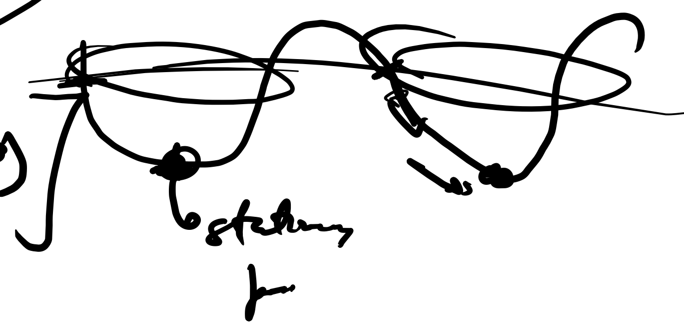
If f is $\in C^1(\mathbb{R}^n)$ then solve $\boxed{\nabla f = 0}$ to get all candidates for minima.

Gradient Descent Method:



$$f(\vec{x}+\vec{d}) \approx \boxed{f(\vec{x})} + (\nabla f(\vec{x}), \vec{d})$$

for small $\vec{d}$



$\nabla f(\vec{x})$ is the direction of maximum increase of f.

&

$$\underset{\|\vec{d}\| = \varepsilon}{\cancel{\max}\ \min} \left[f(\vec{x}) + \langle \nabla f(\vec{x}), \vec{d} \rangle \right]$$

$$|\nabla f(\vec{x})| \, |\vec{d}| \cos\theta$$

$$\vec{d} = \varepsilon \cdot - \frac{\nabla f(\vec{x})}{\|\nabla f(\vec{x})\|}$$

* Initialize $\vec{x_0}$.
* Update $\cancel{x_0 x_1}$ $\vec{x_1} = \vec{x_0} + - \varepsilon \dfrac{\nabla f(\vec{x_0})}{\|\nabla f(\vec{x_0})\|}$

$$f(\vec{x_1}) < f(\vec{x_0})$$

$$\vec{x}_2 = \vec{x}_1 - \epsilon \frac{\nabla f(\vec{x}_1)}{\|\nabla f(\vec{x}_1)\|}$$

$\vdots$

$\vec{x}_k$ is Cauchy

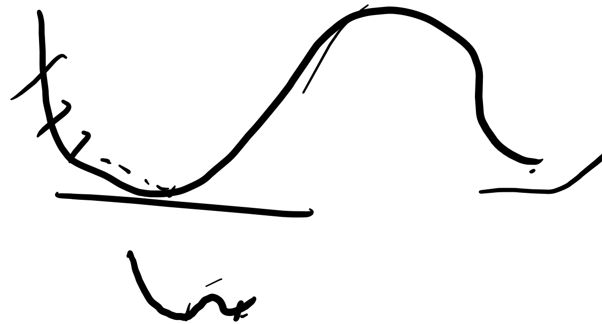
sign in appropriate

$$\boxed{\vec{x}_{k+1} = \vec{x}_k - \epsilon \frac{\nabla f(\vec{x}_k)}{\|\nabla f(\vec{x}_k)\|}}$$

If ϵ is too big,

you may jump to

next minim..



Constrained optimization:

$$\min. \quad f_0(\vec{x})$$

subject to

$$\mathcal{D} \begin{cases} g_i(\vec{x}) \le 0, & i = 1..,m \\ h_j(\vec{x}) = 0 & j = 1..,k \end{cases}$$

(No need to assume f_0 or $\mathcal{D}$ are convex.)

Special Case:

~~Minimize~~ min. $f_0(\vec{x})$

subject to $g(\vec{x}) \leq 0$

(one inequality constraint)

Method of Lagrange multipliers:

$\boxed{f_0, g \in C^1(\mathbb{R}^n)}$. $g(\vec{x}) = 0$ (define a $(n-1)$-dimensional surface)

$$\mathcal{L}(\vec{x}, \lambda) = f_0(\vec{x}) - \boxed{\lambda}\, g(\vec{x})$$

λ → Lagrange multiplier

$\mathcal{L}$ → Lagrange function or Lagrangian.

