Lecture 8: Inequality (mixed) Constraints, Karusch- Kuhn-Tucker Conditions, Convex Programming,Primal-Dual Methods

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 20, 2009

Outline of Lecture 8

- ❑ Lagrange Multipliers and Duality
- ❑ Inequality (Mixed) Constraints
- ❑ Karusch-Kuhn-Tucker (KKT) Conditions
- ❑ Illustrative Examples
- ❑ Convex Programming and Duality
- ❑ Saddle Point Theorem
- ❑ Primal-Dual Methods

Review of Optimality Conditions - 1

$$\min f(\underline{x}) \quad \text{s.t. } \underline{h}(\underline{x}) = \underline{0} \quad \Rightarrow \quad \left.\begin{aligned} &\nabla \underline{f}(\underline{x}^*) + \nabla \underline{h}(\underline{x}^*)\underline{\lambda}^* = \underline{0} \\ &\underline{h}(\underline{x}^*) = \underline{0} \end{aligned}\right\} \text{First order Conditions}$$

- Second-order conditions

$$\underline{y}^T \left[\nabla^2 f(\underline{x}^*) + \sum_{i=1}^{m} \lambda_i \nabla^2 h_i(\underline{x}^*) \right] \underline{y} \geq 0 \quad \ni \nabla \underline{h}_i^T(\underline{x}^*)\underline{y} = 0 \quad \forall i = 1, 2, \ldots, m$$

- What do λs mean ?

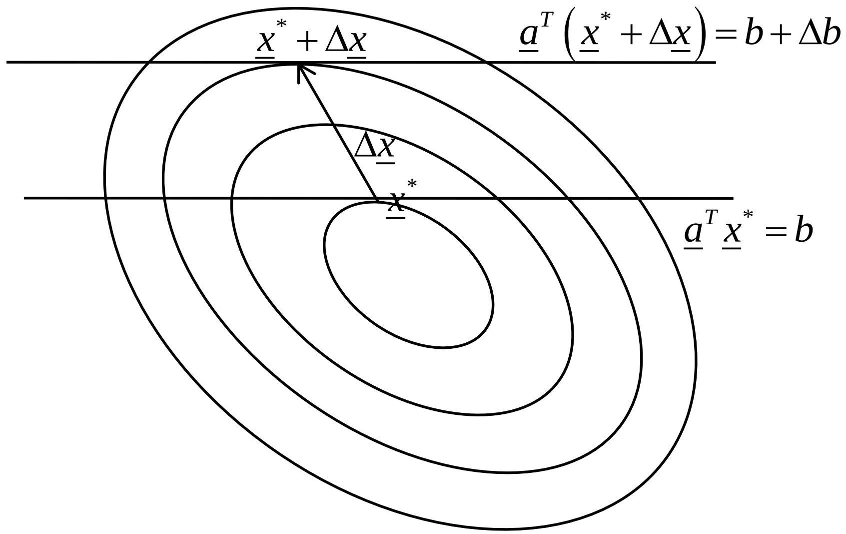
Consider $\min_{\underline{x}} f(\underline{x})$

$$\text{s.t. } \underline{a}^T \underline{x} = b,$$

Suppose $b \rightarrow b + \Delta b$

$$\underline{a}^T \left(\underline{x}^* + \Delta \underline{x} \right) = b + \Delta b$$

$$\Rightarrow \underline{a}^T \Delta \underline{x} = \Delta b$$



Review of Optimality Conditions - 2

$$f\left(\underline{x}^* + \Delta\underline{x}\right) = f(\underline{x}^*) + \nabla \underline{f}^T(\underline{x}^*)\Delta\underline{x}$$

$$\text{know } \nabla f(\underline{x}^*) = -\lambda\underline{a} \Rightarrow \Delta f = -\lambda\underline{a}^T\Delta\underline{x} = -\lambda\Delta b$$

$$\Rightarrow -\frac{\Delta f}{\Delta b} = \lambda \text{ rate of optimal cost decrease as the level of}$$

constraint increases.

In genernal, $\underline{h}\left(\underline{x}^* + \Delta\underline{x}\right) \cong \underline{h}(\underline{x}^*) + \nabla\underline{\underline{h}}^T(\underline{x}^*)\Delta\underline{x} = \Delta\underline{b} \Rightarrow \nabla\underline{\underline{h}}^T(\underline{x}^*)\Delta\underline{x} = \Delta\underline{b}$

$$\Delta f = \nabla f^T(\underline{x}^*)\Delta\underline{x} = -\underline{\lambda}^T\nabla\underline{\underline{h}}^T(\underline{x}^*)\Delta\underline{x} = -\underline{\lambda}^T\Delta\underline{b}$$

$$\Rightarrow \boxed{\underline{\lambda} = -\nabla_b \underline{f}}$$

If we let $p(\underline{u}) = \min\limits_{h(\underline{x})=\underline{u}} f(\underline{x})$

$$\underline{\lambda}(u) = -\nabla_u p(\underline{u}), \qquad -\nabla p(0) = \underline{\lambda}(0) = \underline{\lambda}^*$$

