数理解析研究所講究録
第 1548 巻 2007 年 107-114

無制約最小化問題に対する拡張Barzilai-Borwein法について

東京理科大学大学院・理学研究科 若松 峻彦 (Wakamatsu Takahiko)
東京理科大学大学院・理学研究科 成島 康史 (Narushima Yasushi)
Graduate School of Science,
Tokyo University of Science
東京理科大学・数理情報科学科 矢部 博 (Yabe Hiroshi)
Department of Mathematical Information Science,
Tokyo University of Science

1 はじめに

以下の無制約最小化問題を取り扱う.

$$\min \quad f(x)=\frac{1}{2}x^TAx-b^Tx, \tag{1}$$

但し, $A\in R^{n\times n}$ は正定値対称行列, $b\in R^n$ とする. このとき, 問題(1) に対する最急降下方向を用いた勾配法を考える. その反復式は以下のようなものである.

$$x_{k+1}=x_k-\frac{1}{\alpha_k}g_k, \tag{2}$$

但し, g_k を x_k における f の勾配ベクトル, $\frac{1}{\alpha_k}$ をステップ幅とする. 最急降下法では, 正確な直線探索を行った α_k が採用される. このときの α_k は

$$\alpha_k=\frac{g_k^TAg_k}{g_k^Tg_k} \tag{3}$$

で表わされる. 最急降下法は最適化においてたびたび用いられる方法であるが, ヘッセ行列 A の条件数が大きいときには収束が遅いことが知られている. これを改良するために, いくつかの α_k の選び方が研究されている. 準ニュートン法において, 近似行列に課される条件としてセカント条件

$$Bs_{k-1}=y_{k-1}$$

が知られている. 但し B はヘッセ行列 $\nabla^2f(x_k)$ の近似行列とし, $s_{k-1}=x_k-x_{k-1}$, $y_{k-1}=g_k-g_{k-1}$ とする. Barzilai and Borwein [1] は α_k を決める際に, セカント条件の B を αI に置き換え, 以下のような最小化問題を考えた.

$$\alpha_k=\arg\min_{\alpha}\|\alpha Is_{k-1}-y_{k-1}\|_2.$$

この最小化問題の解は,

$$\alpha_k=\frac{s_{k-1}^Ty_{k-1}}{s_{k-1}^Ts_{k-1}} \tag{4}$$

となる．この α_k を用いた勾配法は, Barzilai-Borwein 法と呼ばれるもので, 実用上では最急降下法に比べて少ない反復回数で収束することが知られている．Barzilai-Borwein 法の収束性については, Raydan [4] が狭義凸 2 次関数に対する大域的収束性を示し, Dai and Liao [2] が R–1 次収束性を示している．さらに近年, Friedlander et al. [3] は, Barzilai-Borwein 法を含む方法として遅延付き勾配法を提案した．遅延付き勾配法における α_k は次のようなものである．

$$
\begin{aligned}
\alpha_k &= \frac{g_{\nu(k)}^T A^{\rho(k)+1} g_{\nu(k)}}{g_{\nu(k)}^T A^{\rho(k)} g_{\nu(k)}}, \qquad (5)\\
\nu(k) &\in \{k, k-1, ..., \max\{0, k-m\}\},\\
\rho(k) &\in \{q_1, ..., q_m\},
\end{aligned}
$$

但し, m は正整数, $q_j (j = 1, ..., m)$ は $q_j \geq -2$ を満たす整数とする．遅延付き勾配法については, Friedlander et al. [3] が狭義凸 2 次関数に対する大域的収束性をしめし, さらに特別な仮定のもとで Q–超 1 次収束性を示している．

2 提案する解法

本論では, 次の α_k を利用する勾配法を提案する．

$$
\begin{aligned}
\alpha_k &= \sum_{i=1}^{\ell} \phi_i \frac{g_{\nu_i(k)}^T A^{\rho_i(k)+1} g_{\nu_i(k)}}{g_{\nu_i(k)}^T A^{\rho_i(k)} g_{\nu_i(k)}}, \qquad (6)\\
\phi_i &\geq 0 \ (i = 1, ..., \ell), \quad \sum_{i=1}^{\ell} \phi_i = 1,\\
\nu_i(k) &\in \{k, k-1, ..., \max\{0, k-m\}\},\\
\rho_i(k) &\in \{q_1, ..., q_m\},
\end{aligned}
$$

但し, ℓ, m は正整数, $q_j (j = 1, ..., m)$ は整数とする．これを拡張 Barzilai-Borwein 法と呼ぶ．また α_k は A のレイリー商の凸結合であり, A が正定値対称行列であることから,

$$0 < \lambda_{min} \leq \alpha_k \leq \lambda_{max} \qquad \text{for all } k$$

が成り立つ．但し, λ_{min} と λ_{max} は, それぞれ行列 A の最小固有値と最大固有値である．ここで, (2) と $g_k = Ax_k - b$ から

$$s_k = -\frac{1}{\alpha_k} g_k \qquad (7)$$

および

$$y_k = As_k \qquad (8)$$

を得る. よって (6) の α_k は, (7), (8) から,

$$\begin{aligned}\alpha_k &= \sum_{i=1}^{\ell}\phi_i \frac{s_{\nu_i(k)}^T A^{\rho_i(k)+1} s_{\nu_i(k)}}{s_{\nu_i(k)}^T A^{\rho_i(k)} s_{\nu_i(k)}} \\ &= \sum_{i=1}^{\ell}\phi_i \frac{y_{\nu_i(k)}^T A^{\rho_i(k)-1} y_{\nu_i(k)}}{y_{\nu_i(k)}^T A^{\rho_i(k)-2} y_{\nu_i(k)}}\end{aligned} \tag{9}$$

