Chapter 21 Problem with Inequality Constraints

1. Karush-Kuhn-Tucker Condition
2. Second-order Conditions

Constrained Optimization

standard form of constrained optimization

$$\min\ f(\boldsymbol{x}), \quad \text{s.t.}\ \ \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0},\ \ \boldsymbol{g}(\boldsymbol{x}) \le \mathbf{0}, \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$ $(m \le n)$, and $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^p$.

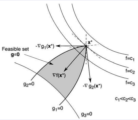
★ For inequality constraint, e.g., $g_i(\boldsymbol{x}) \ge 0$, it can be reformulated as $-g_i(\boldsymbol{x}) \le 0$.

Definition (active/inactive constraint)

An inequality constraint $g_j(\boldsymbol{x}) \le 0$ is said to be

active at $\boldsymbol{x}^*$ if $g_j(\boldsymbol{x}^*) = 0$; and inactive at $\boldsymbol{x}^*$ if $g_j(\boldsymbol{x}^*) < 0$.

Let $J(\boldsymbol{x}^*)$ be the indices of active inequality constraints at $\boldsymbol{x}^*$, i.e., $J(\boldsymbol{x}^*) := \{j \mid g_j(\boldsymbol{x}^*) = 0\}$.



Let $\boldsymbol{x} \in \mathbb{R}^2$, $g_j(\boldsymbol{x}) \le 0$ $(j = 1, 2, 3)$ be inequality constraints. By viewing the left figure, $\boldsymbol{x}^* \in \mathbb{R}^2$ is a minimizer. (why?)
Clearly, the inequality $g_3 \le 0$ is inactive, i.e., $g_3(\boldsymbol{x}^*) < 0$.

Definition (regular point)

Let $\boldsymbol{x}^*$ satisfy $\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}$, $\boldsymbol{g}(\boldsymbol{x}^*) \le \boldsymbol{0}$. $\boldsymbol{x}^* \in \mathbb{R}^n$ is a regular point if the vectors $\{\nabla h_i(\boldsymbol{x}^*)\}_{i=1}^m$, $\{\nabla g_j(\boldsymbol{x}^*)\}_{j\in J(\boldsymbol{x}^*)}$ are linearly independent.

Lemma (Farkas' Lemma)

Let cone $C := \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} = \sum_{j=1}^m \mu_j \boldsymbol{a}_j,\ \mu_j \ge 0,\ j = 1, \cdots, m\}$ *and its polar cone* $C^\circ := \left\{\boldsymbol{y} \in \mathbb{R}^n \mid \boldsymbol{a}_j^\top \boldsymbol{y} \le 0,\ j = 1, \cdots, m\right\}$. *Then,* $\boldsymbol{x}^\top \boldsymbol{y} \le 0$ *for all* $\boldsymbol{x} \in C$ *and* $\boldsymbol{y} \in C^\circ$.

proof. (trivial to check). Let $\boldsymbol{x}$ satisfy $\boldsymbol{x}^\top \boldsymbol{y} \le 0$ for all $\boldsymbol{y} \in C^\circ$, and consider its projection $\bar{\boldsymbol{x}}$ on C.

$$\therefore\ (\boldsymbol{x} - 0)^\top (\boldsymbol{x} - \bar{\boldsymbol{x}}) = \|\boldsymbol{x} - \bar{\boldsymbol{x}}\|^2, \quad (\boldsymbol{x} - \bar{\boldsymbol{x}})^\top \boldsymbol{a}_j \le 0, \quad \forall j = 1, \cdots, m.$$

$\therefore\ (\boldsymbol{x} - \bar{\boldsymbol{x}}) \in C^\circ$, and using the hypothesis,

$$\boldsymbol{x}^\top (\boldsymbol{x} - \bar{\boldsymbol{x}}) \le 0 \Longrightarrow \boldsymbol{x} = \bar{\boldsymbol{x}} \Longrightarrow \boldsymbol{x} \in C.$$

Example (linearly constrained optimization)

$$\min\ f(\boldsymbol{x}),\quad \text{s.t.}\ \ \boldsymbol{a}_j^\top \boldsymbol{x} \le b_j,\ j=1,\cdots,m. \tag{2}$$

If $\boldsymbol{x}^*$ is a local minimizer, then there exists $\boldsymbol{\mu}^* = (\mu_1^*,\cdots,\mu_m^*) \in \mathbb{R}_+^m$ such that $\nabla f(\boldsymbol{x}^*) + \sum_{j=1}^m \mu_j^* \boldsymbol{a}_j = 0,\ \ \mu_j^* = 0,\ \ \forall j \notin J(\boldsymbol{x}^*)$.

★ Property: no need for regularity.

Analysis: As a local minimizer $\boldsymbol{x}^*$ of (2) is also a local minimizer of

$$\min\ f(\boldsymbol{x}),\ \text{s.t.}\ \boldsymbol{a}_j^\top \boldsymbol{x} \le b_j,\ j \in J(\boldsymbol{x}^*).$$

$\therefore \xRightarrow{\text{VI form}} \nabla f(\boldsymbol{x}^*)^\top(\boldsymbol{x}-\boldsymbol{x}^*) \ge 0,\ \forall \boldsymbol{x}$ satisfying $\boldsymbol{a}_j^\top \boldsymbol{x} \le b_j,\ \ j \in J(\boldsymbol{x}^*)$.

