

2 变数関数の極値と Hesse 行列【微分積分】

2 变数関数の極値問題についての初学者向けの解説。1 变数関数の極値の求め方の復習から始め、同じ考え方で 2 变数関数の極値問題にアプローチする。このノートでは Hesse 行列を用いた極値の判定法が適用できる C^2 級関数(2 回微分可能で 2 階(偏)導関数が連続な関数)のみを考えるが、部分的に条件を緩められる主張については脚注で補足する。Taylor 展開(漸近展開)と行列の対角化の知識を仮定する。

1 变数関数の極値

1 变数関数が極値をとるための必要条件 ある点 $x = a$ の近くでの関数 $f(x)$ の増減を考える。¹ $x = a$ からの変位 Δx に対する $f(x)$ の変化を

$$\Delta f(a; \Delta x) \equiv f(a + \Delta x) - f(a)$$

と書こう。このとき、関数 $f(x)$ が極値をとるとは、以下のように定義される。

- 任意の微小変位 $\Delta x (\neq 0)$ に対して $\Delta f(a; \Delta x) > 0$ となるとき、関数 $f(x)$ は $x = a$ で**極小値**をとる。²
- 任意の微小変位 $\Delta x (\neq 0)$ に対して $\Delta f(a; \Delta x) < 0$ となるとき、関数 $f(x)$ は $x = a$ で**極大値**をとる。
- 微小変位 Δx の符号に依存して $\Delta f(a; \Delta x)$ の符号が変わるととき、関数 $f(x)$ は $x = a$ で**極値**をとらない。

関数 $f(x)$ が**極大値・極小値**をとる点のことを、**極大点・極小点**(あわせて**極値点**)と呼ぶ。関数 $f(x)$ の $x = a$ 付近での振舞いは、Taylor 展開(漸近展開)で調べることができる。³

$$f(a + \Delta x) = f(a) + f'(a)\Delta x + o(\Delta x) \quad (\Delta x \rightarrow 0)$$

ここで $f'(a) \neq 0$ であれば、 Δx の符号に応じて $\Delta f(a; \Delta x) = f'(a)\Delta x + o(\Delta x)$ の符号が変化するから、関数 $f(x)$ は $x = a$ で**極値**をとらない。したがって、この対偶をとることで、関数 $f(x)$ が $x = a$ で**極値**をとるならば、

$$f'(a) = 0$$

であることがわかる。⁴ これは必要条件であり、十分条件ではないことに注意。⁵ $f'(a) = 0$ を満たす点のことを**停留点**(stationary point)または**臨界点**(critical point)と呼ぶ。停留点が**極値**をとる点の候補を与える。

1 变数関数が極値をとるための十分条件 $f'(a) = 0$ を満たす停留点 $x = a$ の周りで、

$$f(a + \Delta x) = f(a) + \frac{1}{2}f''(a)\Delta x^2 + o(\Delta x^2)$$

と展開できる。展開の 1 次の項 $f'(a)\Delta x$ は停留点の条件 $f'(a) = 0$ で消えていることに注意。停留点 $x = a$ からの微小変位 Δx に対する $f(x)$ の変化は、

$$\Delta f(a; \Delta x) \equiv f(a + \Delta x) - f(a) = \frac{1}{2}f''(a)\Delta x^2 + o(\Delta x^2)$$

と評価できる。停留点からの微小変化 $\Delta f(a; \Delta x)$ の表式から、 $f(x)$ の**極値**について以下のようないくつかの判定ができる。

- $f''(a) > 0$ ならば、任意の微小変位 $\Delta x (\neq 0)$ に対して $\Delta f(a; \Delta x) > 0$ となるから、 $x = a$ は**極小点**である。
- $f''(a) < 0$ ならば、任意の微小変位 $\Delta x (\neq 0)$ に対して $\Delta f(a; \Delta x) < 0$ となるから、 $x = a$ は**極大点**である。
- $f''(a) = 0$ のときは、 $x = a$ が**極値点**かどうか判定できない。この場合には、無視していた $o(\Delta x^2)$ の部分が重要になる。 $f(x)$ が高階微分可能な場合には、より高次の導関数を調べるなどの方法がある。⁶

