Lecture 7: Constrained Optimization: Necessary and Sufficient Conditions

Prof. Krishna R. Pattipati
Dept. of Electrical and Computer Engineering
University of Connecticut
Contact: **krishna@engr.uconn.edu** **(860) 486-2890**

ECE 6437

Computational Methods for Optimization

Fall 2009
October 13, 2009

Outline of Lecture 7

- Necessary and Sufficient Conditions
- Methods of Specifying constraint set (Ω)
- Basic Result: Necessary Conditions of Optimality
- Examples
- Equality Constraints
- Economical Interpretation of Lagrange Multipliers

Constrained Minimization Problem

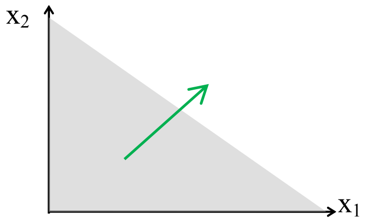
- " Minimize $f(\underline{x})$ subject to $\underline{x} \in \Omega \subset R^n; \Omega =$ constraint set
  - $\underline{x}^*$ is a local minimum of f over Ω if $\exists$ a scalar $\varepsilon > 0 \ni$
    $$f(\underline{x}^*) \le f(\underline{x}) \ \forall\, \underline{x} \in \Omega, \| \underline{x} - \underline{x}^* \| < \varepsilon$$
  - $\underline{x}^*$ is a global minimum of f over Ω if
    $$f(\underline{x}^*) \le f(\underline{x}) \forall \underline{x} \in \Omega$$
  - When f and Ω are convex, local minimum $\Leftrightarrow$ global minimum

- Methods of specifying constraint set Ω
  - Ω is non-empty, convex, and closed subset of R^n
  - $f(\underline{x})$ is continuously differentiable over Ω
  - Difficulty with an open set
    $$\min x^2, \Omega = \{x \mid 0 < x < 1\}; x^* \text{ is undefined}; x^* \to 0$$
  - Difficulty with a non-convex set: Nec. conditions of optimality fail!!

Methods of Specifying Ω - 1

- Set constraints: $\underline{x} \in \Omega =$ convex
  - Non negative orthant constraints

$$\Omega = \{\underline{x} \mid x_i \geq 0; i = 1, 2, \ldots, n\}$$

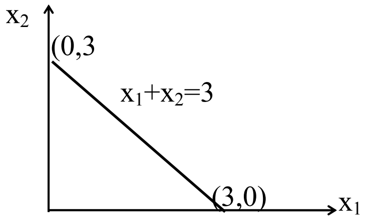


  - Simple bounds

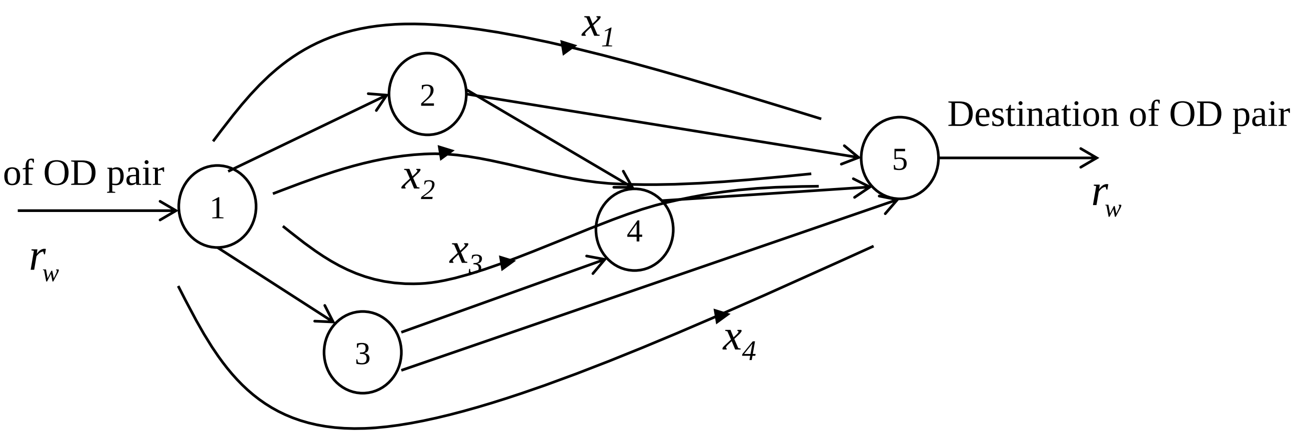
$$\Omega = \{\underline{x} \mid \alpha_i \leq x_i \leq \beta_i; i = 1, 2, \ldots, n\}$$

- Equality constraints: $\Omega = \{\underline{x} \mid h_i(\underline{x}) = 0; i = 1, 2, \ldots, n\}$
  - Simplex constraints

$$\Omega = \{\underline{x} \mid \sum_{i=1}^{n} x_i = r; x_i \geq 0\}$$



  - Some applications: routing, allocation problems



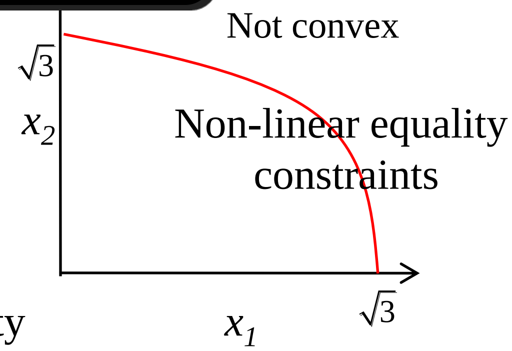
Methods of Specifying Ω - 2

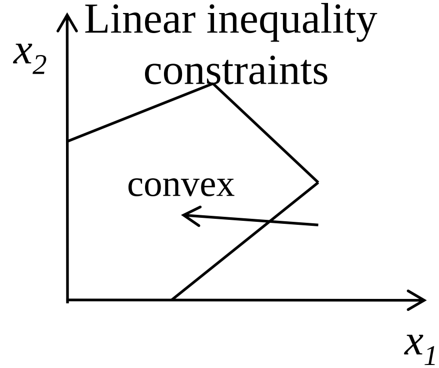
$$\min D(\underline{x}) = \sum_{i,j} D_{ij}\Big(\sum_{\substack{\text{all paths} \\ p \text{ traversing} \\ (i,j)}} x_p\Big)$$

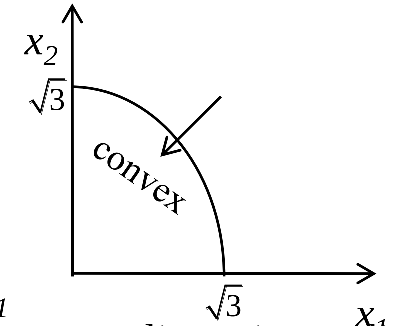
$$F_{ij} = \sum_{\substack{\text{all paths } p \\ \text{traversing } (i,j)}} x_p$$

$$s.t. \sum_{p \in P_w} x_p = r_w \;\forall w \in W$$

$$x_p \geq 0 \;\forall p \in P_w, w \in W$$







Non-linear inequality constraints

- Linear equality constraints:

$$\Omega = \{\underline{x} \mid x_i \geq 0; A\underline{x} = b\}$$

- Nonlinear equality constraints:

$$\Omega = \{\underline{x} \mid x_1^2 + x_2^2 = 3; x_1, x_2 \geq 0\}$$

$\Rightarrow$ For $\underline{x}$ to be convex, equality constraints must be linear

