Lecture 10: Optimality Conditions for Linearly Constrained Problems

Shi Pu

School of Data Science (SDS)
The Chinese University of Hong Kong, Shenzhen

Outline

Outline

- A hyperplane

$$H = \left\{ \mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} = b \right\} \quad (\mathbf{a} \in \mathbb{R}^n \setminus \{\mathbf{0}\}, b \in \mathbb{R})$$

is said to strictly separate a point $\mathbf{y} \notin S$ from S if

$$\mathbf{a}^\top \mathbf{y} > b$$

and

$$\mathbf{a}^\top \mathbf{x} \leq b \text{ for all } \mathbf{x} \in S.$$

Theorem (separation of a point from a closed and convex set)

Let $C \subset \mathbb{R}^n$ be a nonempty closed and convex set, and let $\mathbf{y} \notin C$. Then there exists $\mathbf{p} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ and $\alpha \in \mathbb{R}$ such that

$$\mathbf{p}^\top \mathbf{y} > \alpha \textit{ and } \mathbf{p}^\top \mathbf{x} \leq \alpha \textit{ for all } \mathbf{x} \in C.$$

Proof.

- By the second orthogonal projection theorem, the vector $\bar{\mathbf{x}} = P_C(\mathbf{y}) \in C$ satisfies

$$(\mathbf{y} - \bar{\mathbf{x}})^\top (\mathbf{x} - \bar{\mathbf{x}}) \leq 0 \text{ for all } \mathbf{x} \in C$$

which is the same as

$$(\mathbf{y} - \bar{\mathbf{x}})^\top \mathbf{x} \leq (\mathbf{y} - \bar{\mathbf{x}})^\top \bar{\mathbf{x}} \text{ for all } \mathbf{x} \in C$$

- Denote $\mathbf{p} = \mathbf{y} - \bar{\mathbf{x}} \neq \mathbf{0}$ and $\alpha = (\mathbf{y} - \bar{\mathbf{x}})^\top \bar{\mathbf{x}}$. Then

$$\mathbf{p}^\top \mathbf{x} \leq \alpha \text{ for all } \mathbf{x} \in C$$

- On the other hand,

$$\mathbf{p}^\top \mathbf{y} = (\mathbf{y} - \bar{\mathbf{x}})^\top \mathbf{y} = (\mathbf{y} - \bar{\mathbf{x}})^\top (\mathbf{y} - \bar{\mathbf{x}}) + (\mathbf{y} - \bar{\mathbf{x}})^\top \bar{\mathbf{x}} = \|\mathbf{y} - \bar{\mathbf{x}}\|^2 + \alpha > \alpha.$$

$\square$

Lemma (Farkas' Lemma)

Let $\mathbf{c} \in \mathbb{R}^n$ *and* $\mathbf{A} \in \mathbb{R}^{m \times n}$. *Then exactly one of the following systems has a solution:*

(A) $\mathbf{A}\mathbf{x} \leq \mathbf{0}, \ \ \mathbf{c}^\top \mathbf{x} > 0$

(B) $\mathbf{A}^\top \mathbf{y} = \mathbf{c}, \ \ \mathbf{y} \geq 0$

Another equivalent formulation is the following.

Lemma (Farkas' Lemma - second formulation)

Let $\mathbf{c} \in \mathbb{R}^n$ *and* $\mathbf{A} \in \mathbb{R}^{m \times n}$. *Then the following claims are equivalent:*

1. *The implication* $\mathbf{A}\mathbf{x} \leq \mathbf{0} \Rightarrow \mathbf{c}^\top \mathbf{x} \leq \mathbf{0}$ *holds true.*
2. *There exists* $\mathbf{y} \in \mathbb{R}^m_+$ *such that* $\mathbf{A}^\top \mathbf{y} = \mathbf{c}$.

What does it mean? Example.

$$\mathbf{A} = \begin{pmatrix} 1 & 5 \\ -1 & 2 \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} -1 \\ 9 \end{pmatrix}$$

Proof.

