MA 5037: Optimization Methods and Applications
Optimality Conditions for LCPs

Suh-Yuh Yang (楊肅煜)

Department of Mathematics, National Central University
Jhongli District, Taoyuan City 320317, Taiwan

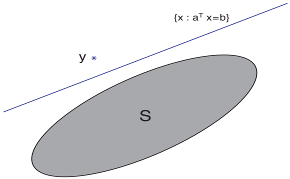
First version: July 29, 2018/Last updated: June 18, 2025

Optimality conditions

- One of the main drawbacks of the concept of *stationarity* is that for most feasible sets, it is rather difficult to validate whether this condition is satisfied or not, and it is even more difficult to use it in order to actually solve the underlying optimization problem.
- Our main objective is to derive an equivalent *optimality condition* that is much easier to handle.
- In this lecture, we will establish the so-called *Karush-Kuhn-Tucker (KKT) conditions* for the special case of linearly constrained problems (LCPs).

Strict separation theorem

- **Definition:** *Given $S \subseteq \mathbb{R}^n$, a hyperplane $H := \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{a}^\top \boldsymbol{x} = b\}$, where $\boldsymbol{a} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ and $b \in \mathbb{R}$, is said to strictly separate a point $\boldsymbol{y} \notin S$ from S if $\boldsymbol{a}^\top \boldsymbol{y} > b$ and $\boldsymbol{a}^\top \boldsymbol{x} \le b$, $\forall\, \boldsymbol{x} \in S$.*



- **Theorem:** *(strict separation theorem) Let $C \subseteq \mathbb{R}^n$ be a closed and convex set and $\boldsymbol{y} \notin C$. Then $\exists\, \boldsymbol{p} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ and $\alpha \in \mathbb{R}$ such that*

$$\boldsymbol{p}^\top \boldsymbol{y} > \alpha \quad \text{and} \quad \boldsymbol{p}^\top \boldsymbol{x} \le \alpha,\ \forall\, \boldsymbol{x} \in C.$$

Proof: By the second projection theorem, the vector $\bar{\boldsymbol{x}} := P_C(\boldsymbol{y}) \in C$ satisfies

$$(\boldsymbol{y} - \bar{\boldsymbol{x}})^\top (\boldsymbol{x} - \bar{\boldsymbol{x}}) \le 0 \quad \forall\, \boldsymbol{x} \in C \Longrightarrow (\boldsymbol{y} - \bar{\boldsymbol{x}})^\top \boldsymbol{x} \le (\boldsymbol{y} - \bar{\boldsymbol{x}})^\top \bar{\boldsymbol{x}} \quad \forall\, \boldsymbol{x} \in C.$$

Denote $\boldsymbol{p} = \boldsymbol{y} - \bar{\boldsymbol{x}} \neq \boldsymbol{0}$ and $\alpha = (\boldsymbol{y} - \bar{\boldsymbol{x}})^\top \bar{\boldsymbol{x}}$. Then we have $\boldsymbol{p}^\top \boldsymbol{x} \le \alpha\ \forall\, \boldsymbol{x} \in C$. On the other hand, $\boldsymbol{p}^\top \boldsymbol{y} = (\boldsymbol{y} - \bar{\boldsymbol{x}})^\top \boldsymbol{y} = (\boldsymbol{y} - \bar{\boldsymbol{x}})^\top (\boldsymbol{y} - \bar{\boldsymbol{x}}) + (\boldsymbol{y} - \bar{\boldsymbol{x}})^\top \bar{\boldsymbol{x}} = \|\boldsymbol{y} - \bar{\boldsymbol{x}}\|^2 + \alpha$. Thus, we have $\boldsymbol{p}^\top \boldsymbol{y} > \alpha$, and the result is established. $\square$

Farkas' lemma: first formulation

- **Farkas' lemma:** *Let $\boldsymbol{c} \in \mathbb{R}^n$ and $\boldsymbol{A} \in \mathbb{R}^{m\times n}$. Then exactly one of the following systems has a solution:*
  (I) $\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{0}, \boldsymbol{c}^\top \boldsymbol{x} > 0.$ (II) $\boldsymbol{A}^\top \boldsymbol{y} = \boldsymbol{c}, \boldsymbol{y} \geq \boldsymbol{0}.$
- **Example:** Consider the following example:

$$\boldsymbol{A} = \begin{bmatrix} 1 & 5 \\ -1 & 2 \end{bmatrix}, \quad \boldsymbol{c} = \begin{bmatrix} -1 \\ 9 \end{bmatrix}.$$

  System (I) is infeasible since the system $\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{0}$ implies the inequality $\boldsymbol{c}^\top \boldsymbol{x} \leq 0$. In practice,

$$\begin{aligned} x + 5y &\leq 0, \\ -x + 2y &\leq 0. \end{aligned}$$

  Then $eqn(1) + 2 \times eqn(2) \Rightarrow -x + 9y \leq 0$, i.e., $\boldsymbol{c}^\top \boldsymbol{x} \leq 0$. *The row vector $\boldsymbol{c}^\top$ can be written as a conic combination of the rows of $\boldsymbol{A}$. In other words, $\boldsymbol{c}$ is a conic combination of the columns of $\boldsymbol{A}^\top$:*

$$\begin{bmatrix} 1 \\ 5 \end{bmatrix} + 2\begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ 9 \end{bmatrix} \text{ or } \underbrace{\begin{bmatrix} 1 & -1 \\ 5 & 2 \end{bmatrix}}_{\boldsymbol{A}^\top} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \underbrace{\begin{bmatrix} -1 \\ 9 \end{bmatrix}}_{\boldsymbol{c}}.$$