- Inequality constraints

$$\Omega = \{\underline{x} \mid g_i(\underline{x}) \leq 0; j = 1, 2, .., r\}$$

- Linear inequality constraints $\Omega = \{\underline{x} \mid \underline{x} \geq 0; A\underline{x} \leq b\}$
- Nonlinear inequality constraints $\Omega = \{\underline{x} \mid x_1^2 + x_2^2 \leq 3; x_i \geq 0\}$

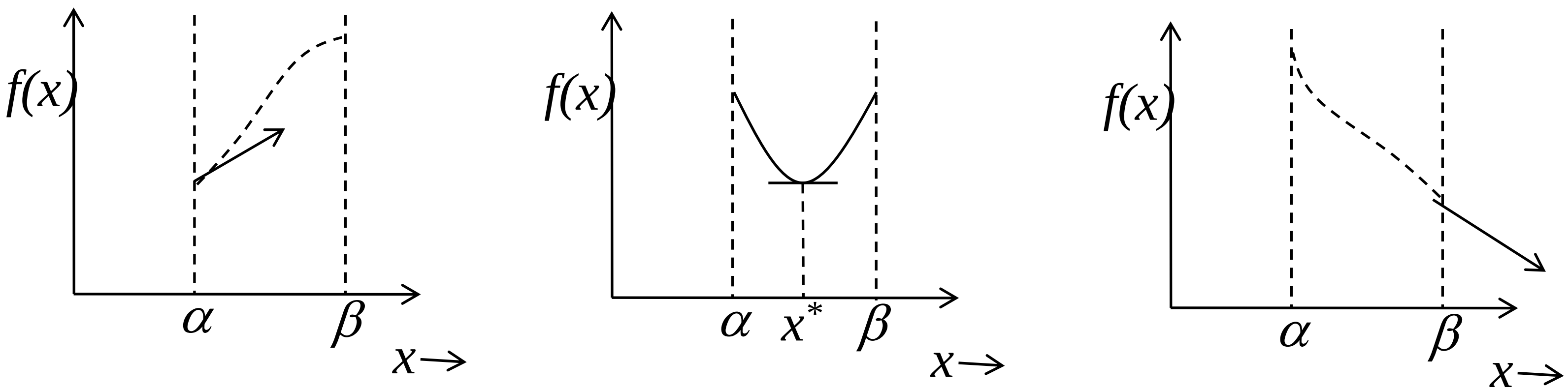
First-order Conditions of Optimality - 1

- Recall unconstrained case: $\nabla f(x^*) = 0$
- What happens when there are constraints?

Necessary: $\nabla^2 f(x^*) \geq 0$;

Sufficiency: $\nabla^2 f(x^*) > 0$

Let us take a simple case where $\Omega = \{\underline{x} | \alpha \leq x_i \leq \beta\}$ and x is scalar



- Case 1: $x^* = \alpha$

$$\Rightarrow \frac{df(x^*)}{dx}\Big|_{\alpha} \geq 0 \; and \; x - x^* \geq 0 \; \forall x \in \Omega \Rightarrow \nabla \underline{f}^T(\underline{x}^*)(\underline{x} - \underline{x}^*) \geq 0$$

- Case 2: If $\alpha < x^* < \beta$ $\frac{df}{dx}\Big|_{x^*} = 0 \; and \; \underline{x} - \underline{x}^*$ can be positive or negative

$$\Rightarrow \nabla \underline{f}^T(\underline{x}^*)(\underline{x} - \underline{x}^*) = 0 \;\; \forall x \in \Omega$$

First-order Conditions of Optimality - 2

- Case 3: $x^* = \beta$

$$\frac{df(x)}{dx}\Big|_{x=x^*} \le 0 \; and \; x - x^* \le 0 \; \forall x \in \Omega$$

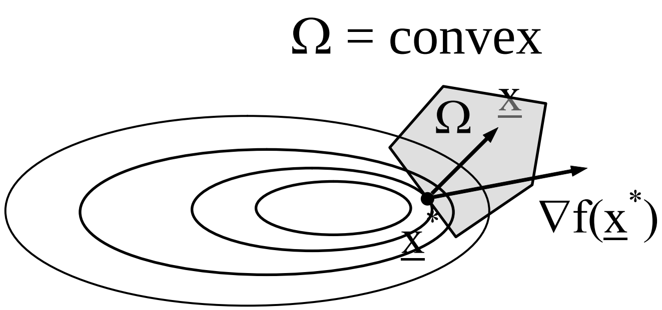
$$\nabla \underline{f}^T(\underline{x}^*)(\underline{x} - \underline{x}^*) \ge 0 \; \forall \underline{x} \in \Omega$$

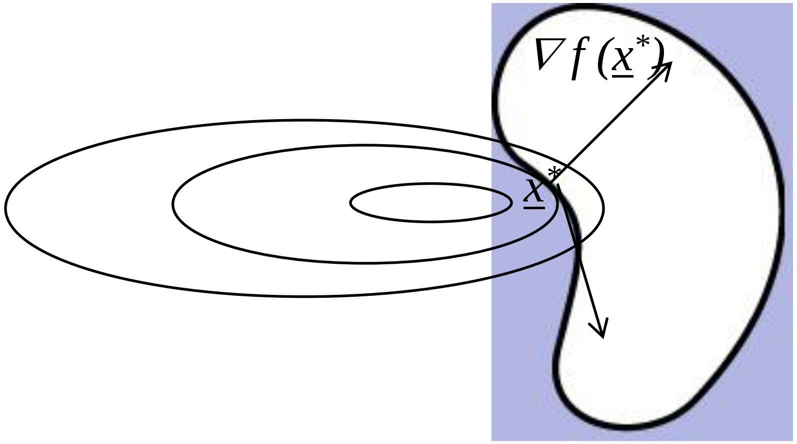
- Basic Result
  - 1. If $\underline{x}^*$ is a local minimum of f over a convex set $\Omega \in R^n$, then

$$\nabla \underline{f}^T(\underline{x}^*)(\underline{x} - \underline{x}^*) \ge 0, \forall \underline{x} \in \Omega \Rightarrow \angle[\nabla f(\underline{x}^*), (\underline{x} - \underline{x}^*)] \le 90^0$$

  - 2. If $f(\underline{x})$ is convex, then the condition is also sufficient

Convexity of Ω is critical





Proof of Basic Result

- Proof
  - 1. Suppose $\underline{x} \in \Omega \ni \nabla \underline{f}^T(\underline{x}^*)(\underline{x}-\underline{x}^*) < 0$. Then, from the mean value theorem, for $\forall \varepsilon\ \exists$ an $\alpha \in [0,1]$ s.t.

$$f(\underline{x}^* + \varepsilon(\underline{x}-\underline{x}^*)) = f(\underline{x}^*) + \varepsilon \nabla \underline{f}^T(\underline{x}^* + \varepsilon\alpha(\underline{x}-\underline{x}^*))(\underline{x}-\underline{x}^*)$$

$$\text{For sufficiently small } \varepsilon, \nabla \underline{f}^T(\underline{x}^* + \varepsilon\alpha(\underline{x}-\underline{x}^*))(\underline{x}-\underline{x}^*) < 0$$

