関数の停留，極大極小をめぐって

2015 年 5 月 5 日

§ 停留点からはじめる

数学的な概念やタームを説明する場合，まず始めににその対象の定性的性質を描き，その性質を満たすものを「『何某』であると定義する」という形がとられることが多いと思う．対象の定性的性質がわかりやすい場合には有効な形式であるけれども，いろいろ込み入ってくるとなかなかそうもいかない．そして往往にして，定性的性質を描く際には，その「『何某』の定義」が暗黙に前提され背景化されている場合が多く，初めて学ぶうぶな時には定性的性質を把握しづらいというのも事実である [1]．ある場合には，その『何某』であるという定義を先に認めてしまい，そこから定性的性質をイメージする努力をするということも効果的な方法なのだ，と考えたりもする．なにより「はっきり」と論が進められるような気になれるのがよい．

関数の停留や極大極小は，筆者にとって定性的性質の把握，記述がなかなかうまくいかない事例であった．もちろん１変数関数であれば，グラフという強力な武器があるのでなんとかはなる．２変数関数でも，まあ，それなりに．しかしながら n 変数関数ともなると，心もとないことこの上はない．それゆえに，ここでは，まず停留の定義を先に了解し，そこからはじめていってみることにする．

§§ 停留の定義

関数 f が「停留する」という概念を，

「f が停留する」とは f の全微分が 0 すなわち $df=0$ となることである．

f の全微分を 0 にする点を「停留点」，その時の関数値を「停留値」とよぶ．

と定義する [2]．

ここで無造作に「点」という言葉をつかっているが，すこし整理しておこう．１変数関数であれば，ことばの「点」と実数軸上のひとつの座標点（１次元の座標点）との対応がすんなり受け入れられる．そしてそれは，２変数関数では２次元座標平面上の点，３変数関数で３次元空間内の点として同様に了解できる．そのように教育を受けてきている。では n 変数関数（多変数関数）ではどうか？日常にはあまり適当なイメージがない．したがって，ここでもそれの定義をしてしまおう．

まず n 変数関数を「独立変数が n 個である関数」という意味であると決める．そしていま，その n 個の独立変数の組を，$(x_1, x_2, \ldots, x_n)$ と表すことにし，この n 個の数の組も「点」と呼ぶことにする．さらにこの点の入れ物を n 次元空間と呼ぶことにする [3]．これで一般の「点」が定義できた.「点」の一般化とも言えようか.

[1] イメージ豊かに生き生きとその定性的性質を思い描ける，というのが「才能」であるのだろう．そこで我々のような凡人は，定性的性質の不完全なイメージと実際の定義をあわせもって微修正しつつ，自らのイメージを作成するという方法をとる．それが学習というものなんだろうと思ったりする．

[2] これを，先に「停留とは云々」と停留の性質を描写し，$df=0$ ならばその性質が満たされる，とやるのもよいのだけれども（正統的？)，停留の性質の描写はものすごくしんどくて，私には上手くできなかった．それゆえ，停留の定義から始める本稿の方法とあいなったのである．

[3] 個数，と，お互い独立である，という考え方を合併したものを「次元」としたにすぎない．

次に，n 変数関数の書き方であるが,「n 変数関数」といえども「関数」であるから，点から数（数値）への写像であって，

$$(x_1, x_2, \ldots, x_n) \quad \mapsto \quad （数値）$$

という関係にある．これを「関数」的に

$$f : (x_1, x_2, \ldots, x_n) \quad \mapsto \quad f(x_1, x_2, \ldots, x_n)$$

と表記することにする．

停留に戻ろう．停留とは，$df = 0$ となるものと定義した．定義はそれで完了している．次に，この定義から，停留についての解析的な条件を導き出そう．全微分の実行からある条件が出てくるのである．停留 $df = 0$ を解析的に表記すれば，

$$df = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 + \cdots + \frac{\partial f}{\partial x_n} dx_n = 0$$

となる．そして，$x_i\ (i = 1, \ldots, n)$ はそれぞれ独立であるから，dx_i もすべて独立であり [4]，結果，$df = 0$ となる必要十分条件として，

$$df = 0 \quad \Longleftrightarrow \quad \frac{\partial f}{\partial x_1} = 0,\ \frac{\partial f}{\partial x_2} = 0,\ \ldots\ ,\ \frac{\partial f}{\partial x_n} = 0$$

が得られるのである．

この式を解いて（解けたとして）得られる点 $(p_1, p_2, \ldots, p_n)$ を関数 f の停留点と言い，そのときの関数値 $f(p_1, p_2, \ldots, p_n)$ を関数 f の停留値と呼ぶ.

