福山大学工学部紀要
第26巻2002年12月

非線形関数における最急降下法の振る舞い

尾関 孝史*

Behavior of the Steepest Descent Method in Minimizing Nonlinear Function

Takashi OZEKI*

ABSTRACT

In this paper, we discuss the limiting behavior of the search direction of the steepest descent method in minimizing general nonlinear function. It is shown that the search direction asymptotically alternates between two directions represented by linear combinations of two eigenvectors of the Hessian matrix of nonlinear function at the minimum point in almost every case. This is similar to the phenomenon in minimizing the quadratic form. However, we also show some special case which is different from the one of the quadratic form. Moreover, we give some necessary conditions in which special cases appear.

キーワード： 最急降下法，非線型関数，反復法，収束

Keywords: Steepest Descent Method, Nonlinear Function, Iterative Method, Convergence

1. まえがき

滑らかな関数の最適値を求める問題に対して、様々な降下法が用いられる。それらのなかでも最急降下法は、もっとも古典的であり、単純な手法である。このため、この手法の性質を調べることは非常に重要である。例えば、最急降下法を二次形式に適用した場合、その探索方向が次第に２つの方向を交互に取るようになることが知られている。従って、最急降下法は本質的には、ある２次元空間内で最適解を探すことになる。そして、その２次元空間は、二次形式を決定する行列の、最大および最小固有値に対応する固有ベクトルによって張られる空間である [1-4]。これらの性質は、Forsythe と Motzkin によって予想され [1]、1959 年に赤池によって証明された [2]。Forsythe は最急降下法を拡張した、S 次元最適降下法に対しても同様な性質が成り立つことを証明しようと試みている [5]。そして、Forsythe はこの論文で、二次形式以外の一般の非線形関数でも同様な性質が証明できるだろうと予想している。実際、著者らは二次形式以外の関数である Rayleigh 商に関して、ほぼ同様な性質が成り立つことを証明した [7]。本論文では、最急降下法が一般の非線形関数に関して二次形式と同様な性質を持つかどうかを議論する。そして、二次形式の場合とは若干異なり、探索方向が２方向を交互に取らない場合が存在することを示す。

2. 最急降下法とその性質

R^n 上で定義された n 変数実数値関数 $f(\boldsymbol{x})$ を２回連続微分可能、即ち C^2 級とする。また，$f(x)$ は孤立した極小点を持つものとする。そして、適当な座標系の設定により、この関数は、原点 $\mathbf{o}$ でこの極小値 0 を取り。関数 $f(\boldsymbol{x})$ の原点でのヘッセ行列 $\nabla^2 f(\mathbf{o})$ が対角行列になるようにする。収束の振る舞いに対しては、この様に勝手な座標系を取っても、一般性を失うことなく議論が可能である。最急降下法は以下のアルゴリズムで与えられる。

最急降下法のアルゴリズム

1. $\boldsymbol{x}_0$ を初期点とし、$k = 0$ とする。
2. 探索方向を最急降下方向 $-\nabla f(\boldsymbol{x}_k)$ とする。また、ステップ幅 α_k を

* 情報処理工学科

$$\frac{\partial f(\boldsymbol{x}_k - \alpha \nabla f(\boldsymbol{x}_k))}{\partial \alpha} = 0 \quad \cdots\cdots\cdots\cdots\cdots\cdots (1)$$

を満たす最小の正数 (Curry の規則 [4]) とする。そして、次の近似解を

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k \nabla f(\boldsymbol{x}_k) \quad \cdots\cdots\cdots\cdots\cdots\cdots (2)$$

とする。

3. 収束条件を評価し、条件を満たさなければ k を増やし、ステップ 2 へ戻る。

定義 2.1 $\boldsymbol{y} \in R^n$ に対して $\Lambda_f(\boldsymbol{y}) = \{\boldsymbol{x} | f(\boldsymbol{x}) \le f(\boldsymbol{y}), \boldsymbol{x} \in R^n\}$ を点 $\boldsymbol{y}$ に対する $f(\boldsymbol{x})$ の準位集合とよぶ。

$f(\boldsymbol{x})$ が連続なら、準位集合は常に閉集合である。また、この時、$\boldsymbol{x}$ が $\Lambda_f(\boldsymbol{y})$ の境界上の点なら $f(\boldsymbol{x}) = f(\boldsymbol{y})$ が成立する。そして、この逆も成立する。

補題 2.1 準位集合 $\Lambda_f(\boldsymbol{a})$ の部分集合のうち、極小点 $\mathbf{o}$ を含む連結集合を $\Lambda_f(\boldsymbol{a}, \mathbf{o})$ で表す。$\Lambda_f(\boldsymbol{a}, \mathbf{o})$ が有界閉集合となる点 $\boldsymbol{a} \in R^n - \{\mathbf{o}\}$ が存在する。

(証明) $\mathbf{o}$ の δ 閉球を

$$B(\delta) = \{\boldsymbol{x} | \|\boldsymbol{x}\| \le \delta, \boldsymbol{x} \in R^n\} \quad \cdots\cdots\cdots\cdots\cdots\cdots (3)$$

で表す。$\nabla^2 f(\boldsymbol{x})$ が連続で、$\nabla^2 f(\mathbf{o}) > 0$ であるから、ある $\delta > 0$ があって、$\boldsymbol{x} \in B(\delta)$ なら $\nabla^2 f(\boldsymbol{x}) > 0$ となる。また、$f(\boldsymbol{x})$ が連続だから、$B(\delta)$ の球面 $S(\delta)$ では、ある 1 点 $\boldsymbol{a} \neq \mathbf{o}$ で最小値 $m = f(\boldsymbol{a}) > 0$ を取る。そこで、

$$\Omega = \{\boldsymbol{x} | f(\boldsymbol{x}) \le m, \boldsymbol{x} \in B(\delta)\} \quad \cdots\cdots\cdots\cdots\cdots (4)$$

とおくと、Ω は $f(\boldsymbol{x})$ の連続性から、明らかに有界閉集合である。また、$B(\delta)$ で $\nabla^2 f(\boldsymbol{x}) > 0$ だから、凸性により、任意の $\boldsymbol{x} \in B(\delta)$ に対して、線分 $[\mathbf{o}, \boldsymbol{x}]$ で $f(\boldsymbol{x})$ は狭義単調増加する。従って、$\boldsymbol{x} \in \Omega$ なら、$[\mathbf{o}, \boldsymbol{x}] \subset \Omega$ となり、Ω は点 $\mathbf{o}$ の連結成分となる。従って、$\Omega \subset \Lambda_f(\boldsymbol{a}, \mathbf{o})$ である。逆に、$\boldsymbol{x} \in \Lambda_f(\boldsymbol{a}, \mathbf{o})$ で $\boldsymbol{x} \notin B(\delta)$ とすると、球面 $S(\delta)$ 上の最小値が $m = f(a)$ であり、$S(\delta)$ で $\nabla^2 f(\boldsymbol{x}) > 0$ だから、どのような曲線で、2 点 $\mathbf{o}$ と $\boldsymbol{x}$ をつないでも途中に $f(\boldsymbol{y}) > m$ となる点が存在する。これは、$\Lambda_f(\boldsymbol{a}, \mathbf{o})$ の連結性に矛盾する。従って、$\Omega = \Lambda_f(\boldsymbol{a}, \mathbf{o})$ が成り立つ。Ω が有界閉集合だから、$\Lambda_f(\boldsymbol{a}, \mathbf{o})$ も有界閉集合となる。■

$\Lambda_f(\boldsymbol{a}, \mathbf{o})$ は凸集合でもある。実際、$\boldsymbol{x}, \boldsymbol{y} \in \Omega$ とすると、明らかに $[\boldsymbol{x}, \boldsymbol{y}] \subset B(\delta)$ である。また、$B(\delta)$ で $f(\boldsymbol{x})$ が凸関数であるから、$0 \le \lambda \le 1$ に対して

$$\begin{aligned} f((1-\lambda)\boldsymbol{x} + \lambda f(\boldsymbol{y})) &\le (1-\lambda) f(\boldsymbol{x}) + \lambda f(\boldsymbol{y}) \\ &\le (1-\lambda) m + \lambda m \\ &= m \quad \cdots\cdots\cdots\cdots\cdots\cdots (5) \end{aligned}$$

従って、$\Lambda_f(\boldsymbol{a}, \mathbf{o})$ は凸集合となる。

最急降下法では以下の収束定理が成立する。

定理 2.1 $\Lambda_f(\boldsymbol{x}_0, \mathbf{o})$ が有界閉集合で、$\Lambda_f(\boldsymbol{x}_0, \mathbf{o})$ で $f(\boldsymbol{x})$ が凸関数なら、点列 $\{\boldsymbol{x}_k\}$ は極小点 $\mathbf{o}$ へ収束する。

(証明) $\boldsymbol{x}_k = \mathbf{o}$ なら明らかに成立するから、全ての k に対して、$\boldsymbol{x}_k \neq \mathbf{o}$ とする。始めに、

$$f(\boldsymbol{x}_{k+1}) < f(\boldsymbol{x}_k) \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (6)$$

が成立することを示す。点 $\boldsymbol{x}_k$ は閉集合 $\Lambda_f(\boldsymbol{x}_k, \mathbf{o})$ の境界点であり、$-\nabla f(\boldsymbol{x}_k)$ は降下方向なため、十分小さな $\alpha > 0$ に対しては

$$f(\boldsymbol{x}_k - \alpha \nabla f(\boldsymbol{x}_k)) < f(\boldsymbol{x}_k) \quad \cdots\cdots\cdots\cdots\cdots\cdots (7)$$

