§ 4.2.3　Method of Steepest Descent

伊 冉 (Andre YI) - 21122200512

2025 年 11 月 21 日

1　Steepest Descent(最急降下方向法)

小さなステップ幅 $\overset{\text{エータ}}{\eta} > 0$ を与えられたとき，点 x^0 において関数 $f : \mathbb{R}^n \to \mathbb{R}$ をもっとも減少させる単位ベクトル ν を求めたい。

テイラー展開の 1 次近似より，

$$f(x^0 + \eta\nu) - f(x^0) = \eta\langle\nabla f(x^0), \nu\rangle + o(\eta^2) \approx \eta\langle\nabla f(x^0), \nu\rangle.$$

スカラー積に対する Cauchy-Schwarz の不等式[1] $|\langle a, b\rangle| \le \|a\|\|b\| \quad (a, b \in \mathbb{R}^n)$ から

$$-\|\nabla f(x^0)\| \cdot \|\nu\| \le \langle\nabla f(x^0), \nu\rangle \le \|\nabla f(x^0)\| \cdot \|\nu\| \quad (\|\nu\| = 1)$$

であり，最小値は

$$\nu^* = -\frac{\nabla f(x^0)}{\|\nabla f(x^0)\|}$$

で達成される。

Ex 1: ν^* の証明

$\nu^* = -\frac{\nabla f(x^0)}{\|\nabla f(x^0)\|}$ となる証明

解 答

Proof. ラグランジュの未定乗数法を用いて示す。制約条件 $\|\nu\|^2 - 1 = 0$ のもとで $\langle\nabla f(x^0), \nu\rangle$ を最小化する。ラグランジュ関数を

$$\mathcal{L}(\nu, \lambda) = \langle\nabla f(x^0), \nu\rangle + \lambda(\|\nu\|^2 - 1)$$

とおく。次に $\mathcal{L}$ の ν に関する偏微分は 0 にする

$$\frac{\partial\mathcal{L}}{\partial\nu} = \nabla f(x^0) + 2\lambda\nu = 0$$

から

$$\nu = -\frac{1}{2\lambda}\nabla f(x^0).$$

制約条件に代入して

$$\|\nu\|^2 = \frac{1}{4\lambda^2}\|\nabla f(x^0)\|^2 = 1$$

を得る。したがって

$$\lambda = \pm\frac{1}{2}\|\nabla f(x^0)\|$$

最小値を与えるのは負号の場合であり，(***i.e.*** $\lambda = -\frac{1}{2}\|\nabla f(x^0)\|$), その λ を ν に代入すると

$$\nu^* = -\frac{\nabla f(x^0)}{\|\nabla f(x^0)\|}$$

■

[1] 証明は Appendix の A.1 を参考してください。

したがって関数値の最大変化は近似的に

$$f(x^0+\eta\nu^*)-f(x^0)=\eta\langle\nabla f(x^0),\nu^*\rangle^2=-\eta\|\nabla f(x^0)\|$$

となる。ここで η を **学習率** *(learning rate)* と呼ぶ。

大域的最小点 x^* の吸引域内の初期点 x^0 から始めて，次の反復を考える：

$$x^{n+1}=x^n+\eta\nu^*=x^n-\eta\frac{\nabla f(x^n)}{\|\nabla f(x^n)\|}. \qquad \textbf{[4.2.3]}$$

このとき

$$f(x^{n+1})-f(x^n)=-\eta\|\nabla f(x^n)\|<0 \quad (\because \eta>0, \|\nabla f(x^n)\|>0)$$

が成り立ち，目的関数は必ず減少する。一方で $\|x^{n+1}-x^n\|=\eta>0$ で一定のため，一般には列 $(x^n)_n$ は収束しない。

Prop. 1.1 (標準的な勾配降下法): ステップ幅を

$$\eta_n=\delta\|\nabla f(x^n)\| \quad (\exists\delta>0)$$

と仮定すると，反復式 (4.2.3) は

$$x^{n+1}=x^n-\delta\nabla f(x^n) \qquad \textbf{[4.2.4]}$$

となる。これは標準的な勾配降下法である。

Ex 2: 同値証明

反復式 (4.2.4) により定義される列 (x^n) が収束することと，勾配列 $\nabla f(x^n)$ が 0 に収束することは同値である。***i.e.*** $(x^n)_n$ が収束すること $\overset{\text{iff}}{\Longleftrightarrow}$ $\nabla f(x^n)\to 0 \quad (x\to\infty)$