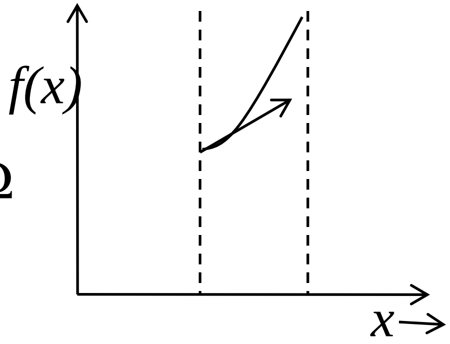
$$\Rightarrow f(\underline{x}^* + \varepsilon(\underline{x}-\underline{x}^*)) < f(\underline{x}^*) \Rightarrow \text{a contradiction}$$

  - 2. Since $f(\underline{x})$ is convex

$$f(\underline{x}) \geq f(\underline{x}^*) + \nabla \underline{f}^T(\underline{x}^*)(\underline{x}-\underline{x}^*); \forall \underline{x} \in \Omega$$

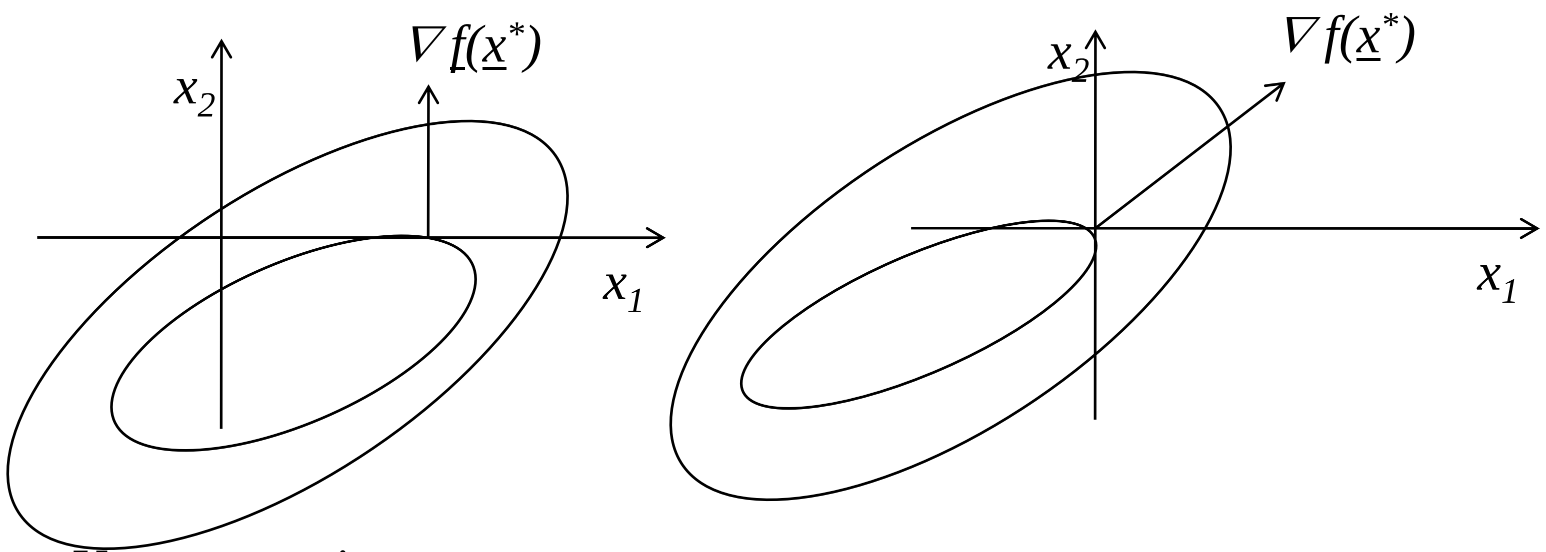
$$\text{Now, since } \nabla \underline{f}^T(\underline{x}^*)(\underline{x}-\underline{x}^*) \geq 0$$

$$f(\underline{x}) \geq f(\underline{x}^*)\ \forall \underline{x} \in \Omega \Rightarrow \underline{x}^* \text{ is a minimum}$$



Specialization of First-order Conditions - 1

- Examples:
  - Example 1: $\Omega = \{\underline{x} \mid x_i \geq 0; \forall i = 1, 2, \ldots, n\}$ non-negative orthant



Know at optimum

$$\nabla \underline{f}^T(\underline{x}^*)(\underline{x} - \underline{x}^*) \geq 0 \; \forall \underline{x} \in \Omega; \sum_{i=1}^{n} \frac{\partial f(\underline{x}^*)}{\partial x_i}(x_i - x_i^*) \geq 0 \; \forall \underline{x} \in \Omega$$

$$x_i^* = 0 \Rightarrow \frac{\partial f}{\partial x_i} \geq 0; Let\ x_k = x_k^* \forall k \neq i$$

$$x_i^* > 0 \Rightarrow \frac{\partial f}{\partial x_i} = 0; To\ show\ this, let\ x_k = x_k^* \forall k \neq i; and\ x_i = \frac{1}{2}x_i^*; x_i = \frac{3}{2}x_i^*$$

Specialization of First-order Conditions - 2

- Example 2: $\Omega=\{\underline{x} \mid \alpha_i \le x_i \le \beta\}$

$$\frac{\partial f}{\partial x_i} = \begin{cases} \ge 0; x_i^* = \alpha_i \\ 0; \alpha_i < x_i^* < \beta \\ \le 0; x_i^* = \beta_i \end{cases}$$

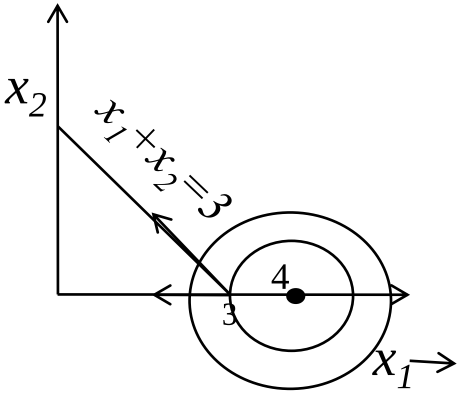
- Example 3: $\Omega=\{\underline{x} \mid x_i \ge 0, \sum_{i=1}^{n} x_i = r\}$ optimization over a simplex

$$\Omega=\{\underline{x} \mid x_i \ge 0, \sum_{i=1}^{n} x_i = r\} \text{ optimization over a simplex}$$

$$\min (x_1-4)^2 + x_2^2 \;\; s.t. x_1 + x_2 = 3 \, and \; x_i \ge 0$$

$$(x_1^*, x_2^*) = (3,0)$$

$$\nabla \underline{f}(\underline{x}^*) = \begin{bmatrix} 2(x_1-4) \\ 2x_2 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}; \frac{\partial f}{\partial x_1} < \frac{\partial f}{\partial x_2}$$



$$\text{In general } \; x_i^* > 0 \Rightarrow \frac{\partial f(\underline{x}^*)}{\partial x_i} \le \frac{\partial f(\underline{x}^*)}{\partial x_j} \forall j$$

All coordinates with positive allocation at the optimum must have equal or minimal partial cost derivative

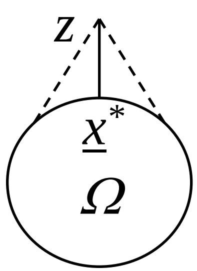
Specialization of First-order Conditions - 3

