後期本読みゼミ第 3 回

最適性の条件〜KKT 条件とその証明〜

高橋 優輝

森田研究室 B4

November 2, 2022

目次

目次

3.0 制約なし最適化問題

制約なし最適化問題を以下のように定義する．

定義：制約なし最適化問題

関数 $f : \mathbb{R}^n \to \mathbb{R}$ を一回微分可能として，

$$\underset{\boldsymbol{x} \in \mathbb{R}^n}{\text{minimize}} \qquad f(\boldsymbol{x}) \tag{3.0}$$

制約なし最適化問題の最適性 1 次の必要条件

制約なし最適化問題の 1 次の最適性の必要条件は以下のようになる．

定理 3.0 (制約なし最適化問題の最適性の 1 次の必要条件)

$\boldsymbol{x}^*$ が問題 (3.0) の局所的最適解 $\Longrightarrow$ $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$

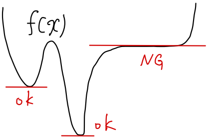
proof.

局所的最適解 $\boldsymbol{x}^*$ から，ある方向 $\boldsymbol{d} \in \mathbb{R}^n (\|\boldsymbol{d}\| = 1)$ に沿って，十分に小さい $\alpha (> 0)$ だけ動いた点を考える．Taylor 展開すると，

$$f(\boldsymbol{x}^* + \alpha \boldsymbol{d}) = f(\boldsymbol{x}^*) + \alpha \nabla f(\boldsymbol{x}^*)^\top \boldsymbol{d} + o(\alpha)$$

ここで，局所的最適解 $\boldsymbol{x}^*$ は $f(\boldsymbol{x}^* + \alpha \boldsymbol{d}) \geq f(\boldsymbol{x}^*)$ を満たすので，$\alpha > 0$ に注意して，

$$\nabla f(\boldsymbol{x}^*)^\top \boldsymbol{d} + \frac{o(\alpha)}{\alpha} \geq 0$$



が成り立つ．ここで，$\alpha \to +0$ とすると，$\frac{o(\alpha)}{\alpha}$ は 0 に収束するので，

$$\nabla f(\boldsymbol{x}^*)^\top \boldsymbol{d} \geq 0$$

が成り立つ．この不等式は任意の方向 $\boldsymbol{d} \in \mathbb{R}^n (\|\boldsymbol{d}\| = 1)$ に対して成り立つので，

$$\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$$

□

制約付き最適化問題ではどうなる？

実行可能解の集合を以下のように定義する．

$$\mathcal{F} = \left\{ (x_1, x_2)^\top \in \mathbb{R}^2 \;\middle|\; |x_1| + \left| x_2 - \frac{1}{2} \right| \leq 1 \right\}$$

<u>A.</u>

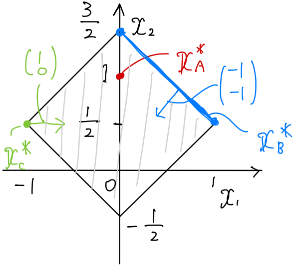
$$\underset{(x_1,x_2)^\top \in \mathcal{F}}{\text{minimize}} \qquad x_1^2 + (x_2 - 1)^2$$

<u>B.</u>

$$\underset{(x_1,x_2)^\top \in \mathcal{F}}{\text{minimize}} \qquad -x_1 - x_2$$

<u>C.</u>

$$\underset{(x_1,x_2)^\top \in \mathcal{F}}{\text{minimize}} \qquad x_1$$



最適解 $\boldsymbol{x}^*$ が内点のとき，$\nabla f(\boldsymbol{x}^*) = \mathbf{0}$.

最適解 $\boldsymbol{x}^*$ が境界上のとき，$\nabla f(\boldsymbol{x}^*) = \mathbf{0}$ は常には成り立たない.(→ 定理 3.0 は使えない．)

→ 定理 3.3 でより一般の場合について考える．

目次

3.1 制約あり最適化問題

制約あり最適化問題を以下のように定義する [*1].

定義：制約あり最適化問題

関数 $f : \mathbb{R}^n \to \mathbb{R}, h_i : \mathbb{R}^n \to \mathbb{R}\ (i = 1, \ldots, m), g_j : \mathbb{R}^n \to \mathbb{R}\ (j = 1, \ldots, r)$ をすべて，微分可能として，

$$
\begin{aligned}
\underset{\boldsymbol{x} \in \mathbb{R}^n}{\text{minimize}} \quad & f(\boldsymbol{x}) \\
\text{subject to} \quad & h_i(\boldsymbol{x}) = 0 \qquad (i = 1, \ldots, m) \\
& g_j(\boldsymbol{x}) \leq 0 \qquad (j = 1, \ldots, r)
\end{aligned}
\tag{3.1}
$$

[*1] 等式制約を取り除いても良いが，教科書に倣って，等式制約ありの問題について考える.

KKT 条件

この章では，以下の制約あり最適化問題の最適性の 1 次条件 (KKT[*2] 条件) について扱う．

定理 3.1(制約あり最適化問題の最適性の 1 次の必要条件;KKT 条件)

Abadie の制約想定 [*3] が成り立つという仮定の下で，
$\boldsymbol{x}^*$ が局所的最適解 $\implies \exists \boldsymbol{\lambda} \in \mathbb{R}^m, \boldsymbol{\mu} \in \mathbb{R}^n \quad s.t.$

$$\begin{cases} \nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(\boldsymbol{x}^*) + \sum_{j=1}^{r} \mu_i \nabla g_j(\boldsymbol{x}^*) = \mathbf{0} & (3.2) \\ h_i(\boldsymbol{x}^*) = 0 \quad (i = 1, \ldots, m) & (3.3) \\ g_j(\boldsymbol{x}^*) \leq 0, \mu_j \geq 0, \mu_j g_j(\boldsymbol{x}^*) = 0 \quad (j = 1, \ldots, r) & (3.4) \end{cases}$$

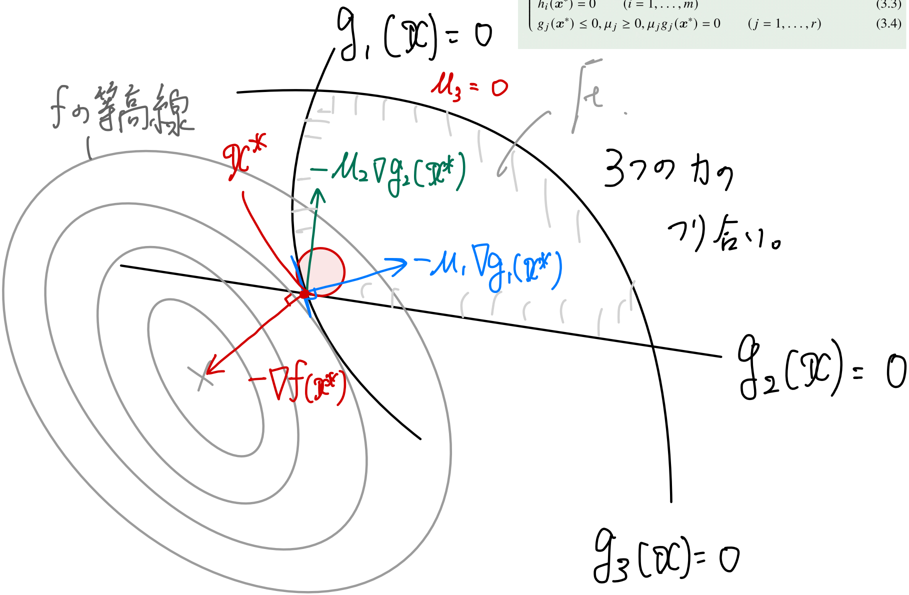
式 (3.4) の等式は，最適解においては，各 j について，$\mu_j = 0$ または，$g_j(\boldsymbol{x}^*) = 0$ が成り立つことを言っており，これを相補性条件という．

[*2] Karush-Kuhn-Tucker

[*3] 3.3 節で扱う．

KKT 条件のイメージ

$$
\begin{cases}
\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(\boldsymbol{x}^*) + \sum_{j=1}^{r} \mu_j \nabla g_j(\boldsymbol{x}^*) = \mathbf{0} & (3.2) \\
h_i(\boldsymbol{x}^*) = 0 \quad (i = 1, \ldots, m) & (3.3) \\
g_j(\boldsymbol{x}^*) \leq 0, \mu_j \geq 0, \mu_j g_j(\boldsymbol{x}^*) = 0 \quad (j = 1, \ldots, r) & (3.4)
\end{cases}
$$



Lagrange 関数を用いた KKT 条件

Lagrange 関数を以下のように定義する．

定義：Lagrange 関数

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

Lagrange 関数を用いると，KKT 条件の式 (3.2)〜(3.4) は，

$$\begin{cases} \nabla_{\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{0} \\ \nabla_{\boldsymbol{\lambda}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{0} \\ \nabla_{\boldsymbol{\mu}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}) \leq \mathbf{0}, \boldsymbol{\mu} \geq \mathbf{0}, \boldsymbol{\mu}^\top \nabla_{\boldsymbol{\mu}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}) = 0 \end{cases} \tag{3.5}$$