Example

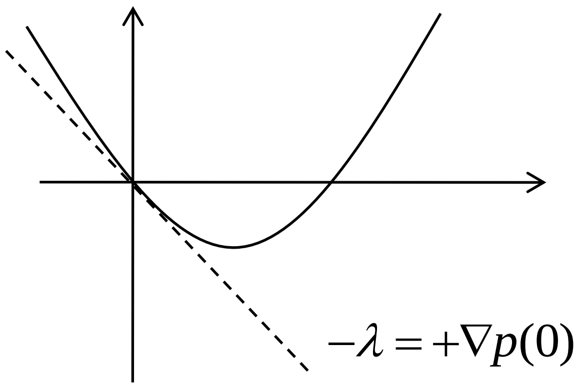
- Example

$$\min f(\underline{x}) = \frac{1}{2}\left(x_1^2 + x_2^2\right) - x_2$$

$$\text{s.t.} \quad x_2 = 0$$

$$p(u) = \min_{x_2 = u} \frac{1}{2}u^2 - u$$

$$L(\underline{x}, \lambda) = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 - x_2 + \lambda x_2$$



Optilmality conditions:

$$2x_1 = 0$$

$$x_2 - 1 + \lambda = 0 \Rightarrow x_2 = 1 - \lambda \Rightarrow \lambda^* = 1$$

$$x_2 = 0$$

Example Solution via Primal-Dual Method

- Primal-Dual Method

1. For a given λ, minimize $L(\underline{x},\lambda)$ with respect to $\underline{x} \Rightarrow q(\lambda) = \min_{\underline{x}} L(\underline{x},\lambda)$
2. $\max q(\lambda)$ with respect to λ

1. $\min_{\underline{x}} L(\underline{x},\lambda) = \frac{1}{2}(1-\lambda)^2 + (\lambda-1)(1-\lambda) = -\frac{1}{2}(1-\lambda)^2$
2. $\max_{\lambda} q(\lambda) \Rightarrow \lambda = 1$

- We will have more to say about this in the context of augmented Lagrangian methods.

Inequality (Mixed) Constraints

$$\min f(\underline{x}) \quad \text{s.t. } h_i(\underline{x}) = 0,\ i = 1, 2, \ldots, m; \quad g_j(\underline{x}) \le 0,\ j = 1, 2, \ldots, r$$

(or)

$$\min f(\underline{x}) \quad \text{s.t. } \underline{h}(\underline{x}) = \underline{0}; \quad \underline{g}(\underline{x}) \le \underline{0}$$

- Def[n]: The inequality constraint $g_j(\underline{x}) \le 0$ is said to be binding or active at the point $\underline{x}^*$ if $g_j(\underline{x}) = 0$; it is nonbinding if $g_j(\underline{x}^*) < 0$

- Set of active(or binding) constraints:

$$\mathcal{A}(\underline{x}^*) = \{ j \mid g_j(\underline{x}^*) = 0,\ j = 1, 2, \ldots, r \}$$

- Idea: If we can identify the active set at the optimum, then solve equality constrained problem & delete the rest of the nonbinding constraints

- Def[n] : $\underline{x}^*$ is a regular point if the gradient vectors $\nabla h_i(\underline{x}^*),\ i = 1, 2, \ldots, m$ and $\nabla g_j(\underline{x}^*),\ j \in \mathcal{A}(\underline{x}^*)$ are linearly independent

Karusch-Kuhn-Tucker (KKT) Conditions -1

- Necessary and sufficient conditions for optimality: Also called Krusch-Kuhn-Tucker conditions. We derive the conditions by converting the problem into an equality constrained problem

$$\min f(\underline{x})$$
$$\text{s.t. } h_1(\underline{x}) = h_2(\underline{x}) = \ldots = h_m(\underline{x}) = 0$$
$$g_1(\underline{x}) + z_1^2 = g_2(\underline{x}) + z_2^2 = \ldots = g_r(\underline{x}) + z_r^2 = 0$$

- Define

$$\left.\begin{aligned} \bar{f}(\underline{x}, \underline{z}) &= f(\underline{x}) \\ \bar{h}_i(\underline{x}, \underline{z}) &= h_i(\underline{x}) \\ \bar{g}_j(\underline{x}, \underline{z}) &= g_j(\underline{x}) + z_j^2 \end{aligned}\right\} \Rightarrow \begin{aligned} &\min \bar{f}(\underline{x}, \underline{z}) \\ &\text{s.t. } \bar{h}_i(\underline{x}, \underline{z}) = 0,\ i = 1, 2, \ldots, m \\ &\qquad \bar{g}_j(\underline{x}, \underline{z}) = 0,\ j = 1, 2, \ldots, r \end{aligned}$$

- $\underline{x}^*$ is a solution of problem 1 $\Leftrightarrow$ $[\underline{x}^*, [-g_1(\underline{x}^*)]^{1/2}, \ldots, [-g_r(\underline{x}^*)]^{1/2}]$ solution to modified problem

Karusch-Kuhn-Tucker (KKT) Conditions - 2

- Binding inequality constraint $\Leftrightarrow$ $z_j^* = 0$

$$\nabla \bar{\underline{f}} = \begin{bmatrix} \nabla \underline{f}(\underline{x}) \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}, \qquad \nabla \underline{h}_i(\underline{x}^*, \underline{z}^*) = \begin{bmatrix} \nabla \underline{h}_i(\underline{x}^*) \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}, \; i = 1, 2, \ldots, m$$

$$\nabla \bar{\underline{g}}_j(\underline{x}^*, \underline{z}^*) = \begin{bmatrix} \nabla \underline{g}_j(\underline{x}^*) \\ 0 \\ \cdot \\ \cdot \\ 2z_j^* \\ 0 \end{bmatrix} \begin{matrix} n \\ \\ \\ \\ n+j \\ \\ \end{matrix} \qquad ; j = 1, 2, \ldots, r$$