を満たす。従って、この時 $\boldsymbol{x}_k - \alpha \nabla f(\boldsymbol{x}_k)$ は $\Lambda_f(\boldsymbol{x}_k, \mathbf{o})$ の内部の点となる。また、$\Lambda_f(\boldsymbol{x}_k, \mathbf{o})$ が有界であるから、α を大きくしていくと、$\boldsymbol{x}_k - \alpha \nabla f(\boldsymbol{x}_k)$ が $\Lambda_f(\boldsymbol{x}_k, \mathbf{o})$ の境界に達して、

$$f(\boldsymbol{x}_k - \alpha \nabla f(\boldsymbol{x}_k)) = f(\boldsymbol{x}_k) \quad \cdots\cdots\cdots\cdots\cdots\cdots (8)$$

となる $\bar{\alpha}$ が存在する。従って、$f(\boldsymbol{x})$ が微分可能だから、

$$\frac{\partial f(\boldsymbol{x}_k - \alpha \nabla f(\boldsymbol{x}_k))}{\partial \alpha} = 0 \quad \cdots\cdots\cdots\cdots\cdots\cdots (9)$$

となる $0 < \alpha_k < \bar{\alpha}$ が存在して、

$$f(\boldsymbol{x}_{k+1}) = f(\boldsymbol{x}_k - \alpha_k \nabla f(\boldsymbol{x}_k)) < f(\boldsymbol{x}_k) \quad \cdots\cdots (10)$$

が成立する。さて、準位集合の定義から、

$$\boldsymbol{x}_{k+1} \in \Lambda_f(\boldsymbol{x}_{k+1}, \mathbf{o}) \subset \Lambda_f(\boldsymbol{x}_k, \mathbf{o}) \quad \cdots\cdots\cdots\cdots (11)$$

が成立する。従って、特に

$$\boldsymbol{x}_{k+1} \in \Lambda_f(\boldsymbol{x}_0, \mathbf{o}) \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (12)$$

が成立する。従って、点列 $\{\boldsymbol{x}_k\}$ は有界閉集合 $\Lambda_f(\boldsymbol{x}_0, \mathbf{o})$ 内の無限点列だから、収束する部分列を持つ。これを $\{\boldsymbol{x}_k\}, k \in K \subset \{0, 1, 2, \cdots\}$ とし、その収束点を $\bar{\boldsymbol{x}}$ とする。また、$\{\boldsymbol{x}_{k+1}\}, k \in K$ も無限点列だから収束する部分列を持つ。これを $\{\boldsymbol{x}_{k+1}\} \in K' \subset K$ とし、その収束点を $\bar{\boldsymbol{x}}'$ とする。即ち、

$$\lim_{k \to \infty, k \in K'} \boldsymbol{x}_k = \bar{\boldsymbol{x}} \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots (13)$$

$$\lim_{k \to \infty, k \in K'} \boldsymbol{x}_{k+1} = \bar{\boldsymbol{x}}' \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots (14)$$

が成り立つ。すると、$\nabla f(\boldsymbol{x})$ の連続性から

$$\lim_{k \to \infty, k \in K'} \nabla f(\boldsymbol{x}_k) = \nabla f(\bar{\boldsymbol{x}}) \quad \cdots\cdots (15)$$

が成立する。ここで、もし、$\nabla f(\bar{\boldsymbol{x}}) \neq \mathbf{o}$ と仮定すると α_k の極限 $\bar{\alpha}$ が存在し、

$$\bar{\boldsymbol{x}}' = \bar{\boldsymbol{x}} - \bar{\alpha} \nabla f(\bar{\boldsymbol{x}}) \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots (16)$$

が成り立つ。すると

$$f(\bar{x}') < f(\bar{x}) \quad \cdots\cdots (17)$$

が成立する。これは、$f(\boldsymbol{x}_k)$ が単調減少して、ある値に収束することに矛盾する。従って、

$$\nabla f(\bar{x}) = 0 \quad \cdots\cdots (18)$$

が成立する。$\Lambda_f(\boldsymbol{x}_0, \mathbf{o})$ で関数 $f(\boldsymbol{x})$ が凸であるから、$\Lambda_f(\boldsymbol{x}_0, \mathbf{o})$ では $\mathbf{o}$ がただ1つの極小値である。従って、$\bar{\boldsymbol{x}} = \mathbf{o}$ が成り立つ。点列 $\{\boldsymbol{x}_k\}$ のすべての集積点に関して同じことが成立するから、結局

$$\lim_{k\to\infty} \boldsymbol{x}_k = \mathbf{o} \quad \cdots\cdots (19)$$

が成立する。 ■

さて、この定理の仮定をみたす $\Lambda_f(\boldsymbol{x}_0, \mathbf{o})$ が存在するかが気になる。しかし、補題2.1で構成した集合 $\Lambda_f(\boldsymbol{a}, \mathbf{o})$ 内に $\boldsymbol{x}_0$ をとれば $\Lambda_f(\boldsymbol{x}_0, \mathbf{o})$ は有界閉集合で、この範囲で $f(\boldsymbol{x})$ が凸関数となることが容易にわかる。以下では、特別に断らないかぎり、$\Lambda_f(\boldsymbol{x}_0, \mathbf{o})$ が有界閉集合で、この範囲で $f(\boldsymbol{x})$ が凸関数であるとする。

定理2.1は、単純に収束性を示しているが、実は、最急降下法は1次収束することが示せる。その証明の準備として、関数 $f(\boldsymbol{x})$ が二次形式の場合を証明する。

補題 2.2 H を $n\times n$ の正値な対角行列とする。二次形式

$$f(\boldsymbol{x}) = \frac{1}{2}(H\boldsymbol{x}, \boldsymbol{x}) \quad \cdots\cdots (20)$$

に対し、最急降下法を行う。この時、ある $0 < c < 1$ が存在し、すべての k に対して

$$\|\boldsymbol{x}_{k+1}\| \le c\|\boldsymbol{x}_k\| \quad \cdots\cdots (21)$$

が成立する。即ち、点列 $\{\boldsymbol{x}_k\}$ は極小点 $\mathbf{o}$ に1次収束する。

(証明) 点 $\boldsymbol{x}$ に最急降下法のアルゴリズムを1回適用したときの、ステップ幅を $\alpha(\boldsymbol{x})$ とすると

$$\nabla f(\boldsymbol{x}) = H\boldsymbol{x} \quad \cdots\cdots (22)$$

と

$$\frac{\partial f(\boldsymbol{x} - \alpha\nabla f(\boldsymbol{x}))}{\partial\alpha} = (\nabla f(\boldsymbol{x} - \alpha\nabla f(\boldsymbol{x})), \nabla f(\boldsymbol{x})) = 0 \quad \cdots\cdots (23)$$

から

$$\alpha(\boldsymbol{x}) = \frac{(\nabla f(\boldsymbol{x}), \nabla f(\boldsymbol{x}))}{(H\nabla f(\boldsymbol{x}), \nabla f(\boldsymbol{x}))} = \frac{(H\boldsymbol{x}, H\boldsymbol{x})}{(H^2\boldsymbol{x}, H\boldsymbol{x})} \quad \cdots\cdots (24)$$

が得られる。従って

$$\boldsymbol{x}_{k+1} = \{I - \alpha(\boldsymbol{x}_k)H\}\boldsymbol{x}_k \quad \cdots\cdots (25)$$

と表せる。そして、

$$\begin{aligned}&\|\boldsymbol{x}_{k+1}\|^2 - \|\boldsymbol{x}_k\|^2 \\ &= \alpha(\boldsymbol{x}_k)[\alpha(\boldsymbol{x}_k)(H\boldsymbol{x}_k, H\boldsymbol{x}_k) - (H\boldsymbol{x}_k, \boldsymbol{x}_k)] \\ &\quad -\alpha(\boldsymbol{x}_k)(H\boldsymbol{x}_k, \boldsymbol{x}_k)\end{aligned} \quad \cdots\cdots (26)$$

となる。ここで、$H = A^2$ と表すと、Cauchy-Schwartz の不等式から

$$\begin{aligned}&\alpha(\boldsymbol{x}_k)(H\boldsymbol{x}_k, H\boldsymbol{x}_k) - (H\boldsymbol{x}_k, \boldsymbol{x}_k) \\ &= \frac{(A^3\boldsymbol{x}_k, A\boldsymbol{x}_k)^4 - \|A^3\boldsymbol{x}_k\|^2\cdot\|A\boldsymbol{x}_k\|^2}{\|A^3\boldsymbol{x}_k\|^2} \\ &\le 0\end{aligned} \quad \cdots\cdots (27)$$

となる。また、$\boldsymbol{x}_k \ne \mathbf{o}$ なら

$$\alpha(\boldsymbol{x}_k)(H\boldsymbol{x}_k, \boldsymbol{x}_k) > 0 \quad \cdots\cdots (28)$$

となることから

$$\|\boldsymbol{x}_{k+1}\| < \|\boldsymbol{x}_k\| \quad \cdots\cdots (29)$$

が成立する。このことから、$\boldsymbol{x} \in R^n - \mathbf{o}$ の実数値関数 $g(\boldsymbol{x})$ を

$$g(\boldsymbol{x}) = \{I - \alpha(x)H\}\boldsymbol{x} \quad \cdots\cdots (30)$$

と定義したとき

$$\frac{\|g(\boldsymbol{x})\|}{\|\boldsymbol{x}\|} < 1 \quad \cdots\cdots (31)$$

が成立する。更に、$\boldsymbol{x}$ を単位球面上に限ると、$g(\boldsymbol{x})$ はコンパクト集合上の実数値連続関数であるから、最大値 $0 < c < 1$ が存在して

$$\frac{\|g(\boldsymbol{x})\|}{\|\boldsymbol{x}\|} < c \quad \cdots\cdots (32)$$

が成立する。ところが、任意の正数 $t > 0$ に対して、$\alpha(t\boldsymbol{x}) = \alpha(\boldsymbol{x})$ だから

$$\frac{\|g(t\boldsymbol{x})\|}{\|t\boldsymbol{x}\|} = \frac{\|g(\boldsymbol{x})\|}{\|\boldsymbol{x}\|} \quad \cdots\cdots (33)$$

