第3章　最適性の条件

前回までで，数理計画問題がどのような問題であるかを説明した．その数理計画問題を計算機で解こうとした場合，どうしたらよいのだろうか？問題が与えられただけでは途方にくれてしまう．計算機ができることは，ある点 $x \in R^n$ が与えられたとき，その点での関数の値 $f(x), g(x), h(x)$ や，それらの微分の値を計算するだけである．そのような大域的な情報が得られない計算機において，大域的最小解を求めることは容易ではない．そこで局所的最小解を求めることになるのだが，それでは，どのような条件を満たした点 x が，局所的最小解とみなすことができるのだろうか？もちろん，その条件は，計算機で容易に判別できるものでなければならない．この節の目的は，そのような最適性の条件を与えることである．

この章では，まず、制約なし最小化問題の最適性の条件を導き，それを用いて，制約つき最小化問題の最適性の条件を与えることにする．

3.1　制約無し最小化問題

制約なし問題の最適性条件は直感的にも理解しやすく、またその証明も簡単である．そこで、この節では、制約なし最小化問題の最適性の条件を実際に導く。

最適性の条件は、次のように必要条件と十分条件の 2 つにわけることができる。

$$\begin{array}{ccc} \text{点 } x^* \text{ が局所的最小解である} & \Rightarrow & x^* \text{ は最適性の必要条件を満たす} \\ \text{点 } x^* \text{ は最適性の十分条件を満たす} & \Rightarrow & x^* \text{ が局所的最小解である} \end{array}$$

もちろん、本来の目的からは、必要十分条件を導くことが望ましい。しかし、最適性の必要十分条件を持つような数理計画問題は、凸計画問題など限られた問題の種類に限られる。そこで、よりよい十分条件を求めることを考える。

十分条件は必要条件を満たさなければならないので、まず必要条件を与える。ここで，x^* は局所的最小解であるとする．このとき，x^* から d 方向に少しずらした点を考える．そのずれを $\alpha \in R$ で表すこととすると，ずらした点は $x^* + \alpha d$ となる．x^* は局所的最適解であるので，α が十分小さければ，

$$f(x^* + \alpha d) - f(x^*) \geq 0$$

が成り立つ．よって，α を正の領域から 0 に近づければ，合成関数の微分を考えれば、

$$0 \leq \lim_{\alpha \downarrow 0} \frac{f(x^* + \alpha d) - f(x^*)}{\alpha} = \nabla f(x^*)^T d$$

を得る．このことがすべての方向 d に対して成り立たなければならないので，$\nabla f(x^*) = 0$ でなければならない．この条件を制約なし最小化問題の __________ と呼ぶ.

次に，f が 2 回微分可能とする．f に対して x^* のまわりで 2 次のテーラー展開をすると，

$$f(x^* + \alpha d) - f(x^*) = \alpha \nabla f(x)^T d + \frac{\alpha^2}{2} d^T \nabla^2 f(x^*) d + o(\alpha^2)$$

を得る．1 次の必要条件より $\nabla f(x^*) = 0$ であり，さらに x^* が局所的最小解であることから，

$$0 \leq \frac{f(x^* + \alpha d) - f(x^*)}{\alpha^2} = \frac{1}{2} d^T \nabla^2 f(x^*) d + \frac{o(\alpha^2)}{\alpha^2}$$

を得る．ここで，$\alpha \to 0$ とすると，

$$0 \leq d^T \nabla^2 f(x^*) d$$

を得る．これは，すべての方向 d で成り立たなければならないので，$\nabla^2 f(x^*)$ は半正定値行列であることを意味している．これを，__________ と呼ぶ.

以上のことをまとめると次の定理を得る．

定理 1 x^* は f の制約なし最小化問題の局所的最小解とする．また，f は x^* の近傍で微分可能とする．このとき，

$$\nabla f(x^*) = 0 \tag{3.1}$$

が成り立つ (**最適性の 1 次の必要条件**)．さらに，f が 2回微分可能とする．このとき，$\nabla^2 f(x^*)$ は半正定値行列である (**最適性の 2 次の必要条件**).

