# Lecture 7: Constrained Optimization: Lagrangian Multipliers, Optimality Conditions

- Equality constrained problems
- Basic Lagrange multiplier theorem
- Inequality constrained problems
- Sensitivity analysis
- Farkas lemma
- Linearly constrained problems

# Equality Constrained Problem

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & h_i(\boldsymbol{x}) = 0, \quad i = 1, ..., m. \end{array}$$

where $f : \mathbb{R}^n \mapsto \mathbb{R}$, $h_i : \mathbb{R}^n \mapsto \mathbb{R}$, $i = 1, ..., m$, are continuously differentiable functions.

## Lagrange Multiplier Theorem

- Let $\boldsymbol{x}^*$ be a local min and a regular point $[\nabla h_i(\boldsymbol{x}^*):$ linearly independent$]$. Then there exist unique scalars $\lambda_1^*, ..., \lambda_m^*$ such that
$$\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(\boldsymbol{x}^*) = 0.$$

- If in addition $f$ and $h$ are twice continuously differentiable,
$$\boldsymbol{y}' \left( \nabla^2 f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla^2 h_i(\boldsymbol{x}^*) \right) \boldsymbol{y} \geq 0, \quad \forall\, \boldsymbol{y} \text{ s.t. } \nabla h(\boldsymbol{x}^*)' \boldsymbol{y} = \mathbf{0}$$

# Proof

- Suppose $\boldsymbol{x}^*$ is a local min satisfying $h(\boldsymbol{x}^*) = \mathbf{0}$ and $\alpha > 0$. Consider
$$f^k(\boldsymbol{x}) = f(\boldsymbol{x}) + k\|\boldsymbol{h}(\boldsymbol{x})\|^2 + \frac{\alpha}{2}\|\boldsymbol{x} - \boldsymbol{x}^*\|^2.$$

- Let $\boldsymbol{x}^k$ be a constrained minimizer of $f^k$ over the ball $\|\boldsymbol{x} - \boldsymbol{x}^*\| \leq 1$. We will show that $\boldsymbol{x}^k$ is an *unconstrained local min* of $f^k$ for all large $k$.

- Taking limit $k \to \infty$ of
$$f^k(\boldsymbol{x}^k) = f(\boldsymbol{x}^k) + k|h(\boldsymbol{x}^k)|^2 + \frac{\alpha}{2}\|\boldsymbol{x}^k - \boldsymbol{x}^*\|^2 \leq f^k(\boldsymbol{x}^*) = f(\boldsymbol{x}^*),$$
along any convergent subsequence of $\{\boldsymbol{x}^k\}$, we get $h(\bar{\boldsymbol{x}}) = \lim_{k\to\infty} h(\boldsymbol{x}^k) = 0$.

- Furthermore, taking limit of $f(\boldsymbol{x}^k) + \frac{\alpha}{2}\|\boldsymbol{x}^k - \boldsymbol{x}^*\|^2 \leq f(\boldsymbol{x}^*)$ shows
$$f(\bar{\boldsymbol{x}}) + \frac{\alpha}{2}\|\bar{\boldsymbol{x}} - \boldsymbol{x}^*\|^2 \leq f(\boldsymbol{x}^*)$$

- Since $h(\bar{\boldsymbol{x}}) = 0$, it follows that $f(\boldsymbol{x}^*) \leq f(\bar{\boldsymbol{x}})$. Thus, we have $\bar{\boldsymbol{x}} = \boldsymbol{x}^*$ and $f(\boldsymbol{x}^*) = f(\bar{\boldsymbol{x}})$.

- Since $\bar{\boldsymbol{x}}$ is any limit point, we have $\boldsymbol{x}^k \to \boldsymbol{x}^*$, so $\|\boldsymbol{x}^k - \boldsymbol{x}^*\| < 1$ for large $k$, $\Rightarrow \boldsymbol{x}^k$ is an unconstrained local min of $f^k$, $\nabla f^k(\boldsymbol{x}^k) = 0,\ \nabla^2 f^k(\boldsymbol{x}^k) \succeq \mathbf{0}$.