$\because\ \boldsymbol{a}_j^\top \boldsymbol{x} \le b_j,\ j \in J(\boldsymbol{x}^*)$ can be written as $\boldsymbol{a}_j^\top(\boldsymbol{x}-\boldsymbol{x}^*) \le 0$, we have

$$\nabla f(\boldsymbol{x}^*)^\top \boldsymbol{y} \ge 0,\ \forall \boldsymbol{y} \text{ satisfying } \boldsymbol{a}_j^\top \boldsymbol{y} \le 0,\ \ j \in J(\boldsymbol{x}^*).$$

By Farkas' lemma, $-\nabla f(\boldsymbol{x}^*) = \sum_{j\in J(\boldsymbol{x}^*)} \mu_j^* \boldsymbol{a}_j$, for some $\mu_j^* \ge 0,\ j \in J(\boldsymbol{x}^*)$.

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

Karush-Kuhn-Tucker (KKT) Condition

Theorem (Karush-Kuhn-Tucker condition)

Let f, $\boldsymbol{h}$, $\boldsymbol{g} \in \mathcal{C}^1$ and let $\boldsymbol{x}^*$ be a regular point. If $\boldsymbol{x}^*$ is a local minimizer of problem (1), then there exist $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ and $\boldsymbol{\mu}^* \in \mathbb{R}^p$ such that

1. primal feasibility: $\boldsymbol{h}(\boldsymbol{x}^*) = \mathbf{0}$, $\boldsymbol{g}(\boldsymbol{x}^*) \leq \mathbf{0}$.
2. dual feasibility: $\boldsymbol{\mu}^* \geq 0$.
3. primal optimality: $Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*\top} D\boldsymbol{h}(\boldsymbol{x}^*) + \boldsymbol{\mu}^{*\top} D\boldsymbol{g}(\boldsymbol{x}^*) = \mathbf{0}^\top$.
4. complementary slackness: $\boldsymbol{\mu}^{*\top}\boldsymbol{g}(\boldsymbol{x}^*) = 0$.

★ $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ and $\boldsymbol{\mu}^* \in \mathbb{R}^p_+$ are called the multipliers associated with equality and inequality constraints, respectively.

★ By defining Lagrangian function $L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^p_+ \to \mathbb{R}$ by

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) := f(\boldsymbol{x}) + \boldsymbol{\lambda}^\top \boldsymbol{h}(\boldsymbol{x}) + \boldsymbol{\mu}^\top \boldsymbol{g}(\boldsymbol{x}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i h_i(\boldsymbol{x}) + \sum_{j=1}^{p} \mu_j g_j(\boldsymbol{x}).$$

★ primal optimality $\iff$ $\min_{\boldsymbol{x} \in \mathbb{R}^n} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$.

★ primal feasibility $\iff$ $\max_{\boldsymbol{\lambda} \in \mathbb{R}^m, \boldsymbol{\mu} \in \mathbb{R}^p_+} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$.

2D Illustration of KKT Condition

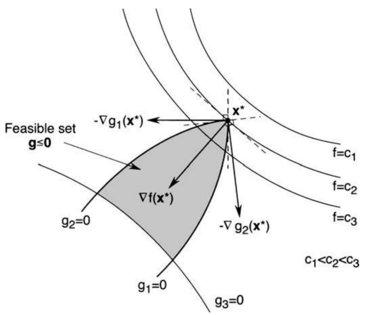
Discussion: $\because$ $\mu_j \geq 0$ and $g_j(\boldsymbol{x}^*) \leq 0$, the complementary slackness

$$\boldsymbol{\mu}^{*\top}\boldsymbol{g}(\boldsymbol{x}^*) = \mu_1^* g_1(\boldsymbol{x}^*) + \mu_2^* g_2(\boldsymbol{x}^*) + \cdots + \mu_p^* g_p(\boldsymbol{x}^*) = 0$$

implies that

- if $g_j(\boldsymbol{x}^*) < 0$, then its multiplier $\mu^* = 0$, i.e., $\mu_j^* = 0$ for all $j \notin J(\boldsymbol{x}^*)$;
- if $g_j(\boldsymbol{x}^*) = 0$, then its multiplier $\mu_j^* \geq 0$, i.e., $\mu_j^* \geq 0$ for all $j \in J(\boldsymbol{x}^*)$.

Let $\boldsymbol{x} \in \mathbb{R}^2$, $g_j(\boldsymbol{x}) \leq 0$ $(j = 1, 2, 3)$ be inequality constraints



By viewing the left figure,
$\boldsymbol{x}^* \in \mathbb{R}^2$ is a minimizer. (why?)
Clearly, the inequality $g_3 \leq 0$ is inactive, i.e., $g_3(\boldsymbol{x}^*) < 0$, $\therefore$ $\mu_3^* = 0$.
By the KKT conditon, we have
$\nabla f(\boldsymbol{x}^*) + \mu_1^* \nabla g_1(\boldsymbol{x}^*) + \mu_2^* \nabla g_2(\boldsymbol{x}^*) = \boldsymbol{0}$,
where $\mu_1^* > 0$ and $\mu_2^* > 0$.
$\therefore$ $\nabla f(\boldsymbol{x}^*)$ is the conic combination of $-\nabla g_1(\boldsymbol{x}^*)$ and $-\nabla g_2(\boldsymbol{x}^*)$ with positive coefficients.

$\because$ $\boldsymbol{x}^*$ is a regular local minimizer of f on $\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0},\ \ \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}\}$,

$\therefore$ $\boldsymbol{x}^*$ is also a regular local minimizer of f on the set

$$\{\boldsymbol{x} \mid \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0},\ \ g_j(\boldsymbol{x}) = 0,\ j \in J(\boldsymbol{x}^*)\}\ .$$

By Lagrangian theorem, there exists $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ and $\boldsymbol{\mu}^* \in \mathbb{R}^p$ such that

$$Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*\top} D\boldsymbol{h}(\boldsymbol{x}^*) + \boldsymbol{\mu}^{*\top} D\boldsymbol{g}(\boldsymbol{x}^*) = \boldsymbol{0}^\top,$$

where $\mu_j^* = 0$ for all $j \notin J(\boldsymbol{x}^*)$.

