Lagrange Multiplier Theory

Lecturer: Pradeep Ravikumar
Co-instructor: Aarti Singh

Convex Optimization 10-725/36-725

Equality Constrained Problems

minimize $f(x)$

subject to $h_i(x) = 0, \qquad i = 1, \ldots, m.$

where $f : \Re^n \mapsto \Re$, $h_i : \Re^n \mapsto \Re$, $i = 1, \ldots, m$, are continuously differentiable functions. (Theory also applies to case where f and h_i are cont. differentiable in a neighborhood of a local minimum.)

Lagrange Multiplier Theorem

- Let x^* be a local min and a regular point [$\nabla h_i(x^*)$: linearly independent]. Then there exist unique scalars $\lambda_1^*, \ldots, \lambda_m^*$ such that

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) = 0.$$

If in addition f and h are twice cont. differentiable,

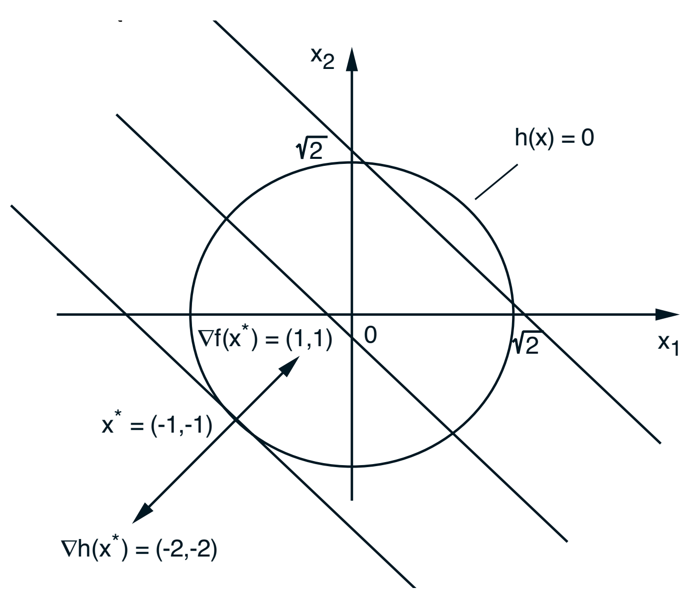
$$y' \left(\nabla^2 f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla^2 h_i(x^*) \right) y \geq 0, \; \forall\, y \text{ s.t. } \nabla h(x^*)' y = 0$$

Lagrange Multiplier Theorem

- Let x^* be a local min and a regular point [$\nabla h_i(x^*)$: linearly independent]. Then there exist unique scalars $\lambda_1^*, \ldots, \lambda_m^*$ such that

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) = 0.$$

Example:



$$\text{minimize} \quad x_1 + x_2$$

$$\text{subject to} \quad x_1^2 + x_2^2 = 2.$$

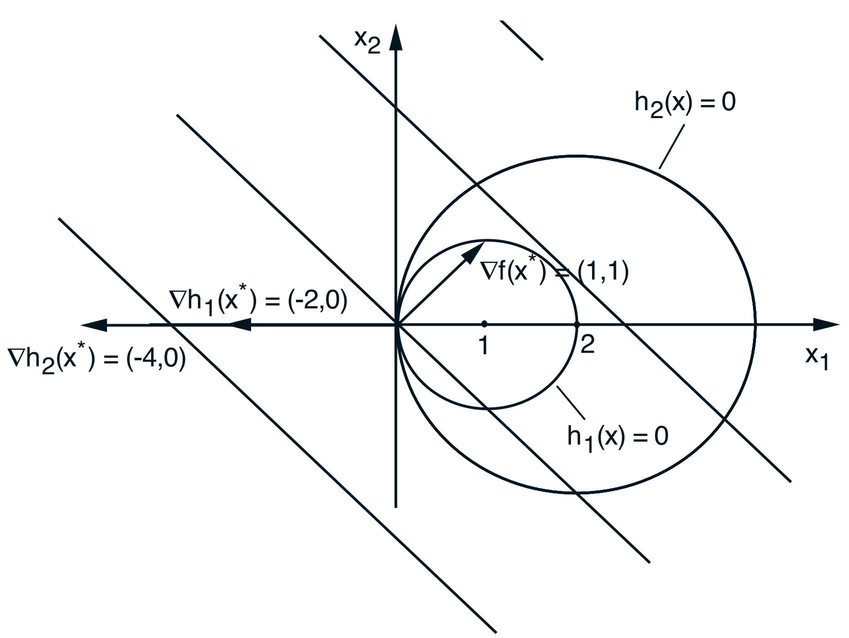
The Lagrange multiplier is $\lambda = 1/2$.