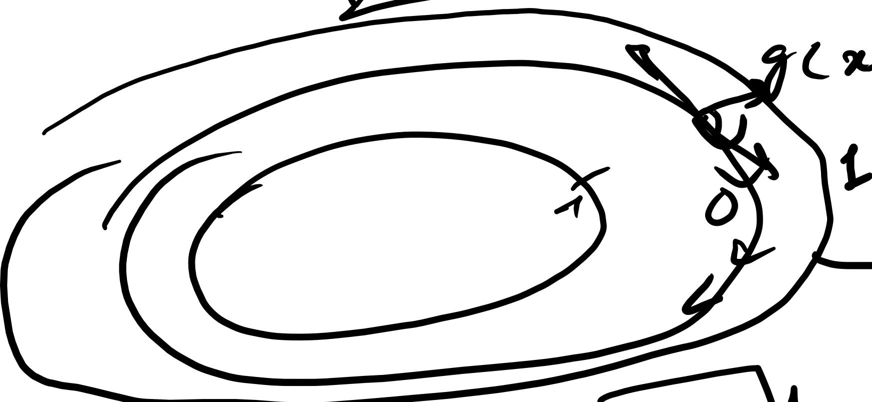
$D = \{ g(\vec{x}) \le 0 \} = \{ g(\vec{x}) < 0 \} \cup \{ g(\vec{x}) = 0 \}$.

nonempty interior

$\nabla f(\vec{x}_0) = 0$ for $\vec{x}_0$ s.t. $g(\vec{x}_0) < 0$.

$g(\gamma(t)) = c$

$\boxed{\nabla g \perp}$ g …

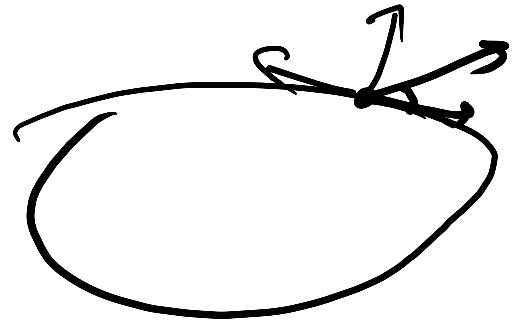


contours of g or level sets of g.

$\boxed{\nabla g \perp}$ contours of g.

Consider the values of f on the level set

$\{g(\bar{x})=0\}$.



∇f

If $\bar{x_0}$ s.t. $g(\bar{x_0})=0$

is a minimal sol^n for f_0 then

$\nabla f_0 \perp$ on $\{g(\bar{x})=0\}$

$\Longleftrightarrow$ $\boxed{\nabla f_0(\bar{x_0}) = \lambda \nabla g(\bar{x_0})}$ for some λ

and $\bar{x_0}$ a feasible pt.

$\min \; \boxed{f_0}$

s.t. $g_1(\bar{x}_0) = 0$

$g_2(\bar{x}_0) = 0$

$$\nabla f(\bar{x}_0) = \lambda_1 \nabla g_1(\bar{x}_0) + \lambda_2 \nabla g_2(\bar{x}_0)$$

$$= [\lambda_1 \;\; \bar{\lambda}_2] \begin{bmatrix} \nabla g_1 \\ \nabla g_2 \end{bmatrix}^T$$

Lagrange multiplier theorem:

Let $f_0 : \mathbb{R}^n \to \mathbb{R}$ be the objective function,

$g : \mathbb{R}^n \to \mathbb{R}^c$ be the (equality)

constraints, both in $C^1(\mathbb{R}^n)$. Let $\vec{x}^*$ be

an optimal sol^n:

$$\boxed{\begin{array}{ll} \text{min.} & f_0(\vec{x}) \\ \text{s.t.} & g(\vec{x}) = \vec{0} \end{array}}$$

(s.t. rank $Dg(\vec{x}^*) = c < n$).

Then there exist unique Lagrange multipliers

$\bar{\lambda}^* \in \mathbb{R}^c$ s.t. $\boxed{Df(\bar{x}^*) = \boxed{\bar{\lambda}^{*T}} Dg(\bar{x}^*)}$

<u>Example</u> : $f_0(x,y) = 8x^2 - 2y$

subject to $\boxed{x^2 + y^2 = 1}$

$$\nabla f_0 = \boxed{(x,y)} = \begin{bmatrix} 16x \\ -2 \end{bmatrix} = \lambda \begin{bmatrix} 2x \\ 2y \end{bmatrix}$$

and $\boxed{x^2 + y^2 = 1}$.

$16x = \lambda 2x \Rightarrow 2x(8-\lambda) = 0$

$-2 = \lambda(2y)$ $\qquad \Rightarrow x = 0$ or $\lambda = 8$

Case I: $x = 0$

$y = \pm 1$. $\qquad (0,-1), (0,1)$

two candidates for minima.

