# 制約つき問題に対する最適性の条件

山下信雄

# 講義内容

1. 制約つき問題に対する最適性の条件
   - 1次の必要条件：Karush-Kuhn-Tucker条件
   - 制約想定
   - 2次の条件

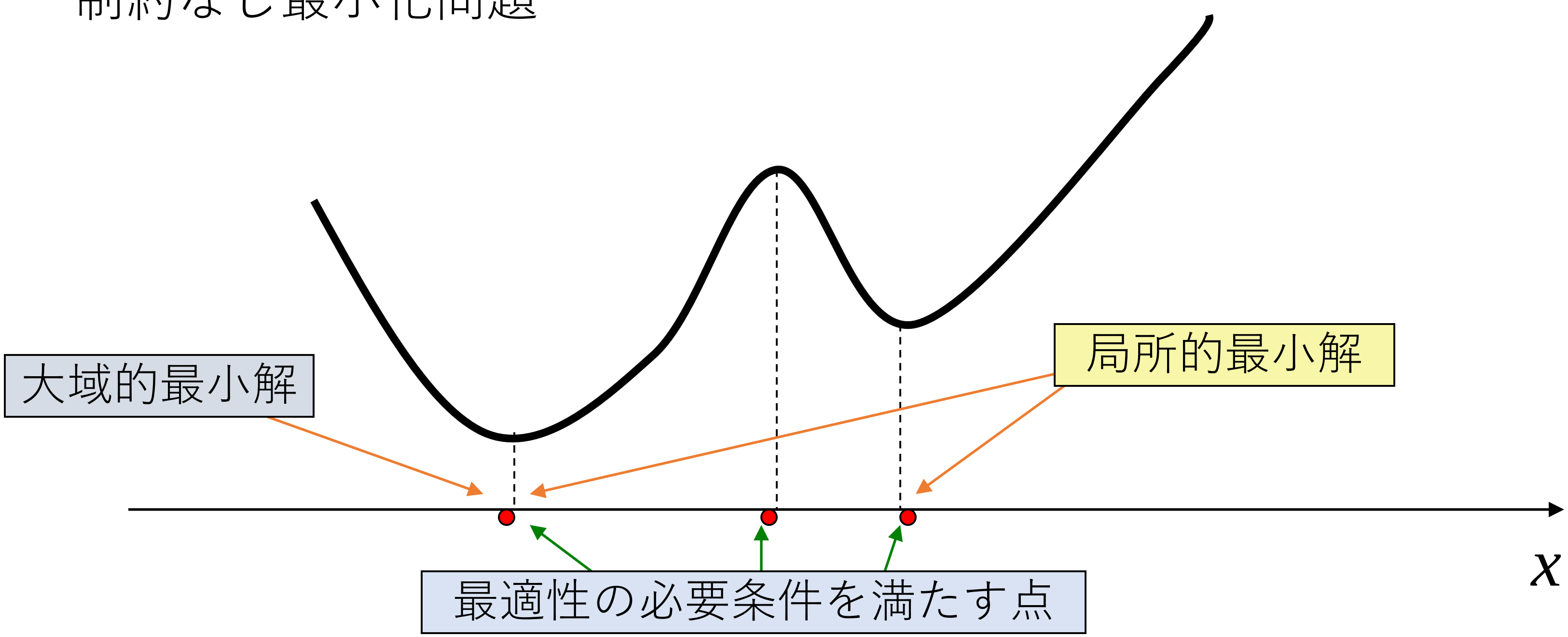
2. 凸最適化におけるKarush-Kuhn-Tucker条件

3. 1次の必要条件の証明

# 最適性の条件

１点の微分や目的関数値などの情報から、その点が最小解であるかどうかを判別する条件　（例：$f'(x) = 0$ ）

制約なし最小化問題



# 制約つき問題の最適性の条件

考える問題：

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & c_i(x) = 0,\ i = 1, \ldots, \ell \\ & c_i(x) \le 0,\ i = \ell + 1, \ldots, m \end{aligned}$$

有効集合(active set)：

$$A(x) = \{ i \in \{\ell + 1, \ldots, m\} \mid c_i(x) = 0 \}$$

# カルーシュ・キューン・タッカー条件（Karush-Kuhn-Tucker条件，KKT条件）

$$\nabla f(x^*) + \sum_{i=1}^{m} \mu_i^* \nabla c_i(x^*) = 0 \tag{1}$$

$$c_i(x^*) = 0 \quad (i = 1, \ldots, \ell) \tag{2}$$

$$c_i(x^*) \le 0, \ \mu_i^* \ge 0, \ \mu_i^* c_i(x^*) = 0 \quad (i = \ell + 1, \ldots, m) \tag{3}$$