Lagrange Multiplier Theorem

- Let x^* be a local min and a regular point [$\nabla h_i(x^*)$: linearly independent]. Then there exist unique scalars $\lambda_1^*, \ldots, \lambda_m^*$ such that

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) = 0.$$

Example:



$$\text{minimize} \quad x_1 + x_2$$

$$\text{s. t.} \quad (x_1 - 1)^2 + x_2^2 - 1 = 0$$

$$(x_1 - 2)^2 + x_2^2 - 4 = 0$$

Lagrange Multiplier Theorem

- Let x^* be a local min and a regular point [$\nabla h_i(x^*)$: linearly independent]. Then there exist unique scalars $\lambda_1^*, \ldots, \lambda_m^*$ such that

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) = 0.$$

When local minimum is not regular, then first-order feasible variation

$$V(x^*) = \left\{ y \mid \nabla h_1(x^*)' y = 0,\ \nabla h_2(x^*)' y = 0 \right\}$$

has larger dimension than true set of feasible variations

$$\{y : h(x^* + y) = 0\}$$

Lagrange Multiplier Theorem

- Let x^* be a local min and a regular point [$\nabla h_i(x^*)$: linearly independent]. Then there exist unique scalars $\lambda_1^*, \ldots, \lambda_m^*$ such that

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) = 0.$$

Optimality of x^* entails that gradient of f at x^* is orthogonal to true set of feasible variations

For a Lagrange Multiplier to exist, gradient of f at x^* must be orthogonal to subspace of first order feasible variations

Lagrangian Function

- Define the Lagrangian function

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

Then, if x^* is a local minimum which is regular, the Lagrange multiplier conditions are written

$$\nabla_x L(x^*, \lambda^*) = 0, \qquad \nabla_\lambda L(x^*, \lambda^*) = 0,$$

Lagrangian Function

- Define the Lagrangian function

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

Then, if x^* is a local minimum which is regular, the Lagrange multiplier conditions are written

$$\nabla_x L(x^*, \lambda^*) = 0, \qquad \nabla_\lambda L(x^*, \lambda^*) = 0,$$

System of $n + m$ equations with $n + m$ unknowns.

Lagrangian Function

- Define the Lagrangian function

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

Then, if x^* is a local minimum which is regular, the Lagrange multiplier conditions are written

$$\nabla_x L(x^*, \lambda^*) = 0, \qquad \nabla_\lambda L(x^*, \lambda^*) = 0,$$

System of $n + m$ equations with $n + m$ unknowns.

- Example

minimize $\frac{1}{2}\left(x_1^2 + x_2^2 + x_3^2\right)$

subject to $x_1 + x_2 + x_3 = 3.$

Necessary conditions

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

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

Example: Portfolio Selection

- Investment of 1 unit of wealth among n assets with random rates of return e_i, and given means $\overline{e}_i$, and covariance matrix $Q = \left[E\{(e_i - \overline{e}_i)(e_j - \overline{e}_j)\}\right]$.
- If x_i: amount invested in asset i, we want to

$$\text{minimize} \quad x'Qx \left(= \text{Variance of return } \sum_i e_i x_i \right)$$

$$\text{subject to} \quad \textstyle\sum_i x_i = 1, \text{ and a given mean } \displaystyle\sum_i \overline{e}_i x_i = m$$

