

Hesse 行列と Jacobi 行列

Ziphil Aleshlas

2014 年 8 月 5 日

1. Hesse 行列

1.1. 2 変数関数の最大最小問題

いきなり多変数関数を扱うのは難しいので、まずは 2 変数関数を考えることにする。

C^2 級、すなわち 2 階連続的微分可能である 2 変数関数 $f(x, y)$ について考える。この最大点および最小点を求めるには、まずその候補となる極大点と極小点を求める必要がある。そして、点 (x_0, y_0) が極大点もしくは極小点であるための必要条件は、

$$\frac{\partial f}{\partial x}(x_0, y_0) = \frac{\partial f}{\partial y}(x_0, y_0) = 0 \quad [1.1]$$

で与えられる。この式 1.1 を満たす点 (x_0, y_0) は、関数 $f(x, y)$ の停留点 (stationary point) もしくは臨界点 (critical point) と呼ばれる。

さて、ここで問題となるのは、式 1.1 が極大点や極小点を与える必要条件にすぎないということである。すなわち、停留点が必ずしも極大点や極小点になるわけではないのである。ここで、関数 $f(x, y)$ についてより深い考察が必要になる。

さて、関数の近似的な振る舞いを調べるのには、Taylor 展開が有効である。そこで、関数 $f(x, y)$ を点 (x_0, y_0) のまわりで Taylor 展開してみると、

$$\begin{aligned} f(x_0 + h, y_0 + k) &= f(x_0, y_0) \\ &+ h \frac{\partial f}{\partial x}(x_0, y_0) + k \frac{\partial f}{\partial y}(x_0, y_0) \\ &+ \frac{1}{2} \left(h^2 \frac{\partial^2 f}{\partial x^2}(x_0, y_0) + 2hk \frac{\partial^2 f}{\partial x \partial y}(x_0, y_0) + k^2 \frac{\partial^2 f}{\partial y^2}(x_0, y_0) \right) \\ &+ o(\rho^2) \end{aligned} \quad [1.2]$$

となる *1. ここで $\rho = \sqrt{h^2 + k^2}$ とした。

*1 最終項 $o(\rho^2)$ は Landau の記号である。ここでは、 $h \rightarrow 0, k \rightarrow 0$ のとき、 $o(\rho^2)$ は ρ^2 が 0 に近づくより急速に 0 に近づく変数であることを表している。 ρ^2 つまり $h^2 + k^2$ と同じ速さで 0 に近づく項は、式 1.2 の第 3 項にすでに存在しているので、これより速く 0 に近づいてしまう $o(\rho^2)$ の項は、 $h \rightarrow 0, k \rightarrow 0$ の極限では無視しても良いということになる。

以下では式 1.1 は成り立っているものとする。このとき、式 1.2 は、

$$f(x_0 + h, y_0 + k) - f(x_0, y_0) = \frac{1}{2} \left(h^2 \frac{\partial^2 f}{\partial x^2}(x_0, y_0) + 2hk \frac{\partial f}{\partial x \partial y}(x_0, y_0) + k^2 \frac{\partial^2 f}{\partial y^2}(x_0, y_0) \right) + o(\rho^2)$$

と書き直せる。点 (x_0, y_0) の付近で、すなわち h, k が十分に小さいとき、この式の右辺が常に正ならば、 $f(x_0 + h, y_0 + k) > f(x_0, y_0)$ が成り立つので、点 (x_0, y_0) の周りの点ではどこでも関数の値が大きくなる。したがって、点 (x_0, y_0) は極小点だと分かる。逆に、右辺が常に正ならば、点 (x_0, y_0) は極大点であるということになる。

h, k が十分に小さいとき、 $o(\rho^2)$ は右辺の符号には影響しないほど小さくなれるので、右辺の符号の鍵を握るのは、

$$X = h^2 \frac{\partial^2 f}{\partial x^2}(x_0, y_0) + 2hk \frac{\partial f}{\partial x \partial y}(x_0, y_0) + k^2 \frac{\partial^2 f}{\partial y^2}(x_0, y_0) \quad [1.3]$$

の部分である。ここで、式の簡略化のために

$$A = \frac{\partial^2 f}{\partial x^2}(x_0, y_0); \quad B = \frac{\partial f}{\partial x \partial y}; \quad C = \frac{\partial^2 f}{\partial y^2}$$

とおけば、式 1.3 は、

$$X = Ah^2 + 2Bhk + Ck^2$$

と書き表すことができる。

さて、 X の符号を考察する。 $\lambda = \frac{h}{k}$ とおいてしまえば、

$$k^2 X = (A\lambda^2 + 2B\lambda + C)$$

と書ける。左辺は λ の 2 次式なので、符号の判別は判別式を用いれば簡単である。すなわち、 $A > 0$ であれば $AC - B^2 > 0$ のとき常に正であり、 $A < 0$ であれば $AC - B^2 > 0$ のとき常に負になる。

このことから、 $A > 0$ かつ $AC - B^2 > 0$ ならば、 X は常に正で、したがって点 (x_0, y_0) は極小点になる。一方、 $A < 0$ かつ $AC - B^2 > 0$ ならば、 X は常に負なので、点 (x_0, y_0) が極大点であることが分かる。