- Suppose that statement 2 holds: $\exists \mathbf{y} \in \mathbb{R}_+^m$ such that $\mathbf{A}^\top \mathbf{y} = \mathbf{c}$.
- To see that the implication 1 holds, suppose that $\mathbf{Ax} \leq \mathbf{0}$ for some $\mathbf{x} \in \mathbb{R}^n$.
- Multiplying this inequality from the left by $\mathbf{y}^\top$:

$$\mathbf{y}^\top \mathbf{Ax} \leq \mathbf{0}.$$

- Hence,

$$\mathbf{c}^\top \mathbf{x} \leq \mathbf{0}.$$

- Suppose that implication 1 is satisfied, and let us show that statement 2 holds. Suppose in contradiction that statement 2 is not true.

Proof Contd.

- Consider the following closed and convex (why?) set

$$S = \left\{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \mathbf{A}^\top \mathbf{y} \text{ for some } \mathbf{y} \in \mathbb{R}^m_+\right\}$$

- $\mathbf{c} \notin S$
- By the separation theorem $\exists \mathbf{p} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ and $\alpha \in \mathbb{R}$ such that $\mathbf{p}^\top \mathbf{c} > \alpha$ and

$$\mathbf{p}^\top \mathbf{x} \leq \alpha \text{ for all } \mathbf{x} \in S \tag{1}$$

- $\mathbf{0} \in S \Rightarrow \alpha \geq 0 \Rightarrow \mathbf{p}^\top \mathbf{c} > 0.$
- (1) is equivalent to $\mathbf{p}^\top \mathbf{A}^\top \mathbf{y} \leq \alpha$ for all $\mathbf{y} \geq \mathbf{0}$ or to $(\mathbf{A}\mathbf{p})^\top \mathbf{y} \leq \alpha$ for all $\mathbf{y} \geq \mathbf{0}$
- Therefore, $\mathbf{A}\mathbf{p} \leq \mathbf{0}$
- Contradiction to the assertion that implication 1 holds.

$\square$

Theorem

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$*. Then exactly one of the following systems has a solution*

(A) $\mathbf{Ax} < \mathbf{0}$

(B) $\mathbf{p} \neq 0, \ \mathbf{A}^\top \mathbf{p} = \mathbf{0}, \ \mathbf{p} \geq \mathbf{0}$

Theorem

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$. *Then exactly one of the following systems has a solution*

(A) $\mathbf{Ax} < \mathbf{0}$

(B) $\mathbf{p} \neq 0, \;\; \mathbf{A}^\top \mathbf{p} = \mathbf{0}, \;\; \mathbf{p} \geq \mathbf{0}$

Proof.

- Suppose that system (A) has a solution.
- Assume in contradiction that (B) is feasible: $\exists \mathbf{p} \neq \mathbf{0}$ satisfying $\mathbf{A}^\top \mathbf{p} = \mathbf{0}, \mathbf{p} \geq \mathbf{0}$.
- Multiplying the equality $\mathbf{A}^\top \mathbf{p} = \mathbf{0}$ from the left by $\mathbf{x}^\top$ yields $(\mathbf{Ax})^\top \mathbf{p} = 0$, which is an impossible equality.

Proof Contd.

- Now suppose that system (A) does not have a solution.
- System (A) is equivalent to (s is scalar) to $\mathbf{A}\mathbf{x} + s\mathbf{e} \leq 0, s > 0$, or to

$$\tilde{\mathbf{A}} \begin{pmatrix} \mathbf{x} \\ s \end{pmatrix} \leq 0, \; \mathbf{c}^\top \begin{pmatrix} \mathbf{x} \\ s \end{pmatrix} > 0$$

  where $\tilde{\mathbf{A}} = (\mathbf{A} \; \mathbf{e})$ and $\mathbf{c} = \mathbf{e}_{n+1}$.
- The infeasibility of (A) is thus equivalent to the infeasibility of system

$$\tilde{\mathbf{A}}\mathbf{w} \leq \mathbf{0}, \; \mathbf{c}^\top \mathbf{w} > 0, \; \mathbf{w} \in \mathbb{R}^{n+1}$$