が成り立つ。結局、任意の $\boldsymbol{x} \in R^n - \mathbf{o}$ に対して

$$\|g(\boldsymbol{x})\| < c\|\boldsymbol{x}\| \quad \cdots\cdots (34)$$

が成り立つ。これは、すべての k に対して

$$\|\boldsymbol{x}_{k+1}\| \le c\|\boldsymbol{x}_k\| \quad \cdots\cdots (35)$$

が成立することを示している。従って、最急降下法は二次形式では1次収束をしている。 ■

定義 2.2 関数 $f(\boldsymbol{x})$ に点 $\boldsymbol{x} \in \Lambda_f(\boldsymbol{x}_0, \mathbf{o}) - \{\boldsymbol{x}\}$ で、最急降下法のアルゴリズムを1回適用したときの、ステップ幅を $\alpha(\boldsymbol{x})$ とし、更新点を $g(\boldsymbol{x})$ で表すことにする。

定義から、明らかに

$$g(\boldsymbol{x}) = \boldsymbol{x} - \alpha(\boldsymbol{x})\nabla f(\boldsymbol{x}) \quad \cdots\cdots (36)$$

の関係がある。

補題 2.3 ある $N > 0$ が存在して、任意の点 $\boldsymbol{x} \in \Lambda_f(\boldsymbol{x}_0, \mathbf{o}) - \{\mathbf{o}\}$ に対し、

$$\|g(\boldsymbol{x})\| \leq N\|\boldsymbol{x}\| \quad \cdots\cdots (37)$$

が成立する。

(証明) $g(\boldsymbol{x}) = \mathbf{o}$ の場合は明らかに成立するので、以後、$g(\boldsymbol{x}) \neq \mathbf{o}$ とする。コンパクト集合 $\Lambda_f(\boldsymbol{x}_0, \mathbf{o})$ で正値対称行列 $\nabla^2 f(\boldsymbol{x})$ は連続だから、最小値 $m > 0$ と最大値 M が存在して、$\boldsymbol{x} \in \Lambda_f(\boldsymbol{x}_0, \mathbf{o})$ なら

$$m \leq \nabla^2 f(\boldsymbol{x}) \leq M \quad \cdots\cdots (38)$$

が成立する。そして、$f(\boldsymbol{x})$ が $\Lambda_f(\boldsymbol{x}_o, \mathbf{o})$ で C^2 級だから、Taylor の定理から、$\boldsymbol{x} \in \Lambda_f(\boldsymbol{x}_o, \mathbf{o})$ ならある $0 < t < 1$ があって

$$f(\boldsymbol{x}) = \left(\nabla^2 f(t\boldsymbol{x})\boldsymbol{x}, \boldsymbol{x}\right) \quad \cdots\cdots (39)$$

が成り立つ。また、$\mathbf{o}, \boldsymbol{x} \in \Lambda_f(\boldsymbol{x}_0, \mathbf{o})$ で、$\Lambda_f(\boldsymbol{x}_0, \mathbf{o})$ が凸集合であるから、$t\boldsymbol{x} \in \Lambda_f(\boldsymbol{x}_0, \mathbf{o})$ が成り立つ。従って

$$m\|\boldsymbol{x}\|^2 \leq f(\boldsymbol{x}) \leq M\|\boldsymbol{x}\|^2 \quad \cdots\cdots (40)$$

が成立する。$g(\boldsymbol{x}) \in \Lambda_f(\boldsymbol{x}_0, \mathbf{o})$ だから、同様にして、

$$m\|g(\boldsymbol{x})\|^2 \leq f(g(\boldsymbol{x})) \leq M\|g(\boldsymbol{x})\|^2 \quad \cdots\cdots (41)$$

が成立する。そして、最急降下法の性質から

$$f(g(\boldsymbol{x})) < f(\boldsymbol{x}) \quad \cdots\cdots (42)$$

が成立するから

$$\|g(\boldsymbol{x})\|^2 \leq \frac{M}{m}\|\boldsymbol{x}\|^2 \quad \cdots\cdots (43)$$

が成立する。■

補題 2.4 $\alpha(\boldsymbol{x})$ は、$\Lambda_f(\boldsymbol{x}_0, \mathbf{o}) - \{\mathbf{o}\}$ で有界である。

補題 2.3から **(証明)** ある $N > 0$ が存在して、任意の点 $\boldsymbol{x} \in \Lambda_f(\boldsymbol{x}_0, \mathbf{o}) - \{\mathbf{o}\}$ に対し、

$$\|\boldsymbol{x} - \alpha(\boldsymbol{x})\nabla f(\boldsymbol{x})\| \leq N\|\boldsymbol{x}\| \quad \cdots\cdots (44)$$

が成立する。従って、三角不等式から

$$\|\alpha(\boldsymbol{x})\nabla f(\boldsymbol{x})\| \leq N\|\boldsymbol{x}\| + \|\boldsymbol{x}\| \quad \cdots\cdots (45)$$

が成り立つ。また、

$$\nabla f(\boldsymbol{x}) = \left\{\int_0^1 \nabla^2 f(t\boldsymbol{x})dt\right\}\boldsymbol{x} \quad \cdots\cdots (46)$$

で、$\boldsymbol{x} \in \Lambda_f(\boldsymbol{x}_0, \mathbf{o}) - \{\boldsymbol{x}\}$ なら

$$m \leq \nabla^2 f(\boldsymbol{x}) \leq M \quad \cdots\cdots (47)$$

が成立するから、

$$m\|\boldsymbol{x}\| \leq \|\nabla f(\boldsymbol{x})\| \leq M\|\boldsymbol{x}\| \quad (48)$$

となる。従って、三角不等式 (45) から

$$0 < \alpha(\boldsymbol{x}) \leq \frac{N+1}{m} \quad \cdots\cdots (49)$$

が成り立つ。■

補題 2.2は以下の定理に発展する。

定理 2.2 定理 2.1の仮定のもとで、点列 $\{\boldsymbol{x}_k\}$ は極小点 $\mathbf{o}$ に 1 次収束する。すなわち、ある $0 < c < 1$ が存在し、十分大きなすべての k に対して

$$\|\boldsymbol{x}_{k+1}\| \leq c\|\boldsymbol{x}_k\| \quad \cdots\cdots (50)$$

が成立する。

(証明)

$$\hat{f}(\boldsymbol{x}) = \frac{1}{2}(\nabla^2 f(\mathbf{o})\boldsymbol{x}, \boldsymbol{x}) \quad \cdots\cdots (51)$$

とおく。すると、

$$\hat{g}(\boldsymbol{x}) = \boldsymbol{x} - \hat{\alpha}(\boldsymbol{x})\nabla \hat{f}(\boldsymbol{x}) \quad \cdots\cdots (52)$$

と表される。ここで

$$\hat{\alpha}(\boldsymbol{x}) = \frac{(\nabla \hat{f}(\boldsymbol{x}), \nabla \hat{f}(\boldsymbol{x}))}{(\nabla^2 f(\mathbf{o})\nabla \hat{f}(\boldsymbol{x}), \nabla \hat{f}(\boldsymbol{x}))} \quad \cdots\cdots (53)$$

である。すると、$\hat{f}(\boldsymbol{x})$ が二次形式だから、補題 2.2から、ある定数 $0 < c < 1$ があって、任意 の $\boldsymbol{x} \in R^n - \{\mathbf{o}\}$ に対し、

$$\|\hat{g}(\boldsymbol{x})\| \leq c\|\boldsymbol{x}\| \quad \cdots\cdots (54)$$

が成立する。

もし、

$$g(\boldsymbol{x}) = \hat{g}(\boldsymbol{x}) + o(\|\boldsymbol{x}\|) \quad \cdots\cdots (55)$$

が成立すれば、

$$\frac{\|g(\boldsymbol{x})\|}{\|\boldsymbol{x}\|} = \frac{\|\hat{g}(\boldsymbol{x})\|}{\|\boldsymbol{x}\|} + \frac{o(\|\boldsymbol{x}\|)}{\|\boldsymbol{x}\|} \leq c + \frac{o(\|\boldsymbol{x}\|)}{\|\boldsymbol{x}\|} \quad \cdots\cdots (56)$$

が成立する。従って、ある $\delta > 0$ と定数 $0 < c < d < 1$ があって $\|\boldsymbol{x}\| < \delta$ なら

$$\|g(\boldsymbol{x})\| \leq d\|\boldsymbol{x}\| \quad \cdots\cdots (57)$$

が成立する。そして、定理 2.1から

$$\lim_{k\to\infty}\|\boldsymbol{x}_k\|=0 \cdots\cdots (58)$$

だから、十分大きなすべての k に対して

$$\|\boldsymbol{x}_{k+1}\|\le d\|\boldsymbol{x}_k\| \cdots\cdots (59)$$

が成立し、定理が証明される。そこで、以下では

$$g(\boldsymbol{x})=\hat{g}(\boldsymbol{x})+o(\|\boldsymbol{x}\|) \cdots\cdots (60)$$