$AC - B^2$ という形が出てきたが、これは行列式を用いてわざわざ分かりにくい書き方をすれば、

$$\begin{vmatrix} A & B \\ B & C \end{vmatrix}$$

とも表せる。 A, B, C をもとに戻せば、

$$\begin{vmatrix} \frac{\partial^2 f}{\partial x^2}(x_0, y_0) & \frac{\partial f}{\partial x \partial y}(x_0, y_0) \\ \frac{\partial f}{\partial y \partial x}(x_0, y_0) & \frac{\partial^2 f}{\partial y^2}(x_0, y_0) \end{vmatrix}$$

という、対称的な形が出てくる^{*2}。この行列を Hesse 行列 (Hessian matrix) と呼び、 $\mathcal{H}f(x_0, y_0)$ と書く。変数の値を明示せず、単に $\mathcal{H}f$ と書くこともある。また、Hesse 行列の行列式を Hesse 行列式 (Hessian determinant) もしくはヘッシアン (hessian) と呼ぶ。

この Hesse 行列を用いると、これまで議論してきた極大極小条件を簡潔に記述することができる。すなわち、 $f_{xx} > 0$ かつ $\det \mathcal{H}f > 0$ を満たす停留点は極小点であり、 $f_{xx} < 0$ かつ $\det \mathcal{H}f > 0$ を満たす停留点は極大点である。

1.2. 一般の Hesse 行列

Hesse 行列式を用いた極大極小の考察は、そのまま多変数関数へ一般化できる。

C^2 級の n 変数関数 $f(x_1, x_2, \dots, x_n)$ の Hesse 行列は、

$$\mathcal{H}f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

で与えられる。1 階導関数が全て 0 であるという条件のもとで、 $\det \mathcal{H}f > 0$ ならば、その点は極値を与えるといえる。

極大点なのか極小点なのかを判別するには、Hesse 行列の左上部分を取り出した行列

$$\mathcal{H}_k f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_k} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_k \partial x_1} & \frac{\partial^2 f}{\partial x_k \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_k^2} \end{pmatrix} \quad (1 \leq k \leq n)$$

について考察する必要がある。 $k = 1, 2, \dots, n-1$ の全てに対し $\det \mathcal{H}_k f$ の符号が正ならば、その点は極小だといえる。一方、 $\det \mathcal{H}_k f$ の符号が $(-1)^k$ の符号と一致していれば、その点は極大だといえる^{*3}。

^{*2} さりげなく B の微分の順序を変えてあるが、関数 $f(x, y)$ は C^2 級であるとして話を進めているので、微分の順序は自由に変えて良く、問題はない。

^{*3} Hesse 行列の左上部分を抜き取った行列が突然出てきて困惑したかもしれないが、これは 2 変数関数の議論のときに f_{xx} の符号を見たところに対応している。実際、 $\mathcal{H}_1 f = f_{x_1 x_1}$ である。

2. Jacobi 行列

2.1. 新しい写像

これまでは、1つの変数から1つの変数の値を定める関数や、複数の変数から1つの変数の値を定める関数を扱ってきたが、複数の変数から複数の変数を定める関数(写像)を考えることもできる。このような、複数の変数を定めるような写像について、微分という概念を考察したい。

2.2. 多変数から多変数への写像の表記

簡単のため、まずは2変数から2変数への写像を考える。

2つの変数 (x, y) があり、これを2つの変数 (s, t) に対応させる写像を考える。この写像を記述するには、

$$s = f(x, y); \quad t = g(x, y)$$

という、2つの2変数関数が必要になる。

さて、これをより簡潔に記したい。そこで、行列のように縦に並べ、さらに関数の引数も縦に並べて、少し気持ち悪いが、

$$\begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} f \begin{pmatrix} x \\ y \end{pmatrix} \\ g \begin{pmatrix} x \\ y \end{pmatrix} \end{pmatrix}$$

と書いてみる。また、 ${}^t(s, t)$ と ${}^t(x, y)$ を縦ベクトル扱いして、それぞれ \mathbf{s} , \mathbf{x} と書くことにすれば、

$$\mathbf{s} = \begin{pmatrix} f(\mathbf{x}) \\ g(\mathbf{x}) \end{pmatrix}$$

と表せる^{*4}。そしてさらに、この式の右辺は関数のベクトルだと考えることもできる。したがって、これを $\mathbf{f}(\mathbf{x})$ と書くことにしてしまえば、

$$\mathbf{s} = \mathbf{f}(\mathbf{x})$$

と、あたかも1変数関数であるかのように表記できてしまう^{*5}。

これと同じようにして、3変数以上の場合も、ベクトルという概念を介して表記を簡明にすることができる。

^{*4} 左上の t は転置を表す。行中に縦ベクトルを書く于行間が無駄になるので、しばしば横ベクトルの転置として縦ベクトルを表記することがある。

^{*5} 太字と細字を混同しないように注意すること。 x は1つの変数だが、 \mathbf{x} はベクトルである。また、関数名である f も太字である。これは、複数の値(ベクトル)を返す関数であることを明示している。なお、文献によっては太字にせずに表記することもある。