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

$$\begin{pmatrix} \mathbf{A}^\top \\ \mathbf{e}^\top \end{pmatrix} \mathbf{z} = \mathbf{c}$$

- $\Leftrightarrow \; \exists \mathbf{z} \in \mathbb{R}^m_+ : \; \mathbf{A}^\top \mathbf{z} = \mathbf{0}, \; \mathbf{e}^\top \mathbf{z} = 1 \Leftrightarrow \; \exists \mathbf{0} \neq \mathbf{z} \in \mathbb{R}^m_+ : \; \mathbf{A}^\top \mathbf{z} = \mathbf{0}$
- $\Rightarrow$ System (B) is feasible. □

Outline

KKT Conditions for Linearly Constrained Problems

Theorem (KKT conditions for linearly constrained problems - necessary optimality conditions)

Consider the minimization problem

$$(P) \qquad \begin{array}{ll} \min & f(\mathbf{x}) \\ s.t. & \mathbf{a}_i^\top \mathbf{x} \leq b_i, \ i = 1, 2, \cdots, m \end{array}$$

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

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i \mathbf{a}_i = \mathbf{0} \tag{2}$$

and

$$\lambda_i(\mathbf{a}_i^\top \mathbf{x}^* - b_i) = 0, \quad i = 1, 2, \cdots, m \tag{3}$$

Proof.

- $\mathbf{x}^*$ is a local minimum $\Rightarrow$ $\mathbf{x}^*$ is a stationary point.
- $\nabla f(\mathbf{x}^*)^\top(\mathbf{x} - \mathbf{x}^*) \geq 0$ for every $\mathbf{x} \in \mathbb{R}^n$ satisfying $\mathbf{a}_i^\top \mathbf{x} \leq b_i$ for any $i = 1, 2, \cdots, m$.
- Denote the set of active constraints by

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

- Making the change of variables $\mathbf{y} = \mathbf{x} - \mathbf{x}^*$, we have $\nabla f(\mathbf{x}^*)^\top \mathbf{y} \geq 0$ for any $\mathbf{y} \in \mathbb{R}^m$ satisfying $\mathbf{a}_i^\top(\mathbf{y} + \mathbf{x}^*) \leq b_i, i = 1, 2, \cdots, m$, or $\nabla f(\mathbf{x}^*)^\top \mathbf{y} \geq 0$ for any $\mathbf{y}$ satisfying

$$\begin{aligned} \mathbf{a}_i^\top \mathbf{y} &\leq 0 \quad i \in I(\mathbf{x}^*) \\ \mathbf{a}_i^\top \mathbf{y} &\leq b_i - \mathbf{a}_i^\top \mathbf{x}^* \quad i \notin I(\mathbf{x}^*) \end{aligned}$$

Proof Contd.

- The second set of inequalities can be removed, that is, we will prove that
$$\mathbf{a}_i^\top \mathbf{y} \leq 0 \text{ for all } i \in I(\mathbf{x}^*) \Rightarrow \nabla f(\mathbf{x}^*)^\top \mathbf{y} \geq 0$$
- Suppose then that $\mathbf{y}$ satisfies $\mathbf{a}_i^\top \mathbf{y} \leq 0$ for all $i \in I(\mathbf{x}^*)$
- Since $b_i - \mathbf{a}_i^\top \mathbf{x}^* > 0$ for all $i \notin I(\mathbf{x}^*)$, it follows that there exists a small enough $\alpha > 0$ for which $\mathbf{a}_i^\top(\alpha \mathbf{y}) \leq b_i - \mathbf{a}_i^\top \mathbf{x}^*$.
- Thus, since in addition $\mathbf{a}_i^\top(\alpha \mathbf{y}) \leq 0$ for any $i \in I(\mathbf{x}^*)$, it follows by the stationarity condition that $\nabla f(\mathbf{x}^*)^\top \mathbf{y} \geq 0$.
- We have shown $\mathbf{a}_i^\top \mathbf{y} \leq 0$ for all $i \in I(\mathbf{x}^*) \Rightarrow \nabla f(\mathbf{x}^*)^\top \mathbf{y} \geq 0$.
- By Farkas' lemma, $\exists \lambda_i \geq 0, i \in I(\mathbf{x}^*)$ s.t. $-\nabla f(\mathbf{x}^*) = \sum_{i \in I(\mathbf{x}^*)} \lambda_i \mathbf{a}_i$.
- Defining $\lambda_i = 0$ for all $i \notin I(\mathbf{x}^*)$ we get that $\lambda_i(\mathbf{a}_i^\top \mathbf{x}^* - b_i) = 0$ for all $i \in \{1, 2, \cdots, m\}$ and $\nabla f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i \mathbf{a}_i = \mathbf{0}$.