Karusch-Kuhn-Tucker (KKT) Conditions - 3

- What do the first and second order conditions yield?
  - 1st order conditions

$$\nabla \underline{\bar{f}}(\underline{x}^*,\underline{z}^*)+\sum_{i=1}^{m}\lambda_i^*\nabla\underline{\bar{h}}_i(\underline{x}^*,\underline{z}^*)+\sum_{j=1}^{r}\mu_j^*\nabla\underline{\bar{g}}_j(\underline{x}^*,\underline{z}^*)=\underline{0}$$

$$\left.\begin{aligned}&\nabla \underline{f}(\underline{x}^*)+\sum_{i=1}^{m}\lambda_i^*\nabla\underline{h}_i(\underline{x}^*)+\sum_{j=1}^{r}\mu_j^*\nabla\underline{g}_j(\underline{x}^*)=\underline{0}\\ &2\mu_j^* z_j^*=2\mu_j^*[-g_j(\underline{x}^*)]^{1/2}=0,\quad j=1,2,\ldots,r\end{aligned}\right\}$$

So, $\quad \mu_j^* g_j(\underline{x}^*)=0,\quad \forall\, j=1,2,\ldots,r$

  - 2nd order conditions

$$[\underline{y}^T\ \ \underline{v}^T]\begin{bmatrix}\overbrace{\nabla^2 f(\underline{x}^*)+\sum_{i=1}^{m}\lambda_i^*\nabla^2 f_i(\underline{x}^*)+\sum_{j=1}^{r}\mu_j^*\nabla^2 g_j(\underline{x}^*)}^{\nabla_{xx}^2 L(\underline{x},\underline{\lambda},\underline{\mu})} & \mathbf{0}\\ \mathbf{0} & \begin{matrix}2\mu_1^* & & & \\ & 2\mu_2^* & & \\ & & \ddots & \\ & & & 2\mu_r^*\end{matrix}\end{bmatrix}\begin{bmatrix}\underline{y}\\ \underline{v}\end{bmatrix}\geq 0$$

Karusch-Kuhn-Tucker (KKT) Conditions - 4

such that

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

$$\nabla \underline{g}_j^T(\underline{x}^*)\underline{y} + 2z_j v_j = 0; \quad j = 1, 2, \ldots, r \qquad (2)$$

- The quadratic term simplifies to

$\underline{y}^T \nabla_{xx}^2 L \underline{y} + 2\sum_{j=1}^{r} \mu_j^* v_j^2 \geq 0, \quad \forall\ \underline{y}\ \&\ \underline{v}$ satisfying (1) and (2)

Suppose $z_j^* = [-g_j(\underline{x}^*)]^{1/2} = 0 \quad \Rightarrow$ active (binding) constraint.

Let $\underline{y} = 0,\ v_j \neq 0,\ v_k = 0,\ k \neq j$. Then, $\mu_j^* v_j^2 \geq 0 \ \Rightarrow\ \mu_j^* \geq 0$ if $g_j(\underline{x}^*) = 0$.

Also, note that $\nabla \underline{g}_j^T(\underline{x}^*)\underline{y} = 0$ for active constraints.

On the other hand, if $z_j^* = [-g_j(\underline{x}^*)]^{1/2} \neq 0$, then $\mu_j^* = 0$

Setting $\underline{y}$ arbitary and $\underline{v} = 0$, we obtain

$$\underline{y}^T \nabla_{xx}^2 L \underline{y} \geq 0 \quad \ni$$

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

$$\nabla \underline{g}_j^T(\underline{x}^*)\underline{y} = 0, \quad \forall j \in \mathcal{A}(\underline{x}^*), \text{ the active constraint set}$$

Karusch-Kuhn-Tucker (KKT) Conditions - 5

- Karusch-Kuhn-Tucker (KKT) necessary conditions

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

$$\mu_j^* \geq 0; \quad \mu_j^* \underline{g}_j(\underline{x}^*) = 0, \quad \forall\, j = 1, 2, \ldots, r$$

NO RESTRICTION ON SIGN OF λ_i^*

$$\underline{y}^T \nabla_{xx}^2 L \underline{y} \geq 0, \text{ for } \underline{y} \ni \nabla \underline{h}_i^T(\underline{x}^*)\underline{y} = 0; \qquad i = 1, 2, \ldots, m$$

$$\nabla \underline{g}_j^T(\underline{x}^*)\underline{y} = 0; \qquad j = \mathcal{A}(\underline{x}^*)$$

- Sufficient conditions: Replace ≥ 0 by > 0

Illustration of Optimality Conditions - 1

- Example

$$\min \frac{1}{2}(x_1^2 + x_2^2 + x_3^2)$$

$$\text{s.t. } x_1 + x_2 + x_3 \le -3$$

Necessary conditions:

$$x_1^* + \mu_1^* = 0$$

$$x_2^* + \mu_1^* = 0$$

$$x_3^* + \mu_1^* = 0$$

$$\mu_1^*(x_1 + x_2 + x_3 + 3) = 0$$

$$\mu_1^* \ge 0$$

case 1: $x_1^* + x_2^* + x_3^* < -3 \Rightarrow \mu_1^* = 0 \Rightarrow x_1^* = x_2^* = x_3^* = 0 > -3$, contradiction

case 2: $x_1^* + x_2^* + x_3^* = -3 \Rightarrow \mu_1^* = 1, x_1^* = x_2^* = x_3^* = -1$

Second order cond^n

$$\underline{y}^T \underline{y} > 0 \quad \forall \underline{y} \ni y_1 + y_2 + y_3 = 0 \Rightarrow y_1^2 + y_2^2 + (y_1 + y_2)^2 > 0 \,\forall \text{non-zero } y_1, y_2$$

so, $\underline{x}^*$ is a strict local minimum (& a global minimum as well)

