

ラグランジュの未定乗数法

- 1. 停留値問題
- 2. 1個の拘束条件を伴う停留値問題
 - 2.1. 従属変数を削除する方法
 - 2.2. 未定乗数を用いる方法
- 3. 未定定数の幾何的解釈
- 4. 複数の拘束条件を持つ停留値問題
- 5. 例題
- 6. Reference

1. 停留値問題

ある多変数関数 $f(x_1, x_2, \dots, x_n)$ の**停留値**問題を考える。極小値(あるいは極大値)を考えることに似ているが、とにかく傾きがゼロになる点を求める(例えば $f(x) = x^3$ における $x = 0$ は極大でも極小でもないが停留値である)。

n 次元を考えるのは大変なので、2変数関数程度のものをイメージすると良い。

停留点においては、わずかに x_i の値を変化させたところで関数の返り値はほとんど変化しない。すなわち、関数 f の停留値からのズレを δf で表すと、

$$\delta f = \frac{\partial f}{\partial x_1} \delta x_1 + \dots + \frac{\partial f}{\partial x_i} \delta x_i + \dots + \frac{\partial f}{\partial x_n} \delta x_n = 0$$

が成立する。このように点を(わずかに)変化させたときの変位を**仮想変位**と呼ぶ。

この式だけでは、例えばいい感じに $\partial f / \partial x_i$ と $\partial f / \partial x_j$ が打ち消し合った結果、 $\delta f = 0$ となっても構わないのだが、停留値であるという条件下においては、いかなる δx についても $\delta f = 0$ が成立する(例えばボウルの底、あるいは鞍点のような点をイメージすると良い)。すなわち、

$$\frac{\partial f}{\partial x_i} = 0$$

がすべての i について成立する。これが停留値となるための条件である。

2. 1個の拘束条件を伴う停留値問題

上記では、すべての変数 x_i が独立であると仮定していた。ここでは、変数が独立ではない場合を考えよう。すなわち、

$$g(x_1, x_2, \dots, x_n) = 0$$

という**拘束条件**が存在するときの f の停留値問題を考える。ここでは、拘束条件が x_1, x_2, \dots, x_n だけで表され、これらの微分にはよらない場合を考えよう。このような条件を**ホロノミック拘束**と呼ぶ。

拘束条件が存在すると、もはや各変数は独立変数ではなく、仮想変位を考えるとときにも制約が生まれてしまう。そうすると停留値問題は単純に

$$\frac{\partial f}{\partial x_i} = 0$$

が満たされればよいというわけにはいかなくなる。

2.1. 従属変数を削除する方法

ホロノミックな拘束条件を用いると、変数の数を減らして考えることができる。すなわち、ある変数 x_k （問題を解く上で都合の良いものを選べばよい）を選び、それを他の $n - 1$ 個の独立変数で表すことができる。ここでは簡単のため、

$$x_n = x_n(x_1, x_2, \dots, x_{n-1})$$

としてみよう。 x_1, x_2, \dots, x_{n-1} は完全に独立な変数となる。
このとき、

$$\frac{\partial f}{\partial x_i} = 0$$

がすべての $i = 1, 2, \dots, n - 1$ に対して満たされているのが f の停留値である。

2.2. 未定乗数を用いる方法

上記の方法は理解しやすい。しかし、従属変数を恣意的に選択する必要がある上、場合によっては式の対称性を失ってしまうことにもなりかねない。また、拘束条件が複雑で、従属変数が簡単に書き下さえない場合も存在するだろう。このような場合に用いたいのが、以下に示す**ラグランジュの未定乗数法**である。

まず、停留値においては拘束条件の有無に関わらず

$$\delta f = \frac{\partial f}{\partial x_1} \delta x_1 + \cdots + \frac{\partial f}{\partial x_i} \delta x_i + \cdots + \frac{\partial f}{\partial x_n} \delta x_n = 0$$

が成立する.

さらに、拘束条件 $g = 0$ から

$$\delta g = \frac{\partial g}{\partial x_1} \delta x_1 + \cdots + \frac{\partial g}{\partial x_i} \delta x_i + \cdots + \frac{\partial g}{\partial x_n} \delta x_n = 0$$

が成立する.