半正定値行列 任意のベクトル $x \in R^n$ に対して、

$$x^T A x \geq 0$$

となる n 行 n 列行列 A を半正定値行列と呼ぶ。さらに、0 でない任意のベクトル $x \in R^n$ に対して、

$$x^T A x > 0$$

が成り立てば、この行列を正定置行列と呼ぶ。たいていの線形代数の教科書で半正定値行列は扱われているが、この行列がどのようなものかイメージするのは難しい。半正定値行列は数理計画法で非常に重要な役割を果たす行列の概念である。これは、後の章で示すように、凸関数のヘシアンは半正定値行列になるからである。また、逆にすべての領域でヘシアンが半正定値行列となる関数は凸関数となる。半正定値性は、スカラーの非負性を、行列に対して一般化したものだと考えられる。実際に $n = 1$ のとき、行列 (スカラー)A が半正定値であるとは $A \geq 0$ を意味しており、また、正定置であるとは、$A > 0$ を意味している。

1 次の必要条件を満たしている点を f の __________ と呼ぶことがある．ここで、次の問題を考えてみよう。

$$\min \quad e^{x_1^2 + x_2^2} \tag{3.2}$$

この問題の一次の必要条件は、

$$\nabla e^{x_1^2+x_2^2} = \begin{pmatrix} 2x_1 e^{x_1^2+x_2^2} \\ 2x_1 e^{x_1^2+x_2^2} \end{pmatrix} = 0$$

となり、この条件を満たす点は、$(x_1, x_2) = (0, 0)$ だけである。一方、あきらかにこの点は、この問題の大域的最適解である。さらに、

$$\nabla^2 e^{x_1^2+x_2^2} = \begin{pmatrix} 2e^{x_1^2+x_2^2} + 4x_1^2 e^{x_1^2+x_2^2} & 4x_1 x_2 e^{x_1^2+x_2^2} \\ 4x_1 x_2 e^{x_1^2+x_2^2} & 2e^{x_1^2+x_2^2} + 4x_2^2 e^{x_1^2+x_2^2} \end{pmatrix}$$

であるので、この点における最適性の 2 次の条件を調べてみると、

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

となり、確かに半正定値行列になっている。この問題に限って言えば、1 次の"必要"条件は"十分"条件にもなっている。しかし、次の例が示すように、局所的最小解の必要条件を与えているだけで，十分条件は与えていないことに注意しよう．実際，$f(x) = -x^4$ という関数を考えてみよう．このとき，$\nabla f(0) = \nabla^2 f(0) = 0$ となるため，$\hat{x} = 0$ は 1 次，2 次両方の必要条件をみたしているが，0 はこの問題の局所的最大値であり，最小値ではない．

それでは，局所的最適解となるための十分条件になる条件は何であろうか？先ほどの例では，$-x^4$ という関数が凹関数 (マイナスにすると凸関数になる関数) であるということに気づいたであろうか？もし，$-x^4$ が凸関数である x^4 であれば，明らかに 0 は最小解であったはずである．このことは，実は一般の凸関数に対しても言えるのである．ここで，f は凸関数で，$\nabla f(x^*) = 0$ だったとしよう．さらに，$\bar{x}$ をこの問題の大域的最小解としよう．すると，凸関数の定義より，

$$f(\alpha\bar{x} + (1-\alpha)x^*) \geq \alpha f(\bar{x}) + (1-\alpha)f(x^*)$$

が成り立つ．この式を変形すると，

$$\frac{f(x^* + \alpha(\bar{x} - x^*) - f(\alpha x^*)}{\alpha} \geq f(\bar{x}) - f(x^*)$$

を得る．この左辺において，$\alpha \to 0$ とすると，左辺は $\nabla f(x^*)$ となり，x^* が停留点であったことより 0 になる．結局，

$$f(x^*) \leq f(\bar{x})$$

を得る．$\bar{x}$ が大域的最適解であったので，x^* も大域的最適解となる．以上のことをまとめると，

定理 2 目的関数 f が凸であるとする．このとき，停留点は大域的最小解である．また，f が狭義の凸関数，つまり，

$$f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha)f(y) \ \ \forall\alpha \in [0,1] \ \ \forall x, y \in R^n \ \ with \ x \neq y$$

を満たせば，停留点 *(*大域的最小解*)* はたかだかひとつである． □

となる．

次に，目的関数が凸関数でない一般の関数の場合の最適性の十分条件を考えてみよう．大域的最適解であることまで保証しなければ，x^* のまわりで f が (狭義の) 凸関数であり，1 次の十分条件が成り立っていれば，上記の定理より，x^* は局所的最小解であるということが期待できる (図 3.1).