Example: Portfolio Selection

- Let λ_1 and λ_2 be the L-multipliers. Have $2Qx^* + \lambda_1 u + \lambda_2 \bar{e} = 0$, where $u = (1, \ldots, 1)'$ and $\bar{e} = (\bar{e}_1, \ldots, \bar{e}_n)'$.

where lambda_1, lambda_2 can be obtained as the solution of:

$$1 = u^T x^* = \frac{\lambda_1}{2} u^T Q^{-1} u + \frac{\lambda_2}{2} u^T Q^{-1} \bar{e}$$

$$m = \bar{e}^T x^* = \frac{\lambda_1}{2} \bar{e}^T Q^{-1} u + \frac{\lambda_2}{2} \bar{e}^T Q^{-1} \bar{e}$$

Sufficiency Conditions

Equality constrained problem

$$\text{minimize} \quad f(x)$$
$$\text{subject to} \quad h_i(x) = 0, \qquad i = 1, \ldots, m.$$

where $f : \Re^n \mapsto \Re$, $h_i : \Re^n \mapsto \Re$, are continuously differentiable. To obtain sufficiency conditions, assume that f and h_i are *twice* continuously differentiable.

Sufficiency Conditions

Second Order Sufficiency Conditions: Let $x^* \in \Re^n$ and $\lambda^* \in \Re^m$ satisfy

$$\nabla_x L(x^*, \lambda^*) = 0, \qquad \nabla_\lambda L(x^*, \lambda^*) = 0,$$

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

Then x^* is a strict local minimum.

Sufficiency Conditions: Example

Example: Minimize $-(x_1x_2 + x_2x_3 + x_1x_3)$ subject to $x_1 + x_2 + x_3 = 3$. We have that $x_1^* = x_2^* = x_3^* = 1$ and $\lambda^* = 2$ satisfy the 1st order conditions. Also

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

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

Sufficiency Conditions: Example

Example: Minimize $-(x_1x_2 + x_2x_3 + x_1x_3)$ subject to $x_1 + x_2 + x_3 = 3$. We have that $x_1^* = x_2^* = x_3^* = 1$ and $\lambda^* = 2$ satisfy the 1st order conditions. Also

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

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

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

Hence, x^* is a strict local minimum.

SENSITIVITY THEOREM

Sensitivity Theorem: Consider the family of problems

$$\min_{h(x)=u} f(x) \qquad (*)$$

parameterized by $u \in \Re^m$. Assume that for $u = 0$, this problem has a local minimum x^*, which is regular and together with its unique Lagrange multiplier λ^* satisfies the sufficiency conditions.

Then there exists an open sphere S centered at $u = 0$ such that for every $u \in S$, there is an $x(u)$ and a $\lambda(u)$, which are a local minimum-Lagrange multiplier pair of problem (*). Furthermore, $x(\cdot)$ and $\lambda(\cdot)$ are continuously differentiable within S and we have $x(0) = x^*$, $\lambda(0) = \lambda^*$. In addition,