解答

Proof. $\Rightarrow$: $(x^n)_n$ が収束すると仮定する。このとき

$$0=\lim_{n\to\infty}\|x^{n+1}-x^n\|=\delta\lim_{n\to\infty}\|\nabla f(x^n)\|$$

が成り立つので，

$$\nabla f(x^n)\longrightarrow 0 \quad (n\to\infty)$$

が従う。

$\Leftarrow$: $\nabla f(x^n)\to 0$ $(n\to\infty)$ を仮定する。任意の $p\geq 1$ に対し，三角不等式より

$$\begin{aligned}\|x^{n+p}-x^n\|&=\left\|(x^{n+p}-x^{n+p-1})+\cdots+(x^{n+1}-x^n)\right\|\\&\leq\|x^{n+p}-x^{n+p-1}\|+\cdots+\|x^{n+1}-x^n\| \quad ^{3}\\&=\delta\sum_{j=0}^{p-1}\left\|\nabla f(x^{n+j})\right\|.\end{aligned}$$

[2] 幾何の視点から考えると、$\langle\nabla f(x^0),\nu^*\rangle=\|\nabla f(x^0)\|\cdot\|\nu^*\|\cdot\cos\theta=-\|\nabla f(x^0)\|$ となる。ここで、θ は $\nabla f(x^0)$ と ν^* のなす角であり、ν^* は $\nabla f(x^0)$ の反対方向を向いているため、$\theta=\pi$ である。

[3] 証明は Appendix の A.2 を参考してください。

したがって,

$$\lim_{n\to\infty}\|x^{n+p}-x^n\| \le \delta \sum_{j=0}^{p-1} \lim_{n\to\infty}\left\|\nabla f(x^{n+j})\right\| = 0.$$

よって $(x^n)_n$ はコーシー列[4]であり，したがって収束する。

■

もし f が連続的に微分可能 (f は C^1級) であれば，$\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ は連続関数である。したがって、$x^n \to x^*$ から $\nabla f(x^n) \to \nabla f(x^*)$ が従う。これは最適性条件と一致している。

Ex．1.2: 峡谷関数の例関数

$$f(x,y) = \frac{1}{2}y^2 - x, \quad (x,y) \in (0,1)\times(-2,2)$$

を考える。これは**峡谷状** *(Canyon Type)* のグラフをもち，点 $(1,0)$ で最小値をとる。勾配は

$$\nabla f(x,y)^{\mathsf{T}} = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) = (-1, y)$$

である。

反復式 (4.2.4) は

$$\begin{cases} x^{n+1} = x^n + \delta, \\ y^{n+1} = (1-\delta)y^n \end{cases}$$

となり，任意の初期点 (x^0, y^0) に対し

$$x^n = x^0 + n\delta, \qquad y^n = (1-\delta)^n y^0$$

と表せる[5]。

次に、$(y^n)_n$ が収束する $\overset{\text{iff}}{\Longleftrightarrow} \lim_{n\to\infty}(1-\delta)^n = 0$ により、$|1-\delta| < 1 \iff 0 < \delta < 2$ である。

この範囲で次の二つの挙動が区別される。

(i) $0 < \delta < 1$ のとき，$1-\delta \in (1,0) \implies |1-\delta| < 1 \implies \lim_{n\to\infty}(1-\delta)^n = 0$。従って、$(y^n)_n$ は収束する。$x^{n+1} - x^n = \delta \implies (x^n)_n$ は公差 δ の等差数列として増加する。***i.e.*** $x^n \le 1 \implies x^n = x^0 + n\delta \le 1 \implies n \le \dfrac{1-x^0}{\delta} \quad (\delta > 0) \implies n = \lfloor \dfrac{1-x^0}{\delta} \rfloor$ [6]。図 1 の **a** を参照してください。