To complete the proof, it suffices to show that: $\mu_j^* \geq 0$ for all $j \in J(\boldsymbol{x}^*)$, which can be proven by contradiction.

Assume that there exists $j \in J(\boldsymbol{x}^*)$ such that $\mu_j^* < 0$. Let us denote

$$\hat{S} := \{\boldsymbol{x} \mid \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0},\ \ g_i(\boldsymbol{x}) = 0,\ \ i \in J(\boldsymbol{x}^*), i \neq j\}\,,$$

$$\hat{T}(\boldsymbol{x}^*) := \{\boldsymbol{y} \mid D\boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{y} = \boldsymbol{0},\ \ Dg_i(\boldsymbol{x}^*)\boldsymbol{y} = 0,\ \ i \in J(\boldsymbol{x}^*), i \neq j\}\,.$$

$\because$ $\boldsymbol{x}^*$ is a regular point,

$\therefore$ $\exists\ \boldsymbol{y} \in \hat{T}(\boldsymbol{x}^*)$ such that $Dg_j(\boldsymbol{x}^*)\boldsymbol{y} \neq 0$ (why?)

Indeed, for all $\boldsymbol{y} \in \hat{T}(\boldsymbol{x}^*)$,
if $Dg_j(\boldsymbol{x}^*)\boldsymbol{y} = \nabla g_j(\boldsymbol{x}^*)^\top \boldsymbol{y} = 0$, then $\nabla g_j(\boldsymbol{x}^*) \in \hat{T}(\boldsymbol{x}^*)^\perp$.

$\therefore$ $\nabla g_j(\boldsymbol{x}^*) \in \operatorname{span}\left[\nabla h_k(\boldsymbol{x}^*), k = 1, \ldots, m, \nabla g_i(\boldsymbol{x}^*), i \in J(\boldsymbol{x}^*), i \neq j\right].$

However, this contradicts the fact that $\boldsymbol{x}^*$ is a regular point.

Assume that there exists $\boldsymbol{y}$ such that $Dg_j(\boldsymbol{x}^*)\boldsymbol{y} < 0$.

The Lagrange condition is

$$Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*\top} D\boldsymbol{h}(\boldsymbol{x}^*) + \mu_j^* Dg_j(\boldsymbol{x}^*) + \sum_{i \neq j} \mu_i^* Dg_i(\boldsymbol{x}^*) = \mathbf{0}^\top.$$

By postmultiplying $\boldsymbol{y}$ and using the fact that $\boldsymbol{y} \in \hat{T}(\boldsymbol{x}^*)$, we have

$$Df(\boldsymbol{x}^*)\boldsymbol{y} = -\mu_j^* Dg_j(\boldsymbol{x}^*)\boldsymbol{y}$$

$\because$ $Dg_j(\boldsymbol{x}^*)\boldsymbol{y} < 0$ and $\mu_j^* < 0$, $\therefore$ $Df(\boldsymbol{x}^*)\boldsymbol{y} < 0$.

$\because$ $\boldsymbol{y} \in \hat{T}(\boldsymbol{x}^*)$,

By Theorem 20.1, we can find a differentiable curve $\{\boldsymbol{x}(t) \mid t \in (a, b)\}$ on S such that there exists $t^* \in (a, b)$ with $\boldsymbol{x}(t^*) = \boldsymbol{x}^*$ and $\dot{\boldsymbol{x}}(t^*) = \boldsymbol{y}$,

$$\frac{d}{dt} f(\boldsymbol{x}(t^*)) = Df(\boldsymbol{x}^*)\boldsymbol{y} < 0$$

This means that there exists $\delta > 0$ such that $\forall\ t \in (t^*, t^* + \delta]$,

$$f(\boldsymbol{x}(t)) < f(\boldsymbol{x}(t^*)) = f(\boldsymbol{x}^*).$$

$\because$ $\frac{d}{dt} g_j(\boldsymbol{x}(t^*)) = Dg_j(\boldsymbol{x}^*)\boldsymbol{y} < 0$

$\therefore$ $\exists \varepsilon > 0$ and $\forall\ t \in [t^*, t^* + \varepsilon]$, $g_j(\boldsymbol{x}(t)) \le 0$.

$\therefore$ $g_j(\boldsymbol{x}(t)) \le 0$, $f(\boldsymbol{x}(t)) < f(\boldsymbol{x}^*)$ for all $t \in (t^*,\ t^* + \min\{\delta, \varepsilon\})$.

This contradicts the assumption that $\boldsymbol{x}^*$ is a local minimizer.

Application of KKT Condition

$$\min \quad -\frac{400R}{(10+R)^2}, \quad \text{s.t.} \; -R \leq 0.$$

Lagrangian function $L : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ reads $L(R, \mu) = -\frac{400R}{(10+R)^2} + \mu(-R)$.

KKT condition (system):
$$\begin{cases} \text{primal feasibility:} \quad -R \leq 0. \\ \text{dual feasibility:} \;\; \mu \geq 0, \\ \text{primal optimality:} \quad -\frac{400(10-R)}{(10+R)^3} - \mu = 0, \\ \text{complementary slackness:} \;\; \mu R = 0, \end{cases} \qquad (*)$$

Solve KKT system (*) by two cases:

- if $\mu > 0$, then $R = 0$ by the complementary slackness, which contradicts the primal optimality.
- if $\mu = 0$, then $R = 10$ by primal optimality.