を証明する。

$\nabla f(\boldsymbol{x})$ は $\mathbf{o}$ の近傍で微分可能だから

$$\begin{aligned}\nabla f(\boldsymbol{x})&=\nabla^2 f(\mathbf{o})\boldsymbol{x}+o(\|x\|)\\&=\nabla\hat{f}(\boldsymbol{x})+o(\|x\|) \cdots\cdots (61)\end{aligned}$$

が成り立つ。また、$\nabla^2 f(\boldsymbol{x})$ は $\mathbf{o}$ の近傍で連続だから

$$\nabla^2 f(\boldsymbol{x})=\nabla^2 f(\mathbf{o})+o(1) \cdots\cdots (62)$$

が成り立つ。更に、

$$\alpha(\boldsymbol{x})=\hat{\alpha}(\boldsymbol{x})+o(1) \cdots\cdots (63)$$

成り立つ。実際、補題 2.4の三角不等式 (45) から

$$o(\|\alpha(\boldsymbol{x})\nabla\boldsymbol{x}\|)=o(\|\boldsymbol{x}\|) \cdots\cdots (64)$$

で、$\nabla f(\boldsymbol{x})$ が微分可能だから

$$\begin{aligned}&\nabla f(\boldsymbol{x}-\alpha(\boldsymbol{x})\nabla f(\boldsymbol{x}))\\&=\nabla f(\boldsymbol{x})-\alpha(\boldsymbol{x})\nabla^2 f(\boldsymbol{x})\nabla f(\boldsymbol{x})+o(\|x\|) \cdots (65)\end{aligned}$$

が成り立つ。ここで、$\alpha(\boldsymbol{x})$ は

$$\begin{aligned}&\frac{\partial f(\boldsymbol{x}-\alpha\nabla f(\boldsymbol{x}))}{\partial\alpha}\\&=\nabla f(\boldsymbol{x}-\alpha\nabla f(\boldsymbol{x}))\nabla f(\boldsymbol{x})=0 \cdots\cdots (66)\end{aligned}$$

の解だから、式 (61)、式 (62) を用いて

$$\alpha(\boldsymbol{x})=\frac{(\nabla\hat{f}(\boldsymbol{x}),\nabla\hat{f}(\boldsymbol{x}))+o(\|\boldsymbol{x}\|^2)}{(\nabla^2 f(\mathbf{o})\nabla\hat{f}(\boldsymbol{x}),\nabla\hat{f}(\boldsymbol{x}))+o(\|\boldsymbol{x}\|^2)} \cdots (67)$$

を得る。これと式 (53) から

$$\alpha(\boldsymbol{x})=\hat{\alpha}(\boldsymbol{x})+o(1) \cdots\cdots (68)$$

を得る。従って、式 (61)、式 (63) から

$$\begin{aligned}&g(\boldsymbol{x})-\hat{g}(\boldsymbol{x})\\&=\{\boldsymbol{x}-\hat{\alpha}(\boldsymbol{x})\nabla\hat{f}(\boldsymbol{x})\}-\{\boldsymbol{x}-\alpha(\boldsymbol{x})\nabla f(\boldsymbol{x})\}\\&=o(\|\boldsymbol{x}\|) \cdots\cdots (69)\end{aligned}$$

が成立する。

なお、ここでは式 (53) から、$\hat{\alpha}(\boldsymbol{x})$ が $R^n-\{\mathbf{o}\}$ で有界であることを用いている。 ■

3. 探索方向の振る舞い

本節では、最急降下法の探索方向の振る舞いに関して論じる。

定義 3.3 点 $\boldsymbol{x}_k$ での最急降下方向 $-\nabla f(\boldsymbol{x}_k)$ を正規化した、

$$\boldsymbol{d}_k=-\frac{\nabla f(\boldsymbol{x}_k)}{\nu_k} \cdots\cdots (70)$$

を点 $\boldsymbol{x}_k$ の探索ベクトルと呼ぶ。ここで、

$$\nu_k=\|\nabla f(\boldsymbol{x}_k)\| \cdots\cdots (71)$$

である。

定理 3.3 最急降下法の連続する 2 つの探索ベクトルは互いの直交している。即ち、$(\boldsymbol{d}_k,\boldsymbol{d}_{k+1})=0$ が成立する。

(証明) ステップ幅 $\alpha(\boldsymbol{x}_k)$ が

$$\begin{aligned}&\frac{\partial f(\boldsymbol{x}_k-\alpha\nabla f(\boldsymbol{x}_k))}{\partial\alpha}\\&=(\nabla f(\boldsymbol{x}_k-\alpha\nabla f(\boldsymbol{x}_k)),\nabla f(\boldsymbol{x}_k))=0 \cdots\cdots (72)\end{aligned}$$

を満たすことから明らか。 ■

さて、$n\times n$ の正値対象行列 A_k を

$$A_k\triangleq\int_0^1\nabla^2 f(t\boldsymbol{x}_{k+1}+(1-t)\boldsymbol{x}_k)dt \cdots\cdots (73)$$

とおくと、

$$\begin{aligned}\nabla f(\boldsymbol{x}_{k+1})&=\nabla f(\boldsymbol{x}_k)+A_k(\boldsymbol{x}_{k+1}-\boldsymbol{x}_k)\\&=\nabla f(\boldsymbol{x}_k)-\alpha_k A_k\nabla f(\boldsymbol{x}_k) \cdots\cdots (74)\end{aligned}$$

が成立する。そして、上式の両辺に $\nabla f(\boldsymbol{x})$ を掛けると、定理 3.3から

$$\begin{aligned}\alpha_k^{-1}&=\frac{(A_k\nabla f(\boldsymbol{x}_k),\nabla f(\boldsymbol{x}_k))}{\|\nabla f(\boldsymbol{x}_k)\|^2}\\&=(A_k\boldsymbol{d}_k,\boldsymbol{d}_k) \cdots\cdots (75)\end{aligned}$$

が成立する。関数 $f(\boldsymbol{x})$ が二次形式の場合、$A_k=\nabla^2 f(\mathbf{o})$ となり、定数行列である。赤池は論文 [2] で、正値対象行列 $\nabla^2 f(\mathbf{o})$ の固有値 $\lambda_i(i=1,\ldots,n)$ を確率変数の値、その確率分布を $\boldsymbol{d}_k$ の各成分の二乗とした確率関数を導入した。そして、α_k^{-1} をその分布の平均と考え、その分散の変化を調べることで、探索ベクトルの振る舞いを解析した。そこで、平均を

$$\bar{\lambda}(\boldsymbol{d}_k)\triangleq(A_k\boldsymbol{d}_k,\boldsymbol{d}_k) \cdots\cdots (76)$$

と定め、二次形式でない場合も同様に、分散

$$V_k\triangleq\|A_k\boldsymbol{d}_k-\bar{\lambda}(\boldsymbol{d}_k)\boldsymbol{d}_k\|^2 \cdots\cdots (77)$$

の変化を調べることにする。式 (74) から、

$$d_{k+1} = -\frac{A_k d_k - \bar{\lambda}(d_k) d_k}{\sqrt{V_k}} \qquad (78)$$

が成り立つ。これから、

$$\sqrt{V_k} = -(A_k d_k, d_{k+1}) \qquad (79)$$

を得る。従って、

$$\begin{aligned}
&(d_{k+2}, d_k) \\
&= -\frac{(A_{k+1} d_{k+1}, d_k)}{\sqrt{V_{k+1}}} \\
&= \frac{\sqrt{V_k} + ((A_k - A_{k+1}) d_{k+1}, d_k)}{\sqrt{V_{k+1}}} \\
&\le 1 \qquad (80)
\end{aligned}$$

が得られる。従って、

$$\sqrt{V_k} \le \sqrt{V_{k+1}} + \|A_{k+1} - A_k\| \qquad (81)$$

を得る。また、

$$0 \le \sqrt{V_k} = |-(A_k d_k, d_{k+1})| \le \|A_k\| \qquad (82)$$

が成り立つ。ここで、$\Lambda_f(x_0, o)$ で $\nabla^2 f(x)$ が有界だから、A_k も有界である。従って、V_k も有界である。更に、V_k に関して次の補題が成立する。

補題 3.5 $\nabla^2 f(x)$ が $\Lambda_f(x_0, o)$ で Lipschitz 連続なら、分散 $\{V_k\}$ はある 0 以上の数に収束する。

(証明) V_k が有界なので、ある M があって、$\sqrt{V_k} \in [0, M]$ となる。そこで、点列 $\{\sqrt{V_k}\}$ が 2 つの集積点 $a, b \in [0, M]$(ただし、$a < b$) を持つと仮定して、矛盾を導く。$\nabla^2 f(x)$ が Lipschitz 連続だから、ある L があって、$x, y \in \Lambda_f(x_0, o)$ なら

$$\|\nabla^2 f(x) - \nabla^2 y\| \le L\|x - y\| \qquad (83)$$

が成り立つ。従って、式 (73) から

$$\begin{aligned}
&\|A_{k+1} - A_k\| \\
&\le L\int_0^1 \|t(x_{k+2} - x_{k+1}) + (1-t)(x_{k+1} - x_k)\| dt \\
&\le L(\|(x_{k+2}\| + 2\|x_{k+1})\| + \|x_k)\|) \qquad (84)
\end{aligned}$$