2.3. 写像の微分

さて、普通の2変数関数の微分可能性(全微分可能性)について、少し復習する。

2変数関数 $f(x, y)$ が点 (x_0, y_0) で微分可能であるというのは、十分小さな数 h, k に対し、 x_0, y_0 によつてのみ決まる数 $f_x(x_0, y_0), f_y(x_0, y_0)$ が存在して、

$$f(x_0 + h, y_0 + k) = \frac{\partial f}{\partial x}(x_0, y_0)h + \frac{\partial f}{\partial y}(x_0, y_0)k + o(\rho) \quad [2.1]$$

と表せることをいうのであった。なお、 $\rho = \sqrt{h^2 + k^2}$ である。

ところで、前項では、多変数をベクトルだとみなして表記を簡略化した。ここでも同じようにして表記を変えてみる。

まず、上式 2.1 は、

$$f \begin{pmatrix} x_0 + h \\ y_0 + k \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial x}(x_0) & \frac{\partial f}{\partial y}(x_0) \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix} + o \begin{vmatrix} h \\ k \end{vmatrix}$$

と書ける。なお、右辺第1項は 1×2 行列と 2×1 行列の積で、右辺第2項の絶対値記号はベクトルの大きさを表すものとする^{*6}。ここでさらに、前項のように ${}^t(x_0, y_0)$ を \mathbf{x}_0 とし、 ${}^t(h, k)$ を \mathbf{h} と書くことにすれば、

$$f(\mathbf{x}_0 + \mathbf{h}) = \begin{pmatrix} \frac{\partial f}{\partial x}(\mathbf{x}_0) & \frac{\partial f}{\partial y}(\mathbf{x}_0) \end{pmatrix} \mathbf{h} + o|\mathbf{h}| \quad [2.2]$$

と書き表すことができる。

さて、こう書いてしまえば、多変数から多変数への写像の微分可能性も容易に定義できる。例えば、前項に示した、2変数から2変数への写像 $\mathbf{f}(\mathbf{x})$ を考えれば、式 2.2 を参考にして、

$$\mathbf{f}(\mathbf{x}_0 + \mathbf{h}) = \begin{pmatrix} \frac{\partial \mathbf{f}}{\partial x}(\mathbf{x}_0) & \frac{\partial \mathbf{f}}{\partial y}(\mathbf{x}_0) \\ \frac{\partial \mathbf{g}}{\partial x}(\mathbf{x}_0) & \frac{\partial \mathbf{g}}{\partial y}(\mathbf{x}_0) \end{pmatrix} \mathbf{h} + o|\mathbf{h}|$$

と表せることを、微分可能性の定義としても良さそうである。このとき、右辺第1項に現れる行列を、Jacobi 行列 (—bian matrix) と呼び、 $\mathcal{J}\mathbf{f}(\mathbf{x}_0)$ と書く^{*7}。単に $\mathcal{J}\mathbf{f}$ とだけ書く場合もある。

^{*6} すなわち ρ である。この表記は行列式と紛らわしいが、行列式は正方行列にしか定義されないことから見分けることは可能であるので、この表記を用いることにする。

^{*7} Jacobi 行列を表す記号は多くある。ここでは \mathcal{J} を用いるが、 \mathcal{D} を用いて表現したり、微分記号を流用することもある。

2.4. 一般の Jacobi 行列

2 変数に限らず多変数の場合の写像でも, Jacobi 行列式は同様に定義される. m 変数から n 変数への写像 $f(\mathbf{x})$ が

$$f(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_n(\mathbf{x}) \end{pmatrix}; \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}$$

によって表されるとき, 写像 $f(\mathbf{x})$ の Jacobi 行列は,

$$\mathcal{J}f = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_m} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_m} \end{pmatrix}$$

で定義される *8.

さて, この Jacobi 行列を用いることで, 多変数から多変数への写像に関しても微分可能性をきれいに定義できるわけである. すなわち, 写像 $f(\mathbf{x})$ を考えるとき, 微小量 \mathbf{h} に対して \mathbf{x}_0 にのみ関係する行列 $\mathcal{J}f(\mathbf{x}_0)$ が存在し,

$$f(\mathbf{x}_0 + \mathbf{h}) = \mathcal{J}f(\mathbf{x}_0)\mathbf{h} + o(|\mathbf{h}|)$$

が成り立つとき, 写像 $f(\mathbf{x})$ は点 \mathbf{x}_0 で微分可能 (differentiable) というのである.

多変数のベクトル表記と Jacobi 行列を用いることで, 多変数から多変数への写像の微分や積分も, 通常の関数と同じように表記し扱うことができるようになる.

2.5. Jacobi 行列式

Jacobi 行列の行列式を, Jacobi 行列式 (—bian determinant) やヤコビアン (jacobian) という. ここでは詳しく述べないが, Jacobi 行列式は重積分の変数変換の際に用いられる.