以上のことをまとめると最適性の十分条件は以下のように与えられる。

定理 3 f を *2* 回微分可能関数とする．このとき，x^* において

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

かつ $\nabla^2 f(x^*)$ が正定値行列とする *(*最適性の十分条件*)*．このとき，x^* は f の制約のない最小化問題の狭義の局所的最小点となる．

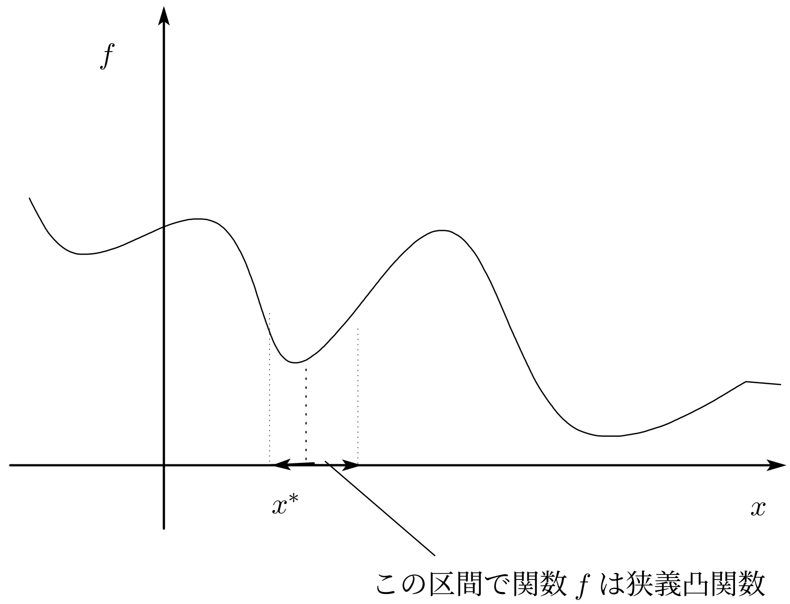

図 3.1: 最適性の十分条件と目的関数の凸性

この定理を証明してみよう。$\nabla^2 f(x^*)$ は正定値行列であるので，次の不等式を満たす正の定数 λ が存在する．

$$d^T \nabla^2 f(x^*) d \geq \lambda \|d\|^2$$

よって，2 次のテーラー展開をすると，

$$f(x^* + d) - f(x^*) = \nabla f(x)^T d + \frac{1}{2} d^T \nabla^2 f(x^*) d + o(\|d\|^2) \geq \frac{\lambda}{2} \|d\|^2 + o(\|d\|^2)$$

を得る．オーダー記号 o の定義より，十分小さい d を考えると，

$$\frac{\lambda}{2} \|d\|^2 + o(\|d\|^2) \leq \frac{\lambda}{4} \|d\|^2$$

となるので

$$f(x^* + d) - f(x^*) \geq \frac{\lambda}{4} \|d\|^2$$

を得る．よって，x^* は f の制約のない最小化問題の狭義の局所的最小解である．

始めに考えた問題 (3.2) では、

$$\nabla^2 f(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

は正定値行列であったので、確かに十分条件が成り立っている。一方、$f(x) = -x^4$ では、$\nabla^2 f(0) = 0$ であり、十分条件が成り立っていない。

このように簡単な問題では、最適性の条件を手計算で求めることによって、最適解を求めることができる場合が多い。一方、複雑な問題でも、後の章で紹介する反復法によって、最適性の 1 次の必要条件を求めることが可能である。そのとき、反復法の終了条件としては、最適性の 1 次の必要条件がどれくらい満たされているか調べることによって可能である。