$$\nabla p(u) = -\lambda(u), \qquad \forall\ u \in S$$

where $p(u)$ is the *primal function*

$$p(u) = f\big(x(u)\big).$$

EXAMPLE

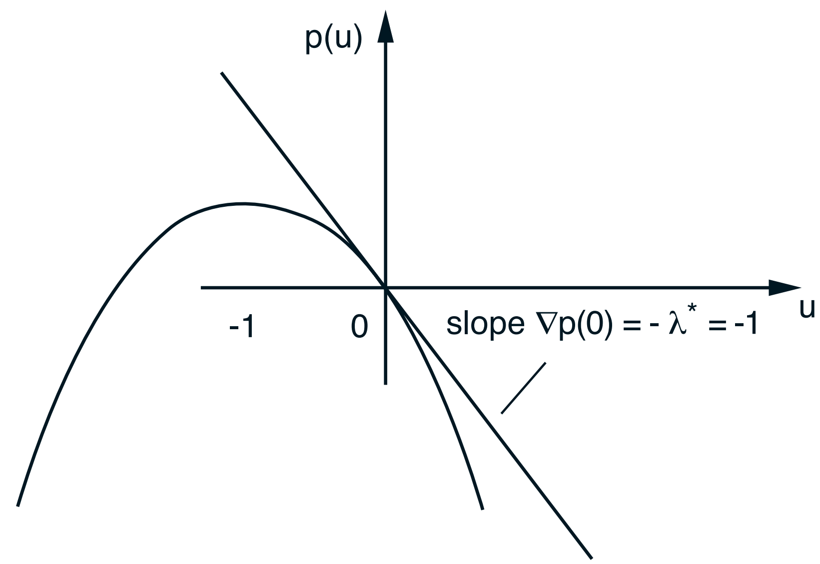


Illustration of the primal function $p(u) = f\big(x(u)\big)$ for the two-dimensional problem

$$\text{minimize} \quad f(x) = \tfrac{1}{2}\big(x_1^2 - x_2^2\big) - x_2$$

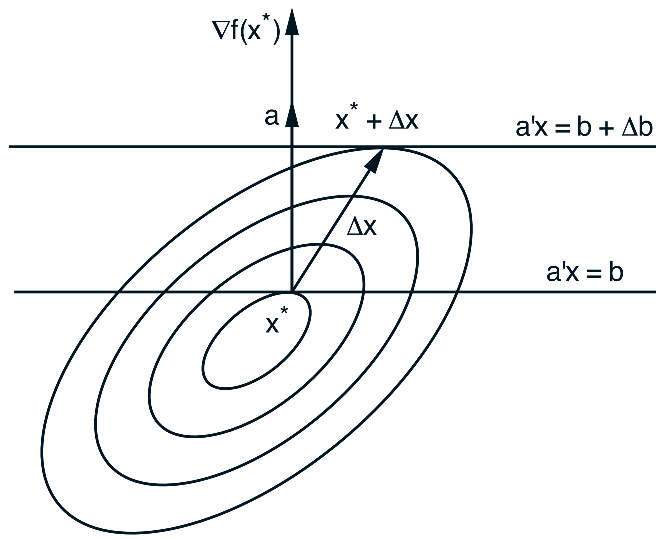
$$\text{subject to} \quad h(x) = x_2 = 0.$$

Here,

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

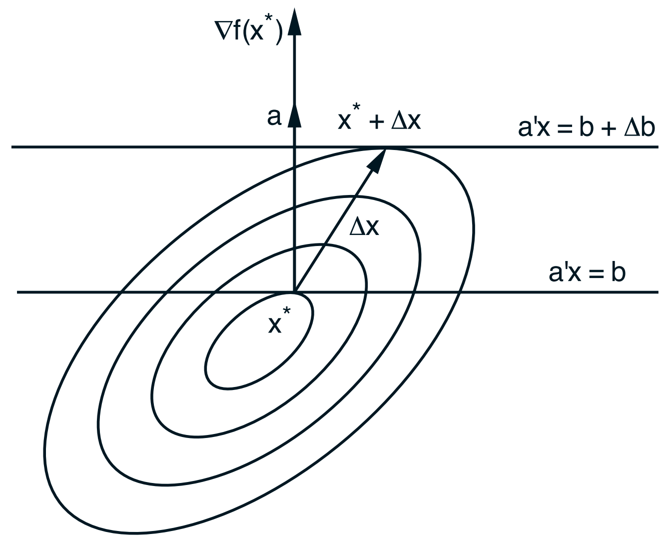
and $\lambda^* = -\nabla p(0) = 1$, consistently with the sensitivity theorem.