¹ $f(x)$ は関数 f の x における値であるが、 $f(x)$ を関数と呼んでしまうことにする。

² もう少し正確には「 a を含むある開区間 I があって、任意の $x \in I \setminus \{a\}$ に対して $f(x) > f(a)$ となるとき、関数 $f(x)$ は $x = a$ で**極小値**をとる」と定義する。極大値についても同様。等号を入れて、 $\Delta f(a; \Delta x) \geq 0$ のとき**極小値**、 $\Delta f(a; \Delta x) \leq 0$ のとき**極大値**と定義することもある。たとえば、定数関数は**極値**をとらないが、広義には任意の点で**極大値**かつ**極小値**をとる。このノートでは狭義の定義を採用する。

³ $f(x)$ が a を含む開区間で C^1 級であれば 1 次の漸近展開ができる。平均値の定理より $f(a + \Delta x) = f(a) + f'(a + \theta\Delta x)\Delta x = f(a) + f'(a)\Delta x + (f'(a + \theta\Delta x) - f'(a))\Delta x$ を満たす $\theta \in (0, 1)$ が存在する。 $f'(x)$ の連続性より $\Delta x \rightarrow 0$ で $f'(a + \theta\Delta x) - f'(a) \rightarrow 0$ なので、 $(f'(a + \theta\Delta x) - f'(a))\Delta x = o(\Delta x)$ である。

⁴ これは $f(x)$ が a で**微分可能**であれば言える。 $f(x)$ が a で**極小値**をとるから、 $f(x)$ の定義域をある開区間 I に制限すると $f(x)$ は a で**最小値**をとる。 $h \neq 0$ に対して $a + h \in I$ であれば $f(a + h) - f(a) > 0$ であるから $\lim_{h \rightarrow +0} (f(a + h) - f(a))/h \geq 0$ かつ $\lim_{h \rightarrow -0} (f(a + h) - f(a))/h \leq 0$ が成り立つので、 $f'(a) = \lim_{h \rightarrow 0} (f(a + h) - f(a))/h = 0$ である。

⁵ $f(x) = x^3$ は $x = 0$ で $f'(0) = 0$ となるが、 $x > 0$ では $f(x) > 0$ 、 $x < 0$ では $f(x) < 0$ であるから、この点で**極小値**も**極大値**もとらない。

⁶ $f(x) = x^3$ の場合には $f''(0) = 0$ となり、 $f'''(0) = 6 \neq 0$ を調べてはじめて $x = 0$ が**極値点**ではないことがわかる。 $f(x) = x^4$ の場合には $f^{(4)}(0) = 24 > 0$ を調べてはじめて $x = 0$ が**極小点**であることがわかる。ただし、前者は脚注⁵の議論、後者は $x \neq 0$ に対して $f(x) = x^4 > 0 = f(0)$ であることを言えば十分で、必ずしも導関数を調べる必要はない。

2変数関数の極値

1変数関数の極値を調べたのと同様の手法で、2変数関数 $f(x, y)$ の極値を調べることができる。途中までの議論は1変数関数の場合と完全に対応しているので、対比させながら読むとよい。

2変数関数の極値点の定義 ある点 $(x, y) = (a, b)$ の近くでの関数 $f(x, y)$ の増減を考える。 (a, b) からの変位 $(\Delta x, \Delta y)$ に対して、

$$\Delta f(a, b) \equiv f(a + \Delta x, b + \Delta y) - f(a, b)$$

と書く。本来なら1変数関数の場合と同様に $\Delta f(a, b; \Delta x, \Delta y)$ などと書くべきだが、煩わしいので $(\Delta x, \Delta y)$ は省略した。このとき、1変数関数の場合と同様に、関数 $f(x, y)$ が極値をとるとは、以下のように定義される。

- 任意の微小変位 $(\Delta x, \Delta y) \neq (0, 0)$ に対して $\Delta f(a, b) > 0$ のとき、 (a, b) は極小点である。⁷
- 任意の微小変位 $(\Delta x, \Delta y) \neq (0, 0)$ に対して $\Delta f(a, b) < 0$ のとき、 (a, b) は極大点である。
- 微小変位 $(\Delta x, \Delta y)$ に依存して $\Delta f(a, b)$ の符号が変わると、 (a, b) は極値点ではない。