Farkas' lemma: second formulation

Let $\boldsymbol{c} \in \mathbb{R}^n$ and $\boldsymbol{A} \in \mathbb{R}^{m\times n}$. Then the following two claims are equivalent:
(A) *The implication $\boldsymbol{A}\boldsymbol{x} \le \boldsymbol{0} \Rightarrow \boldsymbol{c}^\top \boldsymbol{x} \le 0$ holds true.*
(B) $\exists\, \boldsymbol{y} \in \mathbb{R}^m_+$ *such that* $\boldsymbol{A}^\top \boldsymbol{y} = \boldsymbol{c}$.

Proof:

(B) $\Rightarrow$ (A): Assume that system (B) is feasible. Let $\boldsymbol{A}\boldsymbol{x} \le \boldsymbol{0}$ for some $\boldsymbol{x} \in \mathbb{R}^n$. Then $\boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{x} \le 0$. Since $\boldsymbol{c}^\top = \boldsymbol{y}^\top \boldsymbol{A}$, we have $\boldsymbol{c}^\top \boldsymbol{x} \le 0$.

(A) $\Rightarrow$ (B): Suppose in contradiction that system (B) is infeasible. Consider the following closed and convex set $S := \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} = \boldsymbol{A}^\top \boldsymbol{y} \text{ for some } \boldsymbol{y} \in \mathbb{R}^m_+\}$. (The closedness of S follows from Lemma 6.32). Then $\boldsymbol{c} \notin S$. By the strict separation theorem, $\exists\, \boldsymbol{p} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ and $\alpha \in \mathbb{R}$ such that

$$\boldsymbol{p}^\top \boldsymbol{c} > \alpha \quad \text{and} \quad \boldsymbol{p}^\top \boldsymbol{x} \le \alpha,\ \forall\, \boldsymbol{x} \in S.$$

Since $\boldsymbol{0} \in S$, we can conclude that $\alpha \ge 0$ and also $\boldsymbol{p}^\top \boldsymbol{c} > 0 (\Rightarrow \boldsymbol{c}^\top \boldsymbol{p} > 0)$. In addition,

$$\boldsymbol{p}^\top \boldsymbol{x} \le \alpha,\ \forall\, \boldsymbol{x} \in S \Longleftrightarrow \boldsymbol{p}^\top \boldsymbol{A}^\top \boldsymbol{y} \le \alpha,\ \forall\, \boldsymbol{y} \ge \boldsymbol{0} \Longleftrightarrow (\boldsymbol{A}\boldsymbol{p})^\top \boldsymbol{y} \le \alpha,\ \forall\, \boldsymbol{y} \ge \boldsymbol{0},$$

which implies $\boldsymbol{A}\boldsymbol{p} \le \boldsymbol{0}$. We have thus arrived at a contradiction to the assumption that the implication (A) holds (using the vector $\boldsymbol{p}$), and consequently (B) is satisfied. $\square$

Gordan's alternative theorem

Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$. Then exactly one of the following two systems has a solution: (A) $\boldsymbol{Ax} < \boldsymbol{0}$. (B) $\boldsymbol{p} \neq \boldsymbol{0}, \boldsymbol{A}^\top \boldsymbol{p} = \boldsymbol{0}, \boldsymbol{p} \geq \boldsymbol{0}$.

Proof: Assume that system (A) has a solution. Suppose in contradiction that (B) is feasible. Then $\exists\, \boldsymbol{p} \neq \boldsymbol{0}, \boldsymbol{A}^\top \boldsymbol{p} = \boldsymbol{0}, \boldsymbol{p} \geq \boldsymbol{0}. \implies \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{p} = 0 \implies (\boldsymbol{Ax})^\top \boldsymbol{p} = \boldsymbol{0}$. This is impossible since $\boldsymbol{Ax} < \boldsymbol{0}$ and $\boldsymbol{0} \leq \boldsymbol{p} \neq \boldsymbol{0}$.

Now suppose that system (A) does not have a solution. Note that

$$\boldsymbol{Ax} < \boldsymbol{0} \iff \boldsymbol{Ax} + s\boldsymbol{e} \leq \boldsymbol{0}, \text{ for some } s > 0.$$

The latter system can be rewritten as

$$\widetilde{\boldsymbol{A}} \begin{bmatrix} \boldsymbol{x} \\ s \end{bmatrix} \leq \boldsymbol{0}, \quad \boldsymbol{c}^\top \begin{bmatrix} \boldsymbol{x} \\ s \end{bmatrix} > 0,$$

where $\widetilde{\boldsymbol{A}} = [\boldsymbol{A} \quad \boldsymbol{e}]$ and $\boldsymbol{c} = \boldsymbol{e}_{n+1}$. The infeasibility of (A) is thus equivalent to the infeasibility of the system

$$\widetilde{\boldsymbol{A}}\boldsymbol{w} \leq \boldsymbol{0}, \quad \boldsymbol{c}^\top \boldsymbol{w} > 0, \quad \boldsymbol{w} \in \mathbb{R}^{n+1}.$$