$$\|\nabla f(x^k)\| \leq \varepsilon$$

ここで ε は、求めたい”精度”を表している。

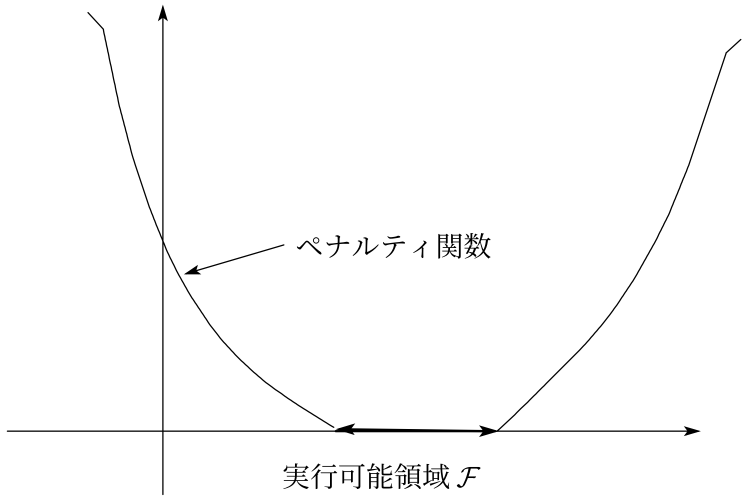

図 3.2: ペナルティ関数

3.2 Karush-Kuhn-Tucker 条件

この節では、制約付き最小化問題の最適性の条件を考える．この条件は ＿＿＿＿＿＿ と呼ばれる数理計画法では非常に重要なものである．この条件を厳密に証明することは，かなりの労力を必要とするので，ここでは，直感的に導出をすることにする．

まず，制約付き最小化問題を，近似的に制約無し最小化問題に置き換えることを考える．そのために，制約をみたしていないときには，その罰として，大きな値を目的関数に付け加えることにする．そのような罰を与える関数を，一般にはペナルティー関数と呼ぶ (図 3.2).

ここでは，等式制約に対するペナルティー関数として $\|h(x)\|^2$ を，不等式制約に対するペナルティー関数として $\sum_{j=1}^m \max\{0, g_j(x)\}^2$ を使うことにする．このとき，ペナルティー関数を加えた，目的関数は，

$$\theta_\gamma(x) := f(x) + \gamma\|h(x)\|^2 + \gamma \sum_{j=1}^m \max\{0, g_j(x)\}^2$$

となる．ここで，γ はペナルティーの大きさを調節するパラメータで、ペナルティーパラメータと呼ばれる．実際,

$$\omega(x) := \lim_{\gamma\to\infty} (\gamma\|h(x)\|^2 + \gamma \sum_{j=1}^m \max\{0, g_j(x)\}^2)$$

とすると,

$$\omega(x) = \begin{cases} 0 & x \text{ が制約を満たすとき} \\ \infty & x \text{ が制約を満たさないとき} \end{cases}$$

となる．このペナルティー関数を用いた次の制約なし最小化問題を考えよう．

$$\min \quad \theta_\gamma(x) \tag{3.3}$$

この問題の最適解は，先ほど述べたペナルティー関数の性質より，$\gamma \to \infty$ とすると，元の問題の最適解と同じになることが予想される．最小化問題 (3.3) は，制約なしの問題であるので，この問題の最適性の 1 次の必要条件は,

$$0 = \nabla\theta_\gamma(x) = \nabla f(x) + \gamma\nabla h(x)h(x) + \gamma\nabla g(x)\hat{g}(x) \tag{3.4}$$

となる．ここで，

$$\hat{g}(x) = \begin{pmatrix} \max\{0, g_1(x)\} \\ \max\{0, g_2(x)\} \\ \vdots \max\{0, g_r(x)\} \end{pmatrix}$$

である．この必要条件に対して，$\gamma \to \infty$ としたら，制約無し最小化問題の必要条件らしきものが出てきそうである．そこで，まず次の仮定が成り立つとしよう。

仮定 1：　$\gamma \to 0$ のとき，(3.4) を満たす $x, \gamma h(x)$ と $\gamma \hat{g}(x)$ それぞれ，x^*, μ^* と λ^* に収束する．

このとき，

$$0 = \nabla f(x^*) + \nabla h(x^*)\mu^* + \nabla g(x^*)\lambda^*$$

となる．さらに，先ほどの議論より，$h(x^*) = 0$ かつ $g(x^*) \le 0$ でなければならない．また，$\hat{g}(x^*)$ の定義より，明らかに $\lambda^* \ge 0$ である．さらに，$g_i(x^*) < 0$ であれば、$\hat{g}_i(x^*) = 0$ となるので，$\lambda_i^* = \lim_{\gamma\to\infty} \gamma \hat{g}_i(x) == 0$ となる．以上のことより，$\lambda_i^* g_i(x^*) = 0$ がすべての i に対して成り立つ．

ところで、これらの性質が成り立つためには、仮定 1 が成り立つ必要があった。一般にはこのような仮定は成り立たないが、以下に定義する __________ が成り立てば、仮定 1 が成り立つことを示すことができる。

まず，不等式制約に対して，集合 A を次のように定義する．

$$A(x) = \{i \mid g_i(x) = 0\}$$

この集合を __________ と呼ぶ．このとき，実行可能点 x に対して，$\nabla h_i(x), i = 1, \ldots, m$ と $\nabla g_j(x), j \in A(x)$ が 1 次独立であるとき，x において 1 次独立の制約想定 (LICQ) が成り立つという．