を得る。また、定理 2.2から、$\sum \|x_k\|$ が収束するので、$\sum \|A_{k+1} - A_k\|$ も収束する。従って、十分大きな N が存在して、$p > q \ge N$ なら

$$\sum_{k=q}^{p-1} \|A_{k+1} - A_k\| < \frac{a-b}{3} \qquad (85)$$

が成立する。また、a, b が $\{\sqrt{V_k}\}$ の集積点であるから、$m > n \ge N$ を満たすある m, n で

$$|\sqrt{V_m} - a| < \frac{b-a}{3} \qquad (86)$$

$$|\sqrt{V_n} - b| < \frac{b-a}{3} \qquad (87)$$

を満たす。すなわち

$$\sqrt{V_m} < a + \frac{b-a}{3} \qquad (88)$$

$$\sqrt{V_n} > b - \frac{b-a}{3} \qquad (89)$$

を満たす。従って、

$$\begin{aligned}
&\sqrt{V_m} + \sum_m^{n-1} \|A_{k+1} - A_k\| \\
&< a + \frac{b-a}{3} + \frac{b-a}{3} < \sqrt{V_n} \qquad (90)
\end{aligned}$$

が成り立つ。一方、不等式 (81) から、

$$\sqrt{V_n} \le \sqrt{V_m} + \sum_m^{n-1} \|A_{k+1} - A_k\| \qquad (91)$$

なので矛盾する。従って、$\{\sqrt{V_k}\}$ は集積点が 2 つ以上存在せず、$\{V_k\}$ は収束する。 ■

系 3.1 分散 V_k の極限値が 0 でなければ、

$$\lim_{k\to\infty} \|d_{k+2} - d_k\|^2 = 0 \qquad (92)$$

が成立する。

(証明) 式 (80) において、k に関して極限を取ると、$\lim_{k\to\infty} \|A_{k+1} - A_k\| = 0$ で、$\{\sqrt{V_k}\}$ が 0 でない値に収束する。従って、$\lim_{k\to\infty} (d_{k+2}, d_k) = 1$ が成立する。これから、

$$\begin{aligned}
\lim_{k\to\infty} \|d_{k+2} - d_k\|^2 &= \lim_{k\to\infty} 2\{1 - (d_{k+2}, d_k)\} \\
&= 0 \qquad (93)
\end{aligned}$$

を得る。 ■

以後、$\nabla^2 f(x)$ が $\Lambda_f(x_0, o)$ で Lipschitz 連続と仮定する。そして、分散の収束値を V で表す。

3.1 分散 V_k の極限値が 0 でない場合

最急降下法を二次形式に適用した時、常に、$A_k = \nabla^2 f(o)$ である。従って、式 (80) から

$$\frac{\sqrt{V_k}}{\sqrt{V_{k+1}}} = (d_{k+2}, d_k) \le 1 \qquad (94)$$

が成立するため、分散 V_k は単調増加する。従って、

$$V = \lim_{k\to\infty} V_k \ge V_0 > 0 \qquad (95)$$

となるから、分散の極限値が 0 でない場合に含まれる。逆に、関数 $f(x)$ が二次形式以外の場合でも、分散 V_k の極限が 0 でない場合は、二次形式の場合 (赤池の定理 [2]) と全く同じ次の定理が得られる。

定理 3.4 分散 V_k の極限が 0 でなければ、最急降下法は、その探索ベクトルが漸近適に互いに直交する 2 つのベクトル間を交互にとるようになる。

図 1、図 2に見られるように、分散 V_k の極限が 0 でなければ、二次形式、二次形式以外のどちらの場合でも、最急降下法はジグザグに最小値に近づく。

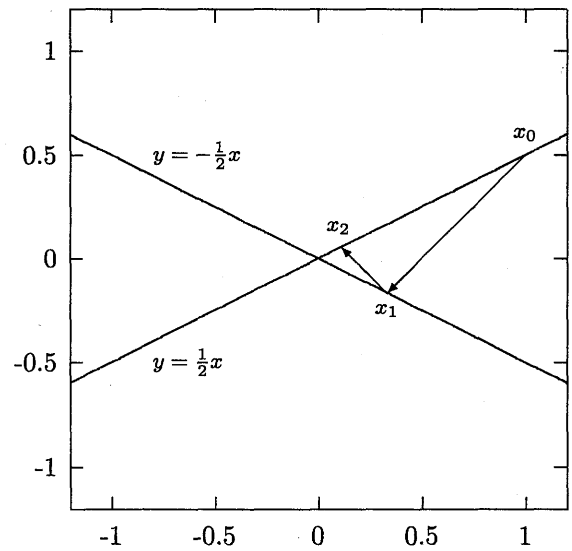

図 1　二次形式の例: $f(x,y)=\frac{1}{2}x^2+y^2$
Fig.1 Example of quadratic form.

以下では、この定理の証明をいくつかの補題を用いて行う。

定義 3.4 探索ベクトルの点列 $\{\boldsymbol{d}_k\}$ は単位球面 S^{n-1} 上の点列であるから S^{n-1} 上に集積点を持つ。これらの集積点の集合を $\{\boldsymbol{r}_\alpha\}$ と表す。

定義 3.5 正値対称行列 $\nabla^2 f(\boldsymbol{x})$ の固有値を $\{\lambda_i \in R : \lambda_1(\boldsymbol{x}) < \lambda_2(\boldsymbol{x}) < \cdots < \lambda_n(\boldsymbol{x})\}$ とおく。そして、その固有ベクトルから生成される適当な正規直交基底を $\{\boldsymbol{u}_i(\boldsymbol{x})\}$ で表す。特に、記号の簡略化のため、$A=\nabla^2 f(\mathbf{o})$、$\boldsymbol{u}_i=\boldsymbol{u}_i(\mathbf{o})$、$\lambda_i=\lambda_i(\mathbf{o})$、とする。なお、この正規直交規定は連続に取るものとする。そして、探索ベクトル $\boldsymbol{d}_k$ と集積点 $\boldsymbol{r}_\alpha$ をこれらの正規直交基底で表した成分をそれぞれ $d_i^{(k)}$ と $r_i^{(\alpha)}$ とする。すなわち、$d_i^{(k)} \triangleq (\boldsymbol{d}_k, \boldsymbol{u}_i(\boldsymbol{x}_k))$, $r_i^{(\alpha)} \triangleq (\boldsymbol{r}_\alpha, \boldsymbol{u}_i)$ とする。

すると、探索方向 $\boldsymbol{d}_k$ と集積点 $\boldsymbol{r}_\alpha$ はそれぞれ

$$\boldsymbol{d}_k = \sum_{i=1}^{n} d_i^{(k)} \boldsymbol{u}_i(\boldsymbol{x}_k), \quad \cdots\cdots (96)$$

$$\boldsymbol{r}_\alpha = \sum_{i=1}^{n} r_i^{(\alpha)} \boldsymbol{u}_i \quad \cdots\cdots (97)$$

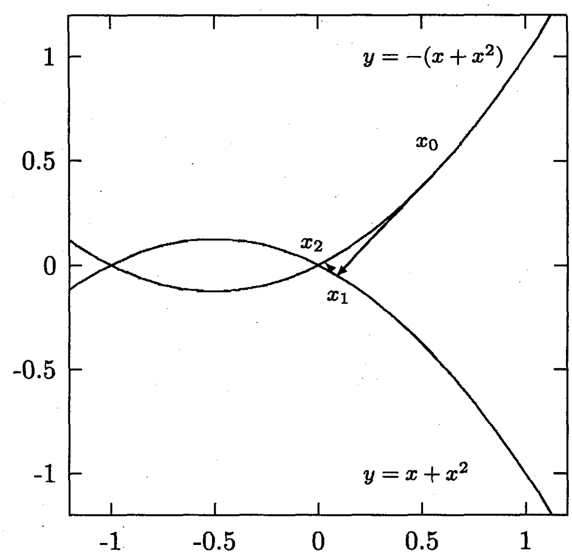

図 2　二次形式でない例：$f(x,y)=\frac{1}{2}x^2+\frac{1}{3}x^3+y^2$
Fig.2 Example of not quadratic form.

と表される。

補題 3.6 任意の集積点 $\boldsymbol{r}_\alpha$ は非零の成分を丁度 2 つ持つ。

(証明) 式 (78) から

$$\|\boldsymbol{d}_{k+2}-\boldsymbol{d}_k\|^2 = \left\| \left\{ \frac{\{\boldsymbol{A}_{k+1}-\bar{\lambda}(\boldsymbol{d}_{k+1})\boldsymbol{I}\}\{\boldsymbol{A}_k-\bar{\lambda}(\boldsymbol{d}_k)\boldsymbol{I}\}}{\sqrt{V_{k+1}\cdot V_k}} - \boldsymbol{I} \right\} \boldsymbol{d}_k \right\|^2$$

が成立する。補題 3.5の系 3.1から

$$\lim_{k\to\infty} \|\boldsymbol{d}_{k+2}-\boldsymbol{d}_k\|^2 = 0 \quad \cdots\cdots (98)$$

が成立するので、集積点 $\boldsymbol{r}_\alpha$ へ収束する $\{\boldsymbol{d}_k\}$ の部分列を考え、式 (98) の極限をとると、

$$\left\| \left\{ \frac{\{\boldsymbol{A}-\bar{\lambda}(\boldsymbol{T}\boldsymbol{r}_\alpha)\boldsymbol{I}\}\{\boldsymbol{A}-\bar{\lambda}(\boldsymbol{r}_\alpha)\boldsymbol{I}\}}{V} - \boldsymbol{I} \right\} \boldsymbol{r}_\alpha \right\|^2 = 0 \quad \cdots\cdots (99)$$