と表せる．

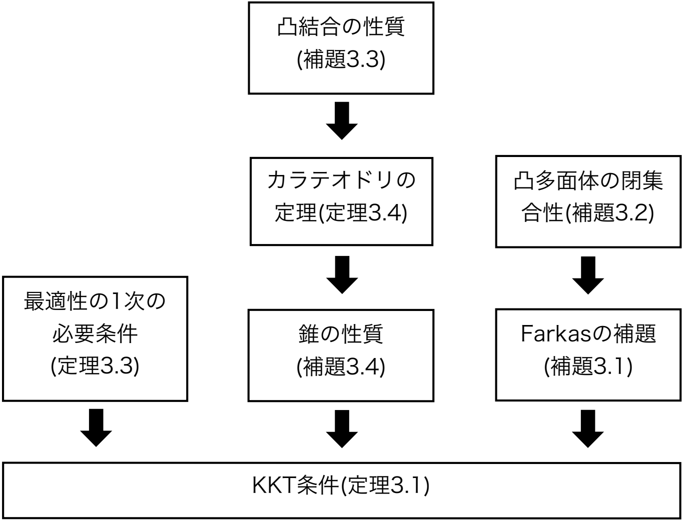
$$\begin{cases} \nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(\boldsymbol{x}^*) + \sum_{j=1}^{r} \mu_i \nabla g_j(\boldsymbol{x}^*) = \mathbf{0} & (3.2) \\ h_i(\boldsymbol{x}^*) = 0 \quad (i = 1, \ldots, m) & (3.3) \\ g_j(\boldsymbol{x}^*) \leq 0, \mu_j \geq 0, \mu_j g_j(\boldsymbol{x}^*) = 0 \quad (j = 1, \ldots, r) & (3.4) \end{cases}$$

KKT 条件を満たす点 $(\boldsymbol{x}^*, \boldsymbol{\lambda}, \boldsymbol{\mu})$ を KKT 点と呼ぶ．
制約あり最適化問題に対する多くのアルゴリズムでは，最適性の 1 次の必要条件を満たす KKT 点を求めることが目的となっている．
KKT 条件は必要条件であり，KKT 条件を満たしているからといって，$\boldsymbol{x}^*$ が局所的最適解となるとは限らない．(後に反例を示す．)
一方，凸計画問題においては，KKT 条件が大域的最適性の必要十分条件となる．(後に示す．)

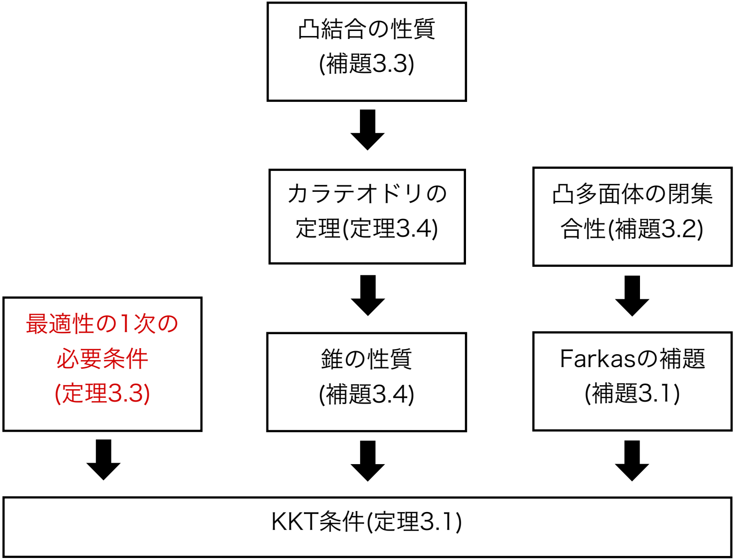
目次

KKT 条件の証明

以下のフローチャートに従って証明を行なう．



定理 3.3(最適性の 1 次の必要条件)



最適性の 1 次の必要条件

定理 3.3(最適性の 1 次の必要条件)

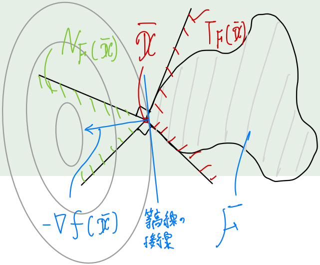
f を連続微分可能な関数，$\bar{\boldsymbol{x}}$ を次の非線形計画問題の局所的最適解とする．

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \mathcal{F} \end{array}$$

このとき，

$$-\nabla f(\bar{\boldsymbol{x}}) \in N_{\mathcal{F}}(\bar{\boldsymbol{x}})$$

が成り立つ．



集合 $\mathcal{S}$ の $\bar{\boldsymbol{x}}$ における接錐 $T_{\mathcal{S}}(\bar{\boldsymbol{x}})$

$$\overset{\text{def}}{\Longleftrightarrow} \; T_{\mathcal{S}}(\bar{\boldsymbol{x}}) = \left\{ \boldsymbol{y} \in \mathbb{R}^n \;\middle|\; \boldsymbol{y} = \lim_{k\to\infty} a_k(\boldsymbol{x}^k - \bar{\boldsymbol{x}}),\ \lim_{k\to\infty} \boldsymbol{x}^k = \bar{\boldsymbol{x}}, \boldsymbol{x}^k \in \mathcal{S}, a_k \geq 0 \right\}$$

集合 $\mathcal{S}$ の $\bar{\boldsymbol{x}}$ における法線錐 $N_{\mathcal{S}}(\bar{\boldsymbol{x}}) \overset{\text{def}}{\Longleftrightarrow} N_{\mathcal{S}}(\bar{\boldsymbol{x}}) = T_{\mathcal{S}}(\bar{\boldsymbol{x}})^*$

proof.

$\boldsymbol{y} \in T_{\mathcal{F}}(\bar{\boldsymbol{x}})$ とする．このとき，接錐の定義より，

$$\boldsymbol{x}^k \in \mathcal{F}, \lim_{k\to\infty} \boldsymbol{x}^k = \bar{\boldsymbol{x}}, \alpha_k \geq 0, \lim_{k\to\infty} \alpha_k(\boldsymbol{x}^k - \bar{\boldsymbol{x}}) = \boldsymbol{y}$$

を満たす列 $\left\{\boldsymbol{x}^k\right\}$ と $\{\alpha_k\}$ が存在する．

f は連続微分可能より，$f(\boldsymbol{x}^k)$ を $\bar{\boldsymbol{x}}$ の周りで Taylor 展開すると，

$$f(\boldsymbol{x}^k) - f(\bar{\boldsymbol{x}}) = \langle \nabla f(\bar{\boldsymbol{x}}), \boldsymbol{x}^k - \bar{\boldsymbol{x}} \rangle + o(\|\boldsymbol{x}^k - \bar{\boldsymbol{x}}\|)$$

が成り立つ．また，$\bar{\boldsymbol{x}}$ は局所的最適解より，十分大きな k に対して，

$f(\boldsymbol{x}^k) - f(\bar{\boldsymbol{x}}) \geq 0$ が成り立つので，

$$\begin{aligned} 0 &\leq \alpha_k \langle \nabla f(\bar{\boldsymbol{x}}), \boldsymbol{x}^k - \bar{\boldsymbol{x}} \rangle + \alpha_k o(\|\boldsymbol{x}^k - \bar{\boldsymbol{x}}\|) \qquad (\because \alpha_k \geq 0, \forall k) \\ &= \langle \nabla f(\bar{\boldsymbol{x}}), \alpha_k(\boldsymbol{x}^k - \bar{\boldsymbol{x}}) \rangle + \|\alpha_k(\boldsymbol{x}^k - \bar{\boldsymbol{x}})\| \frac{o(\|\boldsymbol{x}^k - \bar{\boldsymbol{x}}\|)}{\|\boldsymbol{x}^k - \bar{\boldsymbol{x}}\|} \end{aligned}$$

が成り立つ．さらに，$k \to \infty$ とすると，以下の式が成り立つ．

$$0 \leq \langle \nabla f(\bar{\boldsymbol{x}}), \boldsymbol{y} \rangle \iff 0 \geq \langle -\nabla f(\bar{\boldsymbol{x}}), \boldsymbol{y} \rangle$$

これが任意の $\boldsymbol{y} \in T_{\mathcal{F}}(\bar{\boldsymbol{x}})$ について成り立つので，法線錐の定義から，

$$-\nabla f(\bar{\boldsymbol{x}}) \in N_{\mathcal{F}}(\bar{\boldsymbol{x}})$$

□

$\bar{\boldsymbol{x}}$ を局所的最適解とすると，制約なし最適化問題の場合，$\mathcal{F} = \mathbb{R}^n$ となり，$N_{\mathcal{F}}(\bar{\boldsymbol{x}}) = \{\mathbf{0}\}$ であるので，$\nabla f(\bar{\boldsymbol{x}}) = \mathbf{0}$ となる．
条件 $\nabla f(\hat{\boldsymbol{x}}) \in N_{\mathcal{F}}(\hat{\boldsymbol{x}})$ を満たす $\hat{\boldsymbol{x}}$ を停留点と呼び，この条件を最適性の 1 次の必要条件と呼ぶこともある．
一般には，$N_{\mathcal{F}}(\bar{\boldsymbol{x}})$ を具体的に表すのは困難．
$\longrightarrow$ 制約なし最適化問題では，定理 3.0[*4] を，制約あり最適化問題では，定理 3.1(KKT 条件) を用いて停留点を求めるのが良い．

[*4] $\bar{\boldsymbol{x}}$ が制約なし最適化問題の局所的最適解 $\implies \nabla f(\bar{\boldsymbol{x}}) = \mathbf{0}$

制約付き最適化問題ではどうなる？ (再掲)

実行可能解の集合を以下のように定義する．

$$\mathcal{F} = \left\{ (x_1, x_2)^\top \in \mathbb{R}^2 \;\middle|\; |x_1| + \left| x_2 - \frac{1}{2} \right| \leq 1 \right\}$$

<u>A.</u>

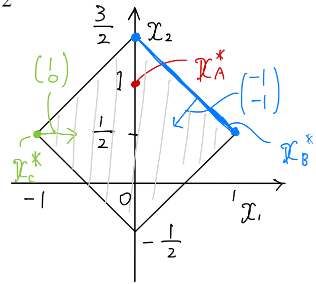
$$\underset{(x_1,x_2)^\top \in \mathcal{F}}{\text{minimize}} \qquad x_1^2 + (x_2 - 1)^2$$