1変数関数の場合の微小変位 Δx の“方向”が正負の2通りしか無かったのに対して、2変数関数の場合の変位 $(\Delta x, \Delta y)$ の方向には、 360° の自由度があることに注意が必要である。

2変数関数が極値をとるための必要条件 ある点 $(x, y) = (a, b)$ の周りでの関数 $f(x, y)$ の振舞いは、Taylor展開(漸近展開)を用いて調べることができる。⁸

$$f(a + \Delta x, b + \Delta y) = f(a, b) + f_x(a, b)\Delta x + f_y(a, b)\Delta y + o(\sqrt{\Delta x^2 + \Delta y^2}) \quad ((\Delta x, \Delta y) \rightarrow (0, 0))$$

もし $f_x(a, b) \neq 0$ であれば、 $(\Delta x, \Delta y) = (\epsilon, 0)$ と $(\Delta x, \Delta y) = (-\epsilon, 0)$ という2つの方向で $\Delta f(a, b)$ の符号が異なるから、そのような (a, b) は極値点ではない。 $f_y(a, b) \neq 0$ の場合も同様である。したがって、この対偶をとることで、関数 $f(x, y)$ が点 $(x, y) = (a, b)$ で極値をとるためには、

$$f_x(a, b) = f_y(a, b) = 0$$

が必要であることがわかる。⁹ 条件 $f_x(a, b) = f_y(a, b) = 0$ を満たす点を停留点または臨界点と呼ぶ。1変数関数の場合と同様、この条件も極値をとるための必要条件であり、十分条件ではない。停留点で極値をとらない場合、その点を鞍点(saddle point)と呼ぶ。¹⁰

停留点付近での関数の振舞い $f_x(a, b) = f_y(a, b) = 0$ を満たす停留点 (a, b) の周りで、関数 $f(x, y)$ は、

$$f(a + \Delta x, b + \Delta y) = f(a, b) + \frac{1}{2}(\bar{f}_{xx}\Delta x^2 + 2\bar{f}_{xy}\Delta x\Delta y + \bar{f}_{yy}\Delta y^2) + o(\Delta x^2 + \Delta y^2) \quad ((\Delta x, \Delta y) \rightarrow (0, 0))$$

と展開できる。ここで、表記の簡略化のために、 (a, b) で評価した値を \bar{f} のように表した。たとえば、 $\bar{f}_{xx} = f_{xx}(a, b)$ である。¹¹ 停留点 $(x, y) = (a, b)$ からの微小変位 $(\Delta x, \Delta y)$ に対する $f(x, y)$ の変化は、

$$\begin{aligned} \Delta f(a, b) &\equiv f(a + \Delta x, b + \Delta y) - f(a, b) \\ &= \frac{1}{2}(\bar{f}_{xx}\Delta x^2 + 2\bar{f}_{xy}\Delta x\Delta y + \bar{f}_{yy}\Delta y^2) + o(\Delta x^2 + \Delta y^2) \quad ((\Delta x, \Delta y) \rightarrow (0, 0)) \end{aligned}$$

と評価できる。1変数関数の場合と異なり、関数の値の変化 $\Delta f(a, b)$ が3つの項(と微小項)の和で表されており、特に $\Delta x\Delta y$ の項の存在により、この表式を見て直ちに $\Delta f(a, b)$ の正負を判定することはできない。 $\Delta f(a, b)$ は2次形式と呼ばれる形をしており、その解析には線形代数の知識を用いることができる。

⁷もう少し正確には「 (a, b) を含むある領域(連結開集合) D があって、任意の $(x, y) \in D \setminus \{(a, b)\}$ に対して $f(x, y) > f(a, b)$ となるとき、関数 $f(x, y)$ は $(x, y) = (a, b)$ で極小値をとる」と定義する。

