準ニュートン法

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

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

を解くことを考える．

ニュートン (Newton) 法では，ニュートン方程式

$$\nabla^2 f(\boldsymbol{x}_k)\, \boldsymbol{d}_k = -\nabla f(\boldsymbol{x}_k) \tag{1}$$

の解として探索方向 $\boldsymbol{d}_k$ を定め，解を

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{d}_k \tag{2}$$

と更新する．準ニュートン法は，(1) の $\nabla^2 f(\boldsymbol{x}_k)$ を適当な正定値行列 B_k で近似し，探索方向 $\boldsymbol{d}_k$ を線形方程式

$$B_k\, \boldsymbol{d}_k = -\nabla f(\boldsymbol{x}_k) \tag{3}$$

の解として定める解法である．

アルゴリズム 1 （準ニュートン法）

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

1: $B_0 \leftarrow I$.
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: 　$\|\nabla f(\boldsymbol{x}_k)\| < \epsilon$ ならば，$\boldsymbol{x}_k$ を解として出力し終了する．
4: 　$B_k\, \boldsymbol{d}_k = -\nabla f(\boldsymbol{x}_k)$ を解いて $\boldsymbol{d}_k$ を求める．
5: 　直線探索によりステップ幅 $\alpha_k > 0$ を定める．
6: 　$\boldsymbol{x}_{k+1} \leftarrow \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$.
7: 　B_{k+1} を生成する．
8: **end for**

アルゴリズム 1 の 7 行目では，$\boldsymbol{x}_k$, $\boldsymbol{x}_{k+1}$, $\nabla f(\boldsymbol{x}_k)$, $\nabla f(\boldsymbol{x}_{k+1})$ を用いて，行列 B_k に修正を加えることで行列 B_{k+1} を生成する．具体的な生成法はいくつか提案されているが，なかでも BFGS 公式 [*1]

$$B_{k+1} = B_k + \frac{\boldsymbol{y}_k \boldsymbol{y}_k^\top}{\boldsymbol{y}_k^\top \boldsymbol{s}_k} - \frac{B_k \boldsymbol{s}_k \boldsymbol{s}_k^\top B_k}{\boldsymbol{s}_k^\top B_k \boldsymbol{s}_k} \tag{4}$$

は最もよく用いられているものの一つである．ただし，記号の簡単のために

$$\boldsymbol{s}_k := \boldsymbol{x}_{k+1} - \boldsymbol{x}_k, \quad \boldsymbol{y}_k := \nabla f(\boldsymbol{x}_{k+1}) - \nabla f(\boldsymbol{x}_k)$$

とおいた．

[*1] C. G. Broyden (1970), R. Fletcher (1970), D. Goldfarb (1970), D. F. Shanno (1970) によって独立に提案されたので，彼らの名前の頭文字をとってこのようによばれている．

(4) で定められる B_{k+1} は，次の性質 (a)–(c) を満たすことが確認できる：

(a) B_{k+1} は $B_{k+1}\boldsymbol{s}_k = \boldsymbol{y}_k$ を満たす（この等式は，セカント条件とよばれている）．

(b) B_k が対称行列ならば，B_{k+1} も対称行列である．

(c) B_k が正定値かつ $\boldsymbol{s}_k^\top \boldsymbol{y}_k > 0$ ならば，B_{k+1} も正定値である．

このうち，性質 (c) に注目する．アルゴリズム 1 の 5 行目では，ステップ幅 α_k がアルミホ (Armijo) の条件

$$f(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k) \le f(\boldsymbol{x}_k) + c_1(\nabla f(\boldsymbol{x}_k)^\top \boldsymbol{d}_k)\alpha_k \tag{5}$$

を満たすように定めることが多い（c_1 は $0 < c_1 < 1$ を満たす定数である）．実は，これに加えて条件

$$\nabla f(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k)^\top \boldsymbol{d}_k \ge c_2(\nabla f(\boldsymbol{x}_k)^\top \boldsymbol{d}_k) \tag{6}$$

も満たすように選べば，$\boldsymbol{s}_k^\top \boldsymbol{y}_k > 0$ が成り立つことを示すことができる（c_2 は $c_1 < c_2 < 1$ を満たす定数である）．なお，条件 (5) と (6) とをあわせたものは，直線探索におけるウルフ (Wolfe) の条件とよばれている．さて，$\boldsymbol{s}_k$ の定義より $\boldsymbol{d}_k = \boldsymbol{s}_k/\alpha_k$ が成り立つが，これを (6) へ代入することで不等式 $\nabla f(\boldsymbol{x}_{k+1})^\top \boldsymbol{s}_k \ge c_2(\nabla f(\boldsymbol{x}_k)^\top \boldsymbol{s}_k)$ が得られる．この不等式と $\boldsymbol{y}_k$, $\boldsymbol{s}_k$ の定義および (3) より

$$\begin{aligned}\boldsymbol{y}_k^\top \boldsymbol{s}_k = (\nabla f(\boldsymbol{x}_{k+1}) - \nabla f(\boldsymbol{x}_k))^\top \boldsymbol{s}_k &\ge (c_2 - 1)\nabla f(\boldsymbol{x}_k)^\top (\alpha_k \boldsymbol{d}_k) \\ &= (c_2 - 1)(-\nabla f(\boldsymbol{x}_k)^\top B_k^{-1} \nabla f(\boldsymbol{x}_k))\alpha_k\end{aligned} \tag{7}$$

が得られる．B_k が正定値であれば B_k^{-1} も正定値であるから，$c_2 < 1$ より (7) の最右辺は正である．したがって，$\boldsymbol{s}_k^\top \boldsymbol{y}_k > 0$ が成り立つ．

行列 B_k の更新公式 (4) を，逆行列 B_k^{-1} に関する式に書き直すこともできる．つまり，$H_k := B_k^{-1}$, $H_{k+1} := B_{k+1}^{-1}$ とおくと，(4) と等価な式として

$$H_{k+1} = H_k - \frac{(H_k \boldsymbol{y}_k \boldsymbol{s}_k) + (H_k \boldsymbol{y}_k \boldsymbol{s}_k)^\top}{\boldsymbol{s}_k^\top \boldsymbol{y}_k} + \left(1 + \frac{\boldsymbol{y}_k^\top H_k \boldsymbol{y}_k}{\boldsymbol{s}_k^\top \boldsymbol{y}_k}\right)\frac{\boldsymbol{s}_k \boldsymbol{s}_k^\top}{\boldsymbol{s}_k^\top \boldsymbol{y}_k} \tag{8}$$

が得られる．この公式は，BFGS 公式の H 公式とよばれている [*2]．

アルゴリズム 2 （BFGS 公式の H 公式を用いた準ニュートン法）

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

1: $H_0 \leftarrow I$.
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: 　$\|\nabla f(\boldsymbol{x}_k)\| < \epsilon$ ならば，$\boldsymbol{x}_k$ を解として出力し終了する．
4: 　$\boldsymbol{d}_k = -H_k \nabla f(\boldsymbol{x}_k)$.
5: 　直線探索によりステップ幅 $\alpha_k > 0$ を定める．
6: 　$\boldsymbol{x}_{k+1} \leftarrow \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$.
7: 　(8) を用いて H_{k+1} を生成する．
8: **end for**

アルゴリズム 2 の 4 行目では，探索方向 $\boldsymbol{d}_k$ は行列とベクトルの積として求められている．つまり，H 公式を用いると，(3) のような連立一次方程式を解く必要がない．

（以上）

[*2] これに対して，(4) は B 公式とよばれることがある．