- Taking limit of

$$\mathbf{0} \quad = \quad \nabla f(\boldsymbol{x}^k) + 2kh(\boldsymbol{x}^k)\nabla h(\boldsymbol{x}^k) + \alpha(\boldsymbol{x}^k - \boldsymbol{x}^*) \tag{1}$$

  Since $\nabla h(\boldsymbol{x}^*)$ has rank $m$, $\nabla h(\boldsymbol{x}^k)$ also has rank $m$ for large $k$, so $\nabla h(\boldsymbol{x}^k)'\nabla h(\boldsymbol{x}^k)$: invertible. Thus, multiplying (1) w/ $\nabla h(\boldsymbol{x}^k)'$ yields

$$kh(\boldsymbol{x}^k) = -\left(\nabla h(\boldsymbol{x}^k)'\nabla h(\boldsymbol{x}^k)\right)^{-1} \nabla h(\boldsymbol{x}^k)' \left(\nabla f(\boldsymbol{x}^k) + \alpha(\boldsymbol{x}^k - \boldsymbol{x}^*)\right).$$

- Taking limit as $k \to \infty$ and $\boldsymbol{x}^k \to x^*$,

$$\{kh(\boldsymbol{x}^k)\} \to \left(\nabla h(\boldsymbol{x}^*)'\nabla h(\boldsymbol{x}^*)\right)^{-1} \nabla h(\boldsymbol{x}^*)'\nabla f(\boldsymbol{x}^*) \equiv \boldsymbol{\lambda}.$$

Taking limit as $k \to \infty$ in Eq. (1), we obtain

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

- **Exercise:** 2nd order L-multiplier condition: Use 2nd order unconstrained condition for $\boldsymbol{x}^k$, and algebra.

# Lagrangian Function

- Define the Lagrangian function

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i h_i(\boldsymbol{x}).$$

Then, if $\boldsymbol{x}^*$ is a local minimum which is regular, the Lagrange multiplier conditions can be written as

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}, \quad \nabla_{\boldsymbol{\lambda}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}) = \mathbf{0}.$$

- System of $n+m$ equations with $n+m$ unknowns.
- 2nd order condition: $\boldsymbol{y}' \nabla^2_{\boldsymbol{xx}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \boldsymbol{y} \geq 0, \quad \forall\, \boldsymbol{y}$ s.t. $\nabla h(\boldsymbol{x}^*)' \boldsymbol{y} = \mathbf{0}$.
- Example:

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\left(x_1^2 + x_2^2 + x_3^2\right) \\ \text{subject to} & x_1 + x_2 + x_3 = 3. \end{array}$$

Necessary conditions

$$x_1^* + \lambda^* = 0, \quad x_2^* + \lambda^* = 0, \quad x_3^* + \lambda^* = 0, \quad x_1^* + x_2^* + x_3^* = 3.$$

# Sufficiency Condition

- Second Order Sufficiency Conditions: Let $\boldsymbol{x}^* \in \mathbb{R}^n$ and $\boldsymbol{\lambda} \in \mathbb{R}^m$ satisfy

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}, \quad \nabla_{\boldsymbol{\lambda}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0},$$

$$\boldsymbol{y}' \nabla^2_{\boldsymbol{xx}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \boldsymbol{y} > 0, \quad \forall\, \boldsymbol{y} \neq \mathbf{0} \text{ with } \nabla h(\boldsymbol{x}^*)' \boldsymbol{y} = \mathbf{0}.$$

Then $\boldsymbol{x}^*$ is a strict local minimum.

- Example:

$$\begin{array}{ll} \text{minimize} & -(x_1x_2 + x_2x_3 + x_1x_3) \\ \text{subject to} & x_1 + x_2 + x_3 = 3. \end{array}$$

We have that $x_1^* = x_2^* = x_3^* = 1$ and $\boldsymbol{\lambda}^* = 2$ satisfy the 1st order conditions. Also

$$\nabla^2_{\boldsymbol{xx}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \begin{bmatrix} 0 & -1 & -1 \\ -1 & 0 & -1 \\ -1 & -1 & 0 \end{bmatrix}$$

We have for all $\boldsymbol{y} \neq \mathbf{0}$ with $\nabla h(\boldsymbol{x}^*)' \boldsymbol{y} = \mathbf{0}$ or $y_1 + y_2 + y_3 = 0$,

$$\boldsymbol{y}' \nabla^2_{\boldsymbol{xx}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \boldsymbol{y} = -y_1(y_2 + y_3) - y_2(y_1 + y_3) - y_3(y_1 + y_2) = y_1^2 + y_2^2 + y_3^2 > 0.$$

Hence, $\boldsymbol{x}^*$ is a strict local minimum.

