

§ 4.2.4 Line Search Method

伊 冉 (Andre YI)

2026 年 1 月 15 日

1 Line Search Method (直線探索法)

通常の最急降下法 (Steepest Descent) では、学習率 (ステップ幅) η を固定値として扱うのが、直線探索法はこの η も最適化の対象とする手法である。つまり、現在の点 x^0 から勾配方向 $-\nabla f(x^0)$ に進む際、「どのくらい進めばその直線上での最小値にたどり着くか」を計算して移動する。

Def. 1.1 (アルゴリズムの定義):

点 x^0 における勾配方向への探索は、以下の最適化問題として定式化される。適切なステップ幅 $\eta_0 > 0$ は、以下の式を満たすものとして選ばれる。

$$\eta_0 = \arg \min_{\eta} f(x^0 - \eta \nabla f(x^0)) \quad [1]$$

これを一般化すると、反復 n 回目における更新式は以下ようになる。

$$\eta_n = \arg \min_{\eta} f(x^n - \eta \nabla f(x^n)) \quad [2]$$

$$x^{n+1} = x^n - \eta_n \nabla f(x^n) \quad [3]$$

Ex 1.1: 直線探索法の直交性の証明

直線探索法において、各反復での勾配ベクトルが直前の勾配ベクトルと直交することを証明せよ。

Proof.

現在の点 x^0 からの移動を η の関数 $g(\eta)$ とおく

$$g(\eta) = f(x^0 - \eta \nabla f(x^0))$$

この関数 $g(\eta)$ が最小となる η_0 を求めるため、 η で微分して 0 とおく

$$\begin{aligned} g'(\eta) &= \frac{d}{d\eta} g(\eta) \\ &= \nabla f(x^0 - \eta \nabla f(x^0))^T \cdot \frac{d}{d\eta} (x^0 - \eta \nabla f(x^0)) \\ &= -\langle \nabla f(x^0 - \eta \nabla f(x^0)), \nabla f(x^0) \rangle \end{aligned}$$

ここで、 $x^1 = x^0 - \eta_0 \nabla f(x^0)$ であるため、最適値 η_0 において $g'(\eta_0) = 0$ ということは、

$$\langle \nabla f(x^1), \nabla f(x^0) \rangle = 0$$

を意味する。*i.e.* 勾配ベクトル $\nabla f(x^1)$ は、今進んできた方向 $\nabla f(x^0)$ と直交 (Normal) する。 ■

¹ $\arg \min$ は「最小値を与える点」を意味する。

Ex 1.2: 変数関数の最小化

実変数 $x \in \mathbb{R}$ の関数 $f(x) = \frac{1}{2}(ax - b)^2$ を考える (ただし $a \neq 0, b \in \mathbb{R}$)。この関数の最小値が $x^* = b/a$ であることは明らかだが、これを最急降下法を用いて導出する。

解答

関数を x で微分する。

$$f'(x) = \frac{d}{dx} \left(\frac{1}{2}(ax - b)^2 \right) = (ax - b) \cdot a = a^2x - ab$$

初期点 x^0 から開始し、ステップ幅 (学習率) を固定値 δ とする。更新式は $x^{n+1} = x^n - \delta f'(x^n)$ となるので、代入する。

$$\begin{aligned} x^{n+1} &= x^n - \delta(a^2x^n - ab) \\ &= (1 - \delta a^2)x^n + \delta ab \end{aligned}$$

ここで、 $\alpha = 1 - \delta a^2, \beta = \delta ab$ と置くと、漸化式は $x^{n+1} = \alpha x^n + \beta$ という線形漸化式になる。この数列の一般項は、 x^0 を用いて以下のように表される。

$$x^n = \alpha^n x^0 + \beta \sum_{k=0}^{n-1} \alpha^k = \alpha^n x^0 + \beta \frac{1 - \alpha^n}{1 - \alpha}$$

ここで、収束するためには公比 $|\alpha| < 1$ である必要がある。 δ が十分に小さく、 $0 < \delta < \frac{1}{a^2}$ であれば、 $0 < \alpha < 1$ となり、 $\alpha^n \rightarrow 0 (n \rightarrow \infty)$ が成立する。よって極限值は：

$$\begin{aligned} x^* &= \lim_{n \rightarrow \infty} x^n = 0 \cdot x^0 + \beta \frac{1 - 0}{1 - \alpha} \\ &= \frac{\beta}{1 - \alpha} = \frac{\delta ab}{1 - (1 - \delta a^2)} \\ &= \frac{\delta ab}{\delta a^2} = \frac{b}{a} \end{aligned}$$