を満たす。ここで、

$$A = \nabla^2 f(\mathbf{o}) \quad \cdots\cdots (100)$$

$$\bar{\lambda}(\boldsymbol{r}_\alpha) = (A\boldsymbol{r}_\alpha, \boldsymbol{r}_\alpha) \quad \cdots\cdots (101)$$

$$\boldsymbol{T}\boldsymbol{r}_\alpha = -\frac{(A-\bar{\lambda}(\boldsymbol{r}_\alpha)I)}{\sqrt{V}}\boldsymbol{r}_\alpha \quad \cdots\cdots (102)$$

である。これから、$1 \le i \le n$ に対して

$$\left\{(\lambda_i-\bar{\lambda}(\boldsymbol{T}\boldsymbol{r}_\alpha))(\lambda_i-\bar{\lambda}(\boldsymbol{r}_\alpha))-V\right\} r_i^{(\alpha)} = 0 \cdot (103)$$

を得る。式 (103) は、λ_i の二次方程式である。従って、集積点 $\boldsymbol{r}_\alpha$ は高々2つの i を除いて、その成分 $r_i^{(\alpha)}$ は零でなければならないことがわかる。また、$\boldsymbol{r}_\alpha$ は単位球 S^{n-1} 上の点列であるのでその集積点は少なくとも1つの成分が非零である。そして、もしある集積点 $\boldsymbol{r}_\alpha$ が唯1つの成分しか持たないとすると、$\boldsymbol{r}_\alpha$ は $\nabla^2 f(\mathbf{o})$ の固有ベクトルとなるので式 (77) と式 (75) から $\lim_{k\to\infty} V_k = 0$ が成立する。これは仮定に矛盾する。結局、任意の集積点 $\boldsymbol{r}_\alpha$ は非零の成分を丁度2つ持つ。 ■

補題 3.7 集積点全体からなる集合 $\{\boldsymbol{r}_\alpha\}$ はただ2つの点からなる。

(証明) 最初に $\{\boldsymbol{r}_\alpha\}$ が有限個の点からなることを示す。補題 3.6から各集積点 $\boldsymbol{r}_\alpha$ はある $i, j (1 \le i < j \le n)$ に対して

$$\boldsymbol{r}_\alpha = r_i^{(\alpha)} \boldsymbol{u}_i + r_j^{(\alpha)} \boldsymbol{u}_j \quad \cdots\cdots (104)$$

$$(r_i^{(\alpha)})^2 + (r_j^{(\alpha)})^2 = 1 \quad \cdots\cdots (105)$$

と表される。ここで $r_i^{(\alpha)}$ と $r_j^{(\alpha)}$ は零でない成分である。

また、集積点 $\boldsymbol{r}_\alpha$ における V_α の値を

$$\begin{aligned} V_\alpha &\triangleq \|\boldsymbol{A}\boldsymbol{r}_\alpha - \bar{\lambda}(\boldsymbol{r}_\alpha)\boldsymbol{r}_\alpha\|^2 \\ &= (\lambda_i - \lambda_j)^2 (r_i^{(\alpha)})^2 (r_j^{(\alpha)})^2 \quad \cdots\cdots (106) \end{aligned}$$

で与える。すると、式 (77) から、V_k は各集積点 $\boldsymbol{r}_\alpha$ で連続となる。従って、補題 3.5から任意の集積点 $\boldsymbol{r}_\alpha$ に対して

$$V_\alpha = \lim_{k\to\infty} V_k = V \ (107)$$

が成り立つ。その結果、式 (106) と式 (107) から

$$(\lambda_i - \lambda_j)^2 (r_i^{(\alpha)})^2 (r_j^{(\alpha)})^2 = V \quad \cdots\cdots (108)$$

を得る。式 (105) と式 (108) を $(r_i^{(\alpha)})^2$ と $(r_j^{(\alpha)})^2$ に関する連立方程式考えれば固定した i と j に対しては高々2つの解しか存在しない。今、$1 \le i < j \le n$ であったから集積点は高々有限個しか存在しないことになる。以後、この有限個の集積点を $\boldsymbol{r}_i (i = 1, \ldots, L)$ と表すことにする。

次に、集積点の集合 $\{\boldsymbol{r}_i\}$ が2点からなることを示す。ϵ を

$$\epsilon \triangleq \frac{1}{3} \min_{1 \le i < j \le L} \|\boldsymbol{r}_i - \boldsymbol{r}_j\| \quad \cdots\cdots (109)$$

で決められる正の数とする。また、$U(\boldsymbol{r}_i : \epsilon)$ を集積点 $\boldsymbol{r}_i$ の ϵ 近傍とする。すなわち、

$$U(\boldsymbol{r}_i; \epsilon) \triangleq \{\boldsymbol{x} \in S^{n-1} : \|\boldsymbol{x} - \boldsymbol{r}_i\| < \epsilon\}$$

とする。すると、これらの近傍は互いに離れている。また、ある十分大きな整数 N_1 が存在して、$k \ge N_1$ なら

$$\boldsymbol{d}_k \in \bigcup_{i=1}^{L} U(\boldsymbol{r}_i; \epsilon) \quad \cdots\cdots (110)$$

が成立する。もしそうでないとすれば、$\{\boldsymbol{r}_i\}$ 以外に集積点が存在することになるからである。一方、補題 3.5 の系 3.1から、ある十分大きな整数 $k \ge N_2$ が存在して $k \ge N_2$ なら

$$\|\boldsymbol{d}_{k+2} - \boldsymbol{d}_k\| < \epsilon \quad \cdots\cdots (111)$$

が成立する。今、$N = \max\{N_1, N_2\}$ と置く。$i \ne j$ なら

$$\|\boldsymbol{r}_i - \boldsymbol{r}_j\| \ge 3\epsilon \quad \cdots\cdots (112)$$

が成立するから、もし、$k \ge N$ を満たすある k が

$$\boldsymbol{d}_k \in U(\boldsymbol{r}_i; \epsilon) \quad \cdots\cdots (113)$$

を満たしていれば、

$$\boldsymbol{d}_{k+2} \in U(\boldsymbol{r}_i; \epsilon) \quad \cdots\cdots (114)$$

が成立する。従って、2つの点列 $\{\boldsymbol{d}_{2k}\}$ と $\{\boldsymbol{d}_{2k+1}\}$ は各々収束する。このことから集積点は高々2点しかないことがわかる。そして、補題 3.3から、$\boldsymbol{d}_k$ と $\boldsymbol{d}_{k+1}$ が互いに直交していることを考慮すれば、集積点が丁度2個存在することがわかる。 ■

更に、式 (78) から一方の集積点 $\boldsymbol{r}_1$ が

$$\boldsymbol{r}_1 = r_i^{(1)} \boldsymbol{u}_i + r_j^{(1)} \boldsymbol{u}_j \quad \cdots\cdots (115)$$

と表されれば、他方の集積点は

$$\boldsymbol{r}_2 = -\frac{\boldsymbol{A}\boldsymbol{r}_1 - \bar{\lambda}(\boldsymbol{r}_1)\boldsymbol{r}_1}{\sqrt{V}} = -r_j^{(1)} \boldsymbol{u}_i + r_i^{(1)} \boldsymbol{u}_j \quad \cdot (116)$$

と表されることがわかる。逆に、$\boldsymbol{r}_1$ は $\boldsymbol{r}_2$ を用いて

$$\boldsymbol{r}_1 = -\frac{\boldsymbol{A}\boldsymbol{r}_2 - \bar{\lambda}(\boldsymbol{r}_2)\boldsymbol{r}_2}{\sqrt{V}} \quad \cdots\cdots (117)$$

と表される。従って、点列 $\{\boldsymbol{d}_k\}$ は漸近的に2つの集積点 $\boldsymbol{r}_1$ と $\boldsymbol{r}_2$ を交互に取るようになるのである。

以上で定理 3.4の証明ができた。

定義 3.6 $\nabla^2 f(\boldsymbol{x})$ が対角行列になるとき、$f(\boldsymbol{x})$ は変数分離型と呼ぶ。

例、

$$f(x, y) = \frac{1}{2}x^2 + y^2 + x^3 \quad \cdots\cdots (118)$$

は

$$\nabla^2 f(x, y) = \begin{pmatrix} 1 + 6x & 0 \\ 0 & 2 \end{pmatrix} \quad \cdots\cdots (119)$$

なので、変数分離型である。一方、

$$f(x, y) = \frac{1}{2}x^2 + y^2 + x^2 y \quad \cdots\cdots (120)$$

は

$$\nabla^2 f(x,y) = \begin{pmatrix} 1+2y & 2x \\ 2x & 2 \end{pmatrix} \cdots\cdots\cdots\cdots (121)$$

となり、対角行列でないので、変数分離型でない。なお、二次形式は明らかに変数分離型である。

次の定理は、分散の極限値が0の収束しないための、十分条件を与える。

定理 3.5 有限回の反復で、厳密解が得られないとする。このとき、$f(\boldsymbol{x})$ が変数分離型なら、分散の極限は0でない。従って、定理3.4が成り立つ。

(証明)

$$\left|\frac{(A_{k+1}\boldsymbol{d}_k, \boldsymbol{d}_{k+1})}{(A_k\boldsymbol{d}_k, \boldsymbol{d}_{k+1})}\right| \le 1 + O(\|\boldsymbol{x}_k\|) \cdots\cdots\cdots\cdots (122)$$

