最適化手法 講義資料（平成 24 年冬学期）

準 Newton 法の概要

連続微分可能な関数 $f : \mathbb{R}^n \to \mathbb{R}$ の無制約最小化問題を考える．簡単のために，記号

$$f_k \equiv f(\boldsymbol{x}_k), \quad \nabla f_k \equiv \nabla f(\boldsymbol{x}_k), \quad \nabla^2 f_k \equiv \nabla^2 f(\boldsymbol{x}_k)$$

を用いる．

Newton 法では，Newton 方程式

$$\nabla^2 f_k \boldsymbol{d} = -\nabla f_k \tag{1}$$

の解として探索方向 $\boldsymbol{d}_k$ を定義する．一方，準 Newton 法では，(1) の $\nabla^2 f_k$ を適当な正定値行列 B_k で近似する．つまり，準 Newton 法における探索方向 $\boldsymbol{d}_k$ は，線形方程式

$$B_k \boldsymbol{d} = -\nabla f_k \tag{2}$$

の解として定められる．準 Newton 法の概要は，次の通りである．

アルゴリズム 1 (準 Newton 法のプロトタイプ)**.**

Step 0: 初期点 $\boldsymbol{x}_0$ および $B_0 \succ O$ を選ぶ．十分小さい $\epsilon > 0$ を定め，$k = 0$ とおく．

Step 1: 停止条件 $\|\nabla f_k\| < \epsilon$ が満たされていれば，$\boldsymbol{x}_k$ を解として終了．

Step 2: 線形方程式 (2) を解いて探索方向 $\boldsymbol{d}_k$ を求める．

Step 3: 直線探索により α_k を定める（Wolfe の条件などを用いる）．

Step 4: $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$ と更新する．

Step 5: $B_{k+1} \succ O$ を生成する（後述）．$k \leftarrow k + 1$ とおいて Step 1 へ．

アルゴリズム 1 の Step 5 では，$B_k, \boldsymbol{x}_k, \boldsymbol{x}_{k+1}, \nabla f_k, \nabla f_{k+1}$ を用いて $\nabla^2 f_{k+1}$ の近似行列 B_{k+1} を生成する．簡単のために

$$\boldsymbol{s}_k \equiv \boldsymbol{x}_{k+1} - \boldsymbol{x}_k = \alpha_k \boldsymbol{d}_k, \quad \boldsymbol{y}_k \equiv \nabla f_{k+1} - \nabla f_k, \quad \rho_k \equiv \frac{1}{\boldsymbol{y}_k^{\mathrm{T}} \boldsymbol{s}_k}$$

とおく．B_{k+1} を生成する方法として，次の二つの公式がよく知られている：

1. BFGS 公式

$$B_{k+1} = B_k - \frac{B_k \boldsymbol{s}_k \boldsymbol{s}_k^{\mathrm{T}} B_k}{\boldsymbol{s}_k^{\mathrm{T}} B_k \boldsymbol{s}_k} + \rho_k \boldsymbol{y}_k \boldsymbol{y}_k^{\mathrm{T}} \tag{3}$$

2. DFP 公式

$$B_{k+1} = (I - \rho_k \boldsymbol{y}_k \boldsymbol{s}_k^{\mathrm{T}}) B_k (I - \rho_k \boldsymbol{s}_k \boldsymbol{y}_k^{\mathrm{T}}) + \rho_k \boldsymbol{y}_k \boldsymbol{y}_k^{\mathrm{T}} \tag{4}$$

(3) や (4) で定められる B_{k+1} は，次の性質を満たすことが確認できる．

(a) B_{k+1} はセカント条件 $B_{k+1}\boldsymbol{s}_k = \boldsymbol{y}_k$ を満たす．

(b) B_k が対称行列ならば B_{k+1} も対称行列である．

(c) $B_k \succ O$ かつ $\boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{y}_k > 0$ ならば $B_{k+1} \succ O$ である．