By Farkas' lemma, $\exists\, \boldsymbol{z} \in \mathbb{R}^m_+$ such that

$$\begin{bmatrix} \boldsymbol{A}^\top \\ \boldsymbol{e}^\top \end{bmatrix} \boldsymbol{z} = \boldsymbol{c} \implies \boldsymbol{A}^\top \boldsymbol{z} = \boldsymbol{0}, \quad \boldsymbol{e}^\top \boldsymbol{z} = 1.$$

Since $\boldsymbol{e}^\top \boldsymbol{z} = 1$, $\boldsymbol{z} \neq \boldsymbol{0}$. We have shown the existence of $\boldsymbol{0} \neq \boldsymbol{z} =: \boldsymbol{p} \in \mathbb{R}^m_+$ s.t. $\boldsymbol{A}^\top \boldsymbol{z} = \boldsymbol{0}$. $\square$

KKT conditions: necessary optimality conditions

Gordan's alternative theorem can be used to establish an optimality criterion that is in fact a special case of the so-called Karush-Kuhn-Tucker (KKT) conditions (Chapter 11). Here we derive the KKT conditions for LCPs by using Farkas' lemma.

Theorem: *Consider the following minimization problem:*

$$(P) \qquad \min f(\boldsymbol{x}) \quad \text{s.t. } \boldsymbol{a}_i^\top \boldsymbol{x} \le b_i, \quad i = 1, 2, \cdots, m,$$

where f is continuously differentiable over $\mathbb{R}^n$, $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_m \in \mathbb{R}^n$, $b_1, b_2, \cdots, b_m \in \mathbb{R}$. Let $\boldsymbol{x}^*$ be a local minimum point of (P). Then there exist $\lambda_1, \lambda_2, \cdots, \lambda_m \ge 0$ such that

$$\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i \boldsymbol{a}_i = \mathbf{0} \quad \text{and} \quad \lambda_i(\boldsymbol{a}_i^\top \boldsymbol{x}^* - b_i) = 0, \; i = 1, 2, \cdots, m.$$

Proof: Since x^* is a local minimum point of (P), x^* is a stationary point. Therefore, $\nabla f(\boldsymbol{x}^*)^\top (\boldsymbol{x} - \boldsymbol{x}^*) \ge 0, \forall\, \boldsymbol{x} \in \mathbb{R}^n$ satisfying $\boldsymbol{a}_i^\top \boldsymbol{x} \le b_i, i = 1, 2, \cdots, m$. Denote the set of active constraints by

$$I(\boldsymbol{x}^*) = \{i : \boldsymbol{a}_i^\top \boldsymbol{x}^* = b_i\}.$$

Making the change of variables $\boldsymbol{y} = \boldsymbol{x} - \boldsymbol{x}^*$, we obtain

$\nabla f(\boldsymbol{x}^*)^\top \boldsymbol{y} \ge 0$ for any $\boldsymbol{y} \in \mathbb{R}^n$ satisfying $\boldsymbol{a}_i^\top (\boldsymbol{y} + \boldsymbol{x}^*) \le b_i$ for $i = 1, 2, \cdots, m$. *(TBC...)*

KKT conditions: necessary optimality conditions (cont'd)

That is, we have

$$\begin{array}{l}\nabla f(\boldsymbol{x}^*)^\top \boldsymbol{y} \geq 0 \text{ for any } \boldsymbol{y} \in \mathbb{R}^n \text{ satisfying } \boldsymbol{a}_i^\top \boldsymbol{y} \leq 0,\ i \in I(\boldsymbol{x}^*) \\ \text{and } \boldsymbol{a}_i^\top \boldsymbol{y} \leq b_i - \boldsymbol{a}_i^\top \boldsymbol{x}^*,\ i \notin I(\boldsymbol{x}^*).\end{array} \qquad (\star)$$

We will show that in fact the second set of inequalities in the latter system can be removed, that is, that the following implication is valid:

$$\text{If } \boldsymbol{a}_i^\top \boldsymbol{y} \leq 0 \text{ for all } i \in I(\boldsymbol{x}^*) \text{ then } \nabla f(\boldsymbol{x}^*)^\top \boldsymbol{y} \geq 0 \quad (\Rightarrow -\nabla f(\boldsymbol{x}^*)^\top \boldsymbol{y} \leq 0).$$

Assume that $\boldsymbol{y}$ satisfies $\boldsymbol{a}_i^\top \boldsymbol{y} \leq 0$ for all $i \in I(\boldsymbol{x}^*)$.

(1) Since $b_i - \boldsymbol{a}_i^\top \boldsymbol{x}^* > 0$ for all $i \notin I(\boldsymbol{x}^*)$, it follows that there exists a small enough $\alpha > 0$ such that $\boldsymbol{a}_i^\top (\alpha \boldsymbol{y}) \leq b_i - \boldsymbol{a}_i^\top \boldsymbol{x}^*$.

(2) In addition, $\boldsymbol{a}_i^\top (\alpha \boldsymbol{y}) \leq 0$ for all $i \in I(\boldsymbol{x}^*)$.

