EECS 127   March 3, 2020   Admin.

Today: Duality.

Unconstrained → find ∇ → set to 0. → Find optimum.

Constrained →

→ Gradient Descent.

→ Duality

$$p^* = \begin{array}{ll} \text{minimize } f_0(\vec{x}) & \\ \text{s.t. } f_i(\vec{x}) \le 0 & \forall\ 1 \le i \le m \\ \quad h_i(\vec{x}) = 0 & \forall\ i \le i < p \end{array}$$

THE PRIMAL

For this: we define the Lagrangian

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

When $\lambda_i \ge 0$

$\vec{\lambda}, \vec{\nu}$ are called Lagrange multipliers.
dual variables.

$$\min_{\vec{x}} L(\vec{x}, \vec{\lambda}, \vec{\nu}) := \underbrace{g(\vec{\lambda}, \vec{\nu})}_{\text{function } \vec{\lambda}, \vec{\nu}}$$

$x = x_1$

$x = x_2$

$\vdots$

$L_1(\vec{\lambda}, \vec{\nu}) = L(\vec{x}_1, \vec{\lambda}, \vec{\nu})$

$L_2(\vec{\lambda}, \vec{\nu}) = L(\vec{x}_2, \vec{\lambda}, \vec{\nu})$

$\vdots$

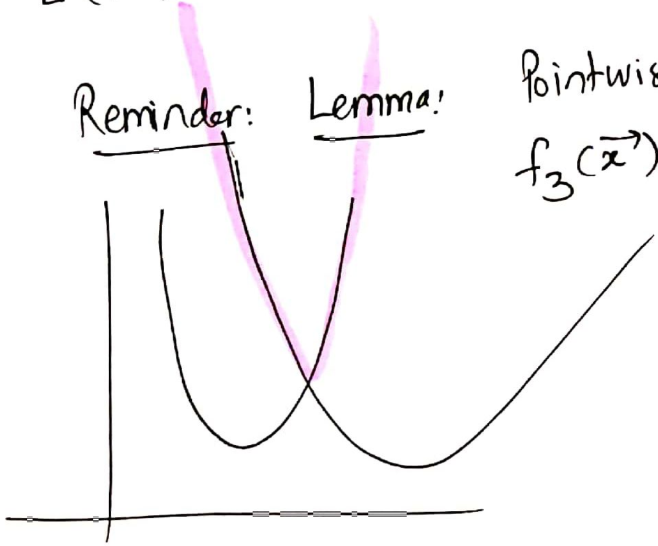
$\underset{\vec{x}}{\overset{\text{inf}}{\min}} \{L_1, L_2 \cdots L\}$.

Observations

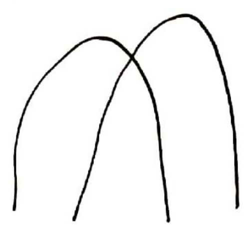
1. g depends on $\vec{\lambda}, \vec{\nu}$.
2. $L(\vec{x}, \vec{\lambda}, \vec{\nu})$ is an affine function of $\vec{\lambda}, \vec{\nu}$.

Reminder: Lemma: Pointwise maximum of convex functions $\Rightarrow$ Convex.

$f_3(\vec{x}) = \max\{f_1(\vec{x}), f_2(\vec{x})\}$.



Pointwise minimum of concave functions $\Rightarrow$ concave.



3. $g(\vec{\lambda}, \vec{\nu})$ is concave !! $\rightarrow \vec{\lambda}, \vec{\nu}$

$\hookrightarrow$ Does not depend on $f_0(x)$ at all.

④ Property of g:

$$g(\vec{\lambda}, \vec{\nu}) \leq p^* \qquad \text{forall } \underbrace{\vec{\lambda} \geq 0}_{\text{element-wise}}, \ \vec{\nu}.$$

Proof:

Consider: $\tilde{x}$ a feasible point for the primal.

$$f_i(\tilde{x}) \leq 0 \quad ① \qquad 1 \leq i \leq m$$

$$h_i(\tilde{x}) = 0 \quad ②$$

$$\lambda_i f_i(\tilde{x}) \leq 0 \quad (1*) \implies \sum \lambda_i f_i(\tilde{x}) \leq 0$$

$$\nu_i h_i(\tilde{x}) = 0 \quad (2*) \implies \sum \nu_i h_i(\tilde{x}) = 0$$

Consider:

$$L(\tilde{x}, \vec{\lambda}, \vec{\nu}) = f_0(\tilde{x}) + \underbrace{\sum \lambda_i f_i(\tilde{x})}_{\leq 0} + \underbrace{\sum \nu_i h_i(\tilde{x})}_{=0}$$

$$\leq f_0(\tilde{x})$$

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

$g(\vec{\lambda}, \vec{\nu})$ lowerbounds all values of f_0 for feasible $\vec{x}$

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

An interpretation:

$$M(\vec{x}) = f_0(\vec{x}) + \sum \mathbb{1}_{neg}\{f_i(x)\} + \sum \mathbb{1}_0(h_i(x)).$$

$$\mathbb{1}_{neg} = \begin{cases} 0 & \text{if } f_i(x) \leq 0 \\ \infty & \text{if } f_i(x) > 0 \end{cases}$$

$$\mathbb{1}_0 = \begin{cases} 0 & \text{if } h_i(x) = 0 \\ \infty & \text{otherwise.} \end{cases}$$

$$\min_{\vec{x}} \; M(\vec{x})$$

→ Hard thesholding penalty for violating a constraint : $M(\vec{x})$

Instead : the Lagrangian gives you only a linear penalty in $\vec{\lambda}$ or $\vec{\nu}$

example:

Minimum norm problem.