Illustration of Optimality Conditions - 2

- Example 2: $\min x^2 + y^2 - 14x - 6y - 7 = (x-7)^2 + (y-3)^2 - 65$

  s.t. $x + y \le 2$

  $x + 2y \le 3$

  Necessary conditions:

  $$2x + \mu_1 + \mu_2 - 14 = 0$$
  $$2y + \mu_1 + 2\mu_2 - 6 = 0$$
  $$\mu_1 (x + y - 2) = 0$$
  $$\mu_2 (x + 2y - 3) = 0$$
  $$\mu_1,\ \mu_2 \ge 0$$

  Case 1: $\mu_1 = 0,\ \mu_2 = 0 \ \Rightarrow\ x = 7,\ y = 3$ Violated constraints

  Case 2: $\mu_1 > 0,\ \mu_2 = 0 \ \Rightarrow$ Solve $\begin{aligned} 2x + \mu_1 &= 14 \\ 2y + \mu_1 &= 6 \\ x + y &= 2 \end{aligned} \ \Rightarrow\ \begin{aligned} \mu_1 &= 8 \\ x &= 3 \\ y &= -1 \end{aligned}$

  so, $x = 3,\ y = -1,\ \mu_1 = 8,\ \mu_2 = 0$

  $x + 2y = 3 - 2 < 3$ OK $\Rightarrow$ optimal solution

Illustration of Optimality Conditions - 3

Check case 3: $\mu_1 = 0,\ \mu_2 > 0$

$$\left.\begin{aligned} 2x+\mu_2 &= 14 \\ 2y+2\mu_2 &= 6 \\ x+2y &= 3 \end{aligned}\right\} \quad \begin{aligned} 4x-2y &= 22 \\ x+2y &= 3 \end{aligned} \Rightarrow \left.\begin{aligned} x &= 5 \\ y &= -1 \end{aligned}\right\} \quad x+y = 4 > 2$$

NO!!

Case 4: $\mu_1 > 0,\ \mu_2 > 0$

$$\left.\begin{aligned} 2x+\mu_1+\mu_2 &= 14 \\ 2y+\mu_1+2\mu_2 &= 6 \\ x+y &= 2 \\ x+2y &= 3 \end{aligned}\right\} \Rightarrow \quad y=1,\ x=1,\ \mu_2=-8,\ \mu_2=20$$

NO GOOD!!

- Second order condition

$$2\left(y_1^2+y_2^2\right) \ni \ y_1+y_2=0$$

$$=4y_1^2>0 \ \ \forall y_1 \neq 0 \Rightarrow \text{ Strict local minimum}$$

Illustration of Optimality Conditions - 4

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

$$L(\underline{x},\lambda,\mu_1,\mu_2) = (x_1-4)^2 + x_2^2 + \lambda(x_1+x_2-3) - \mu_1 x_1 - \mu_2 x_2$$

$$\frac{\partial L}{\partial x_1} = 2(x_1-4)+\lambda-\mu_1 = 0; \frac{\partial L}{\partial x_2} = 2x_2+\lambda-\mu_2 = 0$$

$$\frac{\partial L}{\partial \lambda} = x_1+x_2-3 = 0; \mu_1 x_1 = \mu_2 x_2 = 0$$

$$case1: \mu_1 = 0, \mu_2 = 0 \Rightarrow \lambda = 1, x_1 = 7/2, x_2 = -1/2 \Rightarrow NO\ Good$$

$$case2: \mu_1 = 0, \mu_2 > 0 \Rightarrow \lambda = \mu_2 = 2, x_1 = 3, x_2 = 0 \Rightarrow Optimal$$

$$case3: \mu_1 > 0, \mu_2 = 0 \Rightarrow \lambda = -3, x_1 = 0, x_2 = 3, \mu_1 = -8 \Rightarrow NO\ Good$$

$$case4: \mu_1 > 0, \mu_2 > 0 \Rightarrow x_1 = 0, x_2 = 0 \Rightarrow NO\ Good$$

- Second order condition

$$2\left(y_1^2 + y_2^2\right) \;\; \ni \;\; y_1 + y_2 = 0 \;\&\; y_2 = 0$$

$$= 2y_1^2 > 0 \;\; \forall y_1 \neq 0 \Rightarrow \;\; \text{Strict local minimum}$$

Farkas' Lemma - 1

- For linear inequality constraints, Lagrange multipliers exist even in the absence of regularity condition ⇒ Don't need independence of rows of A