Overall, the only solution to KKT condition is $R = 10$ and $\mu = 0$.

Application of KKT Condition

$$\min \quad -\frac{4000}{(10+R)^2}, \quad \text{s.t.} \ -R \le 0.$$

Lagrangian function $L : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ reads $L(R, \mu) = -\frac{4000}{(10+R)^2} + \mu(-R)$.

KKT condition (system): $\begin{cases} \text{primal feasibility:} \quad -R \le 0. \\ \text{dual feasibility:} \ \ \mu \ge 0, \\ \text{primal optimality:} \ \ \frac{8000}{(10+R)^3} - \mu = 0, \\ \text{complementary slackness:} \ \ \mu R = 0, \end{cases} \qquad (**)$

Solve KKT system (**) by two cases:

1. if $\mu > 0$, then $R = 0$ by the complementary slackness. It follows by the primal optimality that $\mu = 8$, which also satisfies the dual feasibility.
2. if $\mu = 0$, this contradicts the primal optimality.

Overall, the only solution to KKT condition is $R = 0$ and $\mu = 8$.

Application of KKT Condition

$$\min \quad f(x_1,x_2)=x_1^2+x_2^2+x_1x_2-3x_1, \quad \text{s.t.} \quad x_1\ge 0,\ x_2\ge 0.$$

Lagrangian function $L:\mathbb{R}^2\times\mathbb{R}^2_+\to\mathbb{R}$ reads
$L(\boldsymbol{x},\boldsymbol{\mu})=x_1^2+x_2^2+x_1x_2-3x_1-\mu_1x_1-\mu_2x_2.$

KKT system:

$$\begin{cases}\text{primal feasibility: } x_1\ge 0,\ x_2\ge 0.\\ \text{dual feasibility: } \mu_1\ge 0,\ \mu_2\ge 0,\\ \text{primal optimality: } 2x_1+x_2-3-\mu_1=0,\ 2x_2+x_1-\mu_2=0,\\ \text{complementary slackness: } \mu_1x_1=0,\ \mu_2x_2=0,\end{cases} \qquad (*)$$

Solve KKT system (*) by four cases:

1. $\mu_1=0$ and $\mu_2=0$, KKT has no solution (why?).
2. $\mu_1=0$ and $\mu_2>0$, $\mu_1=0,\ \mu_2=\frac{3}{2},\ x_1=\frac{3}{2},\ x_2=0.$
3. $\mu_1>0$ and $\mu_2=0$, KKT has no solution (why?).
4. $\mu_1>0$ and $\mu_2>0$, KKT has no solution (why?).

Overall, the only solution to KKT condition is $\mu_1=0,\ \mu_2=\frac{3}{2},\ x_1=\frac{3}{2},\ x_2=0.$

Application of KKT Condition

$$\min \quad f(\boldsymbol{x}) = -\sum_{i=1}^n \log(x_i + \alpha_i), \quad \text{s.t.} \quad \boldsymbol{x} \geq 0, \ \mathbf{1}^\top \boldsymbol{x} = 1.$$

Lagrangian function $L : \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^n_+ \to \mathbb{R}$ reads

$$\begin{aligned} L(\boldsymbol{x}, \lambda, \boldsymbol{\mu}) &= f(\boldsymbol{x}) + \lambda(\mathbf{1}^\top \boldsymbol{x} - 1) - \boldsymbol{\mu}^\top \boldsymbol{x} \\ &= -\sum_{i=1}^n \log(x_i + \alpha_i) + \lambda(\sum_{i=1}^n x_i - 1) - \sum_{i=1}^n \mu_i x_i. \end{aligned}$$

$$\text{KKT system:} \quad \begin{cases} \text{primal feasibility:} \ \ \boldsymbol{x} \geq 0, \ \mathbf{1}^\top \boldsymbol{x} = 1 \\ \text{dual feasibility:} \ \ \boldsymbol{\mu} \geq 0 \\ \text{primal optimality:} \ \ -\frac{1}{x_i + \alpha_i} + \lambda - \mu_i = 0, \ \ \forall i = 1, \cdots, n \\ \text{complementary slackness:} \ \ \mu_i x_i = 0, \ \ \forall i = 1, \cdots, n \end{cases} \qquad (*)$$

Solve KKT system (*) by four cases:

1. if $\lambda < \frac{1}{\alpha_i}$, then $\mu_i = 0$ and $x_i = \frac{1}{\lambda} - \alpha_i$.
2. if $\lambda \geq \frac{1}{\alpha_i}$, then $\mu_i = \lambda - \frac{1}{\alpha_i}$ and $x_i = 0$.
3. determine λ by $1 = \mathbf{1}^\top \boldsymbol{x} = \sum_{i=1}^n \max\{0, \frac{1}{\lambda} - \alpha_i\}$.

Application of KKT Condition

Example (complementary problem)

$$\boxed{\min \quad f(\boldsymbol{x}), \quad \text{s. t.} \quad \boldsymbol{x} \geq \boldsymbol{0}} \xrightarrow{\text{KKT}} \boxed{\boldsymbol{0} \leq \boldsymbol{x} \perp \nabla f(\boldsymbol{x}) \geq \boldsymbol{0}}$$ (why?)

★ KKT is only necessary condition of minimizer, i.e., the solution of KKT is only a candidate of minimizer.

