最急降下法

微分可能な関数 $f : \mathbb{R}^n \to \mathbb{R}$ の無制約最小化問題

$$\text{Minimize} \quad f(\boldsymbol{x})$$

を考える．この問題に対する最急降下法は，次のように記述できる．

アルゴリズム 1（最急降下法）

Require: $\boldsymbol{x}_0 \in \mathbb{R}^n$, $\epsilon > 0$.

1: **for** $k = 0, 1, 2, \dots$ **do**
2: 　$\|\nabla f(\boldsymbol{x}_k)\| < \epsilon$ ならば，$\boldsymbol{x}_k$ を解として出力し終了する（注意 1 を参照）．
3: 　$\boldsymbol{d}_k \leftarrow -\nabla f(\boldsymbol{x}_k)$.
4: 　ステップ幅 $\alpha_k > 0$ を定める（後述）．
5: 　$\boldsymbol{x}_{k+1} \leftarrow \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$.
6: **end for**

注意 1. 無制約最適化問題の解法では，最適性の 1 次の必要条件 $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$ を満たす点 $\boldsymbol{x}^*$ を求めることを目標にする．この条件は

$$\|\nabla f(\boldsymbol{x}^*)\| = 0 \tag{1}$$

と等価である．しかし，最急降下法をはじめとする反復法では，原理的に，無限回の反復を行わないと（つまり，$k \to \infty$ としないと）条件 (1) は厳密に満たされない．そこで，条件 (1) を少し緩めたものとして，たとえば条件

$$\|\nabla f(\boldsymbol{x}_k)\| < \epsilon$$

を満たす点 $\boldsymbol{x}_k$ が得られたら，それを近似解として出力する．ここで，ϵ はあらかじめ定めた十分に小さい正の定数である． ■

アルゴリズム 1 の 4 行目でステップ幅 α_k を定める代表的な方法は，直線探索とよばれるものである [*1]．まず，関数 ϕ_k を

$$\phi_k(\alpha) = f(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k), \qquad \alpha \geq 0 \tag{2}$$

で定義する．3 行目で $\boldsymbol{d}_k$ が既に得られているので，ϕ_k は 1 変数の関数であることに注意する．点 $\boldsymbol{x}_k$ を通り，向きが $\boldsymbol{d}_k$ の直線を考える．$\phi_k(\alpha)$ は，この直線上において目的関数 f の値がどう変化するかを表している．そこで，$\alpha \geq 0$ のうち $\phi_k(\alpha)$ の値を最小にするものをステップ幅 α_k として採用することを考える．つまり，1 変数の最小化問題

$$\text{Minimize} \quad \phi_k(\alpha) \tag{3a}$$

$$\text{subject to} \quad \alpha \geq 0 \tag{3b}$$

の最適解を α_k とする．この操作は，正確な直線探索とよばれている．

[*1] 直線探索は，最急降下法に限らず，無制約最適化の多くのアルゴリズムにおいて用いられる．

問題 (3) は1変数の最適化問題なので，元の問題（n 変数の最適化問題）よりはずっと扱いやすい．とはいうものの，現実には問題 (3) を厳密に解くことは手間なことが多い．そこで実際には，問題 (3) を厳密に解くのではなく,「より緩い条件」を満たす α をステップ幅 α_k として採用する．代表的な「より緩い条件」の一つは，$\alpha > 0$ が条件

$$\phi_k(\alpha) \leq f(\boldsymbol{x}_k) + c(\nabla f(\boldsymbol{x}_k)^\top \boldsymbol{d}_k)\alpha \tag{4}$$

を満たすならばそれを α_k として採用するというものであり，アルミホ (Armijo) の条件とよばれている．ただし，c は $0 < c < 1$ を満たす定数である．まず，アルミホの条件を満たす $\alpha > 0$ が存在することを確認する．

命題 1. 探索方向 $\boldsymbol{d}_k$ は条件 $\nabla f(\boldsymbol{x}_k)^\top \boldsymbol{d} < 0$ を満たすとする [*2]．また，f は下に有界であることを仮定する．このとき，ある $\hat{\alpha} > 0$ が存在して，$0 < \alpha \leq \hat{\alpha}$ を満たす任意の α が条件 (4) を満たす．

証明. まず，f の Taylor 展開より

$$\phi_k(\alpha) = f(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k) = f(\boldsymbol{x}_k) + (\nabla f(\boldsymbol{x}_k)^\top \boldsymbol{d}_k)\alpha + \mathrm{O}(\alpha^2)$$

が成り立つことと $c < 1$ より，条件 (4) は十分に小さな $\alpha > 0$ に対して成立する．次に，命題の仮定より，任意の $\alpha \geq 0$ に対して，(2) で定義される $\phi_k(\alpha)$ の値は下に有界である．また，1次関数 $l : \mathbb{R} \to \mathbb{R}$ を

$$l(\alpha) = f(\boldsymbol{x}_k) + c(\nabla f(\boldsymbol{x}_k)^\top \boldsymbol{d}_k)\alpha, \qquad \alpha \geq 0$$

で定義すると，$\nabla f(\boldsymbol{x}_k)^\top \boldsymbol{d}_k < 0$ かつ $c > 0$ であることから，$\alpha \to +\infty$ ととると $l(\alpha)$ は下に非有界である．従って，$y = \phi_k(\alpha)$ のグラフと $y = l(\alpha)$ のグラフは $\alpha > 0$ で少なくとも1回交差する．交差する $\alpha > 0$ の値の最小値を $\hat{\alpha}$ とおけば，この命題が得られる． □

命題 1 より，ステップ幅 $\alpha_k > 0$ を十分に小さく選べばアルミホの条件が満たされることがわかる．一方で，α_k の値が小さすぎると，更新後の点 $\boldsymbol{x}_{k+1}$ は現在の点 $\boldsymbol{x}_k$ からほとんど変更されないため，最急降下法が終了するまでに多くの反復回数が必要になってしまう．したがって，α_k はなるべく大きな値とすることが望ましい．そのような α_k の決め方として，バックトラック法とよばれる方法（アルゴリズム 2）がよく用いられている．

アルゴリズム 2 （バックトラック法による直線探索）

Require: 初期値 $\bar{\alpha}$ (> 0), 定数 ρ $(0 < \rho < 1)$

```
α ← ᾱ.
if α が (4) を満たす then
    α_k ← α として終了する.
else
    α ← ρα として 2 へ.
end if
```

つまり，この方法は α を初期値 $\bar{\alpha}$ から徐々に小さくしていき，アルミホの条件を最初に満たした時点での値を α_k として出力するというものである．

（以上）

[*2] このような $\boldsymbol{d}_k$ は，点 $\boldsymbol{x}_k$ における関数 f の降下方向であるという（最急降下法の探索方向 $-\nabla f(\boldsymbol{x}_k)$ が降下方向であることは，すぐに確かめられる）．