- $\mu^*$ を**ラグランジュ乗数**という。
- KKT条件をみたす $(x^*, \mu^*)$ を**KKT点**という。
- 条件（３）を**相補性条件**という。

# KKT条件の意味(1)

**制約なしのとき**：

$$\nabla f(x^*) = 0$$

**等式制約のみのとき**：

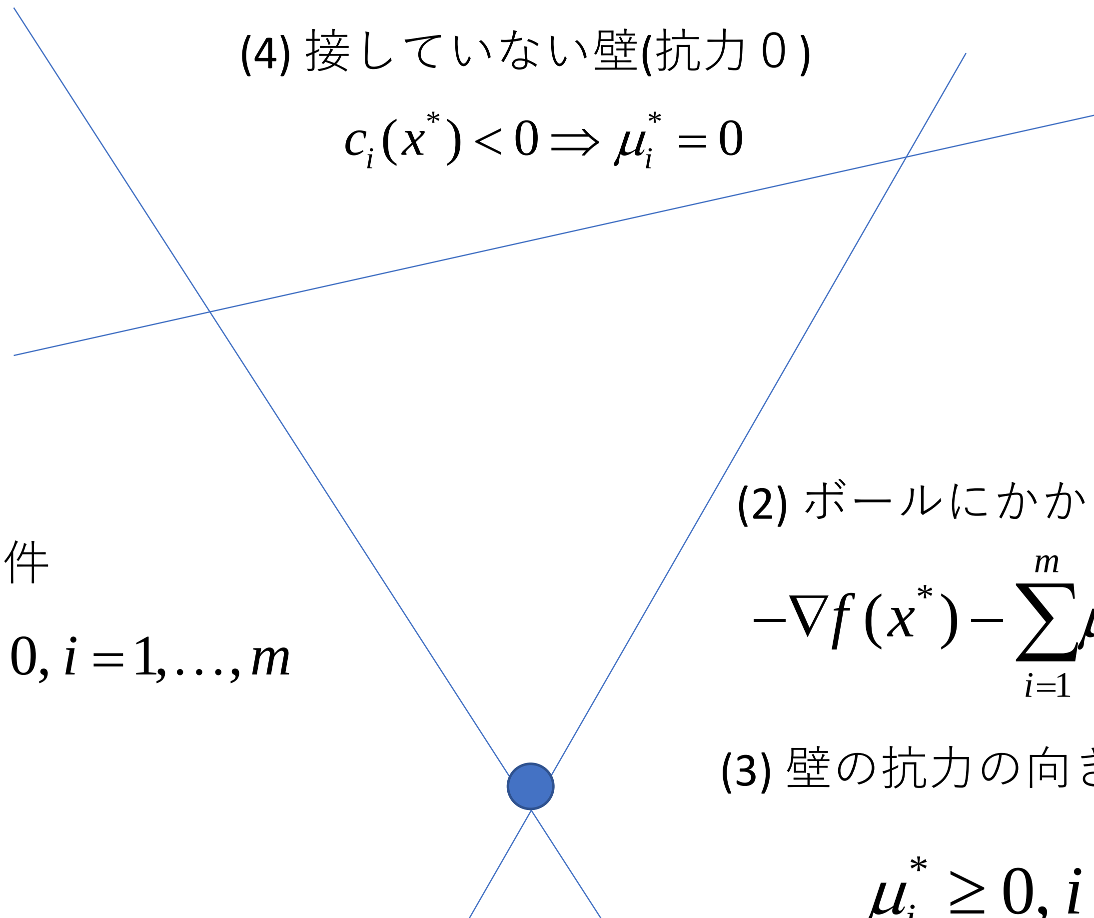
ラグランジュ関数： $L(x, \mu) = f(x) + \sum_{i=1}^{m} \mu_i c_i(x)$

$$\nabla_x L(x^*, \mu^*) = \nabla f(x^*) + \sum_{i=1}^{m} \mu_i^* \nabla c_i(x^*) = 0$$