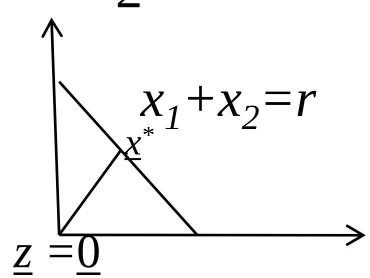
– $$\sum_{i=1}^{n} \frac{\partial f(\underline{x}^*)}{\partial x_i}(x_i - x_i^*) \geq 0 \; \forall x_i \geq 0 \text{ with} \sum_{i=1}^{n} x_i = r$$

$$\text{Suppose } x_i^* > 0. \text{ Pick } x_i = 0, x_j = x_i^* + x_j^* \; s.t. \sum_{i=1}^{n} x_i = r$$

$$\Rightarrow \left(\frac{\partial f(\underline{x}^*)}{\partial x_j} - \frac{\partial f(\underline{x}^*)}{\partial x_i} \right) x_i^* \geq 0 \Rightarrow \frac{\partial f(\underline{x}^*)}{\partial x_i} \leq \frac{\partial f(\underline{x}^*)}{\partial x_j} \; \forall j$$

- Example 4: Projection on a convex set
  - Let Ω be a closed convex set and let $\underline{z}$ be a fixed vector in R^n
  - "projection of $\underline{z}$ onto Ω involves finding a point $\underline{x}^*$ in Ω nearest to $\underline{z}$".
  - Mathematically, $\min_{\underline{x}} \; f(\underline{x}) = \frac{1}{2} \|\underline{z} - \underline{x}\|^2 \; s.t. \; \underline{x} \in \Omega$
  - 4.1





Projection on a Convex Set - 1

– 4.1.

$$\min \frac{1}{2}\|\underline{z}-\underline{x}\|^2 \; s.t. \sum_{i=1}^{n} x_i = r$$

$$\min \frac{1}{2}(z_1-x_1)^2+\frac{1}{2}(z_2-r+x_1)^2 \Rightarrow -z_1+x_1+z_2-r+x_1=0$$

$$\Rightarrow x_1^* = \frac{r}{2}+\frac{z_1-z_2}{2} = \frac{r}{2}+z_1-\bar{z}$$

$$x_2^* = \frac{r}{2}-\frac{z_1-z_2}{2} = \frac{r}{2}+z_2-\bar{z}; \bar{z} = \frac{z_1+z_2}{2}$$

$$\text{In general}, \min \frac{1}{2}\|\underline{z}-\underline{x}\|^2 + \lambda(\underline{x}^T\underline{e}-r) \Rightarrow -(\underline{z}-\underline{x})+\lambda\underline{e}=\underline{0}$$

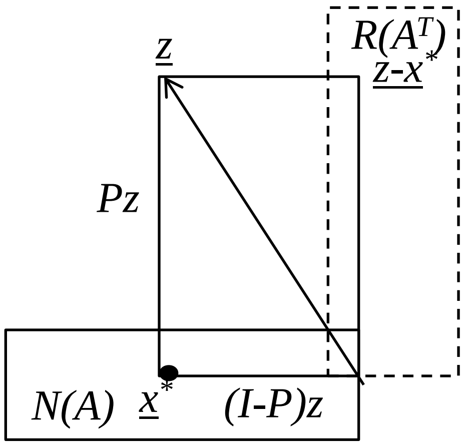
$$\underline{z}-\underline{x}^* = \lambda\underline{e}$$
$$(\underline{z}-\underline{x}^*)^T\underline{x}^* = \lambda\underline{e}^T\underline{x}^*$$
$$\lambda\underline{e}^T\underline{x}^* = \lambda\underline{e}^T(\underline{z}-\lambda\underline{e})$$
$$\Rightarrow (\underline{e}^T\underline{z}-n\lambda) = r$$

$$n\lambda - \underline{e}^T(\underline{z}-\underline{x}^*) = 0 \Rightarrow \lambda = \frac{\underline{e}^T(\underline{z}-\underline{x}^*)}{n} = \frac{\underline{e}^T\underline{z}-r}{n}$$

$$\underline{x}^* = \underline{z}-\lambda\underline{e} = \underline{z}-\frac{\underline{e}^T\underline{z}-r}{n}\underline{e} = \frac{r}{n}\underline{e}+\underline{z}-\bar{z}\underline{e}; \bar{z} = \frac{\underline{e}^T\underline{z}}{n} \Rightarrow \underline{x}^* = \overbrace{(I-\frac{\underline{e}\underline{e}^T}{n})}^{I-P}\underline{z}+\frac{r}{n}\underline{e}$$

Projection on a Convex Set - 2

$$\underline{A} = \underline{e}^T \Rightarrow P = \frac{\underline{e}\,\underline{e}^T}{n}$$

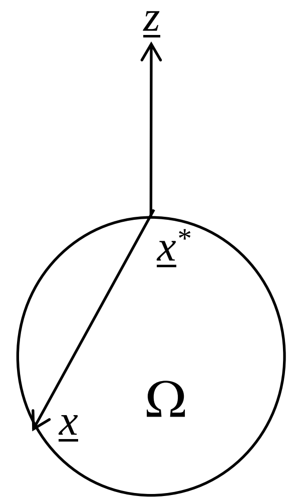


$$\underline{x}^* = \frac{r}{n}\underline{e} + (\underline{z} - \overline{z}\,\underline{e});\ \overline{z} = \frac{\sum_{i=1}^{n} z_i}{n} = \frac{\underline{e}^T \underline{z}}{n}$$

$$= \frac{r}{n}\underline{e} + (I - \frac{\underline{e}\,\underline{e}^T}{n})\underline{z} = \frac{r}{n}\underline{e} + (I - P)\underline{z}$$

$$r = 0 \Rightarrow \underline{x}^* = (I-P)\underline{z} \Rightarrow (\underline{z} - \underline{x}^*) = P\underline{z} \perp \underline{x}^* = (I-P)\underline{z}$$

Definition $\quad \underline{x}^* = \arg\min_{\underline{x}} \|\underline{z} - \underline{x}\|^2 \ s.t.\ \underline{x} \in \Omega$



From necessary conditions

$$\nabla \underline{f}^T(\underline{x}^*)(\underline{x} - \underline{x}^*) \geq 0$$

$$\Rightarrow -(\underline{z}-\underline{x}^*)^T(\underline{x} - \underline{x}^*) \geq 0$$

$$or\ (\underline{z}-\underline{x}^*)^T(\underline{x} - \underline{x}^*) \leq 0$$

Projection on a Convex Set - 3

– 4.2.

$$\min \frac{1}{2}\|\underline{z}-\underline{x}^*\|^2;\ s.t.\ A\underline{x}=\underline{b}$$

A is an mxn matrix
$m<n$
Rank(A)=m

$$L(\underline{x},\underline{\lambda})=\frac{1}{2}(\underline{z}-\underline{x})^T(\underline{z}-\underline{x})+\underline{\lambda}^T(A\underline{x}-\underline{b})$$

$$\frac{\partial L}{\partial \underline{x}}\Big|_{\underline{x}=\underline{x}^*}=\underline{0}\Rightarrow -(\underline{z}-\underline{x}^*)+A^T\underline{\lambda}^*=\underline{0}\Rightarrow \underline{x}^*=\underline{z}-A^T\underline{\lambda}^*$$

$$\text{since } A\underline{x}^*=\underline{b}\Rightarrow \underline{\lambda}^*=(AA^T)^{-1}(A\underline{z}-\underline{b})$$

$$\Rightarrow \underline{x}^*=[I-A^T(AA^T)^{-1}A]\underline{z}+A^T(AA^T)^{-1}\underline{b}$$

$$=(I-P)\underline{z}+A^T(AA^T)^{-1}\underline{b}$$

$$P=A^T(AA^T)^{-1}A=\text{Projection matrix}\Rightarrow P^2=P;P^n=P;n\geq 1$$

$$\text{Recall } \{\underline{x}\mid A\underline{x}=0\}=N(A)$$

$$\underline{x}^*=(I-P)\underline{z}$$

$$\underline{x}^*=\text{Projection of } \underline{z} \text{ onto } N(A)$$

$$P\underline{z}=\underline{z}-\underline{x}^*=\text{Projection of } \underline{z} \text{ onto } R(A^T)$$