# A Useful Lemma

- Let $\boldsymbol{P}$ and $\boldsymbol{Q}$ be two symmetric matrices. Assume that $\boldsymbol{Q} \succeq \mathbf{0}$ and $\boldsymbol{P} \succ \mathbf{0}$ on the nullspace of $\boldsymbol{Q}$, i.e., $\boldsymbol{x}'\boldsymbol{P}\boldsymbol{x} > 0$ for all $\boldsymbol{x} \neq \mathbf{0}$ with $\boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x} = 0$. Then there exists a scalar $c$ such that

$$\boldsymbol{P} + c\boldsymbol{Q}: \text{ positive definite}, \quad \forall\, c > \bar{c}.$$

- Proof: Assume the contrary. Then for every $k$, there exists a vector $\boldsymbol{x}^k$ with $\|\boldsymbol{x}^k\| = 1$ such that

$$(\boldsymbol{x}^k)'\boldsymbol{P}\boldsymbol{x}^k + k(\boldsymbol{x}^k)'\boldsymbol{Q}\boldsymbol{x}^k \leq 0.$$

  Consider a subsequence $\{\boldsymbol{x}^k\}_{k\in\mathcal{K}}$ converging to some $\boldsymbol{x}$ with $\|\boldsymbol{x}\| = 1$. Taking the limit superimum,

$$\boldsymbol{x}'\boldsymbol{P}\boldsymbol{x} + \limsup_{k\to\infty, k\in\mathcal{K}} \left(k(\boldsymbol{x}^k)'\boldsymbol{Q}\boldsymbol{x}^k\right) \leq 0.$$

  We have $(\boldsymbol{x}^k)'\boldsymbol{Q}\boldsymbol{x}^k \geq 0$ (since $\boldsymbol{Q} \succeq \mathbf{0}$), so

$$\{(\boldsymbol{x}^k)'\boldsymbol{Q}\boldsymbol{x}^k\}_{k\in\mathcal{K}} \to 0.$$

  Therefore, $\boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x} = 0$ and using the hypothesis, $\boldsymbol{x}'\boldsymbol{P}\boldsymbol{x} > 0$, a contradiction.

# Proof of Sufficiency Conditions

Consider the *augmented Lagrangian function*

$$L_c(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \boldsymbol{\lambda}'\boldsymbol{h}(\boldsymbol{x}) + \frac{c}{2}\|\boldsymbol{h}(\boldsymbol{x})\|^2,$$

where $c$ is a scalar. We have

$$\nabla_{\boldsymbol{x}} L_c(\boldsymbol{x}, \boldsymbol{\lambda}) = \nabla_{\boldsymbol{x}} L(\boldsymbol{x}, \tilde{\boldsymbol{\lambda}}), \quad \nabla^2_{\boldsymbol{xx}} L_c(\boldsymbol{x}, \boldsymbol{\lambda}) = \nabla^2_{\boldsymbol{xx}} L(\boldsymbol{x}, \tilde{\boldsymbol{\lambda}}) + c\nabla \boldsymbol{h}(\boldsymbol{x})\nabla \boldsymbol{h}(\boldsymbol{x})'$$

where $\tilde{\boldsymbol{\lambda}} = \boldsymbol{\lambda} + c\boldsymbol{h}(\boldsymbol{x})$. If $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ satisfy the sufficiency conditions, we have using the lemma,

$$\nabla_{\boldsymbol{x}} L_c(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}, \quad \nabla^2_{\boldsymbol{xx}} L_c(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) > 0,$$

for sufficiently large $c$. Hence for some $\gamma > 0,\ \epsilon > 0$,

$$L_c(\boldsymbol{x}, \boldsymbol{\lambda}^*) \geq L_c(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) + \frac{\gamma}{2}\|\boldsymbol{x} - \boldsymbol{x}^*\|^2,$$

if $\|\boldsymbol{x} - \boldsymbol{x}^*\| < \epsilon$. Since $L_c(\boldsymbol{x}, \boldsymbol{\lambda}^*) = f(\boldsymbol{x})$ when $\boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}$,