$$\nabla_\mu L(x^*, \mu^*) = (c_1(x^*), \ldots, c_m(x^*))^\top = 0$$

$\Rightarrow$ **ラグランジュの未定乗数法**

# KKT条件の意味(2)



(4) 接していない壁(抗力 0 )

$$c_i(x^*) < 0 \Rightarrow \mu_i^* = 0$$

(2) ボールにかかる力の均衡式

$$-\nabla f(x^*) - \sum_{i=1}^{m} \mu_i^* \nabla c_i(x^*) = 0$$

(1) 壁内の条件

$$c_i(x^*) \leq 0, i = 1, \ldots, m$$

(3) 壁の抗力の向きの条件

$$\mu_i^* \geq 0, i = 1, \ldots, m$$

# 最適性の必要条件

定理１
- $x^*$を局所的最小解とする.
- 制約想定が成り立つとする.

このとき，KKT条件を満たすラグランジュ乗数が存在する.

# 制約想定（Constraint Qualification）
# 実行可能領域と制約関数のよい性質

**線形制約想定**：

すべての制約関数　$c_i \ (i=1,\ldots,m)$　は 1 次関数.

**スレイターの制約想定**：

$c_i \ (i=\ell+1,\ldots,m)$ は凸関数で, $c_i \ (i=1,\ldots,\ell)$ は 1 次関数.

$c_i(x^0)=0 \ (i=1,\ldots,\ell), \quad c_i(x^0)<0 \ (i=\ell+1,\ldots,m)$

となる $x^0$　が存在する.

**１次独立の制約想定（LICQ）**：

$\nabla c_i(x^*) \ \ i \in \{1,2,\ldots,\ell\} \cup A(x^*)$

が 1 次独立

有効集合：

$A(x)=\{i \in \{\ell+1,\ldots,m\} \mid c_i(x)=0\}$

# 制約想定の必要性について

KKT条件が成り立たない例

$$\min \; x_2$$

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

$$(x_1+1)^2 + x_2^{\,2} \le 1$$

実行可能領域 $S = \left\{(0,0)^\top\right\}$

最適解 $x^* = (0,0)^\top$

$$\nabla f(x^*) = \begin{pmatrix} 0 \\ 1 \end{pmatrix},\ \nabla c_1(x^*) = \begin{pmatrix} 2x_1^* - 2 \\ 2x_2^* \end{pmatrix} = \begin{pmatrix} -2 \\ 0 \end{pmatrix},\ \nabla c_1(x^*) = \begin{pmatrix} 2x_1^* + 2 \\ 2x_2^* \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

$\Rightarrow$ KKT条件 $:\nabla f(x^*) + \mu_1^* \nabla c_1(x^*) + \mu_2^* \nabla c_2(x^*) = \begin{pmatrix} -2\mu_1^* + 2\mu_2^* \\ 1 \end{pmatrix} \ne \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

また，スレイターの制約想定もLICQも成り立たない．

# 最適性の2次の必要条件

ラグランジュ関数： $L(x,\mu)=f(x)+\sum_{i=1}^{m}\mu_i c_i(x)$

$x$ に関するヘッセ行列： $\nabla_{xx}^2 L(x,\mu)=\nabla^2 f(x)+\sum_{i=1}^{m}\mu_i\nabla^2 c_i(x)$

接平面： $M(x^*)=\left\{y\in R^n \,\middle|\, \nabla c_i(x^*)^\top y=0\ i\in\{1,\ldots,\ell\}\cup A(x^*)\right\}$

定理 (2次の必要条件)
$x^*$ は局所最適解であり，制約想定が成り立つとする.
このときKKT点 $(x^*,\mu^*)$ が存在し，次の不等式が成り立つ.

$$y^\top\nabla_{xx}^2 L(x^*,\mu^*)y\geq 0\quad \forall y\in M(x^*)$$

# 最適性の2次の十分条件

定理 (2次の十分条件)

$(x^*, \mu^*)$ をKKT点とする．さらに，

$$y^\top \nabla_{xx}^2 L(x^*, \mu^*) y > 0 \quad \forall y \in M(x^*) \text{ such that } y \neq 0$$

が成り立つとき， $x^*$ は局所的最適解である．

# 講義内容

1. 制約つき問題に対する最適性の条件
   - 1次の必要条件：Karush-Kuhn-Tucker条件
   - 制約想定
   - 2次の条件
2. 凸最適化におけるKarush-Kuhn-Tucker条件
3. 一次の必要条件の証明