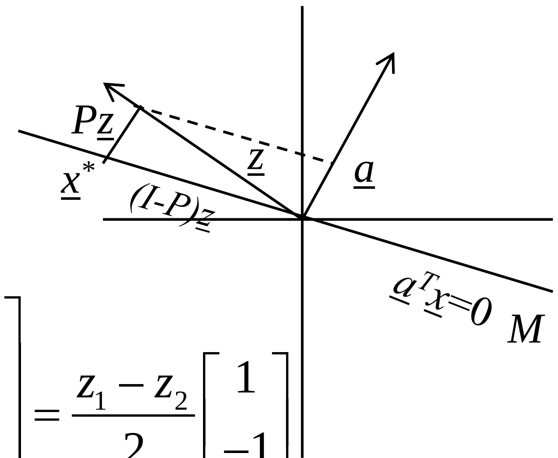
Recall $N(A)\perp R(A^T)$

Projection on a Convex Set - 4

- Suppose m=1 $\Rightarrow$ $\underline{a}^T\underline{x} = 0$ defines an (n-1) dimensional subspace M perpendicular to $\underline{a}$

$$\underline{x}^* = \left[I - \frac{\underline{a}\,\underline{a}^T}{\underline{a}^T\underline{a}} \right] \underline{z} = (I-P)\underline{z};\ P = \frac{\underline{a}\,\underline{a}^T}{\underline{a}^T\underline{a}}$$

$$z - \underline{x}^* = \frac{\underline{a}\,\underline{a}^T}{\underline{a}^T\underline{a}} \underline{z} = P\underline{z} \Rightarrow z - \underline{x}^* \parallel \underline{a}$$



- Suppose $\underline{a}^T = (1,1) \Rightarrow$

$$\underline{x}^* = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{bmatrix} \frac{z_1 - z_2}{2} \\ \frac{z_2 - z_1}{2} \end{bmatrix} = \frac{z_1 - z_2}{2} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$P\underline{z} = \begin{bmatrix} \frac{z_1 + z_2}{2} \\ \frac{z_1 + z_2}{2} \end{bmatrix} = \frac{z_1 + z_2}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

- Recall that a hyperplane H in R^n is defined by $H = \{\underline{x} \mid \underline{a}^T\underline{x} = b\}$

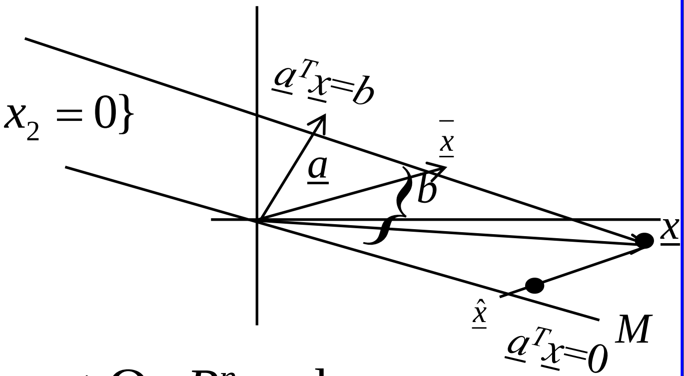
Supporting and Separating Hyperplanes

– Every hyperplane H can be written as $H = \underline{\bar{x}} + M$
  where $\underline{\bar{x}}$ is any vector in H, i.e., $a^T \underline{\bar{x}} = b$

– Example:

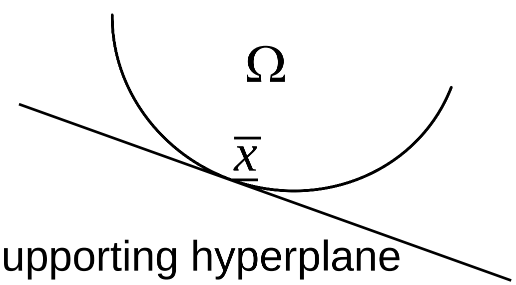
$$H = \{\underline{x} \mid x_1 + x_2 = 1\} = \underbrace{\begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}}_{\bar{x}} + \{\underline{\hat{x}} \mid x_1 + x_2 = 0\}$$

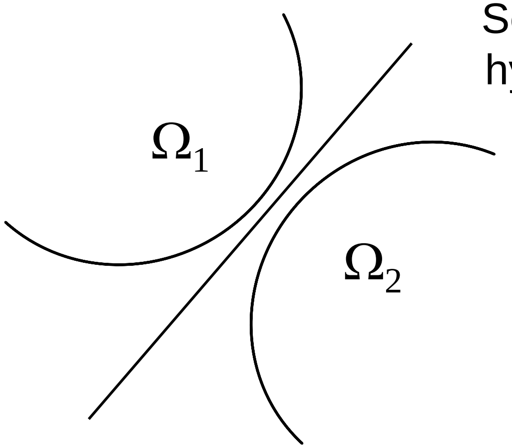
$$\underline{x} = \underline{\bar{x}} + \underline{\hat{x}}$$



- Supporting hyperplane: For every convex set $\Omega \in R^n$ and every boundary point $\underline{\bar{x}}$, $\exists$ a hyperplane that supports Ω at $\underline{\bar{x}}$, i.e.,

$$a^T \underline{\bar{x}} \le a^T \underline{x} \; \forall \underline{x} \in \Omega$$





Equality Constraints

- Separating hyperplane: If Ω_1 and Ω_2 are two disjoint convex sets, then $\exists$ a hyperplane that separates them

$$a^T \underline{x} > a^T \underline{y} \; \forall \underline{x} \in \Omega_1, \underline{y} \in \Omega_2$$

- Equality constraints

$$\min f(\underline{x}) \;\; s.t. \;\; h_i(\underline{x}) = 0, i = 1, 2, \ldots, m$$

$$f(\underline{x}) \text{ and } h_i(\underline{x}) \text{ are differentiable in } R^n$$

Equivalently,

$$\min f(\underline{x})$$

$$s.t. \;\; \underline{h}(\underline{x}) = 0; \;\; h = (h_1, h_2, \ldots, h_m)^T$$

We will provide intuitive proofs first and then provide geometric interpretations later

- Special case:

$$\min f(\underline{x})$$

$$s.t. \;\; A\underline{x} = b$$ A is an m x n matrix

Linear Equality Constraints - 1

– Without loss of generality, assume that the first m columns are independent

$$A = m\begin{bmatrix} \overset{m}{B} & | & \overset{n-m}{N} \end{bmatrix}; \underline{x} = \begin{bmatrix} \underline{x}_B & \underline{x}_N \end{bmatrix}^T; \underline{x}_B \in R^m \ \& \ \underline{x}_N \in R^{n-m}$$

The constraints imply

$$B\underline{x}_B + N\underline{x}_N = \underline{b} \Rightarrow \underline{x}_B = B^{-1}[\underline{b} - N\underline{x}_N]$$

$$\text{Let us look at } f(\underline{x}) = f(\underline{x}_B, \underline{x}_N) = f(B^{-1}(b - N\underline{x}_N), \underline{x}_N)$$