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) + \frac{\gamma}{2}\|\boldsymbol{x} - \boldsymbol{x}^*\|^2, \quad \text{if } \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0},\ \|\boldsymbol{x} - \boldsymbol{x}^*\| < \epsilon.$$

# Sensitivity

- Consider the linearly constrained problem $\min_{\boldsymbol{a}'\boldsymbol{x}=b} f(\boldsymbol{x})$.

- If $b$ is changed to $b+\Delta b$, the minimum $\boldsymbol{x}^*$ will change to $\boldsymbol{x}^*+\Delta\boldsymbol{x}$.

- Since $b+\Delta b = \boldsymbol{a}'(\boldsymbol{x}^*+\Delta\boldsymbol{x}) = \boldsymbol{a}'\boldsymbol{x}^* + \boldsymbol{a}'\Delta\boldsymbol{x} = b + \boldsymbol{a}'\Delta\boldsymbol{x}$, we have $\boldsymbol{a}'\Delta\boldsymbol{x} = \Delta b$. Using the condition $\nabla f(\boldsymbol{x}^*) = -\boldsymbol{\lambda}^*\boldsymbol{a}$,

  $\Delta\text{cost} = f(\boldsymbol{x}^*+\Delta\boldsymbol{x}) - f(\boldsymbol{x}^*) = \nabla f(\boldsymbol{x}^*)'\Delta\boldsymbol{x} + o(\|\Delta\boldsymbol{x}\|) = -\boldsymbol{\lambda}^*\boldsymbol{a}'\Delta\boldsymbol{x} + o(\|\Delta\boldsymbol{x}\|)$

- Thus $\Delta\text{cost} = -\boldsymbol{\lambda}^*\Delta b + o(\|\Delta\boldsymbol{x}\|)$, so up to first order

$$\boldsymbol{\lambda}^* = -\frac{\Delta\text{cost}}{\Delta b}.$$

- For multiple constraints $\boldsymbol{a}_i'\boldsymbol{x} = b_i$, $i = 1, ..., n$, we have

$$\Delta\text{cost} = -\sum_{i=1}^{m} \boldsymbol{\lambda}_i^*\Delta b_i + o(\|\Delta\boldsymbol{x}\|).$$

# Sensitivity Theorem

Consider the family of problems

$$\min_{\boldsymbol{h}(\boldsymbol{x})=\boldsymbol{u}} f(\boldsymbol{x}) \tag{2}$$

parameterized by $\boldsymbol{u} \in \mathbb{R}^m$. Assume that for $\boldsymbol{u} = \mathbf{0}$, this problem has a local minimum $\boldsymbol{x}^*$, which is regular and together with its unique Lagrange multiplier $\boldsymbol{\lambda}^*$ satisfies the sufficiency conditions. Then there exists an open sphere $S$ centered at $\boldsymbol{u} = \mathbf{0}$ such that for every $\boldsymbol{u} \in S$, there is an $\boldsymbol{x}(\boldsymbol{u})$ and a $\boldsymbol{\lambda}(\boldsymbol{u})$, which are a local minimum-Lagrange multiplier pair of problem (2). Furthermore, $\boldsymbol{x}(\cdot)$ and $\boldsymbol{\lambda}(\cdot)$ are continuously differentiable within $S$ and we have

$$\boldsymbol{x}(\mathbf{0}) = \boldsymbol{x}^*, \ \boldsymbol{\lambda}(\mathbf{0}) = \boldsymbol{\lambda}^*.$$

In addition, $\nabla p(\boldsymbol{u}) = -\boldsymbol{\lambda}(\boldsymbol{u}), \quad \forall\ \boldsymbol{u} \in S$ where $p(\boldsymbol{u})$ is the primal function $p(\boldsymbol{u}) = f\left(\boldsymbol{x}(\boldsymbol{u})\right)$.