となることが示せれば、式(80)から

$$\begin{aligned} 1 &\ge \left|\frac{(A_{k+1}\boldsymbol{d}_k, \boldsymbol{d}_{k+1})}{(A_{k+1}\boldsymbol{d}_{k+1}, \boldsymbol{d}_{k+2})}\right| \\ &= \left|\frac{(A_k\boldsymbol{d}_k, \boldsymbol{d}_{k+1})}{(A_{k+1}\boldsymbol{d}_{k+1}, \boldsymbol{d}_{k+2})}\right| \cdot (1 + O(\|\boldsymbol{x}_k\|)) \\ &= \frac{\sqrt{V_k}}{\sqrt{V_{k+1}}} \cdot (1 + O(\|\boldsymbol{x}_k\|)) \cdots\cdots\cdots\cdots (123) \end{aligned}$$

が成り立つ。もし、ある k で $V_k = 0$ となれば、$\boldsymbol{x}_{k+1} = \mathbf{o}$ となり、反復は終了し、仮定に反する。従って、$\lim_{k\to\infty} V_k \ne 0$ が成り立つ。そこで、以下では、不等式(122)を証明する。式(78)から

$$\begin{aligned} & d_i^{(k+1)} \\ &= -\frac{\lambda_i(\boldsymbol{x}_k) - \lambda(\bar{\boldsymbol{d}}_k)}{\sqrt{V_k}} d_i^{(k)} \\ &= -\frac{\sum_{1\le j\le n} (\lambda_i(\boldsymbol{x}_k) - \lambda_j(\boldsymbol{x}_k))(d_j^{(k)})^2}{\sqrt{V_k}} d_i^{(k)} \cdots\cdots (124) \end{aligned}$$

だから

$$\begin{aligned} & (A_k\boldsymbol{d}_k, \boldsymbol{d}_{k+1}) \\ &= -\frac{\sum_{1\le i<j\le n} (\lambda_i(\boldsymbol{x}_k) - \lambda_j(\boldsymbol{x}_k))^2 (d_i^{(k)})^2 (d_j^{(k)})^2}{\sqrt{V_k}} \\ & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (125) \end{aligned}$$

となる。従って、

$$\begin{aligned} & \left|\frac{d_i^{(k)} d_i^{(k+1)}}{(A_k\boldsymbol{d}_k, \boldsymbol{d}_{k+1})}\right| \\ &= -\frac{\sum_{1\le j\le n} |\lambda_i(\boldsymbol{x}_k) - \lambda_j(\boldsymbol{x}_k)| (d_i^{(k)})^2 (d_j^{(k)})^2}{\sum_{1\le i<j\le n} (\lambda_i(\boldsymbol{x}_k) - \lambda_j(\boldsymbol{x}_k))^2 (d_i^{(k)})^2 (d_j^{(k)})^2} \\ &\le \sum_{\substack{1\le j\le n \\ j\ne i}} \frac{1}{|\lambda_i(\boldsymbol{x}_k) - \lambda_j(\boldsymbol{x}_k)|} \cdots\cdots\cdots\cdots (126) \end{aligned}$$

が成り立つ。その結果

$$\begin{aligned} & \left|\frac{(A_{k+1}\boldsymbol{d}_k, \boldsymbol{d}_{k+1})}{(A_k\boldsymbol{d}_k, \boldsymbol{d}_{k+1})}\right| \\ &\le 1 + \left|\frac{((A_{k+1} - A_k)\boldsymbol{d}_k, \boldsymbol{d}_{k+1})}{(A_k\boldsymbol{d}_k, \boldsymbol{d}_{k+1})}\right| \\ &\le 1 + \sum_{1\le i\le n} \|A_{k+1} - A_k\| \left|\frac{d_i^{(k)} d_i^{(k+1)}}{(A_k\boldsymbol{d}_k, \boldsymbol{d}_{k+1})}\right| \\ &= 1 + O(\|\boldsymbol{x}_k\|) \cdots\cdots\cdots\cdots\cdots\cdots (127) \end{aligned}$$

が成立する。 ■

次に、探索ベクトルが次第に収束する2つのベクトル $\boldsymbol{r}_1, \boldsymbol{r}_2$ を構成する2つの成分が、どの固有ベクトルに属しているか調べる。

定義 3.7 ある N があって、$k \ge N$ なら $(\boldsymbol{x}_k, \boldsymbol{u}_i) = 0$ が成り立つとき、固有値 λ_i の成分が縮退するという。

特に、二次形式では $(\boldsymbol{x}_0, \boldsymbol{u}_i) = 0$ なら全ての k に対して、$(\boldsymbol{x}_k, \boldsymbol{u}_i) = 0$ が成り立つ。従って、2つの集積点 $\boldsymbol{r}_1, \boldsymbol{r}_2$ に対しても

$$(\boldsymbol{r}_1, \boldsymbol{u}_i) = (\boldsymbol{r}_2, \boldsymbol{u}_i) = 0$$

が成り立つ。

補題 3.8 最大および最小固有値の成分が縮退しないと仮定する。もし、$f(\boldsymbol{x})$ が変数分離型なら、2つの集積点 $\boldsymbol{r}_1$ と $\boldsymbol{r}_2$ は、どちらも最大固有値 λ_n 及び最小固有値 λ_1 に対応する2つの成分を持つ。

(証明) 集積点 $\boldsymbol{r}_1$ が最大固有値 λ_n の成分を持つことを背理法で示す。集積点 $\boldsymbol{r}_1$ が最大固有値 λ_n の成分を持たないと仮定する。すると、

$$\boldsymbol{r}_1 = r_i^{(1)}\boldsymbol{u}_i + r_j^{(1)}\boldsymbol{u}_j \cdots\cdots\cdots\cdots\cdots\cdots (128)$$

と表すことができる。ここで、$1 \le i < j < n$ かつ $r_i^{(1)}, r_j^{(1)} \ne 0$ である。また、式(116)によって、もう一方の集積点 $\boldsymbol{r}_2$ は

$$\boldsymbol{r}_2 = -r_j^{(1)}\boldsymbol{u}_i + r_i^{(1)}\boldsymbol{u}_j \cdots\cdots\cdots\cdots\cdots\cdots (129)$$

と表されることから

$$\lambda_1 < \bar{\lambda}(\boldsymbol{r}_1),\ \bar{\lambda}(\boldsymbol{r}_2) < \lambda_j < \lambda_n \cdots\cdots\cdots\cdots (130)$$

を得る。点列 $\{\boldsymbol{d}_k\}$ が $\boldsymbol{r}_1$ か $\boldsymbol{r}_2$ に集積するから、ある十分大きな整数 N_1 が存在して、$k \ge N_1$ なら

$$\lambda_1 < \bar{\lambda}(\boldsymbol{d}_k) < \lambda_j < \lambda_n \cdots\cdots\cdots\cdots\cdots\cdots (131)$$

が成立する。$\nabla^2 f(\boldsymbol{x})$ が Lipschitz 連続であるから

$$A_k = A + O(\|\boldsymbol{x}_k\|) \cdots\cdots\cdots\cdots\cdots\cdots (132)$$

である。そこで、式 (78) から $k \geq N_1$ に対して

$$\frac{d_n^{(k+1)}}{d_j^{(k+1)}} = \frac{\left[(\lambda_n - \bar{\lambda}(\boldsymbol{d}_k)) + O(\|\boldsymbol{x}_k\|)\right] d_n^{(k)}}{\left[(\lambda_j - \bar{\lambda}(\boldsymbol{d}_k)) + O(\|\boldsymbol{x}_k\|)\right] d_j^{(k)}}$$
$$\geq \frac{\left[(\lambda_n - \lambda_1 + O(\|\boldsymbol{x}_k\|)\right] d_n^{(k)}}{\left[(\lambda_j - \lambda_1 + O(\|\boldsymbol{x}_k\|)\right] d_j^{(k)}} \cdots\cdots\cdots (133)$$

が成立する。また、$\|\boldsymbol{x}_k\| \to 0$ だから、ある $c > 1$ と十分大きな N_2 があって、$k \geq N_2$ なら

$$\left| \frac{\lambda_n - \lambda_1 + O(\|\boldsymbol{x}_k\|)}{\lambda_j - \lambda_1 + O(\|\boldsymbol{x}_k\|)} \right| > c \geq 1 \cdots\cdots\cdots\cdots\cdots (134)$$

が成り立つ。そこで、

$$\lim_{k\to\infty} |d_j^{(k)}| \geq \min\{|r_j^{(1)}|, |r_j^{(2)}|\} > 0 \cdots\cdots\cdots (135)$$

を考慮すれば

$$\lim_{k\to\infty} d_n^{(k)} = \infty \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (136)$$

を得る。これは $\boldsymbol{d}_k$ が単位球 S^{n-1} 上の点であることに矛盾する。従って、2 つの集積点 $\boldsymbol{r}_1, \boldsymbol{r}_2$ は共に λ_n の成分を持たなければならない。

同様にして、2 つの集積点 $\boldsymbol{r}_1, \boldsymbol{r}_2$ が λ_1 の成分を持つことが示せる。 ■

式 (133) から、十分大きな k に対して、$d_i^{(k)} \neq 0$ なら $d_i^{(k+1)} \neq 0$ であり、以降、縮退が起らないことがわかる。また、この定理では、変数分離型を仮定している。変数分離型でないとき、2 つの集積点がどの固有ベクトルで表されるかは不明である。

3.2 分散 V_k の極限値が 0 の場合

分散が 0 に収束するには、式 (77) から、集積点で

$$\|A\boldsymbol{r}_\alpha - \bar{\lambda}(\boldsymbol{r}_\alpha)\boldsymbol{r}_\alpha\| = 0 \cdots\cdots\cdots\cdots\cdots\cdots\cdots (137)$$