# 凸最適化問題に対する十分性

**定理 2**

$f, c_i\ (i=\ell+1,\dots m)$ を凸関数，$c_i, i=1,\dots,\ell$ を 1 次関数とする．
$(x^*, \mu^*)$ を

$$\begin{aligned} \min\ & f(x) \\ \text{s.t.}\ & c_i(x)=0,\ i=1,\dots,\ell \\ & c_i(x)\le 0,\ i=\ell+1,\dots,m \end{aligned}$$

のKKT点とする．
このとき，$x^*$ は問題の大域的最小解である．

# 定理２の証明

$$\nabla f(x^*) + \sum_{i=1}^{m} \mu_i^* \nabla c_i(x^*) = 0 \qquad (1)$$
$$c_i(x^*) = 0 \quad (i = 1, \ldots, \ell) \qquad (2)$$
$$c_i(x^*) \le 0,\ \mu_i^* \ge 0,\ \mu_i^* c_i(x^*) = 0 \quad (i = \ell+1, \ldots, m) \qquad (3)$$

KKT条件(2)と(3)より，$x^*$ は実行可能解.

よって，任意の実行可能解 $y \in S$ に対して

$$f(x^*) \le f(y)$$

を示せばよい.

いま，

$$g(x) \triangleq L(x, \mu^*) = f(x) + \sum_{i=1}^{m} \mu_i^* c_i(x)$$

とすると，$g$ は凸関数で，KKT条件(1)より

$$\nabla g(x^*) = \nabla f(x^*) + \sum_{i=1}^{m} \mu_i^* \nabla c_i(x^*) = 0$$

$$\nabla f(x^*) + \sum_{i=1}^{m} \mu_i^* \nabla c_i(x^*) = 0 \qquad (1)$$

$$c_i(x^*) = 0 \quad (i = 1, \dots, \ell) \qquad (2)$$

$$c_i(x^*) \le 0,\ \mu_i^* \ge 0,\ \mu_i^* c_i(x^*) = 0 \quad (i = \ell+1, \dots, m) \qquad (3)$$

これは　$x^*$ が

$$\min\ g(x)$$

$$\text{s.t.}\quad x \in R^n$$

の最小解であることを意味している.

よって,

KKT条件(2)

KKT条件(3)

$$f(x^*) = f(x^*) + \sum_{i=1}^{\ell} \mu_i^* c_i(x^*) + \sum_{i=\ell+1}^{m} \mu_i^* c_i(x^*)$$

$$= g(x^*)$$

$$\le g(y)$$

$c_i(y) = 0\ (i = 1, \dots, \ell)$

$c_i(y) \le 0,\ \mu_i^* \ge 0\ (i = 1, \dots, \ell)$

$$= f(y) + \sum_{i=1}^{\ell} \mu_i^* c_i(y) + \sum_{i=\ell+1}^{m} \mu_i^* c_i(y)$$

$$\le f(y)$$

# 講義内容

1. 制約つき問題に対する最適性の条件
   - 1次の必要条件：Karush-Kuhn-Tucker条件
   - 制約想定
   - 2次の条件
2. 凸最適化におけるKarush-Kuhn-Tucker条件
3. 一次の必要条件の証明

# 等式制約のときのみの定理１の証明

(LICQが成り立つとする)

$x^*$ は局所的最適解であるから

$\varepsilon > 0$が存在して

$f(x^*) \le f(x)$

for all $x$ such that $c_i(x) = 0 \ (i = 1, \ldots, m)$

$$\|x - x^*\| \le \varepsilon$$

次の問題 $(P^k)$ を考える.

$$(P^k) \quad \begin{array}{ll} \min & f^k(x) \\ \text{s.t.} & \|x - x^*\| \le \varepsilon \end{array}$$

ただし, $f^k(x) = f(x) + \frac{k}{2} \sum_{i=1}^{m} c_i(x)^2 + \frac{1}{2} \|x - x^*\|^2$

- 目的関数 $f^k$ は連続関数
- 実行可能領域は有界閉集合

$\Rightarrow$ $(P^k)$ は最適解をもつ

$(P^k)$ の最適解のひとつを $x^k$ とする.