⁸ $f(x, y)$ が⁹ (a, b) を含む領域で C^1 級であれば1次の漸近展開ができる。2変数の平均値の定理を用いて脚注³と同様に証明できる。

⁹これは $f(x, y)$ が (a, b) で偏微分可能であれば言える。 $f(x, y)$ が (a, b) で極値をとるので、 $y = b$ を固定した1変数関数 $f(x, b)$ は $x = a$ で極値をとり、 $f_x(a, b) = 0$ である。同様に $f_y(a, b) = 0$ も言える。

¹⁰極大点・極小点に狭義・広義の定義があったことに伴い、鞍点にも狭義・広義の定義がある。

¹¹この表記は一般的ではないので、他では(定義せずに)使ってはならない。

Hesse 行列の導入 これ以降, $\Delta f(a, b)$ の引数を省略して単に Δf と書くことにする. Δf は行列を用いて

$$2\Delta f = [\Delta x \ \Delta y] \begin{bmatrix} \bar{f}_{xx} & \bar{f}_{xy} \\ \bar{f}_{xy} & \bar{f}_{yy} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}$$

と書くことができる. 右辺が見やすいように全体に 2 を掛けた. また, 微小項は無視した. ここに現れる実対称行列

$$H \equiv \begin{bmatrix} \bar{f}_{xx} & \bar{f}_{xy} \\ \bar{f}_{xy} & \bar{f}_{yy} \end{bmatrix}$$

は **Hesse 行列** (Hessian matrix) と呼ばれる. Hesse 行列 H は実対称行列なので, 直交行列 O を用いて

$$O^T H O = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

と直交対角化することができる. λ_1 と λ_2 は Hesse 行列の固有値であり, Hesse 行列が実対称行列であることから, 共に実数である.¹² 直交行列 O を用いて

$$\begin{bmatrix} \Delta X \\ \Delta Y \end{bmatrix} = O^T \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}$$

と変換すると, 以下のように Δf を「2乗の和」の形に書き直すことができる.¹³

$$2\Delta f = [\Delta X \ \Delta Y] \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \Delta X \\ \Delta Y \end{bmatrix} = \lambda_1(\Delta X)^2 + \lambda_2(\Delta Y)^2.$$

$(\Delta x, \Delta y)$ と $(\Delta X, \Delta Y)$ が一対一に対応することに注意すると, Hesse 行列の 2 つの固有値 λ_1, λ_2 の正負と Δf の正負の関係について以下のことが言える.

- 固有値が共に正の場合, 任意の微小変位 $(\Delta X, \Delta Y) \neq (0, 0)$ に対して $\Delta f > 0$ より, (a, b) は極小点である.
- 固有値が共に負の場合, 任意の微小変位 $(\Delta X, \Delta Y) \neq (0, 0)$ に対して $\Delta f < 0$ より, (a, b) は極大点である.
- 固有値の符号が異なる場合, 2 つの方向 $(\Delta X, \Delta Y) = (\epsilon, 0)$ と $(\Delta X, \Delta Y) = (0, \epsilon)$ とで Δf の符号が異なる. つまり, (a, b) は極値点ではなく, 鞍点である.
- 固有値が 0 を含む場合, $(\Delta X, \Delta Y) = (\epsilon, 0)$ または $(\Delta X, \Delta Y) = (0, \epsilon)$ の方向で $\Delta f = 0$ となり, この方向の関数 $f(x, y)$ の増減を判定できない. この場合には無視していた微小項を調べなくてはならない.

Hesse 行列の固有値と関数が極値をとる条件 これまでの議論で, 関数 $f(x, y)$ の停留点 (a, b) が極値点か否かは, Hesse 行列の固有値によって決まることがわかった. Hesse 行列の固有値は, 固有値方程式

$$0 = |\lambda E - H| = \lambda^2 - (\bar{f}_{xx} + \bar{f}_{yy})\lambda + (\bar{f}_{xx}\bar{f}_{yy} - (\bar{f}_{xy})^2) = \lambda^2 - (\text{tr } H)\lambda + \det H,$$

の解として求まる. ここで, $\text{tr } H$ は Hesse 行列のトレース, $\det H$ は行列式である. Hesse 行列の行列式 (Hessian determinant) は, 単にヘッシャン (Hessian) と呼ばれることが多い. 固有値方程式を解いて固有値 λ を求めてよいが, 必要なのは固有値 λ の値そのものではなく λ の正負であるから, 2 次方程式の解と係数の関係を用いればよい.¹⁴

$$\lambda_1 + \lambda_2 = \text{tr } H, \quad \lambda_1 \lambda_2 = \det H,$$

が成り立つから, 以下の条件が得られる.