A is full row rank.

$$p^* = \min \vec{x}^T \vec{x}$$
$$\text{s.t. } A\vec{x} = \vec{b} \Leftrightarrow A\vec{x} - \vec{b} = 0$$

$$L(\vec{x}, \vec{\nu}) = \vec{x}^T \vec{x} + \vec{\nu}^T (A\vec{x} - \vec{b})$$: Convex function $\vec{x}$

$$g(\vec{\nu}) = \min_{\vec{x}} L(\vec{x}, \vec{\nu})$$

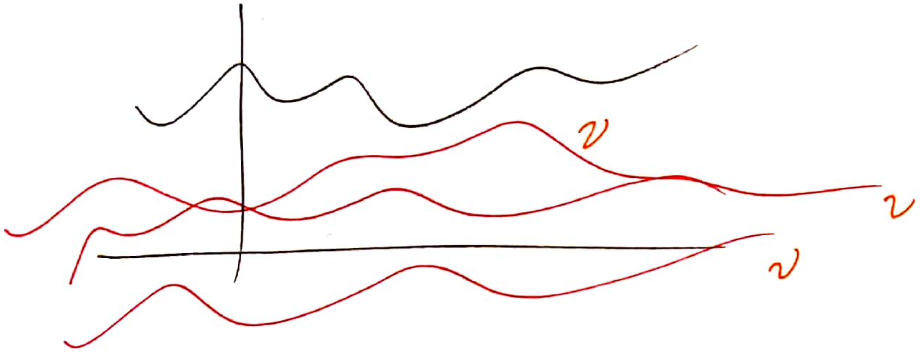
Minimize : $\nabla_{\vec{x}} L = 0.$

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

setting to 0 $\rightarrow$ $\vec{x} = -\frac{1}{2} A^T \vec{\nu}$

$$g(\vec{\nu}) = L(-\tfrac{1}{2} A^T \vec{\nu}, \vec{\nu}) = \frac{1}{4} \vec{\nu}^T A A^T \vec{\nu} + \vec{\nu}^T (A(-\tfrac{1}{2} A^T \vec{\nu}) - \vec{b})$$
$$= -\frac{1}{4} \vec{\nu}^T A A^T \vec{\nu} - \vec{\nu}^T \vec{b}$$

$$p^* \geq g(\vec{\nu})$$



Lagrange Dual Problem!

$$d^* = \max_{\substack{\vec{\lambda} \geq 0 \\ \vec{\nu}}} g(\vec{\lambda}, \vec{\nu}) \quad \Big\} \quad \text{CONVEX PROGRAM.}$$

$$\max \quad g(\vec{\nu}) = d^*$$

$$\nabla_{\vec{\nu}} \, g(\vec{\nu}) = -\frac{1}{4}\left(2AA^T\right)\vec{\nu} - \vec{b}$$

Set $= 0$.

$$\vec{\nu}_* = -2\,(AA^T)^{-1}.\vec{b}$$

$$\vec{x}_1 = -\frac{1}{2}A^T\vec{\nu}_* = A^T(AA^T)^{-1}.\vec{b}.$$

We know:

$d^* \le p^*$ Weak duality.

$\boxed{d^* = p^* \quad \text{Strong duality.}}$

Example:

$$p^* = \min_{\text{s.t. } x_i^2 = 1} \quad \vec{x}^T W \vec{x} \qquad i = 1, \cdots n. \qquad W \in S^n$$

$x_i = +1$ or $x_i = -1$.

W_{ij}

Partioning problem.

$$\vec{x}^T \begin{bmatrix} \nu_1 & & 0 \\ & \nu_2 \ddots & \\ 0 & & \nu_n \end{bmatrix} \vec{x}$$

$$L(x, \nu) = x^T W x + \sum_{i=1}^{n} \nu_i (x_i^2 - 1).$$

$$= \underbrace{x^T (W + \operatorname{diag}(\vec{\nu}))\, x}_{\text{quadratic}} - \sum_{i=1}^{n} \nu_i$$

$x^T Q \cdot x \geq 0$ if Q PSD.

$$g(\vec{\nu}) = \inf_{\vec{x}} L(\vec{x}, \vec{\nu}) = \begin{cases} -\infty & \text{if } W + \operatorname{diag}(\vec{\nu}) \text{ is not PSD.} \\ -\sum_{i=1}^{n} \nu_i & \text{if } W + \operatorname{diag}(\vec{\nu}) \text{ is PSD} \\ & \qquad \succeq 0 \end{cases}$$

$$\max \; g(\vec{\nu})$$

s.t. $W + \operatorname{diag}(\vec{\nu})$ is PSP

Let us pick some value for $\vec{\nu}$.

$$\vec{\nu} = - \begin{bmatrix} \lambda_{min}(W) \\ \lambda_{min}(W) \\ \vdots \\ \lambda_{min}(W) \end{bmatrix}$$

$W - \lambda_{min} I \succeq 0$.

$\nu_i = \lambda_{min}(W)$.

Lowerband $= n \cdot \lambda_{min}(W) = \sum_{i=1}^{n} \nu_i$

$p^* \geq n \cdot \lambda_{min}(W)$.