<u>B.</u>

$$\underset{(x_1,x_2)^\top \in \mathcal{F}}{\text{minimize}} \qquad -x_1 - x_2$$

<u>C.</u>

$$\underset{(x_1,x_2)^\top \in \mathcal{F}}{\text{minimize}} \qquad x_1$$



最適解 $\boldsymbol{x}^*$ が内点のとき，$\nabla f(\boldsymbol{x}^*) = \mathbf{0}$.
最適解 $\boldsymbol{x}^*$ が境界上のとき，$\nabla f(\boldsymbol{x}^*) = \mathbf{0}$ は常には成り立たない.(→ 定理 3.0 は使えない．)

定理 (凸計画問題についての最適性の 1 次の必要条件の十分性)

$\mathcal{F} \subseteq \mathbb{R}^n$ は空でない凸集合，$f : \mathbb{R}^n \to \mathbb{R}$ は点 $\bar{\boldsymbol{x}} \in \mathcal{F}$ において微分可能な凸関数とする．このとき，式 $-\nabla f(\bar{\boldsymbol{x}}) \in N_{\mathcal{F}}(\bar{\boldsymbol{x}})$ は，$\bar{\boldsymbol{x}}$ が定理 3.3 の非線形計画問題の大域的最適解であるための必要十分条件である．

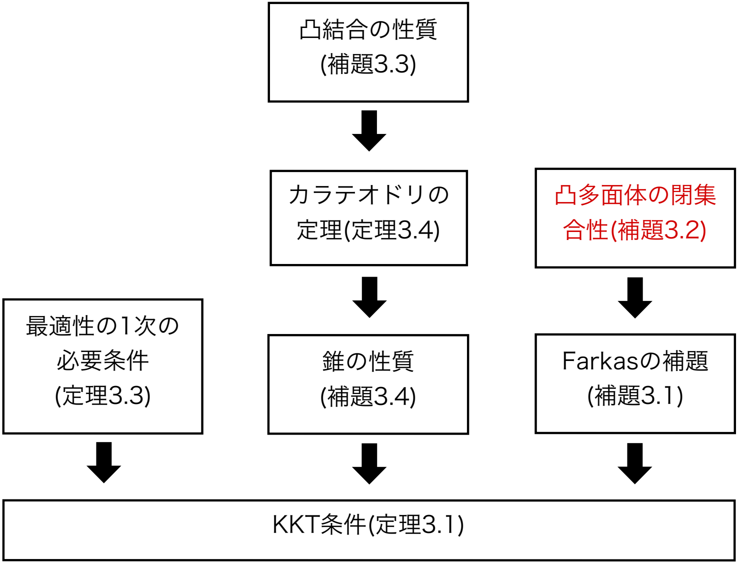
proof.

必要性は定理 3.3 より明らか．
$\bar{\boldsymbol{x}}$ は $-\nabla f(\bar{\boldsymbol{x}}) \in N_{\mathcal{F}}(\bar{\boldsymbol{x}})$ を満たすとして，$\boldsymbol{x} \in \mathcal{F}$ とする．
ここで，$0 < \beta_k < 1$ かつ $\beta_k \to 0\ (as\ k \to \infty)$ を満たす列 $\{\beta_k\}$ を用いて，点列 $\{\boldsymbol{x}^k\}$ を

$$\boldsymbol{x}^k := (1 - \beta_k)\bar{\boldsymbol{x}} + \beta_k \boldsymbol{x}$$

とすると，$\mathcal{F}$ は凸集合であるから，$\{\boldsymbol{x}^k\} \subseteq \mathcal{F}$ が成り立つ．
さらに，$\alpha_k := 1/\beta_k$ として，点列 $\{\alpha_k(\boldsymbol{x} - \bar{\boldsymbol{x}})\}$ を考えると，任意の k について，$\alpha_k(\boldsymbol{x}^k - \bar{\boldsymbol{x}}) = \boldsymbol{x} - \bar{\boldsymbol{x}}$ が成立するので，接錐 $T_{\mathcal{F}}(\bar{\boldsymbol{x}})$ の定義より，$\boldsymbol{x} - \bar{\boldsymbol{x}} \in T_{\mathcal{F}}(\bar{\boldsymbol{x}})$ が成り立つ．
これと，$-\nabla f(\bar{\boldsymbol{x}}) \in N_{\mathcal{F}}(\bar{\boldsymbol{x}})$ より，$\langle -\nabla f(\bar{\boldsymbol{x}}), \boldsymbol{x} - \bar{\boldsymbol{x}} \rangle \leq 0$ が成り立つ．
また，定理 2.5 より，$f(\boldsymbol{x}) \geq f(\bar{\boldsymbol{x}}) + \langle \nabla f(\bar{\boldsymbol{x}}), \boldsymbol{x} - \bar{\boldsymbol{x}} \rangle$ であるので，任意の $\boldsymbol{x} \in \mathcal{F}$ に対して，$f(\boldsymbol{x}) \geq f(\bar{\boldsymbol{x}})$ が成り立つ．よって，十分性が示された． $\square$

補題 3.2(凸多面体の閉集合性)



Farkas の補題の証明の準備

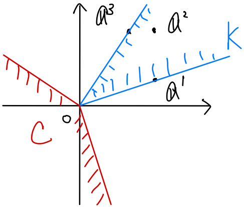
$\boldsymbol{a}^1, \ldots, \boldsymbol{a}^k \in \mathbb{R}^n$ として，集合 C, K を以下のように定義する．

$$C = \left\{ \boldsymbol{y} \in \mathbb{R}^n \mid \langle \boldsymbol{y}, \boldsymbol{a}^i \rangle \leq 0,\ i = 1, \ldots, k \right\}$$

$$K = \left\{ \boldsymbol{x} \in \mathbb{R}^n \,\middle|\, \boldsymbol{x} = \sum_{i=1}^{k} \lambda_i \boldsymbol{a}^i, \lambda_i \geq 0,\ i = 1, \ldots, k \right\}$$

明らかに，C と K は凸錐である．
さらに，C の極錐 C^* は以下のように定義される．

$$C^* = \{ \boldsymbol{x} \mid \langle \boldsymbol{x}, \boldsymbol{y} \rangle \leq 0, \forall \boldsymbol{y} \in C \}$$



Farkas の補題を示す前に，次の補題を示す．

補題 3.2(凸多面体の閉集合性)

すべての $\boldsymbol{x} \in \mathbb{R}^n$ に対して，$\boldsymbol{x}$ に最も近い点 $\boldsymbol{y} \in K$ が存在する．

proof.

帰納法で示す．
$k = 1$ のとき，$K = \left\{ \boldsymbol{x} \mid \boldsymbol{x} = \lambda_1 \boldsymbol{a}^1, \lambda_1 \geq 0 \right\}$ より，明らか．
$k = l - 1$ $(l = 2, 3, \ldots)$ のとき，主張が成り立つと仮定すると，$k = l$ のとき，

$$K = \left\{ \boldsymbol{x} \in \mathbb{R}^n \,\middle|\, \boldsymbol{x} = \sum_{i=1}^{l} \lambda_i \boldsymbol{a}^i, \lambda_i \geq 0,\ i = 1, \ldots, k \right\}$$

である．ここで，$\boldsymbol{a}^j$ のみを除いた $l - 1$ 個の点からなる凸錐 K_j $(j = 1, \ldots, l)$ を以下のように定義する．

$$\begin{aligned} K_j = \{ \boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} = \lambda_1 \boldsymbol{a}^1 + \cdots + \lambda_{j-1} \boldsymbol{a}^{j-1} + \lambda_{j+1} \boldsymbol{a}^{j+1} + \cdots + \lambda_l \boldsymbol{a}^l, \\ \lambda_i \geq 0, i = 1, \ldots, l, i \neq j \} \end{aligned}$$

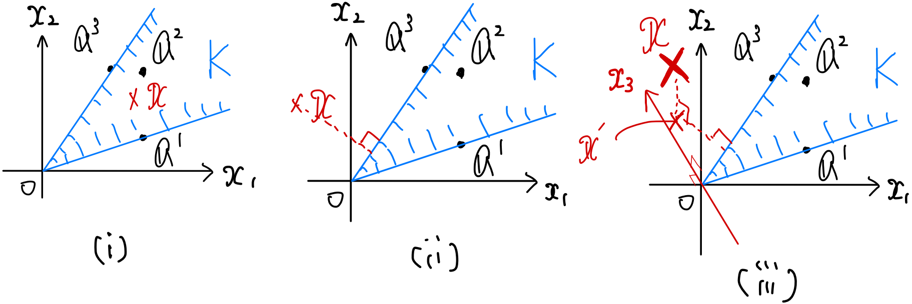
帰納法の仮定より，各 K_j 上で，$\boldsymbol{x}$ に最も近い点 $\boldsymbol{y}^j \in K$ が存在する．
また，$\boldsymbol{a}^1, \ldots, \boldsymbol{a}^l$ によって張られる $\mathbb{R}^n$ の部分空間 L を以下のように定義する．

$$L := \left\{ \boldsymbol{z} \in \mathbb{R}^n \,\middle|\, \boldsymbol{z} = \sum_{i=1}^{l} \alpha_i \boldsymbol{a}^i, \alpha_i \in \mathbb{R}, i = 1, \ldots, l \right\}$$

proof.(続き)

(i)$\boldsymbol{x} \in K$ のとき　(ii)$\boldsymbol{x} \notin K$ かつ $\boldsymbol{x} \in L$ のとき　(iii)otherwise

の 3 つの場合に分けて証明を行なう．



$\boxed{\text{(i). } \boldsymbol{x} \in K \text{ のとき}}$

$\boldsymbol{y} = \boldsymbol{x}$ とすれば良い．

proof.(続き)

