日本オペレーションズ・リサーチ学会
2004年春季研究発表会

無制約最小化問題に対する記憶勾配法の大域的収束性について

02402100　東京理科大学　＊成島　康史　NARUSHIMA Yasushi
01702330　東京理科大学　　矢部　博　　YABE Hiroshi

1. はじめに

以下の無制約最小化問題

$$\min f(x), \quad x \in \mathbb{R}^n$$

を解くための解法を考える．ただし目的関数 $f : \mathbb{R}^n \to \mathbb{R}$ は十分に滑らかであるとする．

上のような無制約最小化問題の有効な解法として準ニュートン法がよく知られている．中でも大規模問題を解くための準ニュートン法として行列を記憶せず過去数本分の探索方向を使用する記憶制限準ニュートン法が近年注目を浴びている．

一方，最急降下法を過去の探索方向を使用して加速する解法を記憶勾配法と呼び Miele ら [1] によって最初のアイディアが提案されている (共役勾配法も記憶勾配法の一種とみなすことができる)．記憶勾配法も記憶制限準ニュートン法と同様に行列を用いず方向を更新するため，大規模問題に適しているといえる．本稿では自動的に降下方向を生成する記憶勾配法を提案し，その大域的収束性について議論する．以下では簡単のために，$g(x) = \nabla f(x)$ とし，$f_k = f(x_k)$，$g_k = g(x_k)$ と表すことにする．

2. 記憶勾配法

無制約最小化問題に対する反復法として，以下のような方法を考える

$$x_{k+1} = x_k + \alpha_k d_k \tag{1}$$

ただし，$\alpha_k > 0$ をステップ幅，d_k を探索方向とし，通常は初期探索方向として初期サイジングを行った最急降下方向，つまり $d_0 = -\gamma_0 g_0$ とする．ただし $\gamma_0 > 0$ はサイジングパラメータである．また，$k \geq 1$ での探索方向は最急降下方向と過去 m 本分の探索方向を利用して

$$d_k = -\gamma_k g_k + \frac{1}{m}\sum_{i=1}^{m} \beta_{ki} d_{k-i} \tag{2}$$

と定義する．ただし，$\beta_{ki} \in \mathbb{R}$ $(i = 1, \ldots, m)$ は適当なパラメータであり，$\gamma_k > 0$ はサイジングパラメータである．

ここで収束性を論じるために，Wolfe 条件と呼ばれるステップ幅 α_k についての条件を与える．

$$\begin{aligned} f(x_k) - f(x_k + \alpha_k d_k) &\geq -\delta \alpha_k g_k^T d_k, & (3)\\ g(x_k + \alpha_k d_k)^T d_k &\geq \sigma g_k^T d_k & (4) \end{aligned}$$

ただし $0 < \delta < \sigma < 1$ とする．

3. 降下方向を生成する記憶勾配法

$g_{k-1}^T d_{k-1} < 0$ を仮定し，次の探索方向が降下方向となるための条件，すなわち

$$g_k^T d_k = -\gamma_k \|g_k\|^2 + \frac{1}{m}\sum_{i=1}^{m} \beta_{ki} g_k^T d_{k-i} < 0 \tag{5}$$

を満たすような d_k を構築する．そのために ψ_{ki} と β_{ki} $(i = 1, \ldots, m)$ を以下のように定義する;

$$\begin{aligned} \psi_{k1} &> \max\{\frac{g_k^T d_{k-1}}{\gamma_k}, 0\}; \quad i = 1 & (6)\\ \psi_{ki} &\geq \max\{\frac{g_k^T d_{k-i}}{\gamma_k}, 0\}; \quad i = 2, \ldots, m & (7)\\ \beta_{ki} &= \|g_k\|^2 \psi_{ki}^{\dagger}. & (8) \end{aligned}$$

ただし $a^{\dagger}$ は

$$a^{\dagger} = \begin{cases} 0 & \text{if } a = 0 \\ \frac{1}{a} & \text{otherwise} \end{cases}$$

により定義されるものとする．ここで，定義より $a^{\dagger} a \leq 1$，$\beta_{k1} > 0$ と $\beta_{ki} \geq 0$ $(i = 2, \ldots, m)$ が成り立つことに注意しよう．このとき (5)，(6)，(7)，(8) より

$$\begin{aligned} g_k^T d_k &= \frac{1}{m}\sum_{i=1}^{m}(-\gamma_k \|g_k\|^2 + \beta_{ki} g_k^T d_{k-i}) \\ &< \frac{1}{m}\sum_{i=1}^{m}(-\gamma_k \|g_k\|^2 + \gamma_k \|g_k\|^2 \psi_{ki}^{\dagger} \psi_{ki}) \\ &\leq 0 & (9) \end{aligned}$$