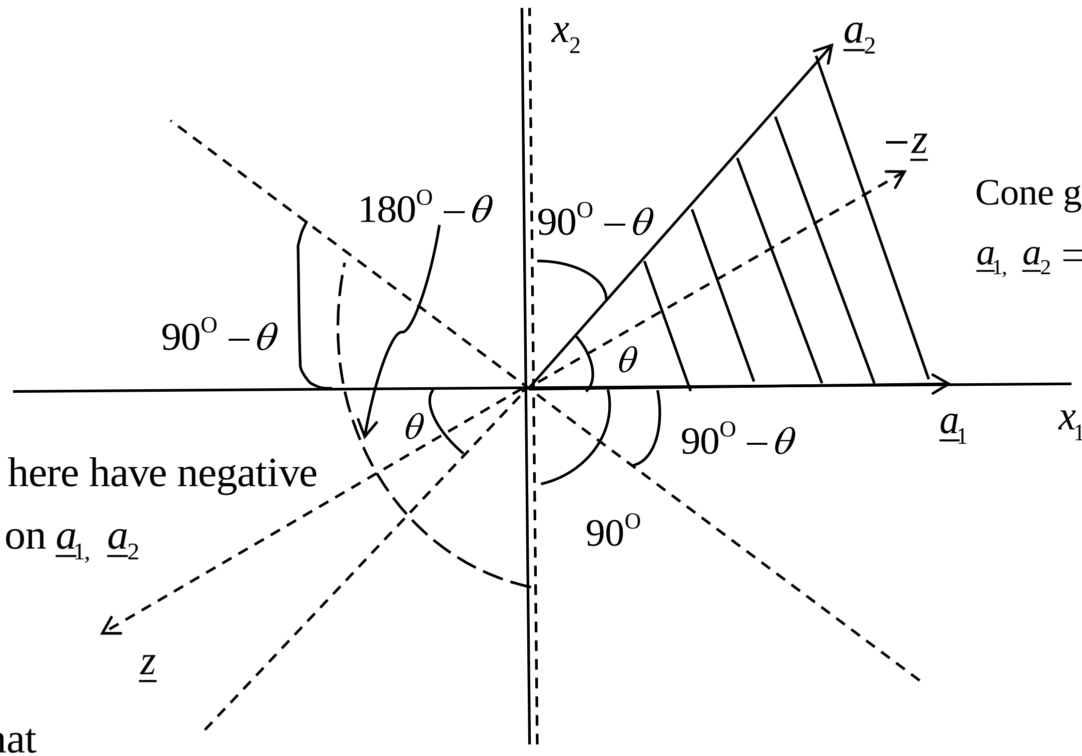
$$\min f(\underline{x})$$
$$\text{s.t. } \underline{a}_j^T \underline{x} \le b_j; \quad j = 1, 2, \ldots, r$$

This proof is based on Farkas' lemma

- Farkas' lemma : Let $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_m \in R^n$. Let $\underline{z}$ be another vector in R^n then

$$\underline{a}_j^T \underline{y} \le 0 \text{ for all } j \Rightarrow \underline{z}^T \underline{y} \ge 0 \Leftrightarrow \underline{z} + \sum_{j=1}^{r} \mu_j \underline{a}_j = 0, \quad \mu_j \ge 0$$

Farkas' Lemma - 2



Cone generated by

$\underline{a}_{1,}\ \underline{a}_2 \Rightarrow \underline{x} = \sum \mu_j \underline{a}_j;\ \mu_j \geq 0$

Vectors in here have negative projection on $\underline{a}_{1,}\ \underline{a}_2$

$\underline{z}$ is such that every vector in the indicated area has positive projection

$-\underline{z}$ lies in the cone $(\underline{a}_i)$

Application of Farkas' Lemma

- Now consider $\min f(\underline{x})$ s.t. $\underline{a}_j^T \underline{x} \le b_j, \; j \in A(\underline{x}^*)$

  From Nec. condition: $\nabla \underline{f}^T(\underline{x}^*)(\underline{x} - \underline{x}^*) \ge 0, \; \forall \; \underline{x} \ni \underline{a}_j^T \underline{x} \le b_j, \; j \in A(\underline{x}^*)$

  since $\underline{a}_j^T \underline{x} \le b_j \Rightarrow \underline{a}_j^T (\underline{x} - \underline{x}^*) \le 0$

  Let $(\underline{x} - \underline{x}^*) = y \Rightarrow \nabla \underline{f}^T(\underline{x}^*)\underline{y} \ge 0, \; \forall \; \underline{y} \ni \underline{a}_j^T \underline{y} \le b_j, \; j \in A(\underline{x}^*)$

  From Farka's lemma, $\nabla \underline{f}(\underline{x}^*) + \sum_{j \in A(\underline{x}^*)} \mu_j^* \underline{a}_j = 0$

  $$\text{since } \mu_j^* = 0 \text{ for } j \notin A(\underline{x}^*) \Rightarrow \nabla \underline{f}^T(\underline{x}^*) + A^T \underline{\mu} = 0$$

- The result extends to equality constraints $\underline{c}_i^T \underline{x} = d_i$ since

  $$\underline{c}_i^T \underline{x} = d_i \quad \Rightarrow \quad \underline{c}_i^T \underline{x} \le d_i$$

  $$\underline{c}_i^T \underline{x} \ge d_i \;\; \text{(or)} \;\; -\underline{c}_i^T \underline{x} \le -d_i$$

Any equality constraint can be re-written as two inequality constraints

Convex Programming and Duality-1

Convex programming problems and Duality

$$\min f(\underline{x})$$
$$\text{s.t. } \underline{x} \in \Omega \text{ and } g_j(\underline{x}) \le 0, \quad j = 1, 2, \ldots, r$$
$$f(\underline{x}) \text{ convex, } g_j(\underline{x}) \text{ is convex and } \Omega \text{ convex}$$

$$\text{Lagrangian } L(\underline{x}, \underline{\mu}) = f(\underline{x}) + \sum_{j=1}^{r} \mu_j g_j(\underline{x})$$

$$\text{Also } \min_{\underline{x} \in \Omega} L(\underline{x}, \underline{\mu}^*) = f(\underline{x}^*)$$

$$\text{Since } \mu_j^* g_j(\underline{x}^*) = 0 \;\Rightarrow\; f(\underline{x}^*) = f(\underline{x}^*) + \sum_{j=1}^{r} \mu_j^* g_j(\underline{x}^*) = \min_{\underline{x} \in \Omega} \left\{ f(\underline{x}) + \sum_{j=1}^{r} \mu_j^* g_j(\underline{x}) \right\}$$