Constrained problem		Unconstrained problem
$\min f(\underline{x})$	$\Rightarrow$	$\min \tilde{f}(\underline{x}_N) = f(B^{-1}(b - N\underline{x}_N), \underline{x}_N)$
$s.t.\ A\underline{x} = b$		$s.t.\ \underline{x}_N \in R^{n-m}$

– Necessary condition for optimality $\nabla\tilde{f}(\underline{x}_N) = 0$

$$\nabla_{\underline{x}_N} \underline{\tilde{f}}(\underline{x}_N) = \nabla_{\underline{x}_N} \underline{f}(\underline{x}^*) + \nabla_{\underline{x}_N} \underline{x}_B \nabla_{\underline{x}_B} \underline{f}(\underline{x}^*)$$

$$= \nabla_{\underline{x}_N} \underline{f}(\underline{x}^*) - N^T (B^{-1})^T \nabla_{\underline{x}_B} \underline{f}(\underline{x}^*) = 0$$

$$\text{We let } (-B^T)^{-1} \nabla_{\underline{x}_B} \underline{f}(\underline{x}^*) = \underline{\lambda}^*$$

Linear Equality Constraints - 2

We have a set of simultaneous equations:

$$\nabla_{\underline{x}_N} \underline{f}(\underline{x}^*) + N^T \underline{\lambda}^* = 0$$

$$\nabla_{\underline{x}_B} \underline{f}(\underline{x}^*) + B^T \underline{\lambda}^* = 0$$

$$\Rightarrow \boxed{\nabla f(\underline{x}^*) + A^T \underline{\lambda}^* = 0}$$

Alternatively, if we write

$\underline{h}(\underline{x}) = A\underline{x} - \underline{b}$, then

$$\nabla \underline{f}(\underline{x}^*) + \nabla \underline{\underline{h}}(\underline{x}^*)\lambda^* = 0$$

$$\nabla \underline{\underline{h}}(\underline{x}^*) = [\nabla \underline{h}_1, \nabla \underline{h}_2, .., \nabla \underline{h}_n] = [row1, row2, ..., row\, m]_{nxm}$$

$$\Rightarrow \nabla \underline{f}(\underline{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla \underline{h}_i(\underline{x}^*) = \underline{0}$$

- At a local minimum GRADIENT OF THE COST FUNCTION = Linear combination of the gradients of CONSTRAINTS & WEIGHTS = Lagrange multipliers

Lagrangian Approach

- Second approach: Form Lagrangian function

$$L(\underline{x},\underline{\lambda}) = f(\underline{x}) + \underline{\lambda}^T \underline{h}(\underline{x}), \underline{\lambda} \in R^m$$

$$\frac{\partial L}{\partial \underline{x}} = \underline{0} \Rightarrow \nabla \underline{f}(\underline{x}^*) + \nabla \underline{\underline{h}}(\underline{x})\underline{\lambda} = 0$$

$$\frac{\partial L}{\partial \lambda} = \underline{0} \Rightarrow \underline{h}(\underline{x}) = \underline{0}$$

Convert an equality constrained problem into an unconstrained problem

when $\underline{h}(\underline{x}) = A\underline{x} - \underline{b} = \underline{0} \Rightarrow \nabla \underline{f}(\underline{x}^*) + A^T \underline{\lambda}^* = \underline{0}; A\underline{x}^* = \underline{b}$

  – Finding stationary $\underline{x}^*$ and $\underline{\lambda}^*$ involves solving (n+m) equations

$$\nabla \underline{f}(\underline{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla \underline{h}_i(\underline{x}^*) = 0$$

$$\underline{h}(\underline{x}^*) = 0$$

- What about second order conditions: Consider a point $\underline{x}^*$ satisfying the 1st order necessary conditions. Suppose we go from $\underline{x}^*$ to $\underline{x}^* + \delta\underline{x}$ another feasible point. To 2nd order

Second Order Conditions - 1

$$\delta f(\underline{x}^*) = f(\underline{x}^* + \delta\underline{x}) - f(\underline{x}^*) = \nabla \underline{f}^T(\underline{x}^*)\delta\underline{x} + \frac{1}{2}\delta\underline{x}^T\nabla^2 f(\underline{x}^*)\delta\underline{x} + \text{higher order}$$

$$\delta h_i(\underline{x}^*) = h_i(\underline{x}^* + \delta\underline{x}) - h_i(\underline{x}^*) = \nabla \underline{h}_i^T(\underline{x}^*)\delta\underline{x} + \frac{1}{2}\delta\underline{x}^T\nabla^2 h_i(\underline{x}^*)\delta\underline{x} + \text{higher order}$$

$$\delta f(\underline{x}^*) + \sum_{i=1}^{m}\lambda_i\delta h_i(\underline{x}^*) = [\nabla \underline{f}(\underline{x}^*) + \sum_{i=1}^{m}\lambda_i\nabla \underline{h}_i(\underline{x}^*)]^T\delta\underline{x}$$

$$+\frac{1}{2}\delta\underline{x}^T[\nabla^2 f(\underline{x}^*) + \sum_{i=1}^{m}\lambda_i\nabla^2 h_i(\underline{x}^*)]\delta\underline{x} + \text{higher order}$$

- For $\underline{x}^* + \delta\underline{x}$ to be feasible need $\delta h_i(\underline{x}^*) = 0\ \forall i = 1,2,..,m$

  $$so,\ \delta f(\underline{x}^*) = \frac{1}{2}\delta\underline{x}^T[\nabla^2 f(\underline{x}^*) + \sum_{i=1}^{m}\lambda_i\nabla^2 h_i(\underline{x}^*)]\delta\underline{x}$$

- For $\underline{x}^*$ to be a local minimum, need $\delta f(\underline{x}^*) \geq 0$ for all feasible $\delta\underline{x}$ around $\underline{x}^*$, i.e., those $\underline{x}^*$ satisfying $h_i(\underline{x}) = 0\ \forall i = 1,2,..,m$

  $$\Rightarrow Need\ \delta\underline{x}^T\underbrace{[\nabla^2 f(\underline{x}^*) + \sum_{i=1}^{m}\lambda_i\nabla^2 h_i(\underline{x}^*)]}_{\nabla^2 L_{xx}(\underline{x}^*)}\delta\underline{x} \geq 0$$

Second Order Conditions - 2

for all $\delta \underline{x}$ satisfying $\delta h_i(\underline{x}) = 0, i = 1, 2, .., m$

since $\delta h_i(\underline{x}) = \nabla \underline{h}_i^T(\underline{x}^*)\delta \underline{x} = 0$

– Necessary conditions:

$$\left.\begin{aligned} &\nabla \underline{f}(\underline{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla \underline{h}_i(\underline{x}^*) = 0 \\ &h_i(\underline{x}^*) = 0; i = 1, 2, .., m \end{aligned}\right\} \text{First order}$$

$$\underline{y}^T \nabla^2 L_{xx} \underline{y} = \underline{y}^T [\nabla^2 f(\underline{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla^2 h_i(\underline{x}^*)]\underline{y} \geq 0 \; \forall \underline{y}$$

$$\nabla \underline{h}_i^T(\underline{x}^*)\underline{y} = 0; i = 1, 2, .., m$$

- Special case: $h_i(\underline{x}) = 0 \Rightarrow \underline{a}_i^T \underline{x} = b_i$
  – Necessary conditions: $$\left.\begin{aligned} \nabla \underline{f}(\underline{x}^*) + A^T \underline{\lambda}^* &= 0 \\ A\underline{x}^* &= \underline{b} \end{aligned}\right\} \text{First order}$$