$\boxed{\text{(ii). } \boldsymbol{x} \notin K \text{ かつ } \boldsymbol{x} \in L}$

$\boldsymbol{y}^1, \ldots, \boldsymbol{y}^l$ の中で，$\boldsymbol{x}$ に最も近い点を $\boldsymbol{y}$ とする．このとき，明らかに $\boldsymbol{y} \in K$ である．
以下では，任意の $\boldsymbol{z} \in K$ に対して，$\|\boldsymbol{x} - \boldsymbol{y}\| \leq \|\boldsymbol{x} - \boldsymbol{z}\|$ となる．すなわち，この $\boldsymbol{y}$ が K 上で最も $\boldsymbol{x}$ に近い点であることを示す．
$\boldsymbol{x} \in L, \boldsymbol{z} \in K$ より，

$$\boldsymbol{x} = \alpha_1 \boldsymbol{a}^1 + \cdots + \alpha_l \boldsymbol{a}^l, \ \boldsymbol{z} = \beta_1 \boldsymbol{a}^1 + \cdots + \beta_l \boldsymbol{a}^l$$

を満たす $\alpha_i \in \mathbb{R}, \beta_i \geq 0 \ (i = 1, \ldots, l)$ が存在する．
いま，$I(\boldsymbol{x}) := \left\{ j \mid \alpha_j < 0 \right\}$ とすると，$\boldsymbol{x} \notin K$ より，$I(\boldsymbol{x})$ は空集合でない．
そのため，

$$t := \min_{j \in I(\boldsymbol{x})} \left\{ \frac{\beta_j}{\beta_j - \alpha_j} \right\}$$

を定義できる．

proof.(続き)

$j \in I(\boldsymbol{x})$ では，$I(\boldsymbol{x})$ の定義と $\beta_i \geq 0$ より，$0 \leq t < 1$ であり，

$$t\alpha_j + (1-t)\beta_j = t(\alpha_j - \beta_j) + \beta_j = (\beta_j - \alpha_j)\left(-t + \frac{\beta_j}{\beta_j - \alpha_j}\right) \geq 0$$

が成り立つ．
ここで，添字 $i \in I(\boldsymbol{x})$ において，$t = \frac{\beta_i}{\beta_i - \alpha_i}$ が成り立つとすると，
$t\alpha_i + (1-t)\beta_i = 0$ となる．
$j \notin I(\boldsymbol{x})$ では，$\alpha_j \geq 0$ であるので，$t\alpha_j + (1-t)\beta_j \geq 0$ が成り立つ．
よって，以下の式が成り立つ．

$$t\boldsymbol{x} + (1-t)\boldsymbol{y} = \sum_{j=1}^{l} \left(t\alpha_j + (1-t)\beta_j\right) \boldsymbol{a}^j \in K_i$$

すなわち，$\boldsymbol{y}$ の定義，$\boldsymbol{y}_i$ の定義，$t\boldsymbol{x} + (1-t)\boldsymbol{y} \in K_i, t \in [0,1)$ に注意すると，

$$\|\boldsymbol{x} - \boldsymbol{y}\| \leq \|\boldsymbol{x} - \boldsymbol{y}^i\| \leq \|\boldsymbol{x} - (t\boldsymbol{x} + (1-t)\boldsymbol{z})\| = (1-t)\|\boldsymbol{x} - \boldsymbol{z}\| \leq \|\boldsymbol{x} - \boldsymbol{z}\|$$

が成り立つ．

proof.(続き)

(iii). $x \notin L$ のとき

L の正規直交基底を $\boldsymbol{e}^1, \dots, \boldsymbol{e}^p$ として，

$$\boldsymbol{x}' := \langle \boldsymbol{x}, \boldsymbol{e}^1 \rangle \boldsymbol{e}^1 + \cdots + \langle \boldsymbol{x}, \boldsymbol{e}^p \rangle \boldsymbol{e}^p$$

とすると，明らかに $\boldsymbol{x}' \in L$ である．
このとき，(i),(ii) より，$\boldsymbol{x}'$ から最も近い点 $\boldsymbol{y} \in K$ が存在する．
以下では，$\boldsymbol{y}$ が $\boldsymbol{x}$ から，K 上で最も近い点であることを示す．
まず，$\{\boldsymbol{e}^i\}$ は直交基底だから，任意の $i = 1, \dots, p$ に対して，

$$\langle \boldsymbol{x} - \boldsymbol{x}', \boldsymbol{e}^i \rangle = \langle \boldsymbol{x}, \boldsymbol{e}^i \rangle - \langle \boldsymbol{x}', \boldsymbol{e}^i \rangle = \langle \boldsymbol{x}, \boldsymbol{e}^i \rangle - \langle \boldsymbol{x}, \boldsymbol{e}^i \rangle = 0$$

が成り立つので，任意の $\boldsymbol{w} \in L$ に対して，

$$\langle \boldsymbol{x} - \boldsymbol{x}', \boldsymbol{w} \rangle = 0$$

が成り立つ．

proof.(続き)

ここで，$\boldsymbol{x}' \in L$ より，$\langle \boldsymbol{x} - \boldsymbol{x}', \boldsymbol{x}' \rangle = 0$ であるから，

$$
\begin{aligned}
&\|\boldsymbol{x} - \boldsymbol{x}'\|^2 + \|\boldsymbol{x}' - \boldsymbol{w}\|^2 \\
&= \|\boldsymbol{x}\|^2 - 2\langle \boldsymbol{x}, \boldsymbol{x}' \rangle + 2\|\boldsymbol{x}'\|^2 - 2\langle \boldsymbol{x}', \boldsymbol{w} \rangle + \|\boldsymbol{w}\|^2 \\
&= \|\boldsymbol{x}\|^2 - 2(\langle \boldsymbol{x} - \boldsymbol{x}', \boldsymbol{x}' \rangle + \langle \boldsymbol{x}', \boldsymbol{x}' \rangle) + 2\|\boldsymbol{x}'\|^2 - 2(\langle \boldsymbol{x} - \boldsymbol{x}', \boldsymbol{w} \rangle + \langle \boldsymbol{x}, \boldsymbol{w} \rangle) + \|\boldsymbol{w}\|^2 \\
&= \|\boldsymbol{x}\|^2 - 2\langle \boldsymbol{x}, \boldsymbol{w} \rangle + \|\boldsymbol{w}\|^2 \qquad (\because \langle \boldsymbol{x} - \boldsymbol{x}', \boldsymbol{w} \rangle = 0, \langle \boldsymbol{x} - \boldsymbol{x}', \boldsymbol{x}' \rangle = 0) \\
&= \|\boldsymbol{x} - \boldsymbol{w}\|^2
\end{aligned}
$$

が成り立つ． ここで，$\boldsymbol{y} \in L$ であるから，

$$\|\boldsymbol{x} - \boldsymbol{y}\|^2 = \|\boldsymbol{x} - \boldsymbol{x}'\|^2 + \|\boldsymbol{x}' - \boldsymbol{y}\|^2$$

が成り立つ．また，任意の $\boldsymbol{z} \in K \subseteq L$ についても，

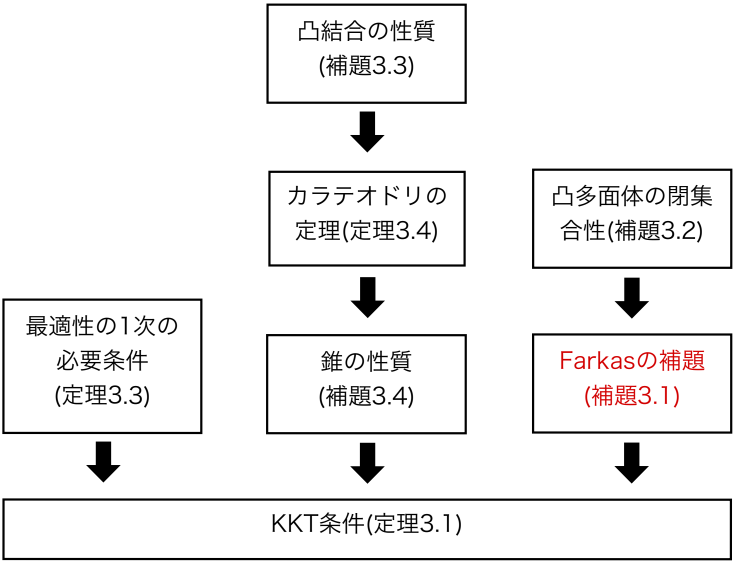
$$\|\boldsymbol{x} - \boldsymbol{z}\|^2 = \|\boldsymbol{x} - \boldsymbol{x}'\|^2 + \|\boldsymbol{x}' - \boldsymbol{z}\|^2$$

が成り立つ．ここで，$\|\boldsymbol{x}' - \boldsymbol{y}\| \leq \|\boldsymbol{x}' - \boldsymbol{z}\|$ に注意すると，

$$\|\boldsymbol{x} - \boldsymbol{y}\|^2 = \|\boldsymbol{x} - \boldsymbol{x}'\|^2 + \|\boldsymbol{x}' - \boldsymbol{y}\|^2 \leq \|\boldsymbol{x} - \boldsymbol{x}'\|^2 + \|\boldsymbol{x}' - \boldsymbol{z}\|^2 = \|\boldsymbol{x} - \boldsymbol{z}\|^2$$

が成り立つ．すなわち，$\boldsymbol{y}$ は $\boldsymbol{x}$ から K 上で最も近い点である．
以上 (i)〜(iii) より，帰納法を用いて示せた． □

補題 3.1(Farkas の補題)



Farkas の補題

補題 3.1(Farkas の補題)

$$K = C^*$$

$\boldsymbol{a}^1, \ldots, \boldsymbol{a}^k \in \mathbb{R}^n$ として，集合 C, K は以下のように定義されている．

$$C = \left\{ \boldsymbol{y} \in \mathbb{R}^n \mid \langle \boldsymbol{y}, \boldsymbol{a}^i \rangle \leq 0,\ i = 1, \ldots, k \right\}$$

$$K = \left\{ \boldsymbol{x} \in \mathbb{R}^n \;\middle|\; \boldsymbol{x} = \sum_{i=1}^{k} \lambda_i \boldsymbol{a}^i, \lambda_i \geq 0,\ i = 1, \ldots, k \right\}$$

C の極錐 C^* は以下の式で与えられる．

$$C^* = \{ \boldsymbol{x} \mid \langle \boldsymbol{x}, \boldsymbol{y} \rangle \leq 0, \forall \boldsymbol{y} \in C \}$$

proof.