Therefore, from $(\star)$, we have $\nabla f(\boldsymbol{x}^*)^\top (\alpha \boldsymbol{y}) \geq 0$ and hence that $\nabla f(\boldsymbol{x}^*)^\top \boldsymbol{y} \geq 0$. By Farkas' lemma (second formulation), $\exists\ \lambda_i \geq 0$, $i \in I(\boldsymbol{x}^*)$, such that

$$-\nabla f(\boldsymbol{x}^*) = \sum_{i \in I(\boldsymbol{x}^*)} \lambda_i \boldsymbol{a}_i.$$

Defining $\lambda_i = 0$ for all $i \notin I(\boldsymbol{x}^*)$, we get that $\lambda_i(\boldsymbol{a}_i^\top \boldsymbol{x}^* - b_i) = 0$ for all $i = 1, 2, \cdots, m$ and

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

as required. $\square$

KKT conditions: sufficient optimality conditions

The KKT conditions are necessary conditions, but when f is convex, they are both necessary and sufficient global optimality conditions.

Theorem: *Consider the minimization problem*

$$(P) \qquad \min f(\boldsymbol{x}) \quad \text{s.t. } \boldsymbol{a}_i^\top \boldsymbol{x} \le b_i, \quad i = 1, 2, \cdots, m,$$

where f is a convex continuously differentiable function over $\mathbb{R}^n$, $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_m \in \mathbb{R}^n$, $b_1, b_2, \cdots, b_m \in \mathbb{R}$. Let $\boldsymbol{x}^*$ be a feasible solution of (P). Then $\boldsymbol{x}^*$ is an optimal solution of (P) if and only if $\exists\, \lambda_1, \lambda_2, \cdots, \lambda_m \ge 0$ such that

$$\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i \boldsymbol{a}_i = \boldsymbol{0} \ \text{ and } \ \lambda_i(\boldsymbol{a}_i^\top \boldsymbol{x}^* - b_i) = 0, \ i = 1, 2, \cdots, m. \quad (\star)$$

Note:

- *The nonnegative scalars $\lambda_1, \lambda_2, \cdots, \lambda_m$ in the KKT conditions are called Lagrange multipliers, where λ_i is the multiplier associated with the ith constraint $\boldsymbol{a}_i^\top \boldsymbol{x} \le b_i$.*
- The conditions $\lambda_i(\boldsymbol{a}_i^\top \boldsymbol{x}^* - b_i) = 0, i = 1, 2, \cdots, m$ are known in the literature as *the complementary slackness (互補鬆弛) conditions.*

Proof of the sufficient optimality conditions

($\Rightarrow$) It has been done in the previous theorem.
($\Leftarrow$) Assume that $\boldsymbol{x}^*$ be a feasible solution of (P) satisfying ($\star$). Let $\boldsymbol{x}$ be any feasible solution of (P). Define the function

$$h(\boldsymbol{x}) := f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i (\boldsymbol{a}_i^\top \boldsymbol{x} - b_i).$$

Then $\nabla h(\boldsymbol{x}^*) = \mathbf{0}$ and since h is convex, it follows that $\boldsymbol{x}^*$ is a minimizer of h over $\mathbb{R}^n$. From ($\star$), we have

$$\begin{aligned} f(\boldsymbol{x}^*) &= f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i (\boldsymbol{a}_i^\top \boldsymbol{x}^* - b_i) = h(\boldsymbol{x}^*) \\ &\leq h(\boldsymbol{x}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i (\boldsymbol{a}_i^\top \boldsymbol{x} - b_i) \leq f(\boldsymbol{x}). \end{aligned}$$

We have thus proven that $\boldsymbol{x}^*$ is a global optimal solution of (P). $\square$

KKT conditions for linearly constrained problems

We can generalize the previous two theorems to the case where linear equality constraints are also present.

Theorem: *Consider the following minimization problem:*

$$(Q) \qquad \begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & \boldsymbol{a}_i^\top \boldsymbol{x} \le b_i, \quad i = 1, 2, \cdots, m, \\ & \boldsymbol{c}_j^\top \boldsymbol{x} = d_j, \quad j = 1, 2, \cdots, p, \end{array}$$

where f is a continuously differentiable function over $\mathbb{R}^n$, $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_m$, $\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_p \in \mathbb{R}^n$, $b_1, b_2, \cdots, b_m, d_1, d_2, \cdots, d_p \in \mathbb{R}$. Then we have

1. **(necessity)** *If $\boldsymbol{x}^*$ is a local minimum point of (Q). Then there exist $\lambda_1, \lambda_2, \cdots, \lambda_m \ge 0$ and $\mu_1, \mu_2, \cdots, \mu_p \in \mathbb{R}$ such that*

$$\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i \boldsymbol{a}_i + \sum_{j=1}^{p} \mu_j \boldsymbol{c}_j = \boldsymbol{0}, \ \ \lambda_i(\boldsymbol{a}_i^\top \boldsymbol{x}^* - b_i) = 0, \ i = 1, 2, \cdots, m. \ (\star)$$

2. **(sufficiency)** *If in addition f is convex over $\mathbb{R}^n$ and $\boldsymbol{x}^*$ is a feasible solution of (Q) for which $\exists\ \lambda_1, \lambda_2, \cdots, \lambda_m \ge 0$, $\mu_1, \mu_2, \cdots, \mu_p \in \mathbb{R}$ such that $(\star)$ are satisfied, then $\boldsymbol{x}^*$ is an optimal solution of (Q).*

Proof of the KKT theorem

The proof is based on the simple observation that a linear equality constraint $\boldsymbol{a}^\top \boldsymbol{x} = b$ can be written as two inequality constraints, $\boldsymbol{a}^\top \boldsymbol{x} \le b$ and $-\boldsymbol{a}^\top \boldsymbol{x} \le -b$.