$\square$

Theorem (KKT conditions for convex linearly constrained problems - necessary and sufficient optimality conditions)

Consider the minimization problem

$$(P) \qquad \begin{array}{ll} \min & f(\mathbf{x}) \\ s.t. & \mathbf{a}_i^\top \mathbf{x} \le b_i, \;\; i = 1, 2, \cdots, m \end{array}$$

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

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \mathbf{a}_i = \mathbf{0} \tag{4}$$

and

$$\lambda_i(\mathbf{a}_i^\top \mathbf{x}^* - b_i) = 0, \quad i = 1, 2, \cdots, m \tag{5}$$

Proof.

- Necessity was proven.
- Suppose that $\mathbf{x}^*$ is a feasible solution of (P) satisfying (4) and (5). Let $\mathbf{x}$ be a feasible solution of (P).
- Define the function

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

- $\nabla h(\mathbf{x}^*) = 0 \Rightarrow \mathbf{x}^*$ is a minimizer of h over $\mathbb{R}^n$.
- $$f(\mathbf{x}^*) = f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i(\mathbf{a}_i^\top \mathbf{x}^* - b_i) \le f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i(\mathbf{a}_i^\top \mathbf{x} - b_i) \le f(\mathbf{x})$$

□

Theorem (KKT conditions for linearly constrained problems)

Consider the minimization problem

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

where f is continuously differentiable, $\mathbf{a}_i, \mathbf{c}_j \in \mathbb{R}^n$, $b_i, d_j \in \mathbb{R}$.

(i) *(necessity of the KKT conditions) If $\mathbf{x}^*$ is a local minimum 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*

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \mathbf{a}_i + \sum_{j=1}^{p} \mu_j \mathbf{c}_j = \mathbf{0} \tag{6}$$

$$\lambda_i(\mathbf{a}_i^\top \mathbf{x}^* - b_i) = 0, \ i = 1, 2, \cdots, m \tag{7}$$

Theorem (KKT conditions for convex linearly constrained problems)

(ii) *(sufficiency in the convex case) If f is convex over $\mathbb{R}^n$ and $\mathbf{x}^*$ is a feasible solution of (Q) for which there exist $\lambda_1, \cdots, \lambda_m \geq 0$ and $\mu_1, \cdots, \mu_p \in \mathbb{R}$ such that (6) and (7) are satisfied, then $\mathbf{x}^*$ is an optimal solution of (Q).*

See Theorem 10.7 of the textbook.

Given the a problem

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

The associated Lagrangian function is

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

The KKT conditions can be written as

$$
\begin{aligned}
& \nabla_{\mathbf{x}} L(\mathbf{x}^*, \lambda, \mu) = \nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\mathbf{x}^*) + \sum_{j=1}^{p} \mu_j \nabla h_j(\mathbf{x}^*) = 0 \\
& \lambda_i g_i(\mathbf{x}^*) = 0,\ \ i = 1, 2, \cdots, m
\end{aligned}
$$

Outline

- $$\begin{aligned} \min \quad & \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) \\ \text{s.t.} \quad & x_1 + x_2 + x_3 = 3 \end{aligned}$$

In class.

Examples

- $$\begin{aligned} \min \quad & x_1^2 + 2x_2^2 + 4x_1x_2 \\ \text{s.t.} \quad & x_1 + x_2 = 1 \\ & x_1, x_2 \geq 0 \end{aligned}$$

In class.