SENSITIVITY - GRAPHICAL DERIVATION



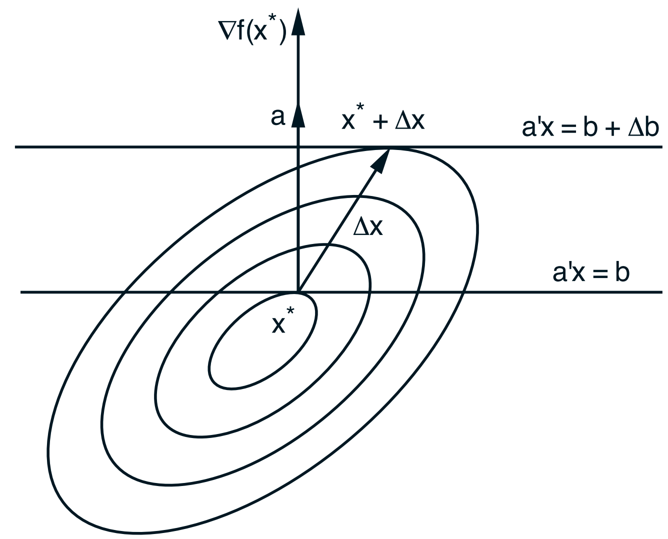
Sensitivity theorem for the problem $\min_{a'x=b} f(x)$. If b is changed to $b+\Delta b$, the minimum x^* will change to $x^*+\Delta x$.

SENSITIVITY - GRAPHICAL DERIVATION



Sensitivity theorem for the problem $\min_{a'x=b} f(x)$. If b is changed to $b+\Delta b$, the minimum x^* will change to $x^*+\Delta x$. Since $b + \Delta b = a'(x^* + \Delta x) = a'x^* + a'\Delta x = b + a'\Delta x$, we have $a'\Delta x = \Delta b$.

SENSITIVITY - GRAPHICAL DERIVATION

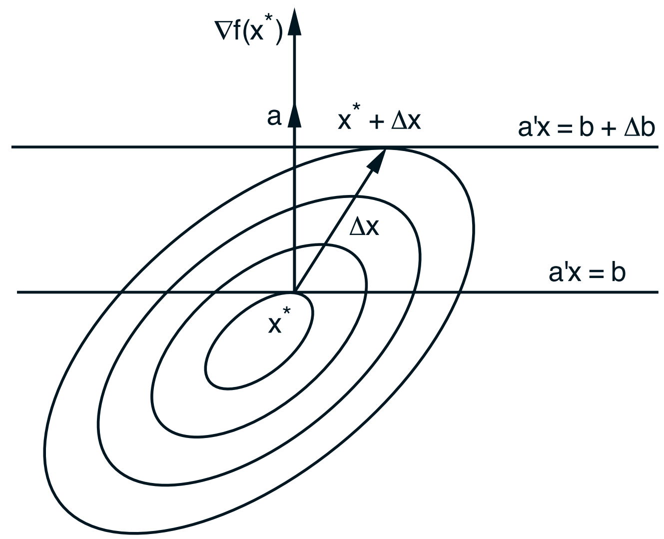


Sensitivity theorem for the problem $\min_{a'x=b} f(x)$. If b is changed to $b+\Delta b$, the minimum x^* will change to $x^*+\Delta x$. Since $b+\Delta b = a'(x^* + \Delta x) = a'x^* + a'\Delta x = b + a'\Delta x$, we have $a'\Delta x = \Delta b$. Using the condition $\nabla f(x^*) = -\lambda^* a$,

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

$$= -\lambda^* a'\Delta x + o(\|\Delta x\|)$$

SENSITIVITY - GRAPHICAL DERIVATION



Sensitivity theorem for the problem $\min_{a'x=b} f(x)$. If b is changed to $b+\Delta b$, the minimum x^* will change to $x^*+\Delta x$. Since $b+\Delta b = a'(x^*+\Delta x) = a'x^* + a'\Delta x = b + a'\Delta x$, we have $a'\Delta x = \Delta b$. Using the condition $\nabla f(x^*) = -\lambda^* a$,

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

$$= -\lambda^* a'\Delta x + o(\|\Delta x\|)$$

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

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

Inequality Constrained Problems

Inequality constrained problem

$$\text{minimize} \quad f(x)$$
$$\text{subject to} \quad h(x) = 0, \qquad g(x) \leq 0$$

where $f : \Re^n \mapsto \Re$, $h : \Re^n \mapsto \Re^m$, $g : \Re^n \mapsto \Re^r$ are continuously differentiable. Here

$$h = (h_1, ..., h_m), \qquad g = (g_1, ..., g_r).$$

TREATING INEQUALITIES AS EQUATIONS

- Consider the **set of active inequality constraints**

$$A(x) = \left\{ j \mid g_j(x) = 0 \right\}.$$

- If x^* is a local minimum:
  - The active inequality constraints at x^* can be treated as equations
  - The inactive constraints at x^* don't matter