- 極小点 $\iff H$ の固有値が共に正 $\iff \lambda_1 \lambda_2 > 0, \lambda_1 + \lambda_2 > 0 \iff \det H > 0, \text{tr } H > 0$.
- 極大点 $\iff H$ の固有値が共に負 $\iff \lambda_1 \lambda_2 > 0, \lambda_1 + \lambda_2 < 0 \iff \det H > 0, \text{tr } H < 0$.
- 極値点でない (鞍点) $\iff H$ の固有値の符号が異なる $\iff \lambda_1 \lambda_2 < 0 \iff \det H < 0$.
- 極値点か否か判定できない $\iff H$ の固有値に 0 を含む $\iff \lambda_1 \lambda_2 = 0 \iff \det H = 0$.

¹² H は実対称行列であるから, Hermite 共役は $H^\dagger \equiv (H^*)^T = H$ を満たす. 固有値を λ , 固有ベクトルを \mathbf{v} ($\neq \mathbf{0}$) として $H\mathbf{v} = \lambda\mathbf{v}$ の Hermite 共役をとると, $\mathbf{v}^\dagger H = \lambda^* \mathbf{v}^\dagger$. 両辺の差をとり \mathbf{v} を右からかけると, $0 = \mathbf{v}^\dagger H\mathbf{v} - \lambda^* \mathbf{v}^\dagger \mathbf{v} = \mathbf{v}^\dagger (\lambda\mathbf{v}) - \lambda^* \mathbf{v}^\dagger \mathbf{v} = (\lambda - \lambda^*) \mathbf{v}^\dagger \mathbf{v}$. ここで, $\mathbf{v} = [v_1, v_2]^T \neq \mathbf{0}$ より $\mathbf{v}^\dagger \mathbf{v} = |v_1|^2 + |v_2|^2 \neq 0$ であるから, $\lambda - \lambda^* = 0$. したがって $\lambda = \lambda^*$ より λ は実数.

¹³ $O O^T = I$ (単位行列) を用いて, $[\Delta x \ \Delta y] H \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = [\Delta x \ \Delta y] O O^T H O O^T \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = [\Delta x \ \Delta y] \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \Delta X \\ \Delta Y \end{bmatrix} = \lambda_1(\Delta X)^2 + \lambda_2(\Delta Y)^2$.

¹⁴ 2 次方程式 $ax^2 + bx + c = 0$ の 2 つの解を α と β とすると, $\alpha + \beta = -b/a, \alpha\beta = c/a$ が成り立つ.

2変数関数の極値点の求め方

関数 $f(x, y)$ に対して

$$f_x(a, b) = f_y(a, b) = 0$$

を満たす停留点 (a, b) を求める。停留点において、Hesse 行列

$$H(a, b) \equiv \begin{bmatrix} f_{xx}(a, b) & f_{xy}(a, b) \\ f_{xy}(a, b) & f_{yy}(a, b) \end{bmatrix}$$

を定義する。Hesse 行列の行列式 (Hessian) とトレースの正負により、以下のように分類できる。

$\det H > 0$	$\text{tr } H > 0$	極小点
	$\text{tr } H < 0$	極大点
$\det H < 0$		極値点でない (鞍点)
$\det H = 0$		この方法では判定できない

ちなみに、 $\det H > 0$ のときには \bar{f}_{xx} と \bar{f}_{yy} が同符号で $\text{tr } H$ の符号と一致するから、¹⁵ \bar{f}_{xx} の方だけを見て、