ここで、**未定定数**と呼ばれるある定数 λ を用いて、

$$\frac{\partial f}{\partial x_1} \delta x_1 + \cdots + \frac{\partial f}{\partial x_i} \delta x_i + \cdots + \frac{\partial f}{\partial x_n} \delta x_n + \lambda \left(\frac{\partial g}{\partial x_1} \delta x_1 + \cdots + \frac{\partial g}{\partial x_i} \delta x_i + \cdots + \frac{\partial g}{\partial x_n} \delta x_n \right) = 0$$

とする. $0 + \lambda \times 0$ という形になっているので、当然 $= 0$ が成立する.

(注) 未定乗数が突然現れたが、動じてはいけな**い**. **ゼロにゼロを足す** というこの謎の操作こそがラグランジュの未定乗数法の核をなすアイデアである.

式が長いので、シグマを用いてまとめると、

$$\sum_{k=1}^n \left(\frac{\partial f}{\partial x_k} + \lambda \frac{\partial g}{\partial x_k} \right) \delta x_k$$

となる.

ここで、中身の決まっていなかった λ の値を

$$\frac{\partial f}{\partial x_n} + \lambda \frac{\partial g}{\partial x_n} = 0$$

を満たすように決定してみよう. このとき、

$$\sum_{k=1}^{n-1} \left(\frac{\partial f}{\partial x_k} + \lambda \frac{\partial g}{\partial x_k} \right) \delta x_k = 0$$

が成立する。任意の δx_k に対してこれが成立するためには,

$$\frac{\partial f}{\partial x_k} + \lambda \frac{\partial g}{\partial x_k} = 0$$

が $k = 1, 2, \dots, n - 1$ について成立する必要がある。

以上をまとめると, $i = 1, 2, \dots, n$ について

$$\frac{\partial f}{\partial x_i} + \lambda \frac{\partial g}{\partial x_i} = 0$$

が成立するのが, 拘束条件を満たした上での停留値であるということになる。

ここから, ホロノミックな拘束条件を含む停留値問題は,

$$\bar{f} = f + \lambda g$$

において x_1, x_2, \dots, x_n がすべて独立変数だとみなしたときの停留値問題に帰着する。

3. 未定定数の幾何的解釈

ここまでの説明は少し抽象的だったので, 具体的なイメージを持っておこう。

拘束条件 $g(x_1, \dots, x_n) = 0$ は, n 次元空間内の $n - 1$ 次元の曲面 (拘束面) を表す。

拘束条件の勾配 $\nabla g = [\partial g / \partial x_1, \dots, \partial g / \partial x_n]^\top$ を取ると, これは拘束面に対して常に垂直である。

これは以下のようにして確かめられる。

拘束面上の2点 x と $x + \varepsilon$ を考えたとき (ε は非常に小さいとする), x の周りのテイラー展開により

$$g(x + \varepsilon) = g(x) + \varepsilon^\top \nabla g(x) + O(\|\varepsilon\|^2)$$

を得る。 x も $x + \varepsilon$ も拘束面上にあることから, $g(x) = g(x + \varepsilon) = 0$ であり, 2次以上の微小項が無視できるほど小さいとき,

$$\varepsilon^\top \nabla g(x)$$

が成立する。これは $\nabla g(x)$ と, ε (拘束面に平行) が垂直であることを意味している。

さて、もとの問題に戻ろう。 $f(x)$ が停留値を取る x^* を拘束面上で探す。

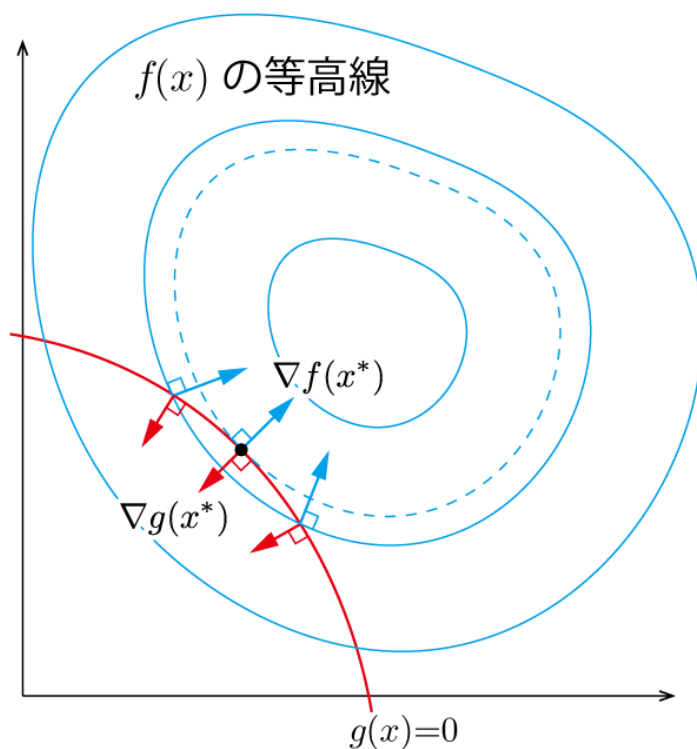
例えば $f(x)$ という山に、 $g(x) = 0$ という道があるというイメージを持とう。このとき $g(x) = 0$ 上を移動しながら、標高が最も高くなる地点 x^* を探す。