補題 $\lim_{k\to\infty} x^k = x^*$

証明：まず，$\lim_{k\to\infty} c_i(x^k) = 0$ を示す．

$\{x^k\}$ は有界であるから，

$$f(x^k) + \frac{1}{2}\left\|x^k - x^*\right\|^2 \geq \gamma \quad \forall k$$

となる $\gamma$ が存在する．

$$f(x^*) = f^k(x^*) \geq f(x^k) + \frac{k}{2}\sum_{i=1}^{m} c_i(x^k) + \frac{1}{2}\left\|x^k - x^*\right\|^2 \geq \frac{k}{2}\sum_{i=1}^{m} c_i(x^k) + \gamma$$

より，

$$\frac{2\left(f(x^*) - \gamma\right)}{k} \geq \sum_{i=1}^{m} c_i(x^k)^2 \geq 0$$

よって，$\lim_{k\to\infty} c_i(x^k) = 0$

$\{x^k\}$ は有界であるから少なくともひとつの集積点をもつ. これを $\bar{x}$ とする.

$\lim_{k\to\infty} c_i(x^k)=0$ より, $\bar{x}$ は実行可能解である.

さらに, $\|\bar{x}-x^*\|\le\varepsilon$ より $f(x^*)\le f(\bar{x})$

$$
\begin{aligned}
f(x^*) &= \lim_{\substack{k\to\infty\\ k\in K}} f^k(x^*)\\
&\ge \lim_{\substack{k\to\infty\\ k\in K}} f^k(x^k)\\
&= \lim_{\substack{k\to\infty\\ k\in K}}\left\{ f(x^k)+\frac{k}{2}\sum_{i=1}^{m} c_i(x^k)^2+\frac{1}{2}\left\|x^k-x^*\right\|^2\right\}\\
&\ge \lim_{\substack{k\to\infty\\ k\in K}}\left\{ f(x^k)+\frac{1}{2}\left\|x^k-x^*\right\|^2\right\}\\
&= f(\bar{x})+\frac{1}{2}\left\|\bar{x}-x^*\right\|^2
\end{aligned}
$$

から $\bar{x}=x^*$

# 定理1の証明

十分大きい $k$ に対しては，$\left\|x^k - x^*\right\| \le \dfrac{\varepsilon}{2}$ となる．
さらに，$\left\|x - x^k\right\| \le \dfrac{\varepsilon}{2}$ となる $x \in R^n$ は

$$\left\|x - x^*\right\| \le \left\|x - x^k\right\| + \left\|x^* - x^k\right\| \le \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

となるから $(P^k)$ の実行可能解．
よって，

$$f^k(x^k) \le f(x) \ \ \forall x \in R^n \ \text{ such that } \left\|x^k - x\right\| \le \varepsilon / 2$$

これは，$x^k$が制約なし問題：

$$\min \ f^k(x)$$

の局所的最小解であることを意味している．

制約なし問題の1次の必要条件より，

$$
\begin{aligned}
0 &= \nabla f^k(x^k) \\
&= \nabla f(x^k) + \sum_{i=1}^{m} k c_i(x^k) \nabla c_i(x^k) + (x^k - x^*)
\end{aligned}
$$

ここで，　$\mu_i^k = k c_i(x^k)$ とおくと

$$
-\nabla f(x^k) - (x^k - x^*) = \sum_{i=1}^{m} \mu_i^k \nabla c_i(x^k) = \nabla c(x^k) \mu^k
$$

ただし，

$$
\nabla c(x) = \left( \nabla c_1(x^k)\ \nabla c_2(x^k) \cdots \nabla c_m(x^k) \right),\ \mu^k = \begin{pmatrix} \mu_1^k \\ \mu_2^k \\ \vdots \\ \mu_m^k \end{pmatrix}
$$

制約なし問題の1次の必要条件より,

$$
\begin{aligned}
0 &= \nabla f^k(x^k) \\
&= \nabla f(x^k) + \sum_{i=1}^{m} k c_i(x^k) \nabla c_i(x^k) + (x^k - x^*)
\end{aligned}
$$

ここで, $\mu_i^k = k c_i(x^k)$ とおくと

$$
-\nabla f(x^k) - (x^k - x^*) = \sum_{i=1}^{m} \mu_i^k \nabla c_i(x^k) = \nabla c(x^k)\mu^k
$$

ただし,

$$
\nabla c(x) = \left(\nabla c_1(x^k)\ \nabla c_2(x^k) \cdots \nabla c_m(x^k)\right),\ \mu^k = \left(\mu_1^k, \mu_2^k, \cdots, \mu_m^k\right)^\top
$$