(1) Consider the equivalent problem

$$(Q') \qquad \begin{array}{ll} \min f(\boldsymbol{x}) & \\ \text{s.t.} & \boldsymbol{a}_i^\top \boldsymbol{x} \le b_i, \quad i = 1, 2, \cdots, m, \\ & \boldsymbol{c}_j^\top \boldsymbol{x} \le d_j, \quad -\boldsymbol{c}_j^\top \boldsymbol{x} \le -d_j, \quad j = 1, 2, \cdots, p. \end{array}$$

Since $\boldsymbol{x}^*$ is an optimal solution of (Q′), it follows that there exist multipliers $\lambda_1, \lambda_2, \cdots, \lambda_m \ge 0$ and $\mu_1^+, \mu_1^-, \mu_2^+, \mu_2^-, \cdots, \mu_p^+, \mu_p^- \ge 0$ such that

$$\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^m \lambda_i \boldsymbol{a}_i + \sum_{j=1}^p \mu_j^+ \boldsymbol{c}_j - \sum_{j=1}^p \mu_j^- \boldsymbol{c}_j = \boldsymbol{0}, \qquad (\star_1)$$

$$\lambda_i(\boldsymbol{a}_i^\top \boldsymbol{x}^* - b_i) = 0, \quad i = 1, 2, \cdots, m, \qquad (\star_2)$$

$$\mu_j^+(\boldsymbol{c}_j^\top \boldsymbol{x}^* - d_j) = 0, \ \mu_j^-(-\boldsymbol{c}_j^\top \boldsymbol{x}^* + d_j) = 0, \quad j = 1, 2, \cdots, p. \qquad (\star_3)$$

We thus obtain that $(\star)$ are satisfied with $\mu_j := \mu_j^+ - \mu_j^-, j = 1, 2, \cdots, p$.

(2) Assume that $\boldsymbol{x}^*$ satisfies $(\star)$. Then it also satisfies $(\star_1)$, $(\star_2)$ and $(\star_3)$ with $\mu_j^+ = \max\{\mu_j, 0\}$, $\mu_j^- = -\min\{\mu_j, 0\}$. By the theorem on page 9, $\boldsymbol{x}^*$ is an optimal solution of (Q′) and thus also an optimal solution of (Q). $\square$

Note: *A feasible point $\boldsymbol{x}^*$ is called a KKT point if there exist multipliers for which $(\star)$ on page 11 are satisfied.*

General nonlinear programming problems

The general setting of general nonlinear programming problems is given by

$$(NLP) \qquad \begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & g_i(\boldsymbol{x}) \leq 0, \quad i = 1, 2, \cdots, m, \\ & h_j(\boldsymbol{x}) = 0, \quad j = 1, 2, \cdots, p, \end{array}$$

where $f, g_1, \cdots, g_m$, and $h_1, \cdots, h_p$ are all continuously differentiable functions over $\mathbb{R}^n$. The associated *Lagrangian function* takes the form

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i g_i(\boldsymbol{x}) + \sum_{j=1}^{p} \mu_j h_j(\boldsymbol{x}).$$

The details of the general NLP will be studied in the next chapter. In the linearly constrained case of problem (Q), the first condition in $(\star)$ on page 11 is the same as

$$\nabla_x L(\boldsymbol{x}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{p} \mu_j \nabla h_j(\boldsymbol{x}^*) = \boldsymbol{0}.$$

The associated Lagrangian function of problem (Q)

Back to problem (Q), we define the matrices $\boldsymbol{A}$ and $\boldsymbol{C}$ and the vectors $\boldsymbol{b}$ and $\boldsymbol{d}$ by

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1^\top \\ \boldsymbol{a}_2^\top \\ \vdots \\ \boldsymbol{a}_m^\top \end{bmatrix}, \quad \boldsymbol{C} = \begin{bmatrix} \boldsymbol{c}_1^\top \\ \boldsymbol{c}_2^\top \\ \vdots \\ \boldsymbol{c}_p^\top \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}, \quad \boldsymbol{d} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_p \end{bmatrix},$$

then the constraints of problem (Q) can be written as

$$\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}, \quad \boldsymbol{C}\boldsymbol{x} = \boldsymbol{d}.$$

The Lagrangian function can be also written as

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \boldsymbol{\lambda}^\top(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) + \boldsymbol{\mu}^\top(\boldsymbol{C}\boldsymbol{x} - \boldsymbol{d}),$$

and the first condition in $(\star)$ takes the form

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \nabla f(\boldsymbol{x}^*) + \boldsymbol{A}^\top \boldsymbol{\lambda} + \boldsymbol{C}^\top \boldsymbol{\mu} = \boldsymbol{0}.$$

Example 1

Consider the problem

$$\min \ \frac{1}{2}(x^2 + y^2 + z^2) \quad \text{s.t. } x + y + z = 3.$$

Since the problem is convex, the KKT conditions are necessary and sufficient. The Lagrangian of the problem is

$$L(x, y, z, \mu) = \frac{1}{2}(x^2 + y^2 + z^2) + \mu(x + y + z - 3).$$

The KKT conditions and the feasibility condition are

$$\frac{\partial L}{\partial x} = x + \mu = 0, \quad \frac{\partial L}{\partial y} = y + \mu = 0,$$
$$\frac{\partial L}{\partial z} = z + \mu = 0, \quad x + y + z = 3.$$

We obtain $x = y = z = 1$ and $\mu = -1$. The unique optimal solution of the problem is $(x, y, z) = (1, 1, 1)$.