この LICQ のもと、制約つき最小化問題の最適性に 1 次の必要条件が以下のように与えられる。ここで、__________ $L : R^{n+m+r}$ を以下のように定義する。

$$L(x, \lambda, \mu) = f(x) + \lambda^T h(x) + \mu^T g(x) \tag{3.5}$$

ここで，λ，μ は，それぞれ，制約 $h(x) = 0$，$g(x) \le 0$ に対する __________ と呼ばれる m, r 次元ベクトルである．

定理 4 x^* を制約つき最小化問題

$$\begin{array}{ll} \min & f(x) \\ \text{subject to} & h(x) = 0 \\ & g(x) \le 0 \end{array}$$

の局所的最小解とし，x^* で *LICQ* が成り立つとする．このとき， 次の等式を満たすラグランジュ乗数 λ^* と μ^* が一意に求まる．

$$\nabla_x L(x^*, \lambda^*, \mu^*) = 0, \nabla_\lambda L(x^*, \lambda^*, \mu^*) = 0 \tag{3.6}$$

$$\mu^* \ge 0,\ \ g(x^*) \le 0,\ \ g(x)^T \mu^* = 0. \tag{3.7}$$

この定理において，(3.6) と (3.7) を合わせて，**Karush-Kuhn-Tucker 条件 (KKT 条件)** と呼ばれている．図 3.3 のような物理現象を考えると，KKT 条件は理解しやすい．この図では，等高線 $f(x) = c$ で示された地形上をボールが転がることを考えている．また，ボールが入ってはい

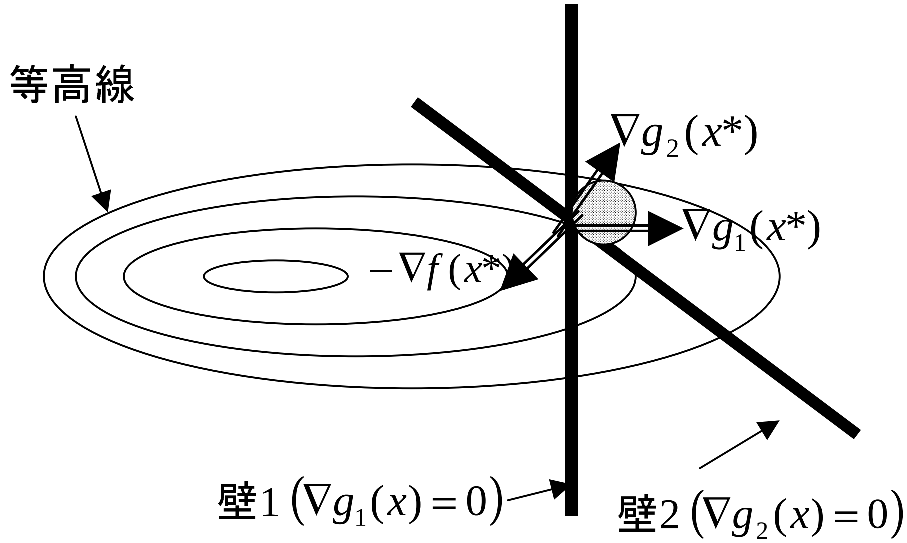

図 3.3: KKT 条件の物理的解釈

けないところは，壁 $g_j(x)=0$ によって塞がれている．このとき，ボールが静止するのは，局所的に窪んでおり，壁に押さえられているところである．そのため，静止しているボールにかかる力は均衡している．この均衡を表すのが，式 (3.6) である．そのとき，$\nabla f(x^*)$ は重力から受ける力に表しており，$\nabla g(x^*)\mu^*$ は壁から受ける反作用の力を表している．(3.7) で $\mu_i^*=0$ となるということは，壁に接していないこと $(g_i(x^*)<0)$ を意味している．