Example (some exercises)

$\min\ x_1 + x_2$, s.t. $2 - x_1^2 - x_2^2 \geq 0$,	$x^* = (-1, -1)$
$\min\ x_1 + x_2$, s.t. $2 - x_1^2 - x_2^2 \geq 0,\ x_2 \geq 0$	$x^* = (-\sqrt{2}, 0)$
$\min\ \frac{1}{2}(x_1^2 + x_2^2)$, s.t. $x_1 - 1 \leq 0$	$x^* = (0,0)$ and $\nu^* = 0$
$\min\ -x_1 x_2$, s.t. $x_1 + x_2^2 - 2 \leq 0, x_1 \geq 0, x_2 \geq 0$	$x^* = (\frac{4}{3}, \sqrt{\frac{2}{3}})$ and $x^* = (0,0)$
$\min\ \boldsymbol{c}^\top \boldsymbol{x}$, s.t. $(\boldsymbol{x} - \boldsymbol{a})^\top \boldsymbol{A} (\boldsymbol{x} - \boldsymbol{a}) \leq b$	$\boldsymbol{x}^* = \boldsymbol{a} - \frac{\boldsymbol{A}^{-1}\boldsymbol{c}}{2\nu}$ with $\nu = \sqrt{\frac{\boldsymbol{c}^\top \boldsymbol{A}^{-1} \boldsymbol{c}}{4b}}$

Second-Order Condition

Definition ($\min\ f(\boldsymbol{x}),\quad \text{s.t.}\quad \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0},\ \ \boldsymbol{g}(\boldsymbol{x}) \le \boldsymbol{0}$)

The Hessian of Lagrangian function with respect to $\boldsymbol{x}$

$$\boldsymbol{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \boldsymbol{F}(\boldsymbol{x}) + [\boldsymbol{\lambda}\boldsymbol{H}(\boldsymbol{x})] + [\boldsymbol{\mu}\boldsymbol{G}(\boldsymbol{x})],$$

where $\boldsymbol{F}(\boldsymbol{x})$ is the Hessian of f at $\boldsymbol{x}$,

$\boldsymbol{\lambda}\boldsymbol{H}(\boldsymbol{x}) = \lambda_1 \boldsymbol{H}_1(\boldsymbol{x}) + \cdots + \lambda_m \boldsymbol{H}_m(\boldsymbol{x})$ with $\boldsymbol{H}_i(\boldsymbol{x})$ as the Hessian of h_i at $\boldsymbol{x}$,
$\boldsymbol{\mu}\boldsymbol{G}(\boldsymbol{x}) = \mu_1 \boldsymbol{G}_1(\boldsymbol{x}) + \cdots + \mu_p \boldsymbol{G}_p(\boldsymbol{x})$ with $\boldsymbol{G}_i(\boldsymbol{x})$ as the Hessian of g_i at $\boldsymbol{x}$.

★ Let $\tilde{J}(\boldsymbol{x}^*, \boldsymbol{\mu}^*) = \{i \mid g_i(\boldsymbol{x}^*) = 0,\ \mu_i^* > 0\}$. The tangent spaces at $\boldsymbol{x}^*$ are

$$T(\boldsymbol{x}^*) = \{\boldsymbol{y} \in \mathbb{R}^n \mid D\boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{y} = \boldsymbol{0}, Dg_j(\boldsymbol{x}^*)\boldsymbol{y} = 0, j \in J(\boldsymbol{x}^*)\},$$
$$\tilde{T}(\boldsymbol{x}^*, \boldsymbol{\mu}^*) = \{\boldsymbol{y} \mid D\boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{y} = \boldsymbol{0}, Dg_i(\boldsymbol{x}^*)\boldsymbol{y} = 0, i \in \tilde{J}(\boldsymbol{x}^*, \boldsymbol{\mu}^*)\},$$

★ $\tilde{J}(\boldsymbol{x}^*, \boldsymbol{\mu}^*) \subset J(\boldsymbol{x}^*) \Longrightarrow T(\boldsymbol{x}^*) \subset \tilde{T}(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$.

Second-Order Conditions

Theorem (second-order necessary conditions, SONC)

Assume that f, $\boldsymbol{h}$ *and* $\boldsymbol{g} \in \mathcal{C}^2$ *for optimization* (1). *Let* $\boldsymbol{x}^*$ *be a local minimizer. If* $\boldsymbol{x}^*$ *is regular, then there exists* $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ *and* $\boldsymbol{\mu}^* \in \mathbb{R}^p$ *such that*

❶ $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_+^p$ *satisfies the KKT conditions.*

❷ $\boldsymbol{y}^\top \boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)\boldsymbol{y} \geq 0$ *for all* $\boldsymbol{y} \in T(\boldsymbol{x}^*)$.

proof. ❶ is the result of KKT condition.

❷ $\because$ $\boldsymbol{x}^*$ is a local minimizer over $\{\boldsymbol{x} \mid \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}\}$,

$\therefore$ it is also a local minimizer over $\{\boldsymbol{x} \mid \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, g_j(\boldsymbol{x}) = \boldsymbol{0}, j \in J(\boldsymbol{x}^*)\}$.

$\therefore$ the SONC for equality constraints can be applicable.

Theorem (second-order sufficient conditions, SOSC)

Assume that f, $\boldsymbol{h}$ *and* $\boldsymbol{g} \in \mathcal{C}^2$ *for optimization* (1). *If there exists a feasible point* $\boldsymbol{x}^* \in \mathbb{R}^n$ *and vectors* $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ *and* $\mu^* \in \mathbb{R}_+^p$ *such that*

❶$(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_+^p$ *satisfies the KKT conditions.*

❷$\boldsymbol{y}^\top \boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)\boldsymbol{y} > 0$ *for all* $\boldsymbol{0} \neq \boldsymbol{y} \in \tilde{T}(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$.