Second Order Conditions - 3

$$\Rightarrow \nabla \underline{f}(\underline{x}^*) + \sum_{i=1}^{m} \lambda_i \underline{a}_i = 0; A = \begin{pmatrix} \underline{a}_1^T \\ \underline{a}_2^T \\ . \\ \underline{a}_m^T \end{pmatrix}$$

$$\underline{y}^T \nabla^2 L_{xx} \underline{y} = \underline{y}^T \nabla^2 f(\underline{x}^*) \underline{y} \geq 0 \, \forall \underline{y} \Rightarrow A\underline{y} = 0$$

Recall that for any $\underline{y} \in R^n$, $A\underline{y} = 0$, defines the null space of A, $N(A)$

Using SVD representation

$$A = U \Sigma V^T = \begin{pmatrix} \underline{u}_1 & \underline{u}_2 & .. & \underline{u}_m \end{pmatrix} \begin{pmatrix} \sigma_1 & & & & . & \\ & . & & & . & 0 \\ & & . & & . & \\ & & & \sigma_m & . & \end{pmatrix} \begin{pmatrix} \underline{v}_1^T \\ . \\ \underline{v}_m^T \\ \underline{v}_{m+1}^T \\ . \\ \underline{v}_n^T \end{pmatrix} = \sum_{i=1}^{m} \underline{u}_i \underline{v}_i^T \sigma_i$$

Second Order Conditions - 4

$$A\underline{y} = 0 \Rightarrow \underline{y} = \sum_{i=1} \alpha_i \underline{v}_{m+i} = \tilde{V}\underline{\alpha}$$

Linear combination of last $(n-m)$ rows of V

$$\Rightarrow \underline{y}^T \nabla^2 L_{xx} \underline{y} \geq 0 \Rightarrow \underline{\alpha}^T \tilde{V}^T \nabla^2 L_{xx} \tilde{V} \underline{\alpha} = \underline{\alpha}^T \tilde{V}^T \nabla^2 f(\underline{x}^*) \tilde{V} \underline{\alpha} \geq 0$$

Alternately, since $(I-P)\underline{y} = 0$ where $P = A^T (AA^T)^{-1} A^T$

$$\Rightarrow Rank[(I-P)\nabla^2 L_{xx}(I-P)] = n-m$$

Infact, any orthonormal basis of $N(A)$ will do to define $\tilde{V}$

- Sufficient conditions:

$$\underline{y}^T \nabla^2 L_{xx} \underline{y} > 0 \, \forall \underline{y} \ni \nabla \underline{h}_i^T(\underline{x})\underline{y} = 0; i = 1, 2, .., m$$

- If $\nabla \underline{h}_i(\underline{x}^*)$ are linearly independent, then the subspace

  Special case: Linear constraints

$$M(\underline{x}^*) = \{\underline{y} \,|\, A\underline{y} = 0\} = N(A)$$

Linear independence $\Rightarrow R(A)$ spans m dimensional subspace$=R^m$

- When $\nabla \underline{h}_i(\underline{x}^*)$ are independent, the constraints are said to be regular or qualified

Illustrative Examples - 1

- Example 1: $\min \frac{1}{2}(x_1^2 + x_2^2 + x_3^2)\ s.t.\ x_1 + x_2 + x_3 = 3$

First order conditions:

$$\left.\begin{aligned} x_1 + \lambda &= 0 \\ x_2 + \lambda &= 0 \\ x_3 + \lambda &= 0 \\ x_1 + x_2 + x_3 &= 3 \end{aligned}\right\} \Rightarrow \begin{aligned} x_1^* = x_2^* = x_3^* &= 1 \\ \lambda^* &= -1 \end{aligned}$$

Second order conditions:

$$\nabla^2 L_{xx} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{aligned} y^T \nabla^2 L_{xx} y &= y_1^2 + y_2^2 + y_3^2 \ni y_1 + y_2 + y_3 = 0 \\ &= y_1^2 + y_2^2 + (y_1 + y_2)^2 > 0 \forall y_1, y_2 \neq 0 \end{aligned}$$

$\Rightarrow$ strict local minimum

Illustrative Examples - 2

Alternately, $a^T = (1,1,1)$ or $A = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}$

Orthonormal basis for $\tilde{V}^T$

$$\underline{v}_2^T = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 & -1 \end{pmatrix}$$

$$\underline{v}_3^T = \frac{1}{\sqrt{6}}\begin{pmatrix} 1 & -2 & 1 \end{pmatrix}$$

$$\tilde{V}^T \nabla^2 L_{xx} \tilde{V} = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ 0 & -\frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} > 0$$

Illustrative Examples - 3

$$\text{Orthogonal Projection: } I-P=\begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{pmatrix};$$

$$(I-P)\nabla^2 L_{xx}(I-P)=(I-P)^2=(I-P)$$

Rank=2 $\Rightarrow$ spans 2-dimensional subspace

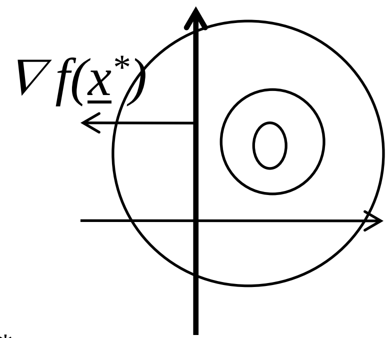
- If $\underline{x}^*$ is a local minimum but not a regular point, $\exists$ no or infinite number of Lagrange multipliers

Example 2:

$$f(\underline{x})=(x_1-1)^2+(x_2-1)^2 \;\; s.t.\, x_1^{\,2}=0$$

$$\min \text{ at } x_1=0 \,\&\, x_2=1 \text{ unique}$$

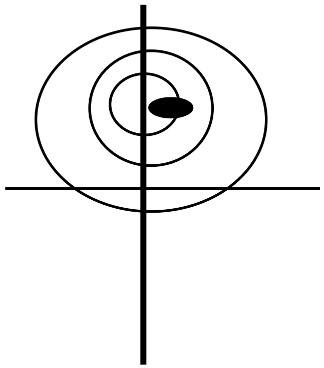
$$\nabla f(\underline{x}^*)+\lambda^*\nabla h(\underline{x}^*)=\begin{pmatrix}-2\\0\end{pmatrix}+\lambda^*\begin{pmatrix}0\\0\end{pmatrix}\neq\begin{pmatrix}0\\0\end{pmatrix}\forall\lambda^*$$



Never satisfied

Illustrative Examples - 4

- Example 3: $\min f(\underline{x}) = x_1^2 + (x_2 - 1)^2 \;\; s.t.\, x_1^2 = 0$



$$\nabla f(\underline{x}^*) + \lambda^* \nabla h(\underline{x}^*) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \lambda^* \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\text{Solution } x_1^* = 0, x_2^* = 1, \lambda^* = \text{anything}$$

- For linear constraints, Lagrange multipliers exist even if the constraints are not regular. We will discuss this later in the context of convex programming problems
- Example 4: A continuous random variable *x*. Don't know the density *p(x)*

$$Given\, E(x) = m = \int_{-\infty}^{\infty} xp(x)dx \qquad (1)$$

$$E(x^2) = m^2 + \sigma^2 = \int_{-\infty}^{\infty} x^2 p(x)dx \quad (2)$$

Illustrative Examples - 5

$$\left.\begin{aligned} Need \quad \int_{-\infty}^{\infty} p(x)dx &= 1 \\ p(x) &\geq 0 \end{aligned}\right\} \qquad (3)$$