Geometric interpretation of Lagrange multiplier vector

$\underline{\mu}^*$ is a Lagrange multiplier vector if and only if the set $S \subset R^{r+1}$ of all possible pairs of $\{\underline{g}(\underline{x}), f(\underline{x})\}$ as $\underline{x}$ ranges over Ω

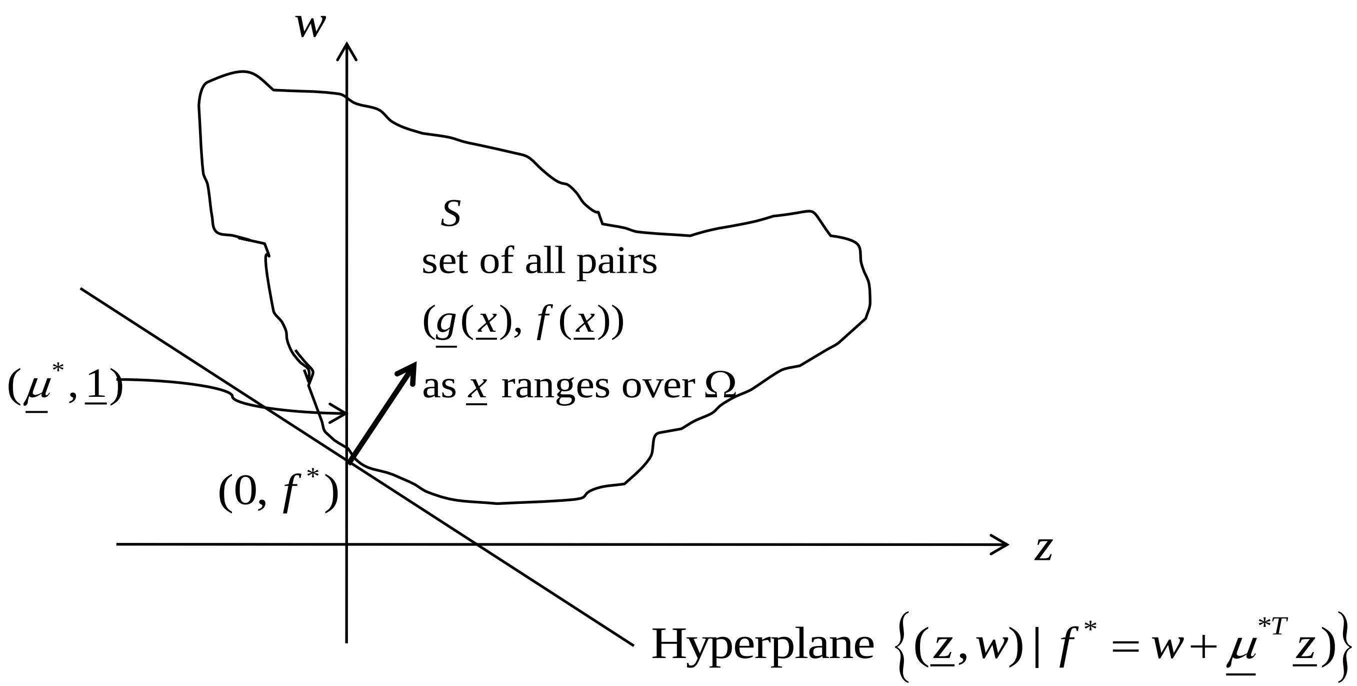
$$S = \{(\underline{z}, w) \mid \underline{z} = \underline{g}(\underline{x}),\ w = f(\underline{x}),\ \underline{x} \in \Omega\}$$

Convex Programming and Duality-2

lies in the half space $\{(\underline{z}, w) \mid f^* \le w + \underline{\mu}^{*T} \underline{z}\}$

The hyperplane $\{(\underline{z}, w) \mid f^* = w + \underline{\mu}^{*T} \underline{z}\}$ passes through $(0, f^*)$

and is the supporting hyperplane of S



Saddle Point Theorem - 1

- Saddle Point Theorem

$$L(\underline{x}^*, \underline{\mu}) \leq L(\underline{x}^*, \underline{\mu}^*) \leq L(\underline{x}, \underline{\mu}^*), \quad \forall\ \underline{x} \in \Omega \text{ and } \underline{\mu} \geq \underline{0}$$

Recall $L(\underline{x}^*, \underline{\mu}^*) = f(\underline{x}^*) = \min_{\underline{x}\in\Omega} L(\underline{x}, \underline{\mu}^*) \leq L(\underline{x}, \underline{\mu}^*)$

$$= f(\underline{x}) + \sum_{j=1}^{r} \mu_j^* g_j(\underline{x}) \leq f(\underline{x}) \text{ since } g_j(\underline{x}) \leq 0$$

Also $L(\underline{x}^*, \underline{\mu}) = f(\underline{x}^*) + \sum_{j=1}^{r} \mu_j g_j(\underline{x}^*) = q(\underline{\mu}) \leq f(\underline{x}^*) = L(\underline{x}^*, \underline{\mu}^*)$ since $\mu_j \geq 0$

- Note that

$$f(\underline{x}^*) = \min_{\underline{x}\in\Omega} \max_{\underline{\mu}\geq 0} L(\underline{x}, \underline{\mu})$$

Since

$$\max_{\underline{\mu}\geq 0} L(\underline{x}, \underline{\mu}) = \max_{\underline{\mu}\geq 0} \left\{ f(\underline{x}) + \sum_{j=1}^{r} \mu_j g_j(\underline{x}) \right\} \leq f(\underline{x})$$

$$= \begin{cases} f(\underline{x}) \text{ if } g_j(\underline{x}) \leq 0;\ j = 1, 2, \ldots, r \\ \infty \qquad \text{otherwise} \end{cases}$$