Case II: $\lambda = 8$

$y = -\frac{1}{8} \qquad x = \pm\sqrt{\frac{63}{64}} = \pm\frac{3\sqrt{7}}{8}$

$(3\sqrt{7}/8, -\frac{1}{8}), (-3\sqrt{7}/8, -\frac{1}{8})$

(No appeal to convexity has been made so far).

<u>Defn</u>: Given the optimization problem

$$\min \; f_0(\vec{x})$$

$$\text{subject to} \quad g_i(\vec{x}) \overset{\geq}{\leq} 0, \quad 1 \leq i \leq m$$

$$h_j(\vec{x}) = 0, \quad 1 \leq j \leq p$$

the Lagrangian associated with this optimization problem:

$$\boxed{L(\vec{x}, \underset{\in \mathbb{R}^m}{\vec{\lambda}}, \underset{\in \mathbb{R}^p}{\vec{\nu}}) = f_0(\vec{x}) + \sum_{i=1}^{m} \lambda_i g_i(\vec{x}) + \sum_{j=1}^{p} \nu_j h_j(\vec{x})}$$

The variables $\lambda_1, \dots, \lambda_m$, $\nu_1, \dots, \nu_p$ are called the <u>Lagrange multipliers</u>.

<u>Defn.</u> Given a Lagrangian $L(\vec{x}, \vec{\lambda}, \vec{\nu})$ of some optimization problem over domain $\mathcal{D}$, the **Lagrangian dual** is the function:

$$F(\vec{\lambda}, \vec{\nu}) = \inf_{\vec{x} \in \mathcal{D}} L(\vec{x}, \vec{\lambda}, \vec{\nu})$$

$$\leq \inf_{\vec{x} \in \mathcal{D}} \left[f(\vec{x}) + \underbrace{\sum_{i=1}^{m} \lambda_i g_i(\vec{x})}_{\leq 0} + \underbrace{\sum_{j=1}^{p} \nu_j h_j(\vec{x})}_{0} \right]$$

suppose $\vec{x}$ is feasible given $\vec{\lambda}, \vec{\nu}$,

$$\leq \inf_{\vec{x} \in \mathcal{D}} f(\vec{x})$$

Since the dual function is the pointwise infimum of a family of affine functions, it is a concave function.

<u>Lower bound property</u> (Weak duality theorem)

If $\boxed{\vec{\lambda} \geq \vec{0}}$ then $L(\vec{\lambda}, \vec{\nu}) \leq p^*$

where p^* is the optimal value of the original optimization problem.

Proof.

$g_i(\vec{x}) \le 0 \qquad 1 \le i \le m$

$\lambda_i \, g_i(\vec{x}) \le 0 \qquad$ since $\lambda_i \ge 0$

$h_j(\vec{x}) = 0 \qquad 1 \le j \le p.$

Hence $\forall \vec{x} \in \mathcal{D}$,

$$f(\vec{x}) + \sum_{i=1}^{m} \lambda_i \, g_i(\vec{x}) + \sum_{j=1}^{p} \nu_j \, h_j(\vec{x})$$

$$\le f(\vec{x})$$

$$\left[F(\vec{\lambda}, \vec{\nu}) \le \inf_{\vec{x} \in \mathcal{D}} f(\vec{x}) = p^* \right]$$

* Gives lower bounds for optimal sol'n.

<u>Examples</u>:

(1) Least (l^2-norm) solution of linear equations,

min. $\|\vec{x}\|_2$ → Same as minimize $\langle \vec{x}, \vec{x} \rangle$

subject to $A\vec{x} = \vec{b}$ → $\mathbb{R}^p$

A: $p \times n$ matrix

$\vec{x} \in \mathbb{R}^n$

<u>Lagrangian</u>:

$$L(\vec{x}, \vec{\nu}) = \langle \vec{x}, \vec{x} \rangle + \vec{\nu}^T (A\vec{x} - \vec{b})$$

$\langle \vec{\nu}, A\vec{x} \rangle$

Note that the Lagrangian is a strictly convex function in terms of $\vec{x}$ (unique minimum).

$$\nabla_{\vec{x}} L(\vec{x}, \vec{\nu}) = 2\vec{x} + A^T \vec{\nu} = \vec{0}$$

$$\Rightarrow \vec{x} = \boxed{-\frac{1}{2} A^T \vec{\nu}.}$$

$$F(\vec{\nu}) = \inf_{\vec{x} \in \mathcal{D}} L(\vec{x}, \vec{\nu})$$

$$= L\left(-\frac{1}{2} A^T \vec{\nu}, \vec{\nu}\right)$$

$$= -\frac{1}{2} \vec{\nu}^T A A^T \vec{\nu} - \langle \vec{b}, \vec{\nu} \rangle$$

(a concave function)

as $-\frac{1}{2} A A^T$ is negative semidefinite.