まず，$C^* \subseteq K$，すなわち，任意の $\boldsymbol{x} \in C^*$ について，$\boldsymbol{x} \in K$ を示す．
補題 3.2 より，$\boldsymbol{x}$ に最も近い K 上の点 $\boldsymbol{y}$ が存在する．このとき，
$\langle \boldsymbol{a}^j, \boldsymbol{x} - \boldsymbol{y} \rangle > 0 \ (j = 1, \ldots, k)$ であると仮定すると，十分小さい $t \in (0, 1)$ に対して，

$$
\begin{aligned}
\|\boldsymbol{x} - (\boldsymbol{y} + t\boldsymbol{a}^j)\|^2 &= \|\boldsymbol{x} - \boldsymbol{y} - t\boldsymbol{a}^j\|^2 \\
&= \|\boldsymbol{x} - \boldsymbol{y}\|^2 - 2t\langle \boldsymbol{a}^j, \boldsymbol{x} - \boldsymbol{y} \rangle + t^2\|\boldsymbol{a}^j\|^2 \\
&< \|\boldsymbol{x} - \boldsymbol{y}\|^2
\end{aligned}
$$

となるが，$\boldsymbol{y} + t\boldsymbol{a}^j \in K$ であり，$\boldsymbol{y}$ が $\boldsymbol{x}$ に最も近い K 上の点であることに矛盾する．また，同様に，$\langle -\boldsymbol{y}, \boldsymbol{x} - \boldsymbol{y} \rangle > 0$ と仮定すると，十分小さい $t \in (0, 1)$ に対して，

$$
\|\boldsymbol{x} - (\boldsymbol{y} - t\boldsymbol{y})\|^2 = \|\boldsymbol{x} - \boldsymbol{y}\|^2 - 2t\langle -\boldsymbol{y}, \boldsymbol{x} - \boldsymbol{y} \rangle + t^2\|\boldsymbol{y}\|^2 < \|\boldsymbol{x} - \boldsymbol{y}\|^2
$$

となるが，$\boldsymbol{y} - t\boldsymbol{y} \in K$ であり，$\boldsymbol{y}$ が $\boldsymbol{x}$ に最も近い K 上の点であることに矛盾する．

proof.(続き)

以上より，任意の $\boldsymbol{x} \in C^*$ と $\boldsymbol{x}$ に最も近い K 上の点 $\boldsymbol{y}$ について，

$$\langle \boldsymbol{a}^j, \boldsymbol{x} - \boldsymbol{y} \rangle \le 0, \qquad (j = 1, \ldots, k) \tag{3.16}$$

$$\langle -\boldsymbol{y}, \boldsymbol{x} - \boldsymbol{y} \rangle \le 0 \tag{3.17}$$

が必要となる．

式 (3.16) と C の定義より，$\boldsymbol{x} - \boldsymbol{y} \in C$ となる．これと，$\boldsymbol{x} \in C^*$ から，

$$\langle \boldsymbol{x}, \boldsymbol{x} - \boldsymbol{y} \rangle \le 0$$

が成り立つ．この不等式と式 (3.17) を足すと，

$$\begin{aligned} &\langle \boldsymbol{x}, \boldsymbol{x} - \boldsymbol{y} \rangle + \langle -\boldsymbol{y}, \boldsymbol{x} - \boldsymbol{y} \rangle = \|\boldsymbol{x} - \boldsymbol{y}\|^2 \le 0 \\ \iff &\ \boldsymbol{x} = \boldsymbol{y} \end{aligned}$$

よって，任意の $\boldsymbol{x} \in C^*$ について，$\boldsymbol{x} \in K$ であるので，$C^* \subseteq K$．

proof.(続き)

次に，$C^* \supseteq K$ を示す．

$\boldsymbol{x} \in K$ であるとき，C の定義と，$\lambda_i \geq 0\ (i = 1, \ldots, k)$ に注意すると，任意の $\boldsymbol{z} \in C$ に対して，

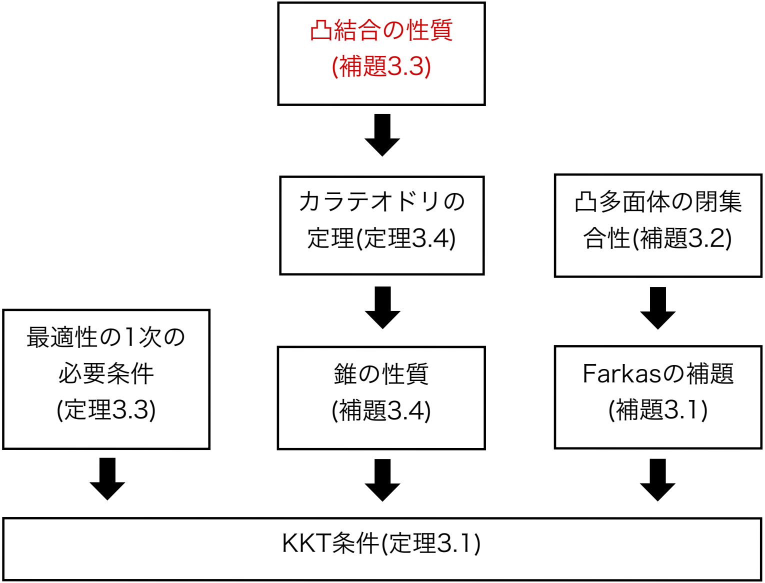
$$\langle \boldsymbol{x}, \boldsymbol{z} \rangle = \sum_{i=1}^{k} \lambda_i \langle \boldsymbol{a}^i, \boldsymbol{z} \rangle \leq 0$$

が成り立つので，C^* の定義から，$\boldsymbol{x} \in C^*$ である．よって，任意の $\boldsymbol{x} \in K$ について，$\boldsymbol{x} \in C^*$ であるので，$C^* \supseteq K$．

以上より，$C^* \subseteq K$，かつ，$C^* \supseteq K \iff C^* = K$

□

補題 3.3(凸結合の性質)

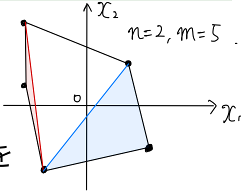

凸結合の性質
(補題3.3)
カラテオドリの
定理(定理3.4)
凸多面体の閉集
合性(補題3.2)
最適性の1次の
必要条件
(定理3.3)
錐の性質
(補題3.4)
Farkasの補題
(補題3.1)
KKT条件(定理3.1)

凸結合の性質

補題 3.3(凸結合の性質)

$m \geq n+2$ とする．点 $\boldsymbol{x} \in \mathbb{R}^n$ が m 個の点の凸結合で表されているとき，$\boldsymbol{x}$ は，m 個の点からたかだか $n+1$ 個の点を選んで，それらの点の凸結合として表すことができる．

proof.



$$\boldsymbol{x} = \sum_{i=1}^{m} \alpha_i \boldsymbol{x}^i, \alpha_i > 0, \sum_{i=1}^{m} \alpha_i = 1$$

と表されているとする．また，$m-1$ 個のベクトル $\boldsymbol{y}_i$ を $\boldsymbol{y}_i := \boldsymbol{x}^i - \boldsymbol{x}^m \ (i = 1, \ldots, m-1)$ とする．
いま，$m-1 > n$ より，$\{\boldsymbol{y}^1, \boldsymbol{y}^2, \ldots, \boldsymbol{y}^{m-1}\}$ は 1 次独立でないので，少なくとも 1 つは正であるような $\beta_i \in \mathbb{R} \ (i = 1, 2, \ldots, m-1)$ が存在して，

$$\boldsymbol{0} = \sum_{i=1}^{m-1} \beta_i \boldsymbol{y}^i = \sum_{i=1}^{m-1} \beta_i \boldsymbol{x}^i - \sum_{i=1}^{m-1} \beta_i \boldsymbol{x}^m$$

が成り立つ．

proof.(続き)

さらに，$\beta_m = -\sum_{i=1}^{m-1} \beta_i$ とすると，

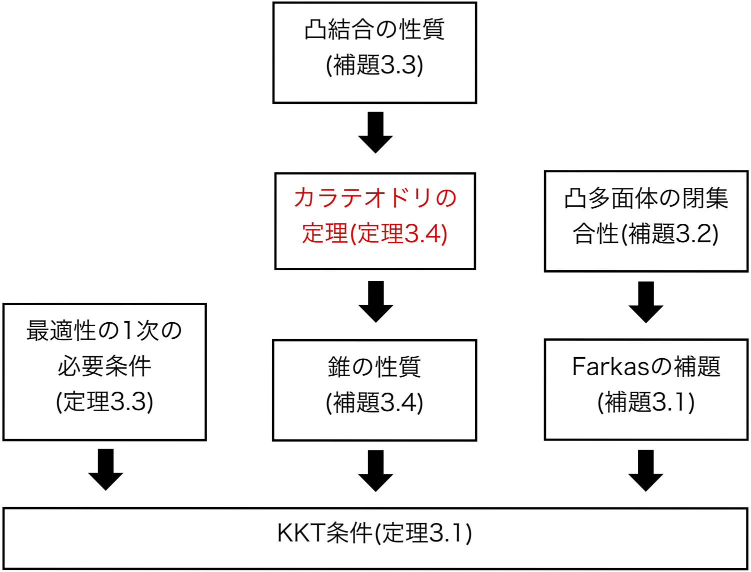
$$\sum_{i=1}^{m} \beta_i = 0, \sum_{i=1}^{m} \beta_i \boldsymbol{x}^i = \sum_{i=1}^{m-1} \beta_i \boldsymbol{x}^i + \beta_m \boldsymbol{x}^m = \mathbf{0}$$