Saddle Point Theorem - 2

If we let

$$q^* = \max_{\underline{\mu} \geq 0} \min_{\underline{x} \in \Omega} L(\underline{x}, \underline{\mu})$$

Also

$$\min_{\underline{x} \in \Omega} L(\underline{x}, \underline{\mu}) \leq L(\underline{z}, \underline{\mu}) \qquad \forall \underline{z} \in \Omega \text{ and } \underline{\mu} \geq 0$$

Taking max over $\underline{\mu}$

$$q^* = \max_{\underline{\mu}} \min_{\underline{x} \in \Omega} L(\underline{x}, \underline{\mu}) \leq \max_{\underline{\mu}} L(\underline{z}, \underline{\mu}) = L(\underline{z}, \underline{\mu}^*)$$

Taking min over $\underline{z} \in \Omega$

$$q^* \leq f^*$$

- Optimal dual solution q^* = optimal primal solution f^*

$$\left.\begin{array}{l} f^* = \inf_{\underline{x} \in \Omega} L(\underline{x}, \underline{\mu}^*) \leq q^* \\ \quad \text{since } q^* \leq f^* \end{array}\right\} \Rightarrow f^* = q^*$$

Primal-Dual Problems - 1

- Primal-Dual problems

$$\min f(\underline{x}) \quad \text{s.t. } g_j(\underline{x}) \le 0,\quad \underline{x} \in \Omega \quad \Leftrightarrow \quad \max_{\underline{\mu} \ge 0} q(\underline{\mu})$$

$$\text{where } q(\underline{\mu}) = \min_{\underline{x}\in\Omega}\left[f(\underline{x}) + \sum_{j=1}^{r} \mu_j g_j(\underline{x}) \right]$$

- Linear programming (LP) with inequality constraints & unconstrained $\underline{x}$

$$\min \underline{c}^T \underline{x} \quad \text{s.t. } \underline{a}_j^T \underline{x} - b_j \ge 0,\quad b_j - \underline{a}_j^T \underline{x} \le 0 \quad \Leftrightarrow \quad \max_{\underline{\mu} \ge 0} q(\underline{\mu})$$

$$\text{where } q(\underline{\mu}) = \min_{\underline{x}} \left\{ \underline{c}^T \underline{x} + \sum_{j=1}^{r} \mu_j \left(-\underline{a}_j^T \underline{x} + b_j \right) \right\}$$

$$q(\underline{\mu}) > -\infty \text{ only if } \underline{c} = \sum_{j=1}^{r} \underline{a}_j \mu_j = A\underline{\mu}$$

<u>PRIMAL</u>

$$\min \underline{c}^T \underline{x}$$

$$\text{s.t. } \underline{a}_j^T \underline{x} - b_j \ge 0,\ j = 1, 2, \ldots r$$

<u>DUAL</u>

$$\max_{\underline{\mu} \ge 0} \underline{b}^T \underline{\mu}$$

$$\text{s.t. } A\underline{\mu} = \underline{c}$$

Primal-Dual Problems - 2

- LP with inequality constraints and non-negative $\underline{x}$

PRIMAL

$$\min \underline{c}^T \underline{x}$$
$$\text{s.t. } A^T \underline{x} \geq \underline{b}$$
$$\underline{x} \geq \underline{0}$$

DUAL

$$\max_{\underline{\mu} \geq \underline{0}} q(\underline{\mu}) = \underline{b}^T \underline{\mu}$$
$$\text{s.t. } A\underline{\mu} \leq \underline{c}$$
$$\underline{\mu} \geq \underline{0}$$

$$\text{Proof}: q(\underline{\mu}) = \min_{\underline{x} \geq 0} \left\{ \underline{c}^T \underline{x} - \underline{\mu}^T (A^T \underline{x} - \underline{b}) \right\}$$

- $$\min \frac{1}{2} \underline{x}^T Q \underline{x} - \underline{b}^T \underline{x}$$
$$\text{s.t. } \underline{a}^T \underline{x} \leq c$$

$$\text{Dual}: \max_{\underline{\mu} \geq \underline{0}} q(\underline{\mu})$$

$$q(\mu) = \min_{\underline{x}} \left\{ \frac{1}{2} \underline{x}^T Q \underline{x} - \underline{b}^T \underline{x} + \mu (\underline{a}^T \underline{x} - c) \right\}$$

$$\Rightarrow \min \text{ at } \underline{x} = Q^{-1} (\underline{b} - \underline{a} \mu)$$

$$q(\mu) = -\frac{1}{2} \mu^2 \underline{a}^T Q^{-1} \underline{a} - \mu d - \frac{1}{2} \underline{b}^T Q^{-1} \underline{b}$$

$$d = c - \underline{a}^T Q^{-1} \underline{b}, \quad \text{Optimal } \mu^* = \max \left(\frac{-d}{\underline{a}^T Q^{-1} \underline{a}}, 0 \right)$$

Primal-Dual Problems - 3

- Production allocation

A unit of electric power must be produced by n units. x_i is the amount produced by unit i. $f_i(x_i)$ cost of producing x_i units by ith unit. Assumed to be convex.

$$\left.\begin{aligned}&\min \sum_{i=1}^{n} f_i(x_i)\\ &\text{s.t.} \quad \sum_{i=1}^{n} x_i = A\\ &\qquad \alpha_i \le x_i \le \beta_i\end{aligned}\right\} \text{ Separable programming problem}$$

$$\text{Dual problem } \max_{\lambda} q(\lambda) = \sum_{i=1}^{n} \overbrace{\min_{\alpha_i \le x_i \le \beta_i} \left[f_i(x_i) + \lambda x_i \right]}^{q_i(\lambda)} - \lambda A$$