Then, $\boldsymbol{x}^*$ *is a strict local minimizer of* (1).

Example ($\min \quad x_1 x_2, \quad \mathrm{s.t.} \quad x_1 + x_2 \geq 2, \quad x_2 \geq x_1$)

(a) write the KKT condition of this problem.

(b) find all solutions of KKT condition and check whether the solution is regular.

(c) find all points in part (b) which also satisfy the SONC.

(d) find all points in part (c) which also satisfy the SOSC.

(e) find all points in part (c) which are local minimizers.

Ans: Lagrangian function $L : \mathbb{R}^2 \times \mathbb{R}^2_+ \to \mathbb{R}$ reads

$$L(\boldsymbol{x}, \boldsymbol{\mu}) = x_1 x_2 - \mu_1(x_1 + x_2 - 2) - \mu_2(x_2 - x_1).$$

KKT system:

$$\begin{cases} \text{primal feasibility: } x_1 + x_2 \geq 2, \;\; x_2 \geq x_1, \\ \text{dual feasibility: } \mu_1 \geq 0, \; \mu_2 \geq 0, \\ \text{primal optimality: } x_2 - \mu_1 + \mu_2 = 0, \; x_1 - \mu_1 - \mu_2 = 0, \\ \text{complementary slackness: } \mu_1(x_1 + x_2 - 2) = 0, \; \mu_2(x_2 - x_1) = 0. \end{cases} \quad (*)$$

Second-Order Conditions

Solve KKT system (*) by four cases:

1. $\mu_1 > 0,\ \mu_2 > 0$: ✗
2. $\mu_1 > 0,\ \mu_2 = 0$: ✓ $\Longrightarrow \mu_1^* = 1, \mu_2^* = 0,\ x_1^* = x_2^* = 1.$
3. $\mu_1 = 0,\ \mu_2 > 0$: ✗
4. $\mu_1 = 0,\ \mu_2 = 0$: ✗

$\because \begin{cases} x_1^* = 1 \\ x_2^* = 1 \end{cases} \Rightarrow \{g_i\}_{i=1}^2$ are active $\Rightarrow \begin{cases} Dg_1(\boldsymbol{x}^*) = [-1, -1] \\ Dg_2(\boldsymbol{x}^*) = [1, -1] \end{cases} \Rightarrow \boldsymbol{x}^*$ is regular.

$\therefore\ T(\boldsymbol{x}^*) = \{\boldsymbol{0}\}$, which implies that the SONC is satisfied (why?).

$\because\ \boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\mu}^*) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and

$\tilde{T}(\boldsymbol{x}^*, \tilde{\boldsymbol{\mu}}^*) = \{\boldsymbol{y} \mid [-1, -1]\boldsymbol{y} = 0\} = \{\boldsymbol{y} \mid y_1 = -y_2\}$.

By taking $\boldsymbol{y} = [1, -1]^\top \in \tilde{T}(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$, we have $\boldsymbol{y}^\top \boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\mu}^*)\boldsymbol{y} = -2 < 0$, which means that the SOSC fails.

- Indeed, by drawing the constraints and level sets of the objective function. Moving in the feasible direction $[1, 1]^\top$, the objective value increase; Moving in the feasible direction $[-1, 1]^\top$, the objective value decrease.

$\therefore\ \boldsymbol{x}^*$ is not a local minimizer.

Example

$$\min\ f(\boldsymbol{x}) = (x_1-1)^2 + x_2 - 2,$$
$$\text{s.t. } h(\boldsymbol{x}) = x_2 - x_1 - 1 = 0,\ g(\boldsymbol{x}) = x_1 + x_2 - 2 \le 0.$$

$\because$ $\nabla h(\boldsymbol{x})$ and $\nabla g(\boldsymbol{x})$ are linearly independent, $\therefore$ all feasible points are regular.

Lagrangian function $L : \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ reads

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

$$\text{KKT system: } \begin{cases} \text{primal feasibility: } x_2 - x_1 - 1 = 0,\ \ x_1 + x_2 - 2 \le 0, \\ \text{dual feasibility: } \mu \ge 0, \\ \text{primal optimality: } [2x_1 - 2 - \lambda + \mu, 1 + \lambda + \mu] = \mathbf{0}^\top, \\ \text{complementary slackness: } \mu(x_1 + x_2 - 2) = 0. \end{cases} \qquad (*)$$

Solve KKT system (*) by two cases:

1. If $\mu > 0$, $(*) \iff \begin{cases} 2x_1 - 2 - \lambda + \mu = 0 \\ 1 + \lambda + \mu = 0 \\ x_2 - x_1 - 1 = 0 \\ x_1 + x_2 - 2 = 0 \end{cases} \Longrightarrow \begin{cases} x_1 = \frac{1}{2} \\ x_2 = \frac{3}{2} \\ \lambda = -1 \\ \mu = 0 \to ✗ \end{cases}$

2. If $\mu=0$, $(*)$ $\iff \begin{cases} 2x_1-2-\lambda=0 \\ 1+\lambda=0 \\ x_2-x_1-1=0 \\ x_1+x_2-2\le 0 \end{cases} \Rightarrow \begin{cases} x_1=\frac{1}{2} \\ x_2=\frac{3}{2} \\ \lambda=-1 \\ \mu=0 \end{cases} \Rightarrow$ $\therefore$ $\boldsymbol{x}^*$ is a solution of KKT, also a candidate of minimizer.