Lower bound property:

$$-\frac{1}{4}\vec{v}^T A A^T \vec{v} - \langle \vec{b}, \vec{v} \rangle \leq p^*$$

$$\forall \vec{v} \in \mathbb{R}^r.$$

(2) Standard form LPP,

$$\text{min } \langle \vec{c}, \vec{x} \rangle$$

$$\text{subject to } \boxed{A\vec{x} = \vec{b}, \quad \vec{x} \geq \vec{0}}$$

$$-x_i^+ \leq 0 \qquad 1 \leq i \leq n.$$

Lagrangian =

$$L(\vec{x}, \vec{\lambda}, \vec{\nu})$$

$$= \langle \vec{c}, \vec{x} \rangle + \langle \vec{\lambda}, -\vec{x} \rangle$$

$$+ \vec{\nu}^T (A\vec{x} - \vec{b})$$

Dual function =

$$\Gamma(\vec{\lambda}, \vec{\nu}) = \inf_{\vec{x} \in D} L(\vec{x}, \vec{\lambda}, \vec{\nu})$$

$$= -\langle \vec{b}, \vec{\nu} \rangle + \left[\inf_{\vec{x} \in D} \left\langle \boxed{\vec{c} - \vec{\lambda} + A^T \vec{\nu}}, \vec{x} \right\rangle \right]$$

Lower bound property.

$$\boxed{A^T \vec{v} = -\vec{c}}$$

If $\quad A^T \vec{v} + \vec{c} = \vec{0}, \; \vec{\lambda}$

$$\boxed{-\langle \vec{b}, \vec{v} \rangle} \le p^*.$$

$$F(0)^{\vec{v}} \le p^*.$$

(3) Two-way partitioning.

$$\min. \ \langle \vec{x}, W\vec{x} \rangle \qquad (\vec{x} \in \mathbb{R}^n)$$

subject to $x_i = \pm 1$, $i = 1, \dots, n$.

(non-convex problem; feasible set has 2^n discrete points.)

interpretation: partition $\{1, \dots, n\}$ into two sets.

W_{ij} is the cost of assigning i, j to the same set, $-W_{ij}$ is the cost of assigning i, j to ~~two~~ different sets.

Lagrangian $L(\vec{x}, \vec{\nu}) = \langle \vec{x}, W\vec{x} \rangle + \sum_{i=1}^{n} \nu_i (x_i^2 - 1)$

Dual: $F(\vec{\nu}) = \inf_{\vec{x} \in \mathbb{R}^n} \left(\langle \vec{x}, W\vec{x} \rangle + \sum_{i=1}^{n} \nu_i (x_i^2 - 1) \right)$

$\langle \vec{x}, \operatorname{diag}(\vec{\nu}) \vec{x} \rangle$

$= \inf_{\vec{x} \in \mathbb{R}^n} \left(\langle \vec{x}, (W + \operatorname{diag}(\vec{\nu})) \vec{x} \rangle - \vec{1}^T \vec{\nu} \right)$

$= \begin{cases} -\vec{1}^T \vec{\nu}, & \text{if } W + \operatorname{diag}(\vec{\nu}) \succeq 0 \\ -\infty, & \text{otherwise.} \end{cases}$

Lower-bound property:

$$p^* \geq -\vec{1}^T \nu,$$

if $\boxed{W + \mathrm{diag}(\vec{\nu}) \succeq 0}$.

$\nu = -\lambda_{min}(W)\vec{1}$ gives the following lower bound,

$$\boxed{p^* \geq n\,\lambda_{min}(W)}$$

smallest eigenval. of W.

(nontrivial lower bound for the partition problem)

$$\boxed{F(\vec{\lambda}, \vec{\nu}) \leq p^*}$$

We want best possible lower bound from planning this strategy.

$$\boxed{\begin{array}{ll} \text{maximize} & F(\vec{\lambda}, \vec{\nu}) \\ \text{subject to} & \vec{\lambda} \geq \vec{0} \end{array}}$$

min $-F$

Dual problem