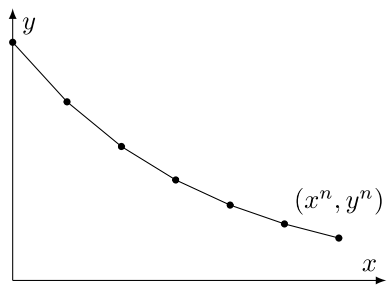
(ii) $1 < \delta < 2$ のとき，$1-\delta < 0$ と $|1-\delta| < 1$ より、y^n は符号を交互に変えながら 0 に振動収束する。これは峡谷の両側を行き来しながら谷底を行き過ぎる挙動に対応する。図 1 の **b** を参照してください。

---

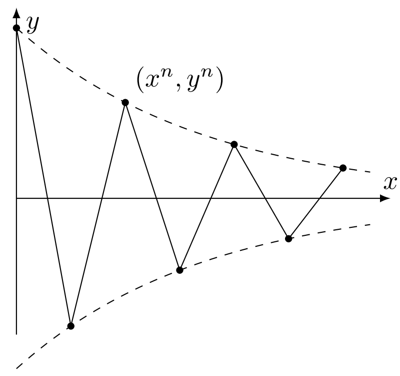
[4] 数列 a_n がコーシー列であるとは，$\lim_{n,m\to\infty}|a_n - a_m| = 0$ ***i.e.*** 任意の $\epsilon > 0$ に対して，ある N が存在して，$m,n \ge N$ のとき $\|x^m - x^n\| < \epsilon$ となる列のことをいう。

[5] 数学的帰納法により、
$x^1 = x^0 + \delta, x^2 = x^1 + \delta = x^0 + 2\delta, \cdots, x^n = x^0 + n\delta$
$y^1 = (1-\delta)y^0, y^2 = (1-\delta)y^1 = (1-\delta)^2 y^0, \cdots, y^n = (1-\delta)^n y^0$

[6] $\lfloor \dfrac{1-x^0}{\delta} \rfloor$ とは、$\dfrac{1-x^0}{\delta}$ を超えない最大の整数である床関数。**例:** $\lfloor 3.7 \rfloor = 3, \lfloor -2.3 \rfloor = -3$



a



b

図 1: *Iterations*: **a.** Case $0 < \delta < 1$. **b.** Case $1 < \delta < 2$.

Appendix A
証明問題

A.1

Cauchy-Schwarz の不等式

Ex 3: Cauchy-Schwarz の不等式の証明

$a, b \in \mathbb{R}^n$ に対して、次の不等式を証明せよ。

$$|\langle a, b \rangle| \leq \|a\|\|b\|$$

等号成立の条件も述べよ。

解 答

Proof. $a, b \in \mathbb{R}^n$ に対して、$t \in \mathbb{R}$ とすると、次の式が成り立つ。

$$0 \leq \|a - tb\|^2 = \langle a - tb, a - tb \rangle = \|a\|^2 - 2t\langle a, b \rangle + t^2\|b\|^2$$

右辺は t に関する二次式であり、常に非負であるため、判別式 $\Delta \leq 0$ が必要がある。すなわち、

$$(-2\langle a, b \rangle)^2 - 4\|b\|^2\|a\|^2 \leq 0$$

これを整理すると、Cauchy-Schwarz の不等式が得られる。

$$|\langle a, b \rangle| \leq \|a\|\|b\|$$

等号成立の条件は、a と b が線形従属である場合、すなわち、あるスカラー $k \in T \subset \{\mathbb{R} \cup \mathbb{C}\}$ が存在して $a = kb$ または $b = ka$ が成り立つ場合である。

■

A.2 三角不等式の証明

Ex 4: 三角不等式の証明

任意のベクトル $a, b \in \mathbb{R}^n$ に対して、次の不等式を証明せよ。

$$\|a+b\| \leq \|a\| + \|b\|$$

解答

Proof. ベクトル $a, b \in \mathbb{R}^n$ に対して、次の式が成り立つ。

$$\begin{aligned} \|a+b\|^2 &= \langle a+b, a+b \rangle \\ &= \|a\|^2 + 2\langle a, b \rangle + \|b\|^2 \\ &\leq \|a\|^2 + 2\|a\|\|b\| + \|b\|^2 \quad \text{(Cauchy-Schwarz の不等式より)} \\ &= (\|a\| + \|b\|)^2 \end{aligned}$$

両辺の平方根を取ると、三角不等式が得られる。

$$\|a+b\| \leq \|a\| + \|b\|$$

■