One way of finding $p(x)$ is to maximize entropy $H = -E\{\log_e p(x)\}$

$$\max_{p(x)\geq 0} - E\{\log_e p(x)\} = \max_{p(x)\geq 0} - \int_{-\infty}^{\infty} p(x)\log_e p(x)dx$$

such that (1), (2), (3) satisfies

$$L(p,\lambda,\mu,\delta) = \int_{-\infty}^{\infty} \{-p(x)\ln p(x) - \lambda x p(x) - \mu x^2 p(x) - \delta p(x)\}dx$$

$$+ \lambda m + \mu(m^2 + \sigma^2) + \delta$$

$$\frac{\partial L}{\partial p(x)} = 0 \Rightarrow -\ln p(x) - 1 - \lambda x - \mu x^2 - \delta = 0$$

$$\Rightarrow p(x) = c\exp(-\delta - 1 - \lambda x - \mu x^2)$$

Illustrative Examples - 6

$$= c^{'} \exp(-\mu(x+\frac{\lambda}{2\mu})^2); c^{'} = c\exp(\mu(\frac{\lambda}{2\mu})^2 - \delta - 1)$$

$$\Rightarrow Normal\ density\ \ p(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

$$-\frac{\lambda}{2\mu} = m$$

$$\mu = \frac{1}{2\sigma^2}$$

For only 1st moment

$$p(x) = c\exp(-\delta - 1 - \lambda x)$$

$$\left.\begin{array}{l} \int_{-\infty}^{\infty} p(x)dx = 1 \Rightarrow \frac{ce^{-(\delta+1)}}{\lambda} = 1 \\ \int_{-\infty}^{\infty} xp(x)dx = m \Rightarrow \frac{ce^{-(\delta+1)}}{\lambda^2} = m \end{array}\right\} \Rightarrow m = \frac{1}{\lambda}$$

Illustrative Examples - 7

$$ce^{-(\delta+1)} = \lambda$$

$$\Rightarrow p(x) = \lambda e^{-\lambda x} = \frac{1}{m} e^{-\frac{1}{m}x} \Rightarrow \text{ exponential density}$$

$$\frac{\partial^2 L}{\partial p^2(x)} = \frac{-1}{p(x)} < 0$$

- Example 5:

$$\max xyz$$

$$s.t. (xy + yz + zx) = \frac{c}{2}$$

$$\text{Solution:} x = y = z = \sqrt{\frac{c}{6}}$$

$$\Rightarrow \text{a cube} \Rightarrow \text{volume} = \frac{c}{6}\sqrt{\frac{c}{6}}$$

Interpretation of Lagrange Multipliers - 1

- Sensitivity interpretation

$$\min f(x_1,x_2)\; s.t.\; h_1(x_1,x_2)=b_1$$

$$L(\underline{x},\lambda_1)= f(\underline{x})+\lambda_1(h_1(\underline{x})-b_1)$$

Suppose x_1^*, x_2^*, λ_1 is optimal

$$0=\frac{\partial L}{\partial x_1}\Rightarrow \frac{\partial f}{\partial x_1}+\lambda_1\frac{\partial h_1}{\partial x_1}=0 \;\&\; \frac{\partial f}{\partial x_2}+\lambda_1\frac{\partial h_1}{\partial x_2}=0$$

$$\frac{\partial f}{\partial b_1}=\frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial b_1}+\frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial b_1}=\nabla_{b_1}\underline{x}\nabla \underline{f}(\underline{x}) \quad (1)$$

$$h_1(\underline{x}^*)-b_1=0\Rightarrow [\frac{\partial h_1}{\partial x_1}\frac{\partial x_1}{\partial b_1}+\frac{\partial h_1}{\partial x_2}\frac{\partial x_2}{\partial b_1}-1]=0 \quad (2)$$

$(1)+\lambda_1*(2)$ gives

$$\frac{\partial f}{\partial b_1}=-\lambda_1+\sum_{i=1}^{2}\underbrace{(\frac{\partial f}{\partial x_i}+\lambda_i\frac{\partial h_1}{\partial x_i})}_{0}\frac{\partial x_i}{\partial b_1}$$

$$\lambda_1=-\frac{\partial f}{\partial b_1}$$

Rate of change of f wrt level of constraint changes (or)
Impact on the cost function if additional resources are added

Interpretation of Lagrange Multipliers - 2

$$In\ general,\ \nabla_b \underline{f}(\underline{x}^*(\underline{b})) = -\underline{\lambda}(\underline{b})$$

Consider $\quad \min f(\underline{x})$

$$s.t.\ \ \underline{h}(\underline{x}) = \underline{b}$$

Necessary condition:

$$\nabla \underline{f}(\underline{x}) + \nabla \underline{\underline{h}}(\underline{x})\underline{\lambda} = 0$$

$$\underline{h}(\underline{x}) = \underline{b}$$

Jacobian at $\underline{x}^*, \underline{\lambda}^*$

$$J = \begin{bmatrix} \overbrace{\nabla^2 f(\underline{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i^2(\underline{x}^*)}^{\nabla^2 L_{xx}} & | & \nabla \underline{\underline{h}}(\underline{x}^*) \\ - & | & - \\ \nabla \underline{\underline{h}}^T(\underline{x}^*) & | & 0 \end{bmatrix}$$

Interpretation of Lagrange Multipliers - 3

J is non-singular (why?)

If not $\exists \begin{pmatrix} \underline{y} \\ \underline{z} \end{pmatrix} \ni J \begin{pmatrix} \underline{y} \\ \underline{z} \end{pmatrix} = \underline{0}$

$$\Rightarrow \nabla^2 L_{xx} \underline{y} + \nabla h(\underline{x}^*) \underline{z} = \underline{0}$$

$$\nabla \underline{\underline{h}}^T (\underline{x}^*) \underline{y} = \underline{0}$$

Premultiply 1st equation by $\underline{y}^T$

$\Rightarrow \underline{y}^T \nabla^2 L_{xx} \underline{y} = 0 \Rightarrow \underline{x}^*, \underline{\lambda}^*$ is not a strict local minimum

Contradiction

$\therefore \underline{y} = 0 \Rightarrow \underline{z} = 0$ since $\nabla \underline{\underline{h}}(\underline{x}^*)$ has rank m

For nearby $\underline{b}$, we have

$$\nabla \underline{f}(\underline{x}(\underline{b})) + \nabla \underline{\underline{h}}(\underline{x}(b)) \underline{\lambda}^*(\underline{b}) = 0$$

$$\underline{h}(\underline{x}(\underline{b})) = \underline{b}$$

Interpretation of Lagrange Multipliers - 4

Taking gradient of Lagrangian wrt $\underline{b}$

$$\nabla_b \underline{x}(\underline{b}) \nabla \underline{f}(\underline{x}(\underline{b})) + \overbrace{\nabla_b \underline{x}(\underline{b}) \nabla \underline{\underline{h}}(\underline{x}(\underline{b}))}^{I} \underline{\lambda}(\underline{b}) = \underline{0}$$

$$\nabla_b \underline{f}(\underline{x}(b)) + \underline{\lambda}(b) = \underline{0} \Rightarrow \quad \boxed{\underline{\lambda}(b) = -\nabla_b \underline{f}(\underline{x}(b))}$$

Summary

- Necessary and Sufficient Conditions
- Methods of Specifying constraint set (Ω)
- Basic Result: Necessary Conditions of Optimality
- Examples
- Equality Constraints
- Economical Interpretation of Lagrange Multipliers