x^* においては $\nabla f(x^*)$ も拘束面に対して垂直でなければならない。なぜなら、仮に垂直でないとする
と、拘束面に沿ってさらに大きい $f(x)$ の値が取れるから。

したがって、 ∇f と ∇g は（向きが逆の場合も含めて）平行なベクトルとなる。すなわち、

$$\nabla f + \lambda \nabla g = 0$$

が停留条件として書き直せることになる。ここで λ はラグランジュの未定定数であり、正の値でも負の値でもありえる。



ここで、ラグランジュ関数

$$L(x, \lambda) = f(x) + \lambda g(x)$$

を定義すると、停留条件は $\nabla_x L = 0$ と表される。また、 $\partial L / \partial \lambda = 0$ も成立する。

以上をまとめると、拘束条件 $g(x) = 0$ のもとで $f(x)$ の停留値をもとめる問題では、ラグランジュ関数 $L(x, \lambda)$ の x と λ の両方に対する停留点を求める問題に帰着する。

x が n 次元のベクトルとすると、 $n + 1$ 本の方程式が得られるので、それらを解けば x^* と λ が求められる。

4. 複数の拘束条件を持つ停留値問題

拘束条件が

$$\begin{aligned}g_1(x_1, x_2, \dots, x_n) &= 0 \\g_2(x_1, x_2, \dots, x_n) &= 0 \\&\vdots \\g_m(x_1, x_2, \dots, x_n) &= 0\end{aligned}$$

で表されるときも同様に考えて, $i = 1, 2, \dots, n$ について

$$\frac{\partial f}{\partial x_i} + \lambda_1 \frac{\partial g_1}{\partial x_i} + \lambda_2 \frac{\partial g_2}{\partial x_i} \cdots + \lambda_m \frac{\partial g_m}{\partial x_i} = 0$$

が成立するのが, 拘束条件を満たした上での停留値であるということになる.

5. 例題

イメージがつかない可能性があるので, 例題を問いておこう.

$x^2 + y^2 = 1$ の単位円上で, $f(x) = x^2 + xy + y^2$ が取りうる値の極値を求めよう.

拘束条件は

$$g(x, y) = x^2 + y^2 - 1 = 0$$

と書ける.

未定乗数 λ を用いて,

$$\begin{aligned}\frac{\partial f}{\partial x} + \lambda \frac{\partial g}{\partial x} &= (2 + 2\lambda)x + y = 0 \\ \frac{\partial f}{\partial y} + \lambda \frac{\partial g}{\partial y} &= x + (2 + 2\lambda)y = 0\end{aligned}$$

ここから, $\lambda = -1/2, -3/2$ が求められる.

$\lambda = -1/2$ に対して

$$x + y = 0$$

となり, $x^2 + y^2 = 1$ と連立して, $x = \pm\sqrt{2}, y = \mp\sqrt{2}$ を得る. このとき f は極小値 -1 を取る.

$\lambda = -3/2$ に対して

$$-x + y = 0$$

となり, $x^2 + y^2 = 1$ と連立して, $x = \pm\sqrt{2}, y = \pm\sqrt{2}$ を得る. このとき f は極大値3を取る.

6. Reference

C. Lanczos, "The variational principles of mechanics (Fourth edition)," Dover Publication, 1986.

C. M. Bishop, "Pattern Recognition and Machine Learning," Springer, 2006.