Example 2

Consider the problem

$$\min\ x^2 + 2y^2 + 4xy \quad \text{s.t. } x + y = 1, \quad x \geq 0, \quad y \geq 0.$$

The problem is nonconvex, since the matrix associated with the quadratic objective function $\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 2 & 2 \end{bmatrix}$ is indefinite. Thus, the KKT conditions are necessary optimality conditions. The Lagrangian of the problem is

$$L(x, y, \mu, \lambda_1, \lambda_2) = x^2 + 2y^2 + 4xy + \mu(x + y - 1) - \lambda_1 x - \lambda_2 y,$$

where $\lambda_1, \lambda_2 \in \mathbb{R}_+$, $\mu \in \mathbb{R}$. The KKT conditions with the feasibility conditions are

$$\frac{\partial L}{\partial x} = 2x + 4y + \mu - \lambda_1 = 0, \quad \frac{\partial L}{\partial y} = 4x + 4y + \mu - \lambda_2 = 0,$$
$$\lambda_1 x = 0, \quad \lambda_2 y = 0, \quad x + y = 1, \quad x \geq 0, \quad y \geq 0, \quad \lambda_1, \lambda_2 \geq 0.$$

Example 2 (cont'd)

Case 1: $\lambda_1 = \lambda_2 = 0$. In this case we obtain the three equations

$$2x + 4y + \mu = 0, \quad 4x + 4y + \mu = 0, \quad x + y = 1,$$

whose solution is $(x, y, \mu) = (0, 1, -4)$. $(x, y) = (0, 1)$ is a KKT point.

Case 2: $\lambda_1 > 0$, $\lambda_2 > 0$. By the complementary slackness conditions, we have $x = y = 0$, which contradicts the constraint $x + y = 1$.

Case 3: $\lambda_1 > 0$, $\lambda_2 = 0$. By the complementary slackness conditions, we have $x = 0 \Rightarrow y = 1$, which was already shown to be a KKT point.

Case 4: $\lambda_1 = 0$, $\lambda_2 > 0$. By the complementary slackness conditions, we have $y = 0 \Rightarrow x = 1 \Rightarrow 2 + \mu = 0$, $4 + \mu - \lambda_2 = 0 \Rightarrow \mu = -2$, $\lambda_2 = 2$. We thus obtain that $(x, y) = (1, 0)$ is also a KKT point.

Since the problem consists of minimizing a continuous function over a compact set it follows from the Weierstrass theorem that it has a global optimal solution. Since $f(1, 0) = 1$ and $f(0, 1) = 2$, $(x, y) = (1, 0)$ is the global optimal solution of the problem.

Orthogonal projection onto an affine space

Let C be the affine space $C := \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}$, where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$. We assume that the rows of $\boldsymbol{A}$ are linearly independent. Given $\boldsymbol{y} \in \mathbb{R}^n$, the optimization problem is

$$\min \ \|\boldsymbol{x} - \boldsymbol{y}\|^2 \quad \text{s.t. } \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}.$$

This is a convex optimization problem, so the KKT conditions are necessary and sufficient. The Lagrangian function is

$$\begin{aligned} L(\boldsymbol{x}, \boldsymbol{\lambda}) &= \|\boldsymbol{x} - \boldsymbol{y}\|^2 + (2\boldsymbol{\lambda})^\top(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) \qquad (\text{Note } \boldsymbol{\mu} := 2\boldsymbol{\lambda}) \\ &= \|\boldsymbol{x}\|^2 - 2(\boldsymbol{y} - \boldsymbol{A}^\top\boldsymbol{\lambda})^\top\boldsymbol{x} - 2\boldsymbol{\lambda}^\top\boldsymbol{b} + \|\boldsymbol{y}\|^2, \quad \boldsymbol{\lambda} \in \mathbb{R}^m. \end{aligned}$$

The KKT conditions are

$$\begin{aligned} & 2\boldsymbol{x} - 2(\boldsymbol{y} - \boldsymbol{A}^\top\boldsymbol{\lambda}) = \boldsymbol{0}, \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ \Longrightarrow\ & \boldsymbol{x} = \boldsymbol{y} - \boldsymbol{A}^\top\boldsymbol{\lambda} \ \Longrightarrow\ \boldsymbol{A}(\boldsymbol{y} - \boldsymbol{A}^\top\boldsymbol{\lambda}) = \boldsymbol{b} \\ \Longrightarrow\ & \boldsymbol{A}\boldsymbol{A}^\top\boldsymbol{\lambda} = \boldsymbol{A}\boldsymbol{y} - \boldsymbol{b} \ \Longrightarrow\ \boldsymbol{\lambda} = (\boldsymbol{A}\boldsymbol{A}^\top)^{-1}(\boldsymbol{A}\boldsymbol{y} - \boldsymbol{b}), \end{aligned}$$

$\boldsymbol{A}\boldsymbol{A}^\top$ is nonsingular since the rows of $\boldsymbol{A}$ are linearly independent. *We obtain the optimal solution:* $P_C(\boldsymbol{y}) = \boldsymbol{y} - \boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{A}^\top)^{-1}(\boldsymbol{A}\boldsymbol{y} - \boldsymbol{b})$.