を満たさなければならない。従って、探索ベクトルが行列 A の固有ベクトルに収束することが必要条件である。

この性質を利用して、分散が 0 に収束し、かつ定理 3.4 が成り立たない例をあげる。

例、

$$f(x,y) = \frac{1}{2}x^2 + y^2 - x^3y + xy^3 \cdots\cdots\cdots\cdots (138)$$

は

$$\nabla^2 f(x,y) = \begin{pmatrix} 1+6xy & 3(x^2-x^2) \\ 3(x^2-y^2) & 2-6xy \end{pmatrix} \cdots (139)$$

なので、変数分離型ではない。(x,y) は R^2 内の点列であるので、$\boldsymbol{d}_k$ が行列 $A = \nabla^2 f(0,0)$ の固有ベクトルなら、それに直交する $\boldsymbol{d}_{k+1}$ も固有ベクトルである。従って、$\{\boldsymbol{d}_k\}$ は常に固有ベクトルである。そこで、

$$\nabla f(x,y) = \begin{pmatrix} x - y^3 + 3x^2y \\ 2y - 3xy^2 + x^3 \end{pmatrix} \cdots\cdots\cdots\cdots (140)$$

から、$2y - 3xy^2 + x^3 = 0$ を満たす点 $(1, -\frac{1}{3})$ を $\boldsymbol{x}_0$ とする。すると

$$\begin{aligned} \lim_{k\to\infty} V_k &= \lim_{k\to\infty} -(A_k\boldsymbol{d}_k, \boldsymbol{d}_{k+1}) \\ &= \lim_{k\to\infty} -(A\boldsymbol{d}_k, \boldsymbol{d}_{k+1}) \\ &= 0 \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (141) \end{aligned}$$

が成立する。また、2 次元であることから、$\boldsymbol{d}_{k+2} = \boldsymbol{d}_k$ または $\boldsymbol{d}_{k+2} = -\boldsymbol{d}_k$ が成り立つが、実際、$\boldsymbol{d}_k$ は

$$(-1,0),\ (0,1),\ (1,0),\ (0,-1),\ (-1,0),\ \ldots$$

と循環する。これは、定理3.4とは異なり、4 つの探索ベクトルを循環し、渦巻き状に収束する。このようなことは、二次形式では起らなかったことである (図 3参照)。

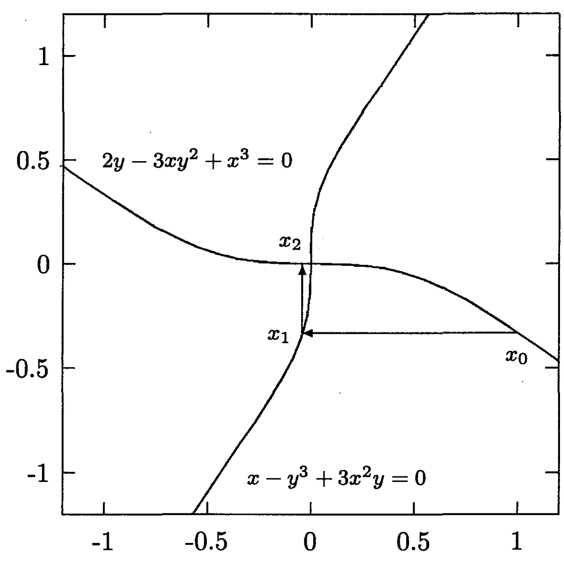

図 3 探索ベクトルが 2 方向にならない例
Fig.3 Example of search directions do not alternate two directions.

しかし、分散が 0 に収束するからといって、必ずしも定理 3.4が成り立たないわけではない。例えば

$$f(x,y) = \frac{1}{2}x^2 + \frac{3}{2}y^2 + x^2y + xy^2 \cdots\cdots\cdots (142)$$

では、$\boldsymbol{x}_0 = (-\frac{1}{2}, -\frac{1}{8})$ とすると、探索ベクトルは行列 A の固有ベクトルで、2 方向を交互に取り、やはりジグザグに解に近づく (図 4参照)。

なお、3 次元以上で、分散が 0 に収束する場合が存在するかどうかは未解決である。最後に、収束速度に関する定理を 1 つ述べる。

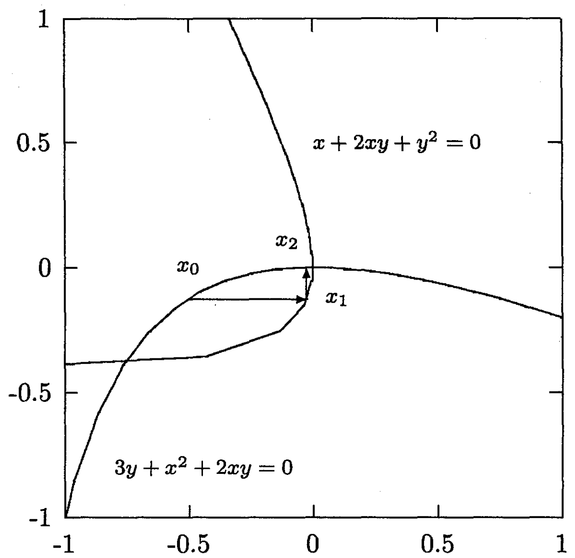

図 4　探索ベクトルが 2 方向になる例
Fig.4 Example of search directions alternate two directions.

定理 3.6 分散が 0 に収束するとき、最急降下法は超 1 次収束する。

（証明） 式 (48) から、$\boldsymbol{x} \in \lambda_f(\boldsymbol{x}_0, \boldsymbol{o})$ で、$m\|\boldsymbol{x}\| \leq \|\nabla f(\boldsymbol{x})\| \leq M\|\boldsymbol{x}\|$ となる m, M が存在するから

$$\frac{\|\boldsymbol{x}_{k+1}\|}{\|\boldsymbol{x}_k\|} \leq \frac{M\|\nabla f(\boldsymbol{x}_{k+1})\|}{m\|\nabla f(\boldsymbol{x}_k)\|} \quad \cdots\cdots (143)$$

となる。ここで、式 (74) と式 (78) から

$$\begin{aligned}\nabla f(\boldsymbol{x}_{k+1}) &= \nu_k\alpha_k(A_k - \alpha_k^{-1}I)\boldsymbol{d}_k \\ &= \nu_k\alpha_k\sqrt{V_k}\boldsymbol{d}_{k+1} \quad \cdots\cdots (144)\end{aligned}$$

だから

$$\frac{\|\boldsymbol{x}_{k+1}\|}{\|\boldsymbol{x}_k\|} \leq \frac{M\alpha_k\sqrt{V_k}}{m} \quad \cdots\cdots (145)$$

となる。α_k が有界で、$\lim_{k\to\infty} V_k = 0$ だから

$$\lim_{k\to\infty} \frac{\|\boldsymbol{x}_{k+1}\|}{\|\boldsymbol{x}_k\|} = 0 \quad \cdots\cdots (146)$$

となり、$\{\boldsymbol{x}_k\}$ は超 1 次収束する。 ■

4. むすび

本論文では、最急降下法が一般の非線形関数に関して二次形式と同様な性質を持つかどうかを議論した。二次形式の場合とは異なり、非線形関数はそのヘッセ行列が、点の位置により異なる。赤池が二次形式の場合に行ったように、この変化するヘッセ行列の固有値を確率変数値、探索ベクトルの各成分の 2 乗を確率分布と考えたとき、その分散がどうなるかを考察した。その結果、分散が零以外に収束する場合は、二次形式と同様に、最急降下法は、その探索ベクトルが漸近適に互いに直交する 2 つのベクトル間を交互にとるようになり、特に変数分離型では、その探索ベクトルは、最大固有値と最小固有値に対応する 2 つの固有ベクトルの張る 2 次元空間内に収束することが示せた。しかし、二次形式の場合とは異なり、分散が 0 に収束する場合が存在し、探索方向が必ずしも、2 方向を交互に取らない例がある。その例として、2 変数の場合を示したが、3 変数以上でこのような場合が存在するかという問題が残されている。

参考文献

[1] G. E. Forsythe and T. S. Motzkin: "Asymptotic Properties of the Optimum Gradient Method," *Bull. AMS*, vol. 57, p. 183, 1951(Abstract).

[2] H. Akaike: "On a Successive Transformation of Probability Distribution and Its Application to the Analysis of the Optimum Gradient Method," *Ann. Inst. Stat. Math.*, vol. 11, pp. 1-16, 1959.

[3] J. Kowalik and M. R. Osborne: *Methods for Unconstrained Optimization Problems*, American Elsevier Publishing Company, New York, 1968.

[4] 今野浩，山下浩：非線形計画法，日科技連，1978.

[5] G. E. Forsythe: "On the Asymptotic Directions of the s-Dimensional Optimum Gradient Method," *Numer. Math.*, vol. 11, pp. 57-76, 1968.

[6] H. Akaike: "On a Computation Method for Eigenvalue Problems and Its Application to Statistical Analysis," *Ann. Inst. Stat. Math.*, vol. 10, pp. 1-20, 1958.

[7] T. Ozeki and T. Iijima: "Behavior of the Steepest Descent Method in Minimizing Rayleigh Quotient," *IEICE Trans. Fundamentals*, vol. E80-A, no. 1, pp. 176-182, Jan. 1997.

[8] D. K. Faddeev and V. N. Faddeeva: *Computational Methods of Linear Algebra*, W. H. Freeman and Company, 1963.

[9] 杉浦光夫：解析入門 I，東京大学出版，1980.

[10] 田村明久，松村正和：最適化法，共立出版，2002.

[11] 茨木俊秀，福島雅夫:最適化の手法，共立出版，1993.