が成り立つ．よって，任意の $\tau \in \mathbb{R}$ について，

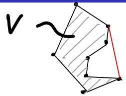
$$\begin{aligned} \boldsymbol{x} &= \sum_{i=1}^{m} \alpha_i \boldsymbol{x}^i = \sum_{i=1}^{m} \alpha_i \boldsymbol{x}^i - \tau \sum_{i=1}^{m} \beta_i \boldsymbol{x}^i = \sum_{i=1}^{m} (\alpha_i - \tau\beta_i) \boldsymbol{x}^i \\ \sum_{i=1}^{m} (\alpha_i - \tau\beta_i) &= \sum_{i=1}^{m} \alpha_i = 1 \end{aligned} \qquad (\star)$$

が成り立つ．今，$\bar{\tau} = \min_i \left\{ \dfrac{\alpha_i}{\beta_i} \mid \beta_i > 0 \right\}$ とすると，少なくとも 1 つの i に対して，$\alpha_i - \bar{\tau}\beta = 0$ であり，その他の i については，$\alpha_i - \bar{\tau}\beta > 0$ となる．よって，式 $(\star)$ から，$\boldsymbol{x}$ はたかだか $m-1$ 個の点の凸結合として表される．この操作を $m = n+2$ となるまで繰り返すことができるので，示せた． □

定理 3.4(カラテオドリの定理)



カラテオドリの定理



定義：凸包

集合 $V \subseteq \mathbb{R}^n$ を含む凸集合 $\bar{V} \supset V$ で，集合 $\bar{V}$ に真に含まれるような V を含む凸集合が存在しないとき，$\bar{V}$ を V の凸包と定義し，$\mathrm{co}\, V$ と表す．

定理 3.4(カラテオドリの定理)

$S \subseteq \mathbb{R}^n$ とする．このとき，

$$\mathrm{co}\, S = S^{n+1}$$

が成り立つ．ここで，

$$S^k = \left\{ \boldsymbol{x} \;\middle|\; \boldsymbol{x} = \sum_{i=1}^{k} \alpha_i \boldsymbol{x}^i, \boldsymbol{x}^i \in S, \alpha_i \geq 0, \sum_{i=1}^{k} \alpha_i = 1 \right\}$$

proof.

$S \subseteq S^{n+1} \subseteq \mathrm{co}\, S$，かつ，$S^{n+1}$ が凸集合であることを示せばよい．
まず，$S^{n+1} \subseteq \mathrm{co}\, S$ を帰納法で示す．$S^1 = S$ より，明らかに $S^1 \subseteq \mathrm{co}\, S$．
$k \in \mathbb{N}$ において，$S^k \subseteq \mathrm{co}\, S$ が成り立つと仮定する．このとき，$\boldsymbol{x} \in S^{k+1}$ とすると，S^{k+1} の定義より，

$$\boldsymbol{x} = \sum_{i=1}^{k} \alpha_i \boldsymbol{x}^i + \alpha_{k+1} \boldsymbol{x}^{k+1}, \alpha_i \geq 0, \sum_{i=1}^{k+1} \alpha_i = 1$$

を満たすような $\boldsymbol{x}^i \in S$ と α_i $(i = 1, \ldots, k+1)$ が存在する．
$\alpha_{k+1} = 1$ のとき，$\boldsymbol{x} = \boldsymbol{x}^{k+1} \in S$ より，明らかに $\boldsymbol{x} \in S$.
$\alpha_{k+1} < 1$ のとき，

$$\boldsymbol{x} = (1 - \alpha_{k+1}) \left(\sum_{i=1}^{k} \frac{\alpha_i}{1 - \alpha_{k+1}} \boldsymbol{x}^i \right) + \alpha_{k+1} \boldsymbol{x}^{k+1}$$

と表せる．

proof.(続き)

ここで，$\alpha_i \geq 0$, かつ,$\sum_{i=1}^{k+1} \alpha_i = 1$ より，$\frac{\alpha_i}{1-\alpha_{k+1}} \geq 0$．これと，

$$\sum_{i=1}^{k} \frac{\alpha_i}{1-\alpha_{k+1}} = \frac{1-\alpha_{k+1}}{1-\alpha_{k+1}} = 1$$

であることに注意すると，

$$\sum_{i=1}^{k} \frac{\alpha_i}{1-\alpha_{k+1}} \boldsymbol{x}^i \in S^k \subseteq \mathrm{co}\ S$$

が成り立つ．ここで，$\boldsymbol{x}^{k+1} \in \mathrm{co}\ S$ で $\mathrm{co}\ S$ は凸集合であるので，凸集合の定義より，

$$\boldsymbol{x} = (1-\alpha_{k+1}) \left(\sum_{i=1}^{k} \frac{\alpha_i}{1-\alpha_{k+1}} \boldsymbol{x}^i \right) + \alpha_{k+1} \boldsymbol{x}^{k+1} \in \mathrm{co}\ S$$

が成り立つ．よって，$S^{k+1} \subseteq \mathrm{co}\ S$ が成り立つ．以上より，数学的帰納法から，$S^{n+1} \subseteq \mathrm{co}\ S$.

proof.(続き)

次に，$S \subseteq S^{n+1}$ を示す．

$\alpha_1 = 1, \alpha_2 = \alpha_3 = \cdots = \alpha_{n+1} = 0$ と限定した S^{n+1} の要素の集合は S に一致するので，$S \subseteq S^{n+1}$.

さらに，S^{n+1} が凸集合であることを示す．

$\boldsymbol{x}, \boldsymbol{y} \in S^{n+1}$ とすると，S^{n+1} の定義より，

$$\begin{cases} \boldsymbol{x} = \sum_{i=1}^{n+1} \alpha_i \boldsymbol{x}^i, \alpha_i \geq 0, \sum_{i=1}^{n+1} \alpha_i = 1 \\ \boldsymbol{y} = \sum_{i=1}^{n+1} \beta_i \boldsymbol{y}^i, \beta_i \geq 0, \sum_{i=1}^{n+1} \beta_i = 1 \end{cases}$$

を満たす $\boldsymbol{x}^i, \boldsymbol{y}^i \in S$ と α_i, β_i $(i = 1, \ldots, n+1)$ が存在する．$\gamma \in [0, 1]$ とすると，

$$\gamma \boldsymbol{x} + (1-\gamma)\boldsymbol{y} = \sum_{i=1}^{n+1} \gamma \alpha_i \boldsymbol{x}^i + \sum_{i=1}^{n+1} (1-\gamma)\beta_i \boldsymbol{y}^i$$

となる．

proof.(続き)

ここで，$\gamma\alpha_i \geq 0, (1-\gamma)\beta_i \geq 0,$

$$\sum_{i=1}^{n+1} \gamma\alpha_i + \sum_{i=1}^{n+1} (1-\gamma)\beta_i = 1$$

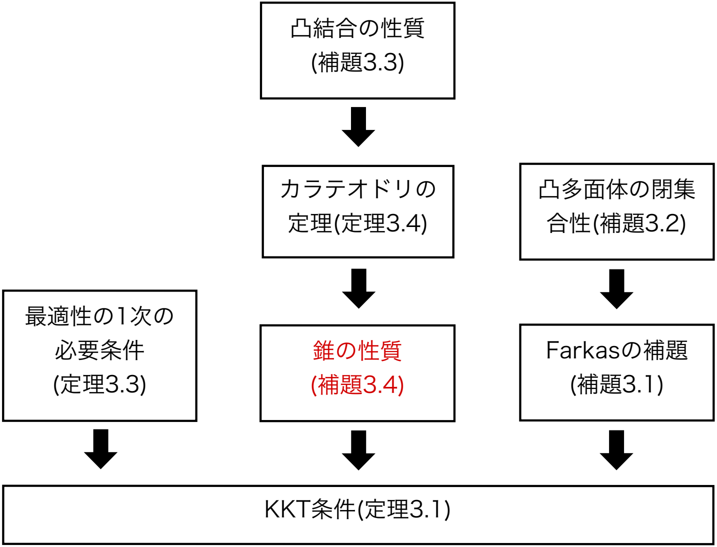
であることに注意すると，補題 3.3[*5] より，$\gamma\boldsymbol{x} + (1-\gamma)\boldsymbol{y}$ は $\{\boldsymbol{x}^i\}$ と $\{\boldsymbol{y}^i\}$ のうち，たかだか $n+1$ 個選んだ点の凸結合で表される．
すなわち，

$$\gamma\boldsymbol{x} + (1-\gamma)\boldsymbol{y} \in S^{n+1}, \forall \boldsymbol{x}, \boldsymbol{y} \in S^{n+1}, \forall\gamma \in [0,1]$$

が成り立つので，S^{n+1} は凸集合である．
以上より，S^{n+1} は $S \subseteq S^{n+1} \subseteq \mathrm{co}\, S$ を満たす凸集合であるが，$\mathrm{co}\, S$ は凸包の定義より，S を含む最小の凸集合なので，$\mathrm{co}\, S = S^{n+1}$. □

[*5] 点 $\boldsymbol{x} \in \mathbb{R}^n$ が $m (\geq n+2)$ 個の点の凸結合で表されているとき，$\boldsymbol{x}$ は，m 個の点からたかだか $n+1$ 個の点を選んで，それらの点の凸結合として表すことができる．

補題 3.4(錐の性質)


凸結合の性質
(補題3.3)
カラテオドリの
定理(定理3.4)
凸多面体の閉集
合性(補題3.2)
最適性の1次の
必要条件
(定理3.3)
錐の性質
(補題3.4)
Farkasの補題
(補題3.1)
KKT条件(定理3.1)

錐の性質

カラテオドリの定理を用いて，錐の性質を示す．

補題 3.4(錐の性質)

C と D を錐とする．このとき，次の命題が成り立つ．
(a). $C \subseteq D \implies C^* \supseteq D^*$
(b). $C^* = (\mathrm{co}\ C)^*$

proof(A).
$C \subseteq D$ とすると，極錐 D^* の定義より，

$$\boldsymbol{y} \in D^* \implies \langle \boldsymbol{y}, \boldsymbol{x} \rangle \leq 0, \forall \boldsymbol{x} \in C(\subseteq D)$$

が成り立つ．また，極錐 C^* の定義より，

$$\langle \boldsymbol{y}, \boldsymbol{x} \rangle \leq 0, \forall \boldsymbol{x} \in C \iff \boldsymbol{y} \in C^*$$

であるので，

$$\boldsymbol{y} \in D^* \implies \boldsymbol{y} \in C^*$$

すなわち，

$$C \subseteq D \implies C^* \supseteq D^*$$

□

proof(B).