これで停留，停留点，停留値の定義は完了．そしてその定義から導かれる停留の（解析的）必要十分条件も導出できた．

§ 極大，極小

停留を定義し，それを与える必要十分条件を導き，それを解くことにより停留点と停留値をもとめる道筋ができた．ここでこの停留点というものの分類に興味をうつしてみる．分類の観点は，停留値がその関数においてもつ属性のひとつで，極大値か極小値か，というものである．

ここで今度は，極大値，極小値の意味を与えねばならない．極大値とは，関数の「局所的な最大値」をあらわすものとして取り扱われ，極小値も「局所的な最小値」として使われる．となると，こんどは「局所的」とは何か，と芋づる式発散が立ち現れる．位相空間論をかじれば,「点の近傍で最大（最小）」であり，今度は「近傍」とは，と再度芋づる発散である．

まともに位相空間論に深入りせずに,「近傍」を適当に（ある意味いい加減に）定義してみる．まず n 次空間を S_n と表記し，その空間内の点（一般化された「点」）を p_n と表すとする．この時，p_n の近傍 $V_n(p_n)$ をつぎのように定義する：

$$V_n(p_n) \text{ は開集合},$$
$$p_n \in V_n(p_n),$$
$$V_n(p_n) \subset S_n.$$

[4] これは筆者的には，直観的にあきらか，としたいのであるが，真面目に厳密さをもとめると，そうもいかないものなのだろうか？

大雑把に言えば，p_n を含む開集合で n 次元空間の一部（部分集合）が近傍である，イメージとしては，縁境界のない傘を思い浮かべてもいいかもしれない．p_n で「傘」をさし，濡れない範囲が近傍である[5].「傘」の大きさは，定義の条件には関係がないので，近傍は一つに定まるわけでない．したがって，上記定義に加えて $V_n(p_n) \subseteq U$ なる U を近傍という場合もあるようだ[6].

とにかく，以上から再度極大極小を考えれば，

$$\text{点 } p_n \text{ で極大} \iff p_n \text{ が停留点であり，かつ，} f(p_n) \text{ が最大値となる近傍が存在する．}$$

$$\text{点 } p_n \text{ で極小} \iff p_n \text{ が停留点であり，かつ，} f(p_n) \text{ が最小値となる近傍が存在する．}$$

となる．

関数が極大となる点を極大点，極小となる点を極小点，そのときのそれぞれの関数値を極大値，極小値と呼ぶ．そして極大値極小値をあわせて極値と呼んだりもする．

さらにいえば，停留点であっても上に書いた極大（極小）となる近傍が存在しない場合もあるだろう．したがって，停留点は，極大値を与える点（極大点）と極小値を与える点（極小点），そのどちらでもないただの停留点の３個に分類できるのである．

§ この先は・・・・

停留からはじめて，その必要十分条件，極大極小の意味，停留点の分類が曲がりなりにも n 変数関数にたいして定義／導出できた．今一度まとめてみると，n 変数関数　$f(x_1, x_2, \ldots, x_n)$ の停留点とは f の全微分において，

$$df = \frac{\partial f}{\partial x_1}dx_1 + \frac{\partial f}{\partial x_2}dx_2 + \cdots + \frac{\partial f}{\partial x_n}dx_n = 0$$

を満たすものであり，dx_i の独立性から，

$$df = 0 \quad \iff \quad \frac{\partial f}{\partial x_1} = 0,\ \frac{\partial f}{\partial x_2} = 0,\ \ldots\ ,\ \frac{\partial f}{\partial x_n} = 0$$

