最急降下法とニュートン法のまとめ

清家大嗣

2018 年 9 月 6 日

1　最急降下法とニュートン法について

1.1　最急降下法

$\boldsymbol{x}$ を n 次のベクトル、f をスカラー関数として $y = f(\boldsymbol{x})$ が最小の値を取る $\boldsymbol{x}$ を計算機により計算する方法を考える。この問題を解くための最も簡単な方法が、最急降下法あるいは勾配法と呼ばれる方法である。

ここで、関数 f の勾配ベクトルは 1 式で与えられる。勾配ベクトルは、関数 f の「等高面」、すなわち $f(\boldsymbol{x}) = C$(定数) となる曲面の法線ベクトルとなる。そして、このベクトルの向きに微小に進んだ場合、関数 f は必ず増加する (プラス × プラスもマイナス × マイナスもプラスとなるため)。

$$\nabla f = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, ..., \frac{\partial f}{\partial x_n}\right] \tag{1}$$

例えば、図 1 は、$f(\boldsymbol{x}) = x_1^2 + x_2^2$ とした場合に、定数となる曲面と法線ベクトル $\boldsymbol{d} = [2x_1, 2x_2]$ の例である。

つまり、先程と同様な原理で、法線ベクトルとは逆向きに微小に進めば、関数 f は必ず減少する (プラス × マイナスもマイナス × プラスもマイナスとなるため)。よって、以下のように、(1) で初期化をしたのちに、(2)-(4) の繰り返しを行うことでこの関数の最小値を求めることを期待できる (最小値がないケースや、局所的な最小値になるケースは除く)。

- (1) 初期値 $\boldsymbol{x}_0$ を適当に定め、また $k = 0$ とする。
- (2) $\nabla f(\boldsymbol{x}_k) = 0$ なら繰り返しを終了。
- (3) 探索ベクトル $\boldsymbol{d}_k = \nabla f(\boldsymbol{x}_k)$ とし、適当な微小な正数 α_k を設定し、$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k \boldsymbol{d}_k$ を計算する。
- (4) $k = k + 1$ とする。

このような手法を最急降下法と呼ぶ。最急降下法は, 簡単である, また、各ステップごとの計算量が少ない, 関数が唯一の極小点を持つときには大域的な収束性が保証されるという利点もある。一方, 解の収束が極めて遅いという欠点がある。

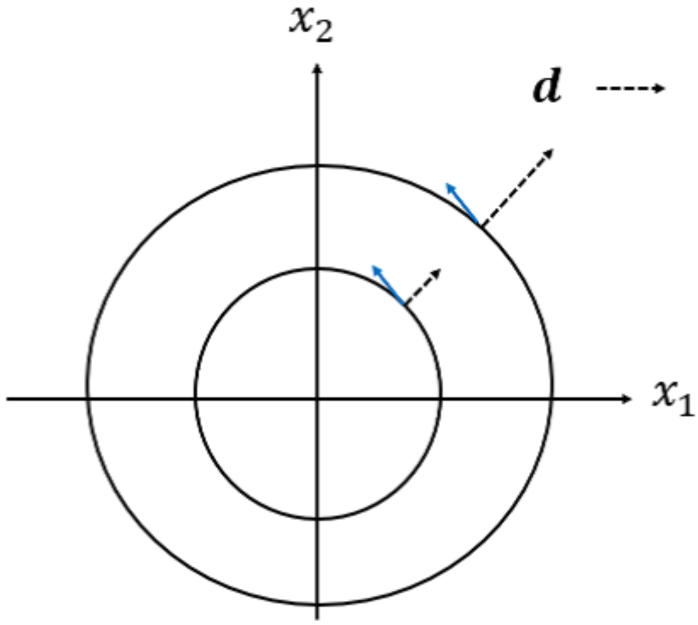

図 1: $f(\boldsymbol{x}) = x_1^2 + x_2^2$ の場合の法線ベクトル $\boldsymbol{d}$ の例

1.2 ニュートン法

最急降下法は、収束の遅さが問題である。そこでその問題を克服する方法のひとつであるニュートン法を考える。 ニュートン法では、まず点 $\boldsymbol{x}_k$ まわりで二次の項まで残した多変数のテイラー展開を考える [1]

$$f(\boldsymbol{x}) \approx f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)(\boldsymbol{x} - \boldsymbol{x}_k) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_k)^T \nabla^2 f(\boldsymbol{x}_k)(\boldsymbol{x} - \boldsymbol{x}_k) \tag{2}$$