$C \subseteq \mathrm{co}\ C$ より，(a) から，$C^* \supseteq (\mathrm{co}\ C)^*$.
任意の $\boldsymbol{x} \in \mathrm{co}\ C$ について，カラテオドリの定理 [*6] から，

$$\boldsymbol{x} = \sum_i \alpha_i \boldsymbol{x}^i, \sum_i \alpha_i = 1$$

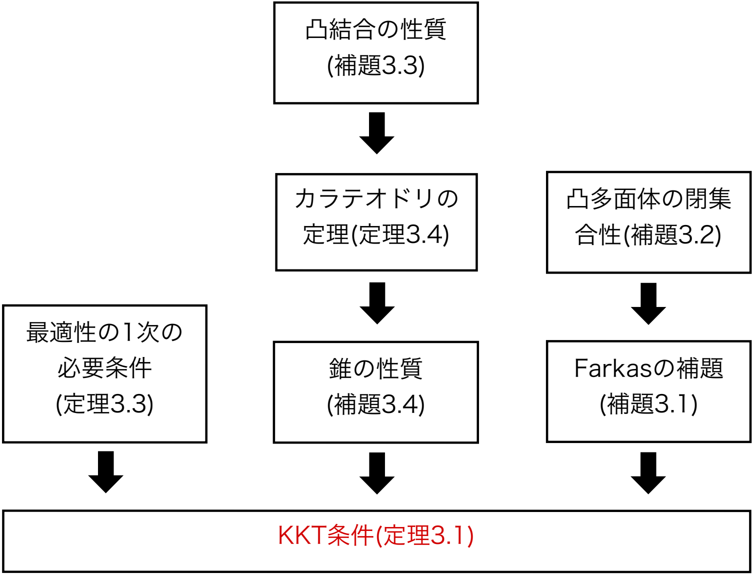
を満たす $\alpha_i \geq 0, \boldsymbol{x}^i \in C$ が存在する．
よって，任意の $\boldsymbol{y} \in C^*$ について，

$$\langle \boldsymbol{y}, \boldsymbol{x} \rangle = \sum_i \alpha_i \langle \boldsymbol{y}, \boldsymbol{x}^i \rangle \leq 0 \qquad (\because \alpha_i \geq 0, \boldsymbol{x}^i \in C, \boldsymbol{y} \in C^*)$$

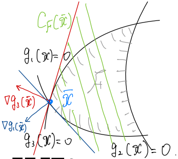
が成り立つ．
これは，極錐の定義から，$\boldsymbol{y} \in (\mathrm{co}\ C)^*$ を表している．よって，
$\boldsymbol{y} \in C^* \implies \boldsymbol{y} \in (\mathrm{co}\ C)^*$. すなわち，$C^* \subseteq (\mathrm{co}\ C)^*$ が成り立つ．
以上より，$C^* \supseteq (\mathrm{co}\ C)^*$ かつ $C^* \subseteq (\mathrm{co}\ C)^* \iff C^* = (\mathrm{co}\ C)^*$ □

[*6] $S \subseteq \mathbb{R}^n$ とする．このとき，$\mathrm{co}\ S = \left\{ \boldsymbol{x} \;\middle|\; \boldsymbol{x} = \sum_{i=1}^{n+1} \alpha_i \boldsymbol{x}^i, \boldsymbol{x}^i \in S, \alpha_i \geq 0, \sum_{i=1}^{n+1} \alpha_i = 1 \right\}$

定理 3.1(KKT 条件)


凸結合の性質
(補題3.3)
カラテオドリの
定理(定理3.4)
凸多面体の閉集
合性(補題3.2)
最適性の1次の
必要条件
(定理3.3)
錐の性質
(補題3.4)
Farkasの補題
(補題3.1)
KKT条件(定理3.1)

KKT 条件の証明



これまでの結果をまとめて，KKT 条件を導く．
以下では，簡単のため，不等式制約のみ [*7] を持つ問題についての KKT 条件を証明する．

定義：有効集合と線形化錐

実行可能解 $\bar{\boldsymbol{x}}$ に対して，$A(\bar{\boldsymbol{x}}) = \{i \mid g_i(\bar{\boldsymbol{x}}) = 0\}$ を有効集合という．また，$\mathcal{F}$ の $\bar{\boldsymbol{x}}$ における線形化錐を以下のように定義する．

$$C_{\mathcal{F}}(\bar{\boldsymbol{x}}) = \{\boldsymbol{y} \in \mathbb{R}^n \mid \langle \nabla g_j(\bar{\boldsymbol{x}}), \boldsymbol{y} \rangle \leq 0, j \in A(\bar{\boldsymbol{x}})\}$$

一般に $C_{\mathcal{F}}(\bar{\boldsymbol{x}})$ は凸集合であるが，$T_{\mathcal{F}}(\bar{\boldsymbol{x}})$ は一般には凸集合でない．
一般に，$T_{\mathcal{F}}(\bar{\boldsymbol{x}}) \subseteq C_{\mathcal{F}}(\bar{\boldsymbol{x}})$ が成り立つ．

[*7] 等式制約 $h_i(\bar{\boldsymbol{x}}) = 0$ は，2 つの不等式制約 $-h_i(\bar{\boldsymbol{x}}) \leq 0, h_i(\bar{\boldsymbol{x}}) \leq 0$ で表すことができる．

定理 (線形化錐と接錐の包含関係)

空でない集合 $\mathcal{F} \subseteq \mathbb{R}^n$ と任意の点 $\bar{\boldsymbol{x}} \in \mathcal{F}$ に対して，$T_{\mathcal{F}}(\bar{\boldsymbol{x}}) \subseteq C_{\mathcal{F}}(\bar{\boldsymbol{x}})$ が成立する．

proof.

$\boldsymbol{y} \in T_{\mathcal{F}}(\bar{\boldsymbol{x}})$ とすると，接錐の定義より，$\alpha_k(\boldsymbol{x}^k - \bar{\boldsymbol{x}}) \to \boldsymbol{y}$ を満たす $\{\boldsymbol{x}^k\} \subseteq \mathcal{F}$ と非負数列 $\{\alpha_k\}$ が存在する．

今，$\boldsymbol{x}^k \in \mathcal{F}$ であるから，全ての i について，

$$g_i(\boldsymbol{x}^k) = g_i(\bar{\boldsymbol{x}}) + \langle \nabla g_i(\bar{\boldsymbol{x}}), \boldsymbol{x}^k - \bar{\boldsymbol{x}} \rangle + o(\|\boldsymbol{x}^k - \bar{\boldsymbol{x}}\|) \leq 0 \qquad (k = 1, 2, \ldots)$$

が成り立つ．特に，$i \in A(\bar{\boldsymbol{x}})$ ならば，$g_i(\bar{\boldsymbol{x}}) = 0$ が成り立つので，

$$\langle \nabla g_i(\bar{\boldsymbol{x}}), \boldsymbol{x}^k - \bar{\boldsymbol{x}} \rangle + o(\|\boldsymbol{x}^k - \bar{\boldsymbol{x}}\|) \leq 0 \qquad (k = 1, 2, \ldots)$$

となる．この左辺に α_k をかけて，$k \to \infty$ の極限を考えると，

$$\begin{aligned}
&\alpha_k \langle \nabla g_i(\bar{\boldsymbol{x}}), \boldsymbol{x}^k - \bar{\boldsymbol{x}} \rangle + \alpha_k o(\|\boldsymbol{x}^k - \bar{\boldsymbol{x}}\|) \\
&= \langle \nabla g_i(\bar{\boldsymbol{x}}), \alpha_k(\boldsymbol{x}^k - \bar{\boldsymbol{x}}) \rangle + \frac{o(\|\boldsymbol{x}^k - \bar{\boldsymbol{x}}\|)}{\|\boldsymbol{x}^k - \bar{\boldsymbol{x}}\|} \|\alpha_k(\boldsymbol{x}^k - \bar{\boldsymbol{x}})\| \\
&\to \langle \nabla g_i(\bar{\boldsymbol{x}}), \boldsymbol{y} \rangle \qquad (as\ k \to \infty)
\end{aligned}$$

となる．よって，任意の $\boldsymbol{y} \in T_{\mathcal{F}}(\bar{\boldsymbol{x}})$ について，$\langle \nabla g_i(\bar{\boldsymbol{x}}), \boldsymbol{y} \rangle \leq 0$ ($i \in A(\bar{\boldsymbol{x}})$)，すなわち，$\boldsymbol{y} \in C_{\mathcal{F}}(\bar{\boldsymbol{x}})$ が成り立つので，$T_{\mathcal{F}}(\bar{\boldsymbol{x}}) \subseteq C_{\mathcal{F}}(\bar{\boldsymbol{x}})$． □

制約想定 $C_{\mathcal{F}}(\bar{\boldsymbol{x}}) \subseteq \mathrm{co}\, T_{\mathcal{F}}(\bar{\boldsymbol{x}})$ が成り立つという仮定の下で，KKT 条件を証明する．

定理 3.5(KKT 条件)