が満たされるものである．

極大値，極小値は，それぞれ局所的に最大，最小を意味するものであり．局所的に最大（最小）とは，その関数値が最大（最小）となる近傍が存在することである．近傍が存在しなければ，局所的に最大（最小）とは言えない．そして，極大点，極小点はまず停留点であるが，停留点であっても極大点でも極小点でもないものも存在する（最大最小をあたえる近傍が存在しない）：

$$\text{（極大点，極小点）} \quad \Longrightarrow \quad \text{（停留点）.}$$

さてこの先は，具体例として実感することがまず必要だろう．本稿の Appendix として簡単な関数系での具体例を付け加えた．１変数関数や２変数関数の極値の判定に Taylor 展開をもちいる方法は秀逸である（変数が多くなると Taylor 展開といえどもそう簡単にはいかないだろう．というか，Taylor 展開自体の複雑さが極端に増してしまうので手に負えない）．

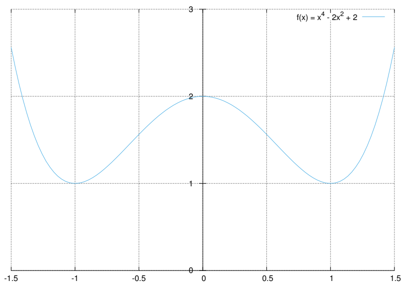
もうひとつは，独立変数間になんらかの関係（制約条件）がある場合の極値をもとめる問題である．ラグランジュの未定乗数法という解法がある．これは別紙になる予定である．

[5] ２次元座標平面では，p_2 を中心とする境界を含まない円内の領域である．
[6] この U は，いろいろな近傍をあつめたもの，と諒解しておいてよいと思う．

A　１変数関数での具体例

停留点と極大極小の関係について，具体例で見て行ってみよう．

A.1　$f(x) = x^4 - 2x^2 + 2$ の場合



$f(x) = x^4 - 2x^2 + 2$ の全微分を計算し，停留点を求めると，

$$
\begin{aligned}
&df = 0 \quad \Longleftrightarrow \quad 4x(x-1)(x+1) = 0 \\
&i.e. \quad x = 0,\ -1,\ 1
\end{aligned}
$$

と３個の停留点が求まる．ここで停留点の周りを調べてみる．

A.1.1　停留点 $x = 0$ の場合

$f(0) = 2$ であることはあきらか．ここで $\varepsilon > 0$ という量をもちいて，$f(0+\varepsilon)$ を考えると，

$$f(0+\varepsilon) = (\varepsilon^2 - 1)^2 + 1$$

であるから，$(\varepsilon^2 - 1)^2 < 1$ の時は $f(0+\varepsilon) < f(0)$ となる．同様に，$\delta > 0$ という量をもちいて，$f(0-\delta)$ を考えると，やはり

$$f(0-\delta) = (\delta^2 - 1)^2 + 1$$

であって，$(\delta^2 - 1)^2 < 1$ の時は $f(0-\delta) < f(0)$ となる．さて，このような量 ε, δ は存在するだろうか？ε の条件式を素直に解いてみると，

$$(\varepsilon^2 - 1)^2 < 1 \quad \Longleftrightarrow \quad \varepsilon^4 - 2\varepsilon^2 < 0 \quad \Longleftrightarrow \quad \varepsilon^2(\varepsilon - \sqrt{2})(\varepsilon + \sqrt{2}) < 0$$

であり，$\varepsilon > 0$ と仮定したのだから，$0 < \varepsilon < \sqrt{2}$ となる ε を選べばよく，確かにこれは存在する．δ についても同様である [7]．したがって，

$$0 < \varepsilon < \sqrt{2},\ 0 < \delta < \sqrt{2} \text{ なる } \varepsilon, \delta \text{ については常に } f(0+\varepsilon) < f(0),\ f(0-\delta) < f(0)$$

ということが成り立つのである．つまり局所的に最大（極大）で，$x = 0$ は極大値 2 を与える極大点である．こういう ε, δ が存在すること，そして $-\delta < x < \varepsilon$ なる近傍で最大値となっていることが「局所的最大」のこころである．