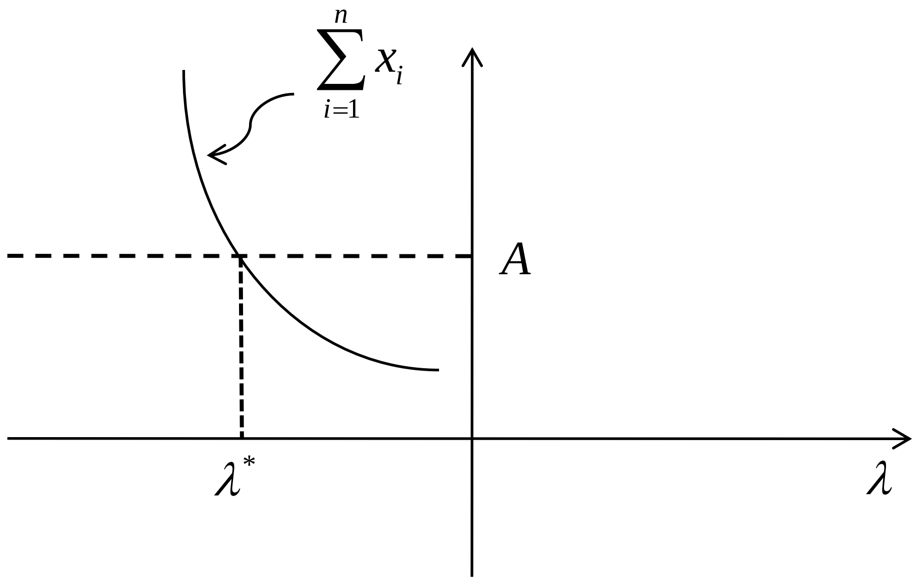
For a given λ, the minimization for each x_i can be carried out independently.

At the solution

Primal-Dual Problems - 4

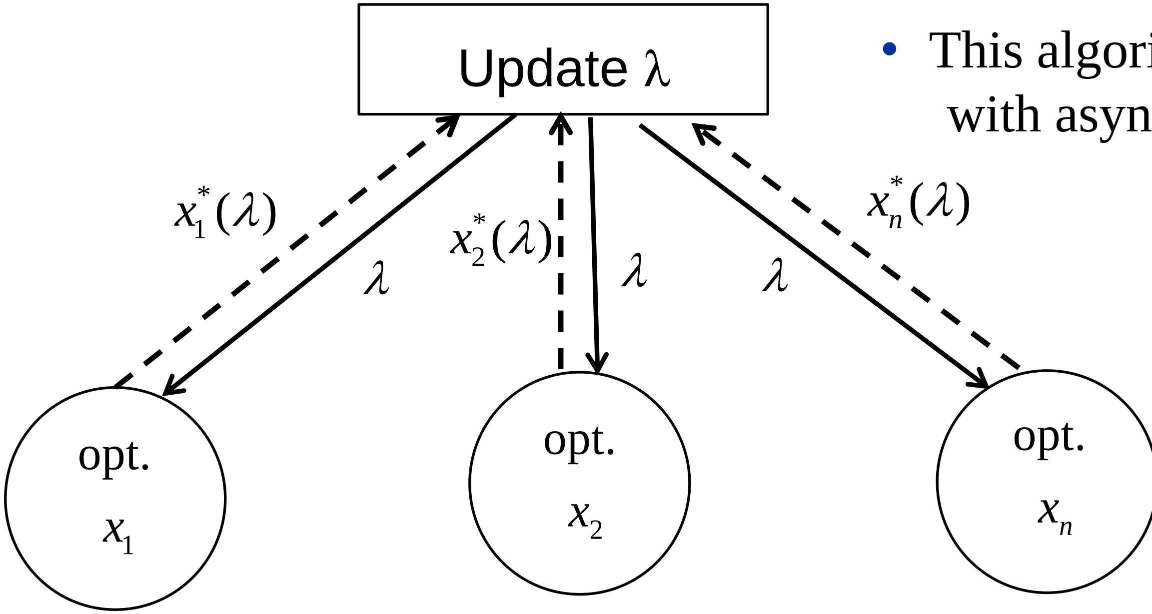
$$\text{Marginal production cost} = \left.\frac{df_i(x_i)}{dx_i}\right|_{x_i^*(\lambda)} = \begin{cases} -\lambda & \text{if } \alpha_i < x_i^*(\lambda) < \beta_i \\ \geq -\lambda & \text{if } x_i^*(\lambda) = \alpha_i \\ \leq -\lambda & \text{if } x_i^*(\lambda) = \beta_i \end{cases}$$

- Note that $\max_{\lambda} q(\lambda)$ is a one-dimensional search problem.



Primal-Dual Problems - 5

- Algorithm is well-suited for parallel implementation



- This algorithm works even with asynchronous updates

Constrained Optimization Algorithms

- Now, we have the machinery to develop algorithms

1. Penalty and augmented Lagrangian methods ….. Relation to primal-dual methods
2. Feasible direction methods…..

Manifold sub-optimization methods

Gradient projection

Reduced gradient

Quadratic programming

3. Solving the necessary conditions of optimality

Summary

- ❑ Lagrange Multipliers and Duality
- ❑ Inequality (Mixed) Constraints
- ❑ Karusch-Kuhn-Tucker (KKT) Conditions
- ❑ Illustrative Examples
- ❑ Convex Programming and Duality
- ❑ Saddle Point Theorem
- ❑ Primal-Dual Methods