となり，d_k は降下方向になることがわかる．よって，すべての k において (6)，(7) が満たされるならば (2)，(8) を用いた記憶勾配法は常に降下方向を生成する．

上記で提案した記憶勾配法のアルゴリズムをまとめれば，以下のとおりである．

Algorithm 3.1

Step 0. 初期点 $x_0 \in \mathbf{R}^n$ と初期探索方向 $d_0 = -\gamma_0 g_0$, $\gamma_0 > 0$ を与え，$k := 0$ とおく．

Step 1. 直線探索により (3)，(4) を満たすステップ幅 α_k を計算して，(1) により x_{k+1} を求める．

Step 2. 停止条件を満たしているならば終了する．満たしていなければ，*Step 3* へ進む．

Step 3. (6)，(7)，(8) を用いて β_{ki} を計算し，(2) により d_{k+1} を求める．

Step 4. $k := k+1$ とおき，*Step 1* に戻る．

4. 大域的収束性

大域的収束性を示すために，次のような仮定をする．

Assumption 4.1

1. 目的関数 f は $\mathbf{R}^n$ において下に有界，かつ準位集合

$$\mathcal{L} = \{x \in \mathbf{R}^n \mid f(x) \leq f(x_0)\}$$

の近傍 $\mathcal{N}$ において連続微分可能．

2. 勾配ベクトル g は $\mathcal{N}$ において Lipschitz 連続すなわち，任意の $x, y \in \mathcal{N}$ に対し

$$\|g(x) - g(y)\| \leq L\|x - y\| \tag{10}$$

を満たす定数 $L > 0$ が存在する．

この仮定の下で，一般的に以下の Lemma が成立する．

Lemma 4.1 Assumption 4.1 が成立しているとする．d_k は降下方向，α_k が Wolfe 条件 (3)，(4) を満たすとき，(1) によって更新される方法は

$$\sum_{k\geq 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty \tag{11}$$

を満たす．(これを Zoutendijk 条件という)

この Lemma を利用することにより，本稿の目的である大域的収束性に関する定理が得られる．

Theorem 4.1 Assumption 4.1 が成立しているとし，初期点を x_0 とする．$\{x_k\}$ を Algorithm 3.1 によって生成される点列とする．このとき ψ_{ki} を $g_k^T d_{k-1} + \|g_k\|\|d_{k-1}\| < \gamma_k \psi_{k1} (i=1)$ かつ $g_k^T d_{k-i} + \|g_k\|\|d_{k-i}\| \leq \gamma_k \psi_{ki} (i = 2, \ldots, m)$ を満たすように選ぶと Algorithm 3.1 による解法は大域的収束する．つまり

$$\liminf_{k\to\infty} \|g_k\| = 0$$

が成立する．

略証．もしアルゴリズムが有限回で停止しないと仮定するとすべての k に対し $\|g_k\| > 0$ が成立する．また (7) が満たされているので，すべての k に対し $g_k^T d_k < 0$ が成立する．一方 (2)，(8) から

$$|g_k^T d_k| > \frac{1}{m} \sum_{i=2}^{m} (\gamma_k \psi_{ki} - g_k^T d_{k-i}) \beta_{ki} \geq 0 \tag{12}$$

が成立する．さらに (2) と (12) から

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} \leq \left(\frac{\frac{1}{m}\sum_{i=1}^{m} \beta_{ki}\|d_{k-i}\|}{\frac{1}{m}\sum_{i=1}^{m}(\gamma_k\psi_{ki} - g_k^T d_{k-i})\beta_{ki}} \right)^2 + \frac{1}{\|g_k\|^2} \tag{13}$$

が得られる．

一方 $\gamma_k \psi_{ki} \geq g_k^T d_{k-i} + \|g_k\|\|d_{k-i}\|$ から

$$\frac{\sum_{i=1}^{m} \beta_{ki}\|d_{k-i}\|}{\sum_{i=1}^{m} \beta_{ki}(\gamma_k\psi_{ki} - g_k^T d_{k-i})} \leq \frac{1}{\|g_k\|} \tag{14}$$

が得られるので (13) と (14) から，すべての k に対し

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} \leq \frac{2}{\|g_k\|^2} \tag{15}$$

が成立する．ここでもし収束しないと仮定すると，定数 $c_1 > 0$ が存在し，すべての k に対し

$$\|g_k\| \geq c_1 \tag{16}$$

となる．よって (15) と (16) より

$$\sum_{k\geq 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} = \infty.$$

これは Lemma 4.1 に矛盾する．よって定理は証明された．□

References

[1] A. Miele and J. W. Cantrell, Study on a memory gradient method for the minimization of function, Journal of Optimization Theory and Applications, Vol.3, No.6 (1969) pp. 459–470.

[2] J. Nocedal and S. J. Wright, Numerical Optimization, Springer Series in Operations Research, Springer Verlage , NewYork, 1999.