ゆえに，$x = 0$ は極大点であり，極大値は $f(0) = 2$ である．

---

[7] $f(x)$ を偶関数として作成したので，たまたま同じ範囲になっただけである．異なっていても全く問題はない．要はこういう ε, δ が存在すればいいのである．

A.1.2　停留点 $x=1$ の場合

$f(1)=1$ であることはあきらか．ここで $\varepsilon>0$ という量をもちいて，$f(1+\varepsilon)$ を考えると，

$$f(1+\varepsilon)=(\varepsilon(\varepsilon+2))^2+1$$

であるから，どんな $\varepsilon>0$ に対しても $f(1)<f(1+\varepsilon)$ である．同様に，$\delta>0$ という量をもちいて，$f(1-\delta)$ を考えると，やはり

$$f(1-\delta)=(\delta(\delta-2))^2+1$$

であって，どんな $\delta>0$ に対しても $f(1)\leq f(1+\delta)$ であり，等号は $\delta=2$ の時（そしてそれは，停留点 $x=-1$ の場合に相当）．したがって，$0<\delta<2$ と選べば，

$$f(1)<f(1+\varepsilon),\ f(1)\leq f(1-\delta)$$

ということが成り立つのである．つまり局所的に最小（極小）で，$x=1$ は極小値 1 を与える極小点となる．そしてこの ε,δ が存在することは自明．したがって最小値となる近傍の存在が確認された.「局所的」のこころが満たされているのである．

ゆえに，$x=1$ は極小点であり，極小値は $f(1)=1$ である．同様な計算を $x=-1$ に対しても実行すれば，$x=-1$ が極小値 1 を与える極小点であることがわかる．

A.1.3　システマティックに

さて，今見てきたように，近傍を構成する ε,δ に対して，それらが最大／最小になる条件を導き出すというのは，いがいと厄介でもある．ここで，導関数の正負（接戦の傾きの正負）を利用してシステマティックにやったのが高校時代の「増減表」である．この関数の例では，１階導関数は $f'(x)=4x(x-1)(x+1)$ であったから，

x	$\cdots$	-1	$\cdots$	0	$\cdots$	1	$\cdots$
$f'(x)$	$-$	0	$+$	0	$-$	0	$+$
$f(x)$	$\searrow$		$\nearrow$		$\searrow$		$\nearrow$

という表になる．導関数が負になる領域は，接線の傾きも負になる領域で，そこでは関数は減少する．正の場合はその反対で，関数は増加する．減少して増加，または，増加して減少，となるのならば，そこには「頂点」のようなものがあって，それは導関数が 0 となるところ（接線が x 軸に平行となるところ）であり，そこが極大，極小である，ということが一目でわかる便利な表なのである．

さらに極大／極小までどんぴしゃと求めたいのなら，２階導関数を利用する方法もある．１階導関数を 0 にする点での２階導関数の値が 0 より大きい場合は極小値，0 より小さい場合は極大値となるのである．

この事実は，$f''(x),f'(x)$ の増減表を作れば確認できる．いまの例の場合，２階導関数をもとめると，

$$f''(x)=12x^2-4=12\left(x-\frac{1}{\sqrt{3}}\right)\left(x+\frac{1}{\sqrt{3}}\right),$$

$$f''(-1)=8,\ f''(0)=-4,\ f''(1)=8$$

なので，$f''(x)$，　$f'(x)$ の増減表は，

x	$\cdots$	-1	$\cdots$	$-\frac{1}{\sqrt{3}}$	$\cdots$	0	$\cdots$	$\frac{1}{\sqrt{3}}$	$\cdots$	1	$\cdots$
$f''(x)$	$+$	8	$+$	0	$-$	-4	$-$	0	$+$	8	$+$
$f'(x)$	$\nearrow$	0	$\nearrow$		$\searrow$	0	$\searrow$		$\nearrow$	0	$\nearrow$

となる．$x=-1$ の周りに注目すると，$f'(x)$ は，増加，0，増加，であるから，マイナスからプラスになるはずである．$x=0$ の周りに注目すると，$f'(x)$ は，減少，0，減少，であるから，プラスからマイナスになる．$f'(x)$ の値の符合がその点で変わるのであるから，その点がそれぞれ極小点，極大点となるのである．スローガン的に繰り返せば,