このうち，性質 (c) の仮定に注目する．実は，Step 3 において α_k を Wolfe の条件を満たすように選ぶと，条件 $\boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{y}_k > 0$ は自動的に満たされる．というのも，まず，$\boldsymbol{s}_k$ の定義および Wolfe の条件 $\nabla f(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k)^{\mathrm{T}}\boldsymbol{d}_k \geq c_2 \nabla f_k^{\mathrm{T}} \boldsymbol{d}_k$ より $\nabla f_{k+1}^{\mathrm{T}}\boldsymbol{s}_k \geq c_2 \nabla f_k^{\mathrm{T}}\boldsymbol{s}_k$ が得られる．このことと $\boldsymbol{y}_k$ の定義より

$$\boldsymbol{y}_k^{\mathrm{T}}\boldsymbol{s}_k = (\nabla f_{k+1} - \nabla f_k)^{\mathrm{T}}\boldsymbol{s}_k \geq (c_2 - 1)\nabla f_k^{\mathrm{T}}\boldsymbol{s}_k = (c_2-1)\alpha_k \nabla f_k^{\mathrm{T}}\boldsymbol{d}_k \tag{5}$$

が得られる．$B_k \succ O$ ならば，$\boldsymbol{d}_k$ は降下方向である（つまり，$\nabla f_k^{\mathrm{T}}\boldsymbol{d}_k < 0$ を満たす）．このことと $c_2 < 1$ より，(5) の最右辺は正である．従って，$\boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{y}_k > 0$ が成り立つことが分かる．

B_k の更新公式 (3) や (4) に対して Sherman–Morrison–Woodbury の公式 [1]を適用すると，B_{k+1} の逆行列 $H_{k+1} \equiv B_{k+1}^{-1}$ を陽に求めることができる．このようにして得られる H_k の更新公式を，H 公式と呼ぶ．これに対して，(3) および (4) は B 公式と呼ばれることもある．

1. BFGS 公式の H 公式

$$H_{k+1} = (I - \rho_k \boldsymbol{s}_k \boldsymbol{y}_k^{\mathrm{T}}) H_k (I - \rho_k \boldsymbol{y}_k \boldsymbol{s}_k^{\mathrm{T}}) + \rho_k \boldsymbol{s}_k \boldsymbol{s}_k^{\mathrm{T}} \tag{6}$$

2. DFP 公式の H 公式

$$H_{k+1} = H_k - \frac{H_k \boldsymbol{y}_k \boldsymbol{y}_k^{\mathrm{T}} H_k}{\boldsymbol{y}_k H_k \boldsymbol{y}_k} + \rho_k \boldsymbol{s}_k \boldsymbol{s}_k^{\mathrm{T}} \tag{7}$$

アルゴリズム 2 (BFGS 公式の H 公式を用いた準 Newton 法)**.**

Step 0: 初期点 $\boldsymbol{x}_0$ および $H_0 \succ O$ を選ぶ．十分小さい $\epsilon > 0$ を定め，$k = 0$ とおく．

Step 1: 停止条件 $\|\nabla f_k\| < \epsilon$ が満たされていれば，$\boldsymbol{x}_k$ を解として終了．

Step 2: 探索方向を $\boldsymbol{d}_k = -H_k \nabla f_k$ により求める．

Step 3: 直線探索により Wolfe の条件を満たすように α_k を定める．

Step 4: $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$ と更新する．また，$\boldsymbol{s}_k = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k$, $\boldsymbol{y}_k = \nabla f_{k+1} - \nabla f_k$, $\rho_k = 1/\boldsymbol{y}_k^{\mathrm{T}}\boldsymbol{s}_k$ とおく．

Step 5: (6) により H_{k+1} を計算する．$k \leftarrow k+1$ とおいて Step 1 へ．

アルゴリズム 2 では，Step 2 において探索方向 $\boldsymbol{d}_k$ は行列にベクトルを乗じることで求められる．従って，アルゴリズム 2 を実行するには，1 回の反復あたり $O(n^2)$ の計算で済む．つまり，探索方向を求めるために (2) のような線形方程式を解く必要がないことが，H 公式の利点である．

(November 14, 2012, 寒野)

[1]正則な行列 $A \in \mathbb{R}^{n \times n}$ およびベクトル $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$ に対して，行列 $\hat{A} = A + \boldsymbol{a}\boldsymbol{b}^{\mathrm{T}}$ を考える．このとき，$\hat{A}$ が正則ならば $\hat{A}^{-1} = A^{-1} - \dfrac{(A^{-1}\boldsymbol{a})(\boldsymbol{b}^{\mathrm{T}}A^{-1})}{1 + \boldsymbol{b}^{\mathrm{T}}A^{-1}\boldsymbol{a}}$ が成り立つ．