## Examples

- The primal function $p(\boldsymbol{u}) = f\left(\boldsymbol{x}(\boldsymbol{u})\right)$ for the two-dimensional problem

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) = \frac{1}{2}\left(x_1^2 - x_2^2\right) - x_2 \\ \text{subject to} & h(\boldsymbol{x}) = x_2 = 0. \end{array}$$

  is given by

$$p(u) = \min_{h(\boldsymbol{x})=u} f(\boldsymbol{x}) = -\frac{1}{2}u^2 - u$$

  and $\boldsymbol{\lambda}^* = -\nabla p(0) = 1$, consistent with the sensitivity theorem.

- Need for regularity of $\boldsymbol{x}^*$: Change constraint to $h(\boldsymbol{x}) = x_2^2 = 0$. Then $p(u) = -u/2 - \sqrt{u}$ for $u \geq 0$ and is undefined for $u < 0$.

# Proof of Sensitivity Theorem

- Apply implicit function theorem to the system

$$\nabla f(\boldsymbol{x}) + \nabla \boldsymbol{h}(\boldsymbol{x})\boldsymbol{\lambda} = \boldsymbol{0}, \quad \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{u}$$

- For $\boldsymbol{u} = \boldsymbol{0}$ the system has the solution $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$, and the corresponding $(n+m) \times (n+m)$ Jacobian

$$J = \begin{bmatrix} \nabla^2 f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla^2 h_i(\boldsymbol{x}^*) & \nabla h(\boldsymbol{x}^*) \\ \nabla h(\boldsymbol{x}^*)' & \boldsymbol{0} \end{bmatrix}$$

  is shown nonsingular using the sufficiency conditions.

- Hence, for all $\boldsymbol{u}$ in some open sphere $S$ centered at $\boldsymbol{u} = \boldsymbol{0}$, there exist $\boldsymbol{x}(\boldsymbol{u})$ and $\boldsymbol{\lambda}(\boldsymbol{u})$ such that $\boldsymbol{x}(\boldsymbol{0}) = \boldsymbol{x}^*$, $\boldsymbol{\lambda}(\boldsymbol{0}) = \boldsymbol{\lambda}^*$, the functions $\boldsymbol{x}(\cdot)$ and $\boldsymbol{\lambda}(\cdot)$ are continuously differentiable, and

$$\nabla f\left(\boldsymbol{x}(\boldsymbol{u})\right) + \nabla \boldsymbol{h}\left(\boldsymbol{x}(\boldsymbol{u})\right)\boldsymbol{\lambda}(\boldsymbol{u}) = 0, \quad h\left(\boldsymbol{x}(\boldsymbol{u})\right) = \boldsymbol{u}$$

- For $\boldsymbol{u}$ close to $\boldsymbol{u} = \mathbf{0}$, using the sufficiency conditions, $\boldsymbol{x}(\boldsymbol{u})$ and $\boldsymbol{\lambda}(\boldsymbol{u})$ are a local minimum-Lagrange multiplier pair for the problem $\min_{\boldsymbol{h}(\boldsymbol{x})=\boldsymbol{u}} f(\boldsymbol{x})$.
- To derive $\nabla p(\boldsymbol{u})$, differentiate $\boldsymbol{h}\left(\boldsymbol{x}(\boldsymbol{u})\right) = \boldsymbol{u}$, to obtain

$$\boldsymbol{I} = \nabla \boldsymbol{x}(\boldsymbol{u}) \nabla h\left(\boldsymbol{x}(\boldsymbol{u})\right),$$

and combine with the relations

$$\nabla \boldsymbol{x}(\boldsymbol{u}) \nabla f\left(\boldsymbol{x}(\boldsymbol{u})\right) + \nabla \boldsymbol{x}(\boldsymbol{u}) \nabla h\left(\boldsymbol{x}(\boldsymbol{u})\right)\boldsymbol{\lambda}(\boldsymbol{u}) = \mathbf{0}$$

and

$$\nabla p(\boldsymbol{u}) = \nabla_{\boldsymbol{u}}\left\{f\left(\boldsymbol{x}(\boldsymbol{u})\right)\right\} = \nabla \boldsymbol{x}(\boldsymbol{u}) \nabla f\left(\boldsymbol{x}(\boldsymbol{u})\right) = -\boldsymbol{\lambda}(\boldsymbol{u}).$$

# Inequality Constrained Problem

$$\boxed{\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0},\ \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}. \end{array}}$$

where $f : \mathbb{R}^n \mapsto \mathbb{R}$, $\boldsymbol{h} : \mathbb{R}^n \mapsto \mathbb{R}^m$, $\boldsymbol{g} : \mathbb{R}^n \mapsto \mathbb{R}^r$ are continuously differentiable. Here $\boldsymbol{h} = (h_1, ..., h_m)$, $\boldsymbol{g} = (g_1, ..., g_r)$.