ここで、$\nabla^2 f$ は、ヘッセ行列と呼ばれ次式で表される。

$$\nabla^2 f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix} \tag{3}$$

また、上記定義での偏微分は次式で表されることに注意する (下記の場合、関数 f は全ての変数 x_i に対して C^2 級[2]であれば偏微分の順序交換ができる [2])。

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right) \tag{4}$$

つまり、3 式はたいていの場合対称行列と考えてもよい。また、「極小となるで偏導関数の値が全て 0 かつヘッセ行列が正定値」([1]) より、2 式を次式で変形する ($H = \nabla^2 f(\boldsymbol{x}_k)$, $\boldsymbol{p} = \nabla f(\boldsymbol{x}_k)^T$, $\boldsymbol{\xi} = \boldsymbol{x} - \boldsymbol{x}_k$)。

---

[1] ベクトル表記で 2 式のように、二次の項までの多変数関数のテイラー展開を表現できる理由は 1 変数の場合のテイラー展開を考えるとよく分かる。結局のところ、多変数関数のテイラー展開も 1 変数関数のテイラー展開を適応し続けているだけであり、一階の偏微分が加わる度に

[2] 関数 f に対して 2 回偏導関数が存在して、かつ連続であること

$$f(\boldsymbol{x}) \approx f(\boldsymbol{x}_k) + \boldsymbol{p}^T\boldsymbol{\xi} + \frac{1}{2}\boldsymbol{\xi}^T H\boldsymbol{\xi} \tag{5}$$

$$= f(\boldsymbol{x}_k) - \frac{1}{2}\boldsymbol{p}^T H^{-1}\boldsymbol{p}\frac{1}{2}(\boldsymbol{\xi} + H^{-1}\boldsymbol{p})^T H(\boldsymbol{\xi} + H^{-1}\boldsymbol{p}) \tag{6}$$

H が正定値であることより、$\boldsymbol{\xi} = -H^{-1}\boldsymbol{p}$ の場合に 6 式は最小になる。従って、式をもとの形に書き直して、

$$\boldsymbol{d}_k = \left(\nabla^2 f(\boldsymbol{x}_k)\right)^{-1}\left(\nabla f(\boldsymbol{x}_k)\right)^T \tag{7}$$

上式を用いて、下式のようにして、求めたいベクトルを探索していく。

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k\boldsymbol{d}_k \tag{8}$$

次の文は、[3] からの引用である。Newton 法は一般に最急降下法と比べて極めて高速であり, 特に関数 f が 2 次のときには 1 回の反復で最適解が得られるという 利点がある一方で, 初期値が十分最適値に近くないと収束性が保証されない, ヘッセ行列が正定値でないと利用できない, また正確に求められていないと不安定になる, ヘッセ行列の計算に手間がかかる, という欠点がある。 また, 最急降下法と異なり, Newton 法では, ステップ幅 を 1 に固定しても良い収束特性が得られることも多い。 なお, 関数に複数の極小点があるときには求められるのは 極小点のどれかであって最小点とは限らないことは 最急降下法と同じである。

また、実対称行列に対する半正定値の判定条件は [4] などを参考にするとよい。

参考文献

[1] 高校数学の美しい物語「多変数関数の極値判定とヘッセ行列」https://mathtrain.jp/hessian

[2] 高校数学の美しい物語「偏微分の順序交換の十分条件とその証明」https://mathtrain.jp/henbibunexchange

[3] 琉球大学ウェブ解説「Newton 法」 http://dsl4.eee.u-ryukyu.ac.jp/DOCS/nlp/node5.html

[4] 高校数学の美しい物語「半正定値行列の同値な 4 つの定義（性質）と証明」https://mathtrain.jp/positivesemi