両辺に行列 $\nabla c(x^k)^\top$ を掛けると

$$
\nabla c(x^k)^\top \nabla c(x^k)\mu^k = \nabla c(x^k)^\top \left(-\nabla f(x^k) - (x^k - x^*)\right)
$$

LICQより，十分大きい$k$に対して，

$$\nabla c(x^k)^\top \nabla c(x^k)$$

は正則行列になる．

よって，

$$\mu^k = \left(\nabla c(x^k)^\top \nabla c(x^k)\right)^{-1} \nabla c(x^k)^\top \left(-\nabla f(x^k) - (x^k - x^*)\right)$$

を得る．

$k \to \infty$　とすると，右辺は

$$\left(\nabla c(x^*)^\top \nabla c(x^*)\right)^{-1} \nabla c(x^*)^\top \left(-\nabla f(x^*)\right)$$

に収束する．

これは $\{\mu^k\}$ の極限になるから，これを $\mu^*$　とすると，

$$0 = \lim_{k\to\infty} \left\{ \nabla f(x^k) + \sum_{i=1}^m \mu_i^k \nabla c_i(x^k) + (x^k - x^*) \right\} = \nabla f(x^*) + \sum_{i=1}^m \mu_i^* \nabla c_i(x^*)$$

# 不等式制約を持つときの一般化

不等式制約 $c_i(x) \le 0$ は，等式制約

$$\hat{c}_i(x) = 0$$

と等価．ただし， $\hat{c}_i(x) = \max\{0, c_i(x)\}$

ここで，

$$\nabla\left(\hat{c}_i(x^k)^2\right) = 2\max\{0, c_i(x)\}\nabla c_i(x^k)$$

である．

また， $\delta > 0$ が存在して，$\|x - x^*\| \le \delta$

$$c_i(x) < 0 \ \ \forall i \text{ such that } \ i \in \{\ell + 1, \cdots, m\} \text{ and } \ i \notin A(x^*)$$

$x^*$ は局所的最適解であるから，$\varepsilon > 0$ が存在して，

$$f(x^*) \le f(x) \ \text{ for all } x \text{ such that } c_i(x) = 0 \ (i = 1, \ldots, \ell),$$
$$c_i(x) \le 0 \ (i = \ell + 1, \ldots, m), \ \left\| x - x^* \right\| \le \varepsilon$$

となり，さらに，

$$f(x^*) \le f(x) \ \text{ for all } x \text{ such that } c_i(x) = 0 \ (i = 1, \ldots, \ell),$$
$$c_i(x) \le 0 \ (i \in A(x^*)), \ \left\| x - x^* \right\| \le \min\{\varepsilon, \delta\}$$

を得る．これは $x^*$ が

$$\begin{aligned} &\min \ f(x) \\ &\text{s.t.} \quad c_i(x) = 0 \ (i = 1, \ldots, \ell) \\ &\qquad\ \ \hat{c}_i(x) = 0 \ (i \in A(x^*)) \end{aligned}$$

の局所的最適解であることを意味している．

いま，

$$\mu_i^k = \begin{cases} k c_i(x^k) & (i = 1,\ldots,\ell) \\ k \max\left\{0, c_i(x^k)\right\} & (i \in A(x^*)) \end{cases}$$

とおけば，等式制約条件のみのときと同様にして，$\left\{\mu_i^k\right\}$ は極限 $\mu_i^*$ を持ち，$\mu_i^* \geq 0\ (i \in A(x^*))$ と

$$\nabla f(x^*) + \sum_{i=1}^{\ell} \mu_i^* \nabla c_i(x^*) + \sum_{i \in A(x^*)} \mu_i^* \nabla c_i(x^*) = 0$$

を得る．さらに $\mu_i^* = 0\ (i \in \{\ell+1,\cdots,m\} / A(x^*))$ とおけば，$\nabla f(x^*) + \sum_{i=1}^{m} \mu_i^* \nabla c_i(x^*) = 0$ を得る．また，定義より，

$$\mu_i \geq 0\ (i = \ell+1,\ldots,m)$$

$$\mu_i^* c_i(x^*) = 0 \quad i \in \left\{\ell+1,\cdots,m\right\} / A(x^*)$$

$$\mu_i^* c_i(x^*) = 0 \quad i \in A(x^*) = \left\{\ i \in \left\{\ell+1,\cdots,m\right\} \middle|\ c_i(x^*) = 0\right\}$$