- Consider the set of active inequality constraints

$$A(\boldsymbol{x}) = \{j \mid g_j(\boldsymbol{x}) = \boldsymbol{0}\}$$

- If $\boldsymbol{x}^*$ is a local minimum:
  - ⋆ The active inequality constraints at $\boldsymbol{x}^*$ can be treated as equations
  - ⋆ The inactive constraints at $\boldsymbol{x}^*$ do not matter

- Assuming regularity of $\boldsymbol{x}^*$ and assigning zero Lagrange multipliers to inactive constraints, we have

$$\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(\boldsymbol{x}^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(\boldsymbol{x}^*) = \mathbf{0}, \quad \mu_j^* = 0, \ \forall\, j \notin A(\boldsymbol{x}^*).$$

- Extra property: $\mu_j^* \geq 0$ for all $j$.

- Intuitive reason: Relax the $j$-th constraint, $g_j(\boldsymbol{x}) \leq u_j$, and notice

$$\mu_j^* = -\frac{\Delta\text{cost due to } u_j}{u_j}.$$

# Karash-Kuhn-Tucker Conditions

- Let $\boldsymbol{x}^*$ be a local minimum and a regular point. Then there exist unique Lagrange multiplier vectors $\boldsymbol{\lambda}^* = (\lambda_1^*, ..., \lambda_m^*)$, $\boldsymbol{\mu}^* = (\mu_1^*, ..., \mu_r^*)$, such that

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \mu^*) = \mathbf{0}, \quad \mu_j^* \geq 0, \; j = 1, ..., r, \quad \mu_j^* = 0, j \notin A(\boldsymbol{x}^*).$$

- If $f$, $\boldsymbol{h}$, and $\boldsymbol{g}$ are twice continuously differentiable,

$$\boldsymbol{y}' \nabla^2_{\boldsymbol{xx}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \boldsymbol{y} \geq 0, \quad \text{for all } \boldsymbol{y} \in V(\boldsymbol{x}^*),$$

  where

$$V(\boldsymbol{x}^*) = \{\boldsymbol{y} \mid \nabla \boldsymbol{h}(\boldsymbol{x}^*)' \boldsymbol{y} = 0, \; \nabla g_j(\boldsymbol{x}^*)' \boldsymbol{y} = 0, \; j \in A(\boldsymbol{x}^*)\}.$$

- Similar sufficiency conditions and sensitivity results. They require strict complementarity, i.e.,

$$\mu_j^* > 0, \quad \forall \, j \in A(\boldsymbol{x}^*).$$

# Proof of KKT Conditions

- Use equality-constraints result to obtain all the conditions except for $\mu_j^* \geq 0$ for $j \in A(\boldsymbol{x}^*)$.
- Introduce the penalty functions $g_j^+(\boldsymbol{x}) = \max\{0, g_j(\boldsymbol{x})\},\ j = 1, ..., r$, and for $k = 1, 2, ...$, let $\boldsymbol{x}^k$ minimize

$$f(\boldsymbol{x}) + \frac{k}{2}\|\boldsymbol{h}(\boldsymbol{x})\|^2 + \frac{k}{2}\sum_{j=1}^{r}\left(g_j^+(\boldsymbol{x})\right)^2 + \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{x}^*\|^2$$

  over a closed sphere of $\boldsymbol{x}^*$.
- Using the same argument as for equality constraints,

$$\lambda_i^* = \lim_{k\to\infty} kh_i(\boldsymbol{x}^k), \quad i = 1, ..., m,$$
$$\mu_j^* = \lim_{k\to\infty} kg_j^+(\boldsymbol{x}^k), \quad j = 1, ..., r.$$

- Since $g_j^+(\boldsymbol{x}^k) \geq 0$, we obtain $\mu_j^* \geq 0$ for all $j$.