制約想定 $C_{\mathcal{F}}(\bar{\boldsymbol{x}}) \subseteq \mathrm{co}\, T_{\mathcal{F}}(\bar{\boldsymbol{x}})$ が成り立つという仮定の下で，
$\bar{\boldsymbol{x}}$ が局所的最適解 $\Longrightarrow$ $\exists \boldsymbol{\mu} \in \mathbb{R}^r \qquad s.t.$

$$\nabla f(\bar{\boldsymbol{x}}) + \sum_{j=1}^{r} \mu_j \nabla g_j(\bar{\boldsymbol{x}}) = \mathbf{0}$$

$$\mu_j \geq 0, g_j(\bar{\boldsymbol{x}}) \leq 0, \mu_j g_j(\bar{\boldsymbol{x}}) = 0 \qquad (j = 1, \ldots, r)$$

proof.
$\bar{x}$ が局所的最適解であることから，定理 3.3 より，

$$-\nabla f(\bar{x}) \in N_{\mathcal{F}}(\bar{x})$$

が成り立つ．一方，制約想定 $C_{\mathcal{F}}(\bar{x}) \subseteq \mathrm{co}\, T_{\mathcal{F}}(\bar{x})$ と補題 3.4[*8] より，

$$C_{\mathcal{F}}(\bar{x})^* \supseteq (\mathrm{co}\, T_{\mathcal{F}}(\bar{x}))^* = T_{\mathcal{F}}(\bar{x})^* = N_{\mathcal{F}}(\bar{x})$$

が成り立つ．よって，

$$-\nabla f(\bar{x}) \in C_{\mathcal{F}}(\bar{x})^*$$

が成り立つ．

[*8] C と D を錐とする．このとき，次が成り立つ．
(A) $C \subseteq D \implies C^* \supseteq D^*$　　(B) $C^* = (\mathrm{co}\, C)^*$

proof.(続き)

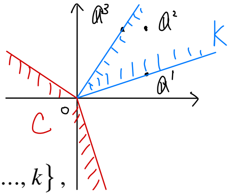
ここで，さらに，補題 3.1[*9] より，

$$
\begin{aligned}
C_{\mathcal{F}}(\bar{\boldsymbol{x}})^* &= \{\boldsymbol{y} \mid \langle \nabla g_i(\bar{\boldsymbol{x}}), \boldsymbol{y} \rangle \leq 0, i \in A(\bar{\boldsymbol{x}})\}^* \\
&= \left\{ \boldsymbol{y} \;\middle|\; \boldsymbol{y} = \sum_{i \in A(\bar{\boldsymbol{x}})} \mu_i \nabla g_i(\bar{\boldsymbol{x}}), \mu_i \geq 0, i \in A(\bar{\boldsymbol{x}}) \right\}
\end{aligned}
$$

であるから，

$$
\nabla f(\bar{\boldsymbol{x}}) + \sum_{j \in A(\bar{\boldsymbol{x}})} \mu_j \nabla g_j(\bar{\boldsymbol{x}}) = \mathbf{0}
$$

を満たす $\mu_j \geq 0, j \in A(\bar{\boldsymbol{x}})$ が存在する．



[*9] $\boldsymbol{a}^i \in \mathbb{R}^n$ $(i = 1, \ldots, k)$ として，$C := \left\{\boldsymbol{y} \in \mathbb{R}^n \mid \langle \boldsymbol{y}, \boldsymbol{a}^i \rangle \leq 0, i = 1, ..., k\right\}$，
$K := \left\{ \boldsymbol{x} \in \mathbb{R}^n \;\middle|\; \boldsymbol{x} = \sum_{i=1}^{k} \lambda_i \boldsymbol{a}^i, \lambda_i \geq 0, i = 1, \ldots, k \right\}$ としたとき，$K = C^*$ が成り立つ．

proof.(続き)

$j \notin A(\bar{\boldsymbol{x}})$ に対しては，$\mu_j = 0$ と置けば，

$$\nabla f(\bar{\boldsymbol{x}}) + \sum_{j=1}^{r} \mu_j \nabla g_j(\bar{\boldsymbol{x}}) = \mathbf{0}$$

が成立する．さらに，$A(\bar{\boldsymbol{x}})$ と μ_j の定義から，μ_j は相補性条件を満たす． □

等式制約ありの問題に対する一般化

不等式制約のみを持つ問題についての KKT 条件が定理 3.5 で与えられることが示された．
等式制約 $h_i(\bar{\boldsymbol{x}}) = 0\ (i = 1, \ldots, m)$ を持つ一般の問題 (3.1) については，線形化錐を以下のように定義すれば良い．

$$C_{\mathcal{F}}(\bar{\boldsymbol{x}}) = \{\boldsymbol{y} \in \mathbb{R}^n \mid \langle \nabla h_i(\bar{\boldsymbol{x}}), \boldsymbol{y}\rangle \leq 0, \langle -\nabla h_i(\bar{\boldsymbol{x}}), \boldsymbol{y}\rangle \leq 0,\ i = 1, \ldots, m^{*10}, \\ \langle \nabla g_j(\bar{\boldsymbol{x}}), \boldsymbol{y}\rangle \leq 0,\ j \in A(\bar{\boldsymbol{x}})\}$$

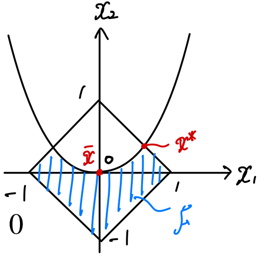
こうすることで，KKT 条件の証明の議論をそのまま拡張することができる．
KKT 条件が成り立つ前提として，制約想定 $C_{\mathcal{F}}(\bar{\boldsymbol{x}}) \subseteq \mathrm{co}\, T_{\mathcal{F}}(\bar{\boldsymbol{x}})$ が必要であったが，この条件が成り立つかどうかのチェックは一般には難しい．
そこで，次節では，この条件が成り立つ十分条件を与える．

[*10] j とは異なり，任意の i は有効集合に含まれることに注意する．

KKT 条件の使用例

以下の問題について考える．

$$\begin{array}{ll} \underset{\boldsymbol{x}\in\mathbb{R}^2}{\text{minimize}} & f(\boldsymbol{x}) = -x_2 \\ \text{subject to} & g_1(\boldsymbol{x}) = -x_1^2 + x_2 \le 0 \\ & g_2(\boldsymbol{x}) = |x_1| + |x_2| - 1 \le 0 \end{array}$$



局所的最適解 $\boldsymbol{x}^* = \left(\frac{-1+\sqrt{5}}{2}, \frac{3-\sqrt{5}}{2}\right)^\top$ について，$\boldsymbol{x}^*$ における有効集合は $A(\boldsymbol{x}^*) = \{1,2\}$, $\nabla g_1(\boldsymbol{x}^*) = \left(1-\sqrt{5}, 1\right)^\top$, $\nabla g_2(\boldsymbol{x}^*) = (1,1)^\top$ より，$\mathcal{F}$ の $\boldsymbol{x}^*$ における線形化錐は，

$C_{\mathcal{F}}(\boldsymbol{x}^*) = \left\{\boldsymbol{y} \in \mathbb{R}^2 \;\middle|\; \left(1-\sqrt{5}\right) y_1 + y_2 \le 0, y_1 + y_2 \le 0\right\}$ となる．

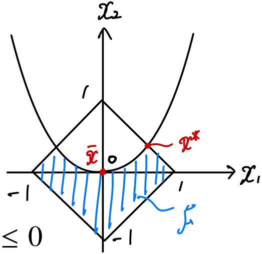
また，$T_{\mathcal{F}}(\boldsymbol{x}^*) = \left\{\boldsymbol{y} \in \mathbb{R}^2 \;\middle|\; \left(1-\sqrt{5}\right) y_1 + y_2 \le 0, y_1 + y_2 \le 0\right\}$ より，$\boldsymbol{x}^*$ は制約想定 $C_{\mathcal{F}}(\boldsymbol{x}^*) \subseteq \mathrm{co}\, T_{\mathcal{F}}(\boldsymbol{x}^*)$ を満足する．

さらに，$\nabla f(\boldsymbol{x}^*) = (0,-1)^\top$ より，$\boldsymbol{\mu} = \left(\frac{1}{\sqrt{5}}, \frac{\sqrt{5}-1}{\sqrt{5}}\right)$ とすれば，$(\boldsymbol{x}^*, \boldsymbol{\mu})$ は KKT 点となる．

KKT 条件を満たすが，局所的最適解でない点の例

以下の問題について考える．(1 枚前と同じ)

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^2}{\text{minimize}} \quad & f(\boldsymbol{x}) = -x_2 \\ \text{subject to} \quad & g_1(\boldsymbol{x}) = -x_1^2 + x_2 \le 0 \\ & g_2(\boldsymbol{x}) = |x_1| + |x_2| - 1 \le 0 \end{aligned}$$



$\bar{\boldsymbol{x}} = (0,0)^\top$ について，$\bar{\boldsymbol{x}}$ における有効集合は $A(\bar{\boldsymbol{x}}) = \{1\}, \nabla g_1(\bar{\boldsymbol{x}}) = (0,1)^\top$ より，$\mathcal{F}$ の $\bar{\boldsymbol{x}}$ における線形化錐は $C_{\mathcal{F}}(\bar{\boldsymbol{x}}) = \left\{\boldsymbol{y} \in \mathbb{R}^2 \mid y_2 \le 0\right\}$ となる．
また，$\mathcal{F}$ の $\bar{\boldsymbol{x}}$ における接錐は $T_{\mathcal{F}}(\bar{\boldsymbol{x}}) = \left\{\boldsymbol{y} \in \mathbb{R}^2 \mid y_2 \le 0\right\}$ なので，$\bar{\boldsymbol{x}}$ は制約想定 $C_{\mathcal{F}}(\bar{\boldsymbol{x}}) \subseteq \mathrm{co}\, T_{\mathcal{F}}(\bar{\boldsymbol{x}})$ を満足する．
さらに，$\nabla f(\bar{\boldsymbol{x}}) = (0,-1)^\top$ より，$\boldsymbol{\mu} = (1,0)^\top$ とすれば，$(\bar{\boldsymbol{x}}, \boldsymbol{\mu})$ は KKT 点となるが，明らかに $\bar{\boldsymbol{x}}$ は局所的最適解ではない．