> ２階導関数を利用する方法もある．１階導関数を 0 にする点での２階導関数の値が 0 より大きい場合は，その点は極小値をあたえる極小点であり，0 より小さい場合は極大値をあたえる極大点となる．

となろう．しかしながら，この方法での増減表は少々くどい気もする．最初の増減表でも十分であろう．さらにもうひとつ．その１階導関数を 0 にする点での２階導関数の値が 0 の場合はどうなるのか？それについては，A.2 の $f(x)=x^3$ のところで述べよう．

A.1.4　Taylor 展開の利用

$f(x)$ の a の周りの Taylor 展開は,

$$f(x)=f(a)+f'(a)(x-a)+\frac{f^{(2)}(a)}{2!}(x-a)^2+O((x-a)^3)$$

であった．これをいま考えている $f(x)=4x^4-2x^2+2$ に使うと，まず，0 の周りのテイラー展開は,

$$f(x)=f(0)+f'(0)x+\frac{f^{(2)}(0)}{2!}x^2+O(x^3)=2-2x^2+O(x^3)$$

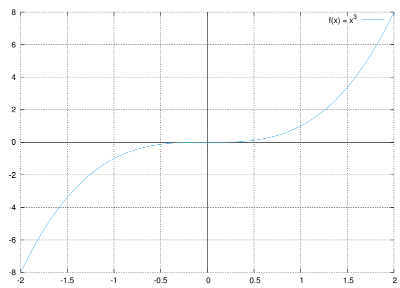
で，$O(x^3)$ を無視すれば，上に凸の放物線つまり，極大となる．$x=1$ の周りでは,

$$f(x)=f(1)+f'(1)(x-1)+\frac{f^{(2)}(1)}{2!}(x-1)^2+O((x-1)^3)=4+4(x-1)^2+O((x-1)^3)$$

で，$O((x-1)^3)$ を無視すれば，下に凸の放物線つまり，極小となる．あたりまえであるが，いままでの結果と整合が取れている．Taylor 展開はこのようにも使える非常に優れたものなのである．

A.2　$f(x)=x^3$ の場合

極大でも極小でもない停留点の例を見てみる．$f(x)=x^3$ の停留点は $x=0$．この時，$f(0)=0$ であり，先と同様に $\varepsilon>0,\delta>0$ という量をもちいれば,



$$f(0+\varepsilon)=\varepsilon^3$$
$$f(0-\delta)=-\delta^3$$

が導かれる．

極大の条件は，$f(0+\varepsilon)<f(0)$，$f(0-\delta)<f(0)$ であった．しかしながらこの場合，$\varepsilon>0$ としたのだから，$\varepsilon^3<0$ を成り立たせる ε は存在しない（δ は存在するけれども）．よってこの停留点は極大ではない．

極小の条件は $f(0+\varepsilon)>f(0)$，$f(0-\delta)>f(0)$ であるが，やはり $\delta>0$ で $-\delta^3>0$ を満たすものは存在しない（こちらの場合は ε は存在する）．よってこの停留点は極小ではない．つまり，最大（最小）を与える近傍が存在しない．ゆえに，この停留点は極大でも極小でもないのである．

これを増減表を使ってみてみよう．２階導関数までいっぺんに書く．$f'(x) = 3x^2$，$f''(x) = 6x$ であるから，

x	$\cdots$	0	$\cdots$
$f''(x)$	$-$	0	$+$
$f'(x)$	$+$	0	$+$
$f(x)$	$\nearrow$		$\nearrow$

$x = 0$ の周りに着目すると，$f'(x)$ は，減少，0，増加であるから，常に 0 以上である．そして，$f(x)$ は増加，増加である．つまり，極大でも極小でもない．そしてそういう場合には，$f''(0)$ は 0 なのである．

$x = 0$ の周りで Taylor 展開をすると，

$$f(x) = f(0) + f'(0)x + \frac{f^{(2)}(0)}{2!}x^2 + O(x^3) = 0$$

であり，これ以上何も言えない．つまり，ただの停留点であるとしか言いようがないのである．