Orthogonal projection onto hyperplanes

1. Consider the hyperplane $H = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{a}^\top \boldsymbol{x} = b\}$, where $\mathbf{0} \neq \boldsymbol{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Since a hyperplane is a spacial case of an affine space, from the last example, we obtain

$$P_H(\boldsymbol{y}) = \boldsymbol{y} - \boldsymbol{a}(\boldsymbol{a}^\top \boldsymbol{a})^{-1}(\boldsymbol{a}^\top \boldsymbol{y} - b) = \boldsymbol{y} - \frac{\boldsymbol{a}^\top \boldsymbol{y} - b}{\|\boldsymbol{a}\|^2}\boldsymbol{a}.$$

2. *(Distance of a point from a hyperplane)*
Let $H = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{a}^\top \boldsymbol{x} = b\}$, where $\mathbf{0} \neq \boldsymbol{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Then

$$d(\boldsymbol{y}, H) = \|\boldsymbol{y} - P_H(\boldsymbol{y})\| = \left\| \boldsymbol{y} - \left(\boldsymbol{y} - \frac{\boldsymbol{a}^\top \boldsymbol{y} - b}{\|\boldsymbol{a}\|^2}\boldsymbol{a} \right) \right\| = \frac{|\boldsymbol{a}^\top \boldsymbol{y} - b|}{\|\boldsymbol{a}\|}.$$

Orthogonal projection onto half-spaces

Let $H^- = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{a}^\top \boldsymbol{x} \le b\}$, where $\boldsymbol{0} \ne \boldsymbol{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Given $\boldsymbol{y} \in \mathbb{R}^n$, the corresponding optimization problem is

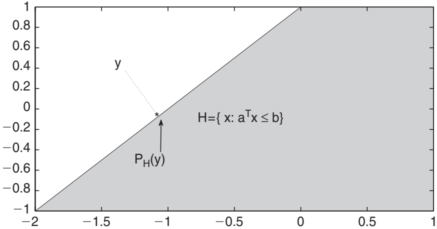
$$\min_{\boldsymbol{x}} \ \|\boldsymbol{x} - \boldsymbol{y}\|^2 \quad \text{s.t. } \boldsymbol{a}^\top \boldsymbol{x} \le b.$$

The Lagrangian of the problem is

$$L(\boldsymbol{x}, \lambda) = \|\boldsymbol{x} - \boldsymbol{y}\|^2 + (2\lambda)(\boldsymbol{a}^\top \boldsymbol{x} - b), \quad \lambda \ge 0,$$

and the KKT conditions with the feasibility condition are

$$2(\boldsymbol{x} - \boldsymbol{y}) + 2\lambda \boldsymbol{a} = \boldsymbol{0}, \quad \lambda(\boldsymbol{a}^\top \boldsymbol{x} - b) = 0, \quad \boldsymbol{a}^\top \boldsymbol{x} \le b, \quad \lambda \ge 0.$$



A vector y and its orthogonal projection onto a half-space

Orthogonal projection onto half-spaces (cont'd)

1. If $\lambda = 0$, then $\boldsymbol{x} = \boldsymbol{y}$ and the KKT conditions are satisfied when $\boldsymbol{a}^\top \boldsymbol{y} \le b$, i.e., $\boldsymbol{y} \in H^-$. Thus, the optimal solution is $P_{H^-}(\boldsymbol{y}) = \boldsymbol{y}$ if $\boldsymbol{y} \in H^-$.
2. Assume that $\lambda > 0$. By the complementary slackness condition we have $\boldsymbol{a}^\top \boldsymbol{x} = b$. Plugging the first equation $\boldsymbol{x} = \boldsymbol{y} - \lambda \boldsymbol{a}$ into $\boldsymbol{a}^\top \boldsymbol{x} = b$, we have

$$\boldsymbol{a}^\top (\boldsymbol{y} - \lambda \boldsymbol{a}) = b \implies \lambda = \frac{\boldsymbol{a}^\top \boldsymbol{y} - b}{\|\boldsymbol{a}\|^2} > 0, \text{ when } \boldsymbol{a}^\top \boldsymbol{y} > b.$$

   The optimal solution is

$$\boldsymbol{x} = \boldsymbol{y} - \frac{\boldsymbol{a}^\top \boldsymbol{y} - b}{\|\boldsymbol{a}\|^2} \boldsymbol{a}.$$

3. To summarize, we have

$$P_{H^-}(\boldsymbol{y}) = \begin{cases} \boldsymbol{y}, & \text{if } \boldsymbol{a}^\top \boldsymbol{y} \le b, \\ \boldsymbol{y} - \dfrac{\boldsymbol{a}^\top \boldsymbol{y} - b}{\|\boldsymbol{a}\|^2} \boldsymbol{a}, & \text{if } \boldsymbol{a}^\top \boldsymbol{y} > b. \end{cases}$$

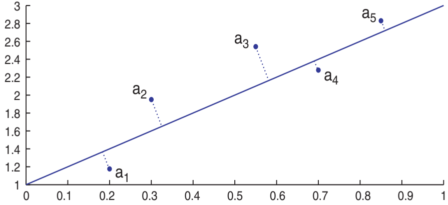
The orthogonal regression problem

Consider the points $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_m \in \mathbb{R}^n$. For a given $\mathbf{0} \neq \boldsymbol{x} \in \mathbb{R}^n$ and $y \in \mathbb{R}$, we define the hyperplane

$$H_{\boldsymbol{x},y} := \{\boldsymbol{a} \in \mathbb{R}^n : \boldsymbol{x}^\top \boldsymbol{a} = y\}.$$