# Linear Constraints

Consider the linearly constrained problem

$$\boxed{\min_{\boldsymbol{a}_j'\boldsymbol{x} \le b_j,\ j=1,\dots,r} f(\boldsymbol{x}).}$$

- Remarkable property: No need for regularity.

- Proposition: If $\boldsymbol{x}^*$ is a local minimum, there exist $\boldsymbol{\mu}^* = (\mu_1^*, \dots, \mu_r^*)$ with $\mu_j^* \ge 0,\ j = 1, \dots, r$, such that

$$\nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{r} \mu_j^* \boldsymbol{a}_j = \boldsymbol{0}, \quad \mu_j^* = 0, \ \ \forall\, j \notin A(\boldsymbol{x}^*).$$

- Proof uses Farkas Lemma: Consider the cones $C$ and $C^\perp$

$$C = \{\boldsymbol{x} \mid \boldsymbol{x} = \sum_{j=1}^{r} \mu_j \boldsymbol{a}_j,\ \mu_j \geq 0\}, \quad C^\perp = \{\boldsymbol{y} \mid \boldsymbol{a}_j'\boldsymbol{y} \leq 0,\ j = 1, ..., r\}$$

  Then

$$\boldsymbol{x} \in C \ \text{ iff } \ \boldsymbol{x}'\boldsymbol{y} \leq 0,\ \forall\, \boldsymbol{y} \in C^\perp.$$

- To see why Farkas' lemma is true, first show that $C$ is closed (nontrivial). Then, let $\boldsymbol{x}$ be such that $\boldsymbol{x}'\boldsymbol{y} \leq 0,\ \forall\, \boldsymbol{y} \in C^\perp$, and consider its projection $\tilde{x}$ on $C$. We have

$$\boldsymbol{x}'(\boldsymbol{x} - \tilde{\boldsymbol{x}}) = \|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|^2, \quad (\boldsymbol{x} - \tilde{\boldsymbol{x}})'\boldsymbol{a}_j \leq 0, \quad \forall j.$$

  Hence, $(\boldsymbol{x} - \tilde{\boldsymbol{x}}) \in C^\perp$, and using the hypothesis,

$$\boldsymbol{x}'(\boldsymbol{x} - \tilde{\boldsymbol{x}}) \leq 0.$$

- From the above two relations, we obtain $\boldsymbol{x} = \tilde{\boldsymbol{x}}$, so $\boldsymbol{x} \in C$.

# Proof of Lagrangian Multiplier Theorem

- The local min $\boldsymbol{x}^*$ of the original problem is also a local min for the problem

$$\min_{\boldsymbol{a}_j'\boldsymbol{x}\le b_j,\ j\in A(\boldsymbol{x}^*)} f(\boldsymbol{x}).$$

Hence

$$\nabla f(\boldsymbol{x}^*)'(\boldsymbol{x}-\boldsymbol{x}^*)\ge 0, \quad \forall\, \boldsymbol{x} \text{ with } \boldsymbol{a}_j'\boldsymbol{x}\le b_j,\ j\in A(\boldsymbol{x}^*).$$

- Since a constraint $\boldsymbol{a}_j'\boldsymbol{x}\le b_j,\ j\in A(\boldsymbol{x}^*)$ can also be expressed as $\boldsymbol{a}_j'(\boldsymbol{x}-\boldsymbol{x}^*)\le 0$, we have

$$\nabla f(\boldsymbol{x}^*)'\boldsymbol{y}\ge 0, \quad \forall\, \boldsymbol{y} \text{ with } \boldsymbol{a}_j'\boldsymbol{y}\le 0,\ j\in A(\boldsymbol{x}^*).$$

- From Farkas' lemma, $-\nabla f(\boldsymbol{x}^*)$ has the form

$$\sum_{j \in A(\boldsymbol{x}^*)} \mu_j^* \boldsymbol{a}_j, \quad \text{for some } \mu_j^* \geq 0,\ j \in A(\boldsymbol{x}^*).$$

- To complete the proof, let $\mu_j^* = 0$ for $j \notin A(\boldsymbol{x}^*)$.