- $\det H = \bar{f}_{xx}\bar{f}_{yy} - (\bar{f}_{xy})^2 > 0$ かつ $\bar{f}_{xx} > 0 \iff (a, b)$ は極小点
- $\det H = \bar{f}_{xx}\bar{f}_{yy} - (\bar{f}_{xy})^2 > 0$ かつ $\bar{f}_{xx} < 0 \iff (a, b)$ は極大点

と書いてある教科書がほとんどである。もちろん \bar{f}_{xx} の代わりに \bar{f}_{yy} の符号を見てもよい。

3変数以上の関数には Hesse 行列の行列式とトレース (あるいは \bar{f}_{xx}) を用いた判定方法は適用できないが、Hesse 行列の固有値を調べることで極値の判定ができる点は同じである。

付録：Hesse 行列を用いない極値の判定方法

$\Delta f(a, b)$ の引数を省略して Δf と書き、微小項は無視する。

$$2\Delta f = \bar{f}_{xx}\Delta x^2 + 2\bar{f}_{xy}\Delta x\Delta y + \bar{f}_{yy}\Delta y^2.$$

(i) $\bar{f}_{xx} = \bar{f}_{yy} = \bar{f}_{xy} = 0$ のとき： $\Delta f = 0$ なので、 (a, b) が極値点か否か判定できない。

(ii) $\bar{f}_{xx} = \bar{f}_{yy} = 0$ かつ $\bar{f}_{xy} \neq 0$ のとき： $\Delta f = \bar{f}_{xy}\Delta x\Delta y$ となるが、 $(\Delta x, \Delta y) = (\epsilon, \epsilon)$ と $(\Delta x, \Delta y) = (\epsilon, -\epsilon)$ とで Δf の符号が異なるので、 (a, b) は極値点ではない。

(iii) $\bar{f}_{xx} \neq 0$ または $\bar{f}_{yy} \neq 0$ のとき： 議論は x と y について対称であるから、 $\bar{f}_{xx} \neq 0$ として一般性を失わない。このとき、 $2\Delta f$ を平方完成すると、

$$2\Delta f = \bar{f}_{xx}\Delta x^2 + 2\bar{f}_{xy}\Delta x\Delta y + \bar{f}_{yy}\Delta y^2 = \bar{f}_{xx} \left(\left(\Delta x + \frac{\bar{f}_{xy}}{\bar{f}_{xx}}\Delta y \right)^2 + \frac{1}{(\bar{f}_{xx})^2}(\bar{f}_{xx}\bar{f}_{yy} - (\bar{f}_{xy})^2)\Delta y^2 \right)$$

となる。したがって、

- $\bar{f}_{xx}\bar{f}_{yy} - (\bar{f}_{xy})^2 > 0$ のとき

$\bar{f}_{xx} > 0$ なら $\Delta f > 0$ であるから極小点、 $\bar{f}_{xx} < 0$ なら $\Delta f < 0$ であるから極大点となる。

- $\bar{f}_{xx}\bar{f}_{yy} - (\bar{f}_{xy})^2 < 0$ のとき

$(\Delta x, \Delta y) = (\epsilon, 0)$ と $(\Delta x, \Delta y) = \epsilon(\bar{f}_{xy}, -\bar{f}_{xx})$ の方向で Δf の符号が異なるので、極値点ではない。

- $\bar{f}_{xx}\bar{f}_{yy} - (\bar{f}_{xy})^2 = 0$ のとき

$(\Delta x, \Delta y) = \epsilon(\bar{f}_{xy}, -\bar{f}_{xx})$ の方向で $\Delta f = 0$ となるから、極値点か否か判定できない。

以上をまとめると、Hesse 行列を用いた場合と同じ結論が得られる。2つの判定方法を見比べるとわかるように、Hesse 行列を用いた議論の方が体系的でわかりやすく、3変数以上の関数への拡張も容易である。

¹⁵ $(\bar{f}_{xy})^2 \geq 0$ であるから、 $\det H = \bar{f}_{xx}\bar{f}_{yy} - (\bar{f}_{xy})^2 > 0$ のためには $\bar{f}_{xx}\bar{f}_{yy} > 0$ が必要であり、 \bar{f}_{xx} と \bar{f}_{yy} は同符号である。