In the orthogonal regression problem, we seek to find $\mathbf{0} \neq \boldsymbol{x} \in \mathbb{R}^n$ and $y \in \mathbb{R}$ such that the sum of squared Euclidean distances between the points $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_m$ to $H_{\boldsymbol{x},y}$ is minimal, i.e.,

$$\min_{\boldsymbol{x},y} \left\{ \sum_{i=1}^{m} d(\boldsymbol{a}_i, H_{\boldsymbol{x},y})^2 : \mathbf{0} \neq \boldsymbol{x} \in \mathbb{R}^n, y \in \mathbb{R} \right\}. \qquad (\spadesuit)$$



The orthogonal regression problem (cont'd)

Theorem: *Let $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_m \in \mathbb{R}^n$ and let $\boldsymbol{A} = [\boldsymbol{a}_1^\top, \boldsymbol{a}_2^\top, \cdots, \boldsymbol{a}_m^\top]^\top$. Then an optimal solution of problem (♠) is given by $\boldsymbol{x}$ that is an eigenvector of the matrix $\boldsymbol{A}^\top(\boldsymbol{I}_m - \frac{1}{m}\boldsymbol{e}\boldsymbol{e}^\top)\boldsymbol{A}$ associated with the minimum eigenvalue and $y = \frac{1}{m}\sum_{i=1}^m \boldsymbol{a}_i^\top \boldsymbol{x}$. Here $\boldsymbol{e}$ is the m-length vector of ones. The optimal function value of problem (♠) is $\lambda_{\min}[\boldsymbol{A}^\top(\boldsymbol{I}_m - \frac{1}{m}\boldsymbol{e}\boldsymbol{e}^\top)\boldsymbol{A}]$.*

Proof: From page 19 (the distance of a point from a hyperplane), the squared Euclidean distance between the point $\boldsymbol{a}_i$ to $H_{\boldsymbol{x},y}$ is given by

$$d(\boldsymbol{a}_i, H_{\boldsymbol{x},y})^2 = \frac{(\boldsymbol{a}_i^\top \boldsymbol{x} - y)^2}{\|\boldsymbol{x}\|^2}, \quad i = 1, 2, \cdots, m.$$

It follows that (♠) is the same as

$$\min_{\boldsymbol{x},y} \left\{ \sum_{i=1}^m \frac{(\boldsymbol{a}_i^\top \boldsymbol{x} - y)^2}{\|\boldsymbol{x}\|^2} : \boldsymbol{0} \neq \boldsymbol{x} \in \mathbb{R}^n, y \in \mathbb{R} \right\}.$$

Fixing $\boldsymbol{x}$ and minimizing first with respect to y we obtain that the optimal y is given by

$$y = \frac{1}{m}\sum_{i=1}^m \boldsymbol{a}_i^\top \boldsymbol{x} = \frac{1}{m}\boldsymbol{e}^\top \boldsymbol{A}\boldsymbol{x}.$$

The orthogonal regression problem (cont'd)

Using the latter expression for y we obtain that

$$\begin{aligned}\sum_{i=1}^{m}\left(\mathbf{a}_i^T\mathbf{x}-y\right)^2 &= \sum_{i=1}^{m}\left(\mathbf{a}_i^T\mathbf{x}-\frac{1}{m}\mathbf{e}^T\mathbf{A}\mathbf{x}\right)^2 \\ &= \sum_{i=1}^{m}(\mathbf{a}_i^T\mathbf{x})^2-\frac{2}{m}\sum_{i=1}^{m}(\mathbf{e}^T\mathbf{A}\mathbf{x})(\mathbf{a}_i^T\mathbf{x})+\frac{1}{m}(\mathbf{e}^T\mathbf{A}\mathbf{x})^2 \\ &= \sum_{i=1}^{m}(\mathbf{a}_i^T\mathbf{x})^2-\frac{1}{m}(\mathbf{e}^T\mathbf{A}\mathbf{x})^2 = \|\mathbf{A}\mathbf{x}\|^2-\frac{1}{m}(\mathbf{e}^T\mathbf{A}\mathbf{x})^2 \\ &= \mathbf{x}^T\mathbf{A}^T\left(\mathbf{I}_m-\frac{1}{m}\mathbf{e}\mathbf{e}^T\right)\mathbf{A}\mathbf{x}.\end{aligned}$$

Therefore, we arrive at the following reformulation of (♠) as a problem consisting of minimizing a Rayleigh quotient:

$$\min_{\boldsymbol{x}}\left\{\frac{\boldsymbol{x}^\top\left[\boldsymbol{A}^\top(\boldsymbol{I}_m-\frac{1}{m}\boldsymbol{e}\boldsymbol{e}^\top)\boldsymbol{A}\right]\boldsymbol{x}}{\|\boldsymbol{x}\|^2} : \boldsymbol{x}\neq\mathbf{0}\right\}.$$

Consequently, an optimal solution of the problem is an eigenvector of the matrix $\boldsymbol{A}^\top(\boldsymbol{I}_m-\frac{1}{m}\boldsymbol{e}\boldsymbol{e}^\top)\boldsymbol{A}$ corresponding to the minimum eigenvalue, and the optimal function value is the minimum eigenvalue $\lambda_{\min}[\boldsymbol{A}^\top(\boldsymbol{I}_m-\frac{1}{m}\boldsymbol{e}\boldsymbol{e}^\top)\boldsymbol{A}]$. □