We now verify $\boldsymbol{x}^*=[\frac{1}{2},\frac{3}{2}]^\top$, $\lambda^*=-1$ and $\mu^*=0$ satisfies the SOSC.

$$\because\ \boldsymbol{L}(\boldsymbol{x}^*,\lambda^*,\mu^*)=\boldsymbol{F}(\boldsymbol{x}^*)+\lambda^*\boldsymbol{H}(\boldsymbol{x}^*)+\mu^*\boldsymbol{G}(\boldsymbol{x}^*)$$
$$=\begin{bmatrix}2&0\\0&0\end{bmatrix}+(-1)\begin{bmatrix}0&0\\0&0\end{bmatrix}+(0)\begin{bmatrix}0&0\\0&0\end{bmatrix}=\begin{bmatrix}2&0\\0&0\end{bmatrix},$$

and $\tilde{T}(\boldsymbol{x}^*,\mu^*)=\{\boldsymbol{y}:Dh(\boldsymbol{x}^*)\boldsymbol{y}=0\}=\left\{[a,a]^\top : a\in\mathbb{R}\right\}$,

$$\therefore\ \boldsymbol{y}^\top\boldsymbol{L}(\boldsymbol{x}^*,\lambda^*,\mu^*)\boldsymbol{y}=[a,a]\begin{bmatrix}2&0\\0&0\end{bmatrix}\begin{bmatrix}a\\a\end{bmatrix}=2a^2.$$

By the SOSC$\Longrightarrow$ $\boldsymbol{x}^*=[\frac{1}{2},\frac{3}{2}]^\top$ is a strict local minimizer.

Quadratic Constrained Quadratic Program

Definition (quadratic constrained quadratic program, QCQP)

$\min\ \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c}^\top \boldsymbol{x} + r$, s.t. $\boldsymbol{x}^\top \boldsymbol{P}_i \boldsymbol{x} + \boldsymbol{q}_i^\top \boldsymbol{x} + r_i \le 0,\ i = 1, \ldots, p$
where $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$, $\boldsymbol{c} \in \mathbb{R}^n$, $\boldsymbol{P}_i \in \mathbb{R}^{n\times n}$, $\boldsymbol{q}_i \in \mathbb{R}^n$,.

properties of QCQP:

QCQP $\iff$ minimize/maximize quadratic function on quadratic constraints.

- if $\boldsymbol{Q} \succeq 0$ and $\boldsymbol{P}_i \succeq 0$, then QCQP is convex optimization.
- if $\boldsymbol{P}_i = 0$, then QCQP reduces to QP.
- if $\boldsymbol{Q} = 0$ and $\boldsymbol{P}_i = 0$, then QCQP reduces to LP.

Second-order Cone Program

Definition (2nd-order cone program, SOCP)

$\min\ \boldsymbol{c}^\top \boldsymbol{x},\ $ s.t. $\|\boldsymbol{A}_i \boldsymbol{x} + \boldsymbol{b}_i\|_2 \leq \boldsymbol{u}_i^\top \boldsymbol{x} + v_i,\ i = 1, \ldots, p$
where $\boldsymbol{c} \in \mathbb{R}^n$, $\boldsymbol{A}_i \in \mathbb{R}^{n_i \times n}$, $\boldsymbol{b}_i \in \mathbb{R}^{n_i}$, $\boldsymbol{u}_i \in \mathbb{R}^n$, $v_i \in \mathbb{R}$.

properties of SOCP:

SOCP $\Longleftrightarrow$ minimize/maximize objective function on second-order cone constraints (a generalization of QCQP and LP).

- constraint is equivalent to $[\boldsymbol{A}_i \boldsymbol{x} + \boldsymbol{b}_i, \boldsymbol{u}_i^\top \boldsymbol{x} + v_i] \in$ second-order cone in $\mathbb{R}^{n_i+1}$.
- if $n_i = 0$, then SOCP reduces to LP; if $\boldsymbol{u}_i = 0$, then SOCP reduces to QCQP.

Remark: Optimality condition I

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable.

FONC: first order necessary condition

SONC: second order necessary condition

SOSC: second order sufficient condition

		$\min\limits_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x})$	$\min\limits_{\boldsymbol{x}\in\Omega} f(\boldsymbol{x})$
$\boldsymbol{x}^*$ is local min	$\xRightarrow{\text{FONC}}$	$\nabla f(\boldsymbol{x}^*) = \mathbf{0}$	$\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) \geq 0, \forall \boldsymbol{d} \in T(\boldsymbol{x}^*)$
	$\xRightarrow{\text{SONC}}$	$\nabla^2 f(\boldsymbol{x}^*) \succeq \mathbf{0}$	$\boldsymbol{d}^\top \nabla^2 f(\boldsymbol{x}^*)\boldsymbol{d} \geq 0, \forall \boldsymbol{d} \in T(\boldsymbol{x}^*)$
	$\xLeftarrow{\text{SOSC}}$	$\begin{cases} \nabla f(\boldsymbol{x}^*) = \mathbf{0} \\ \nabla^2 f(\boldsymbol{x}^*) \succ \mathbf{0} \end{cases}$	$\begin{cases} \boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) \geq 0, \\ \boldsymbol{d}^\top \nabla^2 f(\boldsymbol{x}^*)\boldsymbol{d} > 0, \forall \boldsymbol{d} \in T(\boldsymbol{x}^*) \end{cases}$

★ if f is convex function and Ω is convex set, then $\xRightarrow{\text{FONC}}$ can be $\xLeftrightarrow{\text{FONS}}$

Remark: Optimality condition II

Let $f:\mathbb{R}^n \to \mathbb{R}$ be differentiable.