なお制約想定が成り立たない場合、定理が成り立たない例として以下のものがあげられる。

$$\begin{array}{ll} \min & x \\ \text{subject to} & x^2 = 0 \end{array}$$

この問題の最適解は、$x=0$ である。このとき、$\nabla h(0)=2\times 0=0$ となり、LICQ はなりなっていない。一方、$\nabla f(0)=1$ であり、

$$0=\nabla_x L(x^*,\lambda^*)=1+\lambda^*\times 0=1$$

となり、定理を満たす λ^* は存在しない。

次に、制約付き最小化問題の 2 次の必要条件を与えよう。

	制約なし最小化問題	制約つき最小化問題
ヘシアン	目的関数	ラグランジュ関数
微小方向 d	全空間 R^n	$T(x^*)$

表 3.1: 制約なし最小化問題と制約付き最小化問題の 2 次の必要条件

定理 5 x^* を非線形最小化問題

$$\begin{array}{ll} \min & f(x) \\ \text{subject to} & h(x) = 0 \\ & g(x) \leq 0 \end{array}$$

の局所的最小解とし，x^* で *LICQ* が成り立つとする．さらに f と h が *2* 回微分可能であるならば，

$$y^T \nabla^2_{xx} L(x^*, \lambda^*) y \geq 0 \ \ \forall y \in T(x^*)$$

ここで $T(x^*)$ は

$$T(x^*) = \{y | \nabla h_i(x^*)^T y = 0, i = 1, \ldots, m, \nabla g_j(x^*)^T y = 0 \ \ \forall j \in A(x^*)\}.$$

である．

証明は大変なため省略するが、簡単な意味を説明しておこう。まず，集合 $T(x^*)$ は、x^* から $h(x) = 0$, $g_j(x) = 0, j \in A(x^*)$ を満たすように微小に動かすことができる方向 d の集合である。そして、その方向に対して、ラグランジュ関数は半正定値行列であるということを意味している。このことは、さらに、その方向に対してラグランジュ関数の値が増えないことを意味している。この関係は制約なし最小化問題の最適性の 2 次の必要条件と対応付けて考えると理解しやすい (表 3.1)。

後の章で紹介する制約付き最小化問題に対する多くの解法では，この Karush-Kuhn-Tucker 条件を満たす点 (**KKT 点**) を求めることが目的となっている．もちろん，この条件は必要条件であり，この条件をみたしているからといって，x^* が局所的最小解となるとは限らない．例えば、

$$\begin{array}{ll} \min & -x^4 \\ \text{subject to} & x^2 - 1 \leq 0 \end{array}$$

を考えてみよう。このとき、$x = 0$ は必要条件を満たしているが、この点は局所的最小解ではない。一方，f, g が凸関数, h がアフィン関数で与えられていれば，LICQ のもと，KKT 条件は必要十分条件になることが知られている．

定理 6 f, g が凸関数, h がアフィン関数であるとする。点 x^* 点いおいて *LICQ* が成り立つとする。このとき、x^* が最小化問題の大域的最適解であることの必要十分条件は *Karush-Kuhn-Tucker* 条件を満たす (x^*, λ^*, μ^*) が存在することである。 □

練習問題 1 次の問題の解を求めよ。

$$\begin{array}{ll} \min & x_1^2 + x_2^2 \\ \textit{subject to} & -x_1 - x_2 + 1 \leq 0 \end{array}$$

次に f, g を凸関数と仮定しない場合における十分条件を与える．

定理 7 f と h と g が *2* 回微分可能であるとする．x^* と λ^* と μ^* は次の条件を満たすとする．

$$\nabla_x L(x^*, \lambda^*, \mu^*) = 0, \nabla_\lambda L(x^*, \lambda^*, \mu^*) = 0$$

$$\mu^* \geq 0, \quad g(x^*) \leq 0, \quad g(x^*)^T \mu^* = 0.$$

$$y^T \nabla_{xx}^2 L(x^*, \lambda^*, \mu^*) y > 0 \ \ \forall y \in T(x^*) \ \textit{with} \ y \neq 0$$

ここで，$T(x^*)$ は

$$T(x^*) = \{y | \nabla h_i(x^*)^T y = 0, i = 1, \ldots, m, \nabla g_j(x^*)^T y = 0 \ \ \forall j \in A(x^*)\}.$$

である．さらに，

$$\mu_j^* - g_j(x^*) > 0 \quad \forall j \tag{3.8}$$

が成り立つとする．このとき，x^* は問題の狭義局所的最小解である． □

この定理の意味するところは、制約条件を満たすように変化するときにはラグランジュ関数は増加するということである。つまりラグランジュ関数が狭義の凸関数のようなものになっているということができる。

この定理中の (3.8) を，__________ と呼ぶ．この定理で示された最適性の十分条件のもとで，制約付き最小化問題に対する多くのアルゴリズムが超一次収束することが示されている．