となり、解析的な最小値と一致することが示された。

Ex 1.3: 多変数関数の最小化

$f: \mathbb{R}^k \rightarrow \mathbb{R}$ を $f(x) = \frac{1}{2}\|Ax - b\|^2$ とする。ここで A は $m \times k$ 行列 (rank k)、 $b \in \mathbb{R}^m$ である。この最小値 x^* が正規方程式の解 $x^* = (A^T A)^{-1} A^T b$ に収束することを最急降下法で示す。

証明

まず目的関数を展開し、勾配 $\nabla f(x)$ を求める。

$$\begin{aligned} f(x) &= \frac{1}{2}(Ax - b)^T(Ax - b) \\ &= \frac{1}{2}(x^T A^T Ax - x^T A^T b - b^T Ax + b^T b) \\ &= \frac{1}{2}(x^T A^T Ax - 2\langle A^T b, x \rangle + \|b\|^2) \end{aligned}$$

これを x で微分する (行列微分の公式 $\nabla_x(x^T Qx) = 2Qx$ を使用)。

$$\nabla f(x) = A^T Ax - A^T b \implies \nabla f(x^n) = A^T Ax^n - A^T b$$

最急降下法の更新式 $x^{n+1} = x^n - \delta \nabla f(x^n)$ に代入する。

$$\begin{aligned} x^{n+1} &= x^n - \delta(A^T Ax^n - A^T b) \\ &= (\mathbb{I}_k - \delta A^T A)x^n + \delta A^T b \end{aligned}$$

ここで、 $M = \mathbb{I}_k - \delta A^T A$, \mathbb{I}_k を単位行列と置くと、これは先ほどの1変数の例となる。

$$x^{n+1} = Mx^n + \delta A^T b$$

この漸化式を繰り返し適用すると、以下ようになる。

$$x^n = M^n x^0 + (\mathbb{I}_k - M^n)(\mathbb{I}_k - M)^{-1} \delta A^T b$$

ここで、 $\mathbb{I}_k - M = \delta A^T A \implies (\mathbb{I}_k - M)^{-1} = (\delta A^T A)^{-1}$ $n \rightarrow \infty$ で $M^n \rightarrow 0$ (ゼロ行列) となる条件を考えると、 $x^* = (\mathbb{I}_k - M)^{-1} \delta A^T b = (A^T A)^{-1} A^T b$ となる。

次、 $M^n \rightarrow 0$ となる条件を確認する。これには M のすべての固有値 $\lambda_i(M)$ について $|\lambda_i| < 1$ である必要がある。 $M = \mathbb{I}_k - \delta A^T A$ の固有値 $\lambda_i(M)$ と、 $A^T A$ の固有値 η_i の関係は以下の通りである。

$$\lambda_i(M) = 1 - \delta \eta_i$$

$A^T A$ は正定値行列 (ランク k のため) なので、 $\eta_i > 0$ である。したがって、 δ を十分に小さく選べば、

$$-1 < 1 - \delta \eta_i < 1 \implies |\lambda_i(M)| < 1$$

² $\|v\|^2 = v^T v$ および $\langle u, v \rangle = u^T v = v^T u$ を使用し、 $(Ax - b)^T = (Ax)^T - b^T = x^T A^T - b^T$ と展開した。

を満たすことが可能である。ここで、 $-1 < 1 - \delta\eta_i \implies \delta < \frac{2}{\max \eta_i}$ ³。この条件下で $M^n \rightarrow 0$ となり、極限は以下ようになる。

$$\begin{aligned} x^* &= \lim_{n \rightarrow \infty} x^n = (\mathbb{I}_k - M)^{-1} \delta A^T b \\ &= (\delta A^T A)^{-1} \delta A^T b \\ &= (A^T A)^{-1} A^T b \end{aligned}$$

これにより、最急降下法が正規方程式の解に収束することが証明された。 ■

³ 行列には複数の固有値 η があるが、一番厳しい条件（一番大きい η ）に合わせておく必要がある。そのため、 $\max \eta_i$ となる