Lemma

Let C be the affine space

$$C = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}\}$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Assume the rows of $\mathbf{A}$ are linearly independent. Then

$$P_C(\mathbf{y}) = \mathbf{y} - \mathbf{A}^\top (\mathbf{A}\mathbf{A}^\top)^{-1} (\mathbf{A}\mathbf{y} - \mathbf{b})$$

In class.

Orthogonal Projection onto Hyperplanes

Consider the hyperplane

$$H = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} = b\} \quad (\mathbf{0} \neq \mathbf{a} \in \mathbb{R}^n, b \in \mathbb{R})$$

Then by the previous slide:

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

Lemma (distance of a point from a hyperplane)

Let $H = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} = b\}$*, where* $\mathbf{0} \neq \mathbf{a} \in \mathbb{R}^n$ *and* $b \in \mathbb{R}$*. Then*

$$d(\mathbf{y}, H) = \frac{|\mathbf{a}^\top \mathbf{y} - b|}{\|\mathbf{a}\|}$$

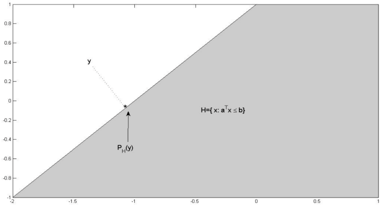
Proof.

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

Orthogonal Projection onto Half-Spaces

Let $H^{-} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} \leq b\}$, where $\mathbf{0} \neq \mathbf{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Then,

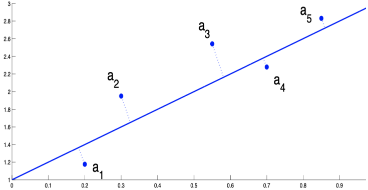
$$P_{H^-}(\mathbf{y}) = \mathbf{y} - \frac{[\mathbf{a}^\top \mathbf{y} - b]_+}{\|\mathbf{a}\|^2}\mathbf{a}$$



In class.

- $\mathbf{a}_1, \cdots, \mathbf{a}_m \in \mathbb{R}^n$
- For a given $\mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n$ and $y \in \mathbb{R}$, we define the hyperplane:

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



- In the orthogonal regression problem we seek to find a nonzero vector $\mathbf{x} \in \mathbb{R}^n$ and $y \in \mathbb{R}$ such that the sum of squared Euclidean distances between the points $\mathbf{a}_1, \cdots, \mathbf{a}_m$ to $H_{\mathbf{x},y}$ is minimal:

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

Orthogonal Regression

- $d(\mathbf{a}_i, H_{\mathbf{x},y})^2 = \frac{(\mathbf{a}_i^\top \mathbf{x} - y)^2}{\|\mathbf{x}\|^2}, \ i = 1, \cdots, m.$
- The Orthogonal Regression problem is the same as

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

- Fixing $\mathbf{x}$ and minimizing first with respect to y we obtain that the optimal y is given by $y = \frac{1}{m}\sum_{i=1}^m \mathbf{a}_i^\top \mathbf{x} = \frac{1}{m}\mathbf{e}^\top \mathbf{A}\mathbf{x}$, where $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m]^\top$.
- Using the above expression for y we obtain that

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

- Therefore, a reformulation of the problem is

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

Proposition

An optimal solution of the orthogonal regression problem is $(\mathbf{x}, y)$, *where* $\mathbf{x}$ *is an eigenvector of* $\mathbf{A}^\top(\mathbf{I}_m - \frac{1}{m}\mathbf{e}\mathbf{e}^\top)\mathbf{A}$ *associated with the minimum eigenvalue and* $y = \frac{1}{m}\sum_{i=1}^m \mathbf{a}_i^\top \mathbf{x}$. *The optimal function value of the problem is* $\lambda_{\min}[\mathbf{A}^\top(\mathbf{I}_m - \frac{1}{m}\mathbf{e}\mathbf{e}^\top)\mathbf{A}]$.

See Lemma 1.12 of the textbook.