TREATING INEQUALITIES AS EQUATIONS

- Consider the **set of active inequality constraints**

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

- If x^* is a local minimum:
  - The active inequality constraints at x^* can be treated as equations
  - The inactive constraints at x^* don't matter

- Assuming regularity of x^* and assigning zero Lagrange multipliers to inactive constraints,

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*) = 0,$$

$$\mu_j^* = 0, \qquad \forall\ j \notin A(x^*).$$

TREATING INEQUALITIES AS EQUATIONS

- Consider the **set of active inequality constraints**

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

- If x^* is a local minimum:
  - The active inequality constraints at x^* can be treated as equations
  - The inactive constraints at x^* don't matter

- Assuming regularity of x^* and assigning zero Lagrange multipliers to inactive constraints,

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*) = 0,$$

$$\mu_j^* = 0, \qquad \forall\ j \notin A(x^*).$$

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

TREATING INEQUALITIES AS EQUATIONS

- Consider the **set of active inequality constraints**

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

- If x^* is a local minimum:
  - The active inequality constraints at x^* can be treated as equations
  - The inactive constraints at x^* don't matter

- Assuming regularity of x^* and assigning zero Lagrange multipliers to inactive constraints,

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*) = 0,$$

$$\mu_j^* = 0, \qquad \forall\ j \notin A(x^*).$$

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

- Intuitive reason: Relax jth constraint, $g_j(x) \leq u_j$.

TREATING INEQUALITIES AS EQUATIONS

- Consider the **set of active inequality constraints**

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

- If x^* is a local minimum:
  - The active inequality constraints at x^* can be treated as equations
  - The inactive constraints at x^* don't matter

- Assuming regularity of x^* and assigning zero Lagrange multipliers to inactive constraints,

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*) = 0,$$

$$\mu_j^* = 0, \qquad \forall\ j \notin A(x^*).$$

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

- Intuitive reason: Relax jth constraint, $g_j(x) \leq u_j$. Since $\Delta\text{cost} \leq 0$ if $u_j > 0$, by the sensitivity theorem, we have

$$\mu_j^* = -(\Delta\text{cost due to } u_j)/u_j \geq 0$$

BASIC RESULTS

Kuhn-Tucker Necessary Conditions: Let x^* be a local minimum and a regular point. Then there exist unique Lagrange mult. vectors $\lambda^* = (\lambda_1^*, \ldots, \lambda_m^*)$, $\mu^* = (\mu_1^*, \ldots, \mu_r^*)$, such that

$$\nabla_x L(x^*, \lambda^*, \mu^*) = 0,$$

$$\mu_j^* \geq 0, \qquad j = 1, \ldots, r,$$

$$\mu_j^* = 0, \qquad \forall\ j \notin A(x^*).$$

If f, h, and g are twice cont. differentiable,

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

where

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

BASIC RESULTS

Kuhn-Tucker Necessary Conditions: Let x^* be a local minimum and a regular point. Then there exist unique Lagrange mult. vectors $\lambda^* = (\lambda_1^*, \ldots, \lambda_m^*)$, $\mu^* = (\mu_1^*, \ldots, \mu_r^*)$, such that

$$\nabla_x L(x^*, \lambda^*, \mu^*) = 0,$$

$$\mu_j^* \geq 0, \qquad j = 1, \ldots, r,$$

$$\mu_j^* = 0, \qquad \forall\ j \notin A(x^*).$$

Complementary Slackness

If f, h, and g are twice cont. differentiable,

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

where

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

BASIC RESULTS

Kuhn-Tucker Necessary Conditions: Let x^* be a local minimum and a regular point. Then there exist unique Lagrange mult. vectors $\lambda^* = (\lambda_1^*, \ldots, \lambda_m^*)$, $\mu^* = (\mu_1^*, \ldots, \mu_r^*)$, such that

$$\nabla_x L(x^*, \lambda^*, \mu^*) = 0,$$

$$\mu_j^* \geq 0, \qquad j = 1, \ldots, r,$$

$$\mu_j^* = 0, \qquad \forall\ j \notin A(x^*).$$

If f, h, and g are twice cont. differentiable,

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

where

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

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

$$\mu_j^* > 0, \qquad \forall\ j \in A(x^*),$$

as well as regularity of x^*.