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

$\boldsymbol{L}(\boldsymbol{x},\boldsymbol{\lambda}) = \nabla^2_{xx} l(\boldsymbol{x},\boldsymbol{\lambda})$

<table>
<tr><td rowspan="4">$\boldsymbol{x}^*$ is local min</td><td></td><td>$\min f(\boldsymbol{x})$, s.t. $\boldsymbol{x}\in\Omega=\{\boldsymbol{x}\in\mathbb{R}^n \mid h_i(\boldsymbol{x})=0, i=1,2,\cdots,m\}$</td></tr>
<tr><td>$\xRightarrow{\text{FONC}}$</td><td>$\begin{cases}\nabla f(\boldsymbol{x}^*) = \sum_{i=1}^m \lambda_i \nabla h_i(\boldsymbol{x}^*) \\ h_i(\boldsymbol{x}^*) = 0,\ i=1,\cdots,m\end{cases}$</td></tr>
<tr><td>$\xRightarrow{\text{SONC}}$</td><td>$\begin{cases}\nabla f(\boldsymbol{x}^*) = \sum_{i=1}^m \lambda_i \nabla h_i(\boldsymbol{x}^*) \\ h_i(\boldsymbol{x}^*) = 0, i=1,\cdots,m\end{cases}$
and $\boldsymbol{d}^\top \boldsymbol{L}(\boldsymbol{x}^*,\boldsymbol{\lambda})\boldsymbol{d} \geq 0,\ \forall \boldsymbol{d}\in T(\boldsymbol{x}^*)$</td></tr>
<tr><td>$\xLeftarrow{\text{SOSC}}$</td><td>$\begin{cases}\nabla f(\boldsymbol{x}^*) = \sum_{i=1}^m \lambda_i \nabla h_i(\boldsymbol{x}^*) \\ h_i(\boldsymbol{x}^*) = 0, i=1,\cdots,m\end{cases}$
and $\boldsymbol{d}^\top \boldsymbol{L}(\boldsymbol{x}^*,\boldsymbol{\lambda})\boldsymbol{d} > 0,\ \forall \boldsymbol{d}\in T(\boldsymbol{x}^*)$</td></tr>
<tr><td colspan="3">★ if f is convex function and Ω is convex set, then $\xRightarrow{\text{FONC}}$ can be $\xLeftrightarrow{\text{FONS}}$
i.e., $h_i(\boldsymbol{x}) = a_i^\top \boldsymbol{x} - b_i$ are linear functions</td></tr>
</table>

Let $f:\mathbb{R}^n \to \mathbb{R}$ be differentiable.

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

$\boldsymbol{L}(\boldsymbol{x},\boldsymbol{\lambda}) = \nabla^2_{xx} l(\boldsymbol{x},\boldsymbol{\lambda})$

<table>
<tr><td colspan="3">$\min f(\boldsymbol{x})$,
s.t. $\boldsymbol{x} \in \Omega = \{\boldsymbol{x} \in \mathbb{R}^n \mid h_i(\boldsymbol{x}) = 0, i = 1, \ldots, m;\ g_j(\boldsymbol{x}) \leqslant 0, j = 1, \ldots, p$</td></tr>
<tr><td rowspan="3">$\boldsymbol{x}^*$ is local min</td><td>$\overset{\text{FONC}}{\Longrightarrow}$</td><td>$\begin{cases} h_i(\boldsymbol{x}^*) = 0, i = 1, \cdots, m; g_j(\boldsymbol{x}^*) \le 0, j = 1, \cdots, p \\ \mu_j \ge 0, j = 1, \cdots, p \\ \nabla f(\boldsymbol{x}^*) \pm \sum_{i=1}^m \lambda_i \nabla h_i(\boldsymbol{x}^*) + \sum_{j=1}^p \mu_j \nabla g_j(\boldsymbol{x}^*) = 0 \\ \mu_j g_j(\boldsymbol{x}^*) = 0, j = 1, \cdots, p \end{cases}$</td></tr>
<tr><td>$\overset{\text{SONC}}{\Longrightarrow}$</td><td>FONC and $\boldsymbol{d}^\top \boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}) \boldsymbol{d} \ge 0, \forall \boldsymbol{d} \in T(\boldsymbol{x}^*)$</td></tr>
<tr><td>$\overset{\text{SOSC}}{\Longleftarrow}$</td><td>FONC and $\boldsymbol{d}^\top \boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}) \boldsymbol{d} > 0,\ \forall \boldsymbol{d} \in \tilde{T}(\boldsymbol{x}^*)$</td></tr>
<tr><td colspan="3">★ if f and g_i are convex and Ω is convex set, then $\overset{\text{FONC}}{\Longrightarrow}$ can be $\overset{\text{FONS}}{\Longleftrightarrow}$
i.e., $h_i(\boldsymbol{x}) = a_i^\top \boldsymbol{x} - b_i$ are linear functions</td></tr>
</table>

Homework

Textbook: 21.1, 21.6, 21.24

Example (some exercises)

min $x_1 + x_2$, s.t. $x_1^2 + x_2^2 - 2 = 0$	$x^* = (-1, -1)$
min $2(x_1^2 + x_2^2 - 1) - x_1$, s.t. $x_1^2 + x_2^2 = 1$	$x^* = (1, 0)$
min $x_1 + x_2$, s.t. $x_1^2 + x_2^2 = 2,\ 0 \leq x_1 \leq 1,\ 0 \leq x_2 \leq 1$	
min $x_1 + x_2$, s.t. $x_1^2 + x_2^2 \leq 1,\ x_1 + x_2 = 3$	
min $x_1 x_2$, s.t. $x_1 + x_2 = 2$	