GENERAL SUFFICIENCY CONDITION

Consider the problem

minimize $f(x)$

subject to $x \in X, \qquad g_j(x) \leq 0, \quad j = 1, \ldots, r.$

Let x^* be feasible and μ^* satisfy

$$\mu_j^* \geq 0, \quad j = 1, \ldots, r, \qquad \mu_j^* = 0, \quad \forall\ j \notin A(x^*),$$

$$x^* = \arg\min_{x \in X} L(x, \mu^*).$$

Then x^* is a global minimum of the problem.

GENERAL SUFFICIENCY CONDITION

Consider the problem

minimize $f(x)$

subject to $x \in X, \qquad g_j(x) \leq 0, \quad j = 1, \ldots, r.$

Let x^* be feasible and μ^* satisfy

$$\mu_j^* \geq 0, \quad j = 1, \ldots, r, \qquad \mu_j^* = 0, \quad \forall\, j \notin A(x^*),$$

$$x^* = \arg\min_{x \in X} L(x, \mu^*).$$

Then x^* is a global minimum of the problem.

Proof: We have

$$f(x^*) = f(x^*) + \mu^{*\prime} g(x^*) = \min_{x \in X} \left\{ f(x) + \mu^{*\prime} g(x) \right\}$$

GENERAL SUFFICIENCY CONDITION

Consider the problem

minimize $f(x)$

subject to $x \in X, \qquad g_j(x) \leq 0, \quad j = 1, \ldots, r.$

Let x^* be feasible and μ^* satisfy

$$\mu_j^* \geq 0, \quad j = 1, \ldots, r, \qquad \mu_j^* = 0, \quad \forall\, j \notin A(x^*),$$

$$x^* = \arg\min_{x \in X} L(x, \mu^*).$$

Then x^* is a global minimum of the problem.

Proof: We have

$$f(x^*) = f(x^*) + \mu^{*\prime} g(x^*) = \min_{x \in X} \Big\{ f(x) + \mu^{*\prime} g(x) \Big\}$$

$$\leq \min_{x \in X,\, g(x) \leq 0} \Big\{ f(x) + \mu^{*\prime} g(x) \Big\} \leq \min_{x \in X,\, g(x) \leq 0} f(x),$$

GENERAL SUFFICIENCY CONDITION

Consider the problem

minimize $f(x)$

subject to $x \in X, \qquad g_j(x) \le 0, \quad j = 1, \ldots, r.$

Let x^* be feasible and μ^* satisfy

$$\mu_j^* \ge 0, \quad j = 1, \ldots, r, \qquad \mu_j^* = 0, \quad \forall\, j \notin A(x^*),$$

$$x^* = \arg \min_{x \in X} L(x, \mu^*).$$

Then x^* is a global minimum of the problem.

Proof: We have

$$f(x^*) = f(x^*) + \mu^{*\prime} g(x^*) = \min_{x \in X} \big\{ f(x) + \mu^{*\prime} g(x) \big\}$$
$$\le \min_{x \in X,\, g(x) \le 0} \big\{ f(x) + \mu^{*\prime} g(x) \big\} \le \min_{x \in X,\, g(x) \le 0} f(x),$$

where the first equality follows from the hypothesis, which implies that $\mu^{*\prime} g(x^*) = 0$, and the last inequality follows from the nonnegativity of μ^*. **Q.E.D.**

- Special Case: Let $X = \Re^n$, f and g_j be convex and differentiable. Then the 1st order Kuhn-Tucker conditions are also sufficient for global optimality.