と書き換えることができる. このとき, (6) において, $\ell = 1$, $\nu_1(k) = k$, $\rho_1(k) = 0$ というパラメータを選んだときは, 最急降下法 (3) となり, (9) において, $\ell = 1$, $\nu_1(k) = \max\{0, k-1\}$, $\rho_1(k) = 0$ というパラメータを選んだときは, Barzilai-Borwein 法 (4) となり, (6) において, $\ell = 1$, $q_j \geq -2$ というパラメータを選んだときは, 遅延付き勾配法 (5) となる. 提案した方法のアルゴリズムは次のようなものである.

アルゴリズム (拡張 Barzilai-Borwein 法)

Step 0. 初期点 x_0 を与え, $k = 0$ とする. 停止条件を満たしていれば終了し, さもなければ Step 1 へ進む.

Step 1. パラメータ α_k を計算する.

Step 2. $x_{k+1} = x_k - \dfrac{1}{\alpha_k} g_k$ により, 点 x_k を更新する.

Step 3. 停止条件を満たしているならば終了する.

Step 4. $k \leftarrow k+1$ とし, Step 1 へ進む.

3 大域的収束性

提案した拡張 Barzilai-Borwein 法の大域的収束性を示す. x_* を最小化問題 (1) の唯一の最小解, $\{x_k\}$ を拡張 Barzilai-Borwein 法によって生成される点列, $e_k = x_* - x_k$ を誤差とする. このとき $b = Ax_*$ より,

$$\begin{aligned}g_k &= Ax_k - b \\ &= Ax_k - Ax_* \\ &= -Ae_k\end{aligned}$$

が成立するので, e_{k+1} と e_k の関係式は,

$$\begin{aligned}e_{k+1} &= x_* - x_{k+1} \\ &= e_k + \frac{1}{\alpha_k} g_k \\ &= \frac{1}{\alpha_k}(\alpha_k I - A) e_k\end{aligned} \tag{10}$$

となる. ここで, 行列 A の固有値を $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ とし, 対応する正規直交化した固有ベクトルを $v_1, v_2, \ldots, v_n$ とする. このとき, 初期誤差 e_0 はある定数 $d_1^0, d_2^0, \ldots, d_n^0$ を用いて

$$e_0 = \sum_{i=1}^{n} d_i^0 v_i$$

と表すことができ, (10) より,

$$\begin{aligned}
e_{k+1} &= \prod_{j=0}^{k} \frac{1}{\alpha_j}(\alpha_j I - A)e_0 \\
&= \left\{\prod_{j=0}^{k} \frac{1}{\alpha_j}(\alpha_j I - A)\right\}\left(\sum_{i=1}^{n} d_i^0 v_i\right) \\
&= \sum_{i=1}^{n} d_i^0 \left\{\prod_{j=0}^{k} \frac{1}{\alpha_j}(\alpha_j I - A)\right\} v_i \\
&= \sum_{i=1}^{n} d_i^0 \left\{\prod_{j=0}^{k} \frac{1}{\alpha_j}(\alpha_j - \lambda_i)\right\} v_i.
\end{aligned}$$

を得る. このことから e_{k+1} と d_i^{k+1} はそれぞれ

$$e_{k+1} = \sum_{i=1}^{n} d_i^{k+1} v_i$$

および

$$d_i^{k+1} = d_i^0 \prod_{j=0}^{k} \left(\frac{\alpha_j - \lambda_i}{\alpha_j}\right)$$

と表すことができる. これらを利用すれば, 以下の補題が成立する.

補題 1
数列 $\{d_1^k\}$ は, 収束因子 $1 - \lambda_1/\lambda_n$ で 0 に 1 次収束する.

補題 2
ある整数 p $(2 \leq p \leq n)$ に対して, 数列 $\{d_1^k\}, \{d_2^k\}, \ldots, \{d_{p-1}^k\}$ がすべて 0 に収束するならば,

$$\liminf_{k\to\infty} |d_p^k| = 0$$

が成立する.

これらの補題から以下の収束定理が導かれる.

定理 1
最小化問題 (1) に対して, $\{x_k\}$ を拡張 Barzilai-Borwein 法によって生成される点列, x_* を唯一の最小解とする. このとき, 点列 $\{x_k\}$ は x_* に大域的収束する.

4 数値実験

この節では, 数値実験結果を報告する. テスト問題として扱ったのは以下の狭義凸2次関数である.

$$f(x) = \frac{1}{2}x^T Diag(\lambda_1, ..., \lambda_n)x, \quad x \in R^n$$

但し, 次元 n, ヘッセ行列の条件数 $cond$ はそれぞれ $10^2, 10^3, 10^4$ として, ヘッセ行列の固有値 $\lambda_i (i = 1, ..., n)$ は $\lambda_1 = 1$, $\lambda_n = cond$ とし, $\lambda_1 < \lambda_i < \lambda_n (i = 2, .., n-1)$ とする. 今回は $\lambda_i (i = 2, ..., n-1)$ をランダムに発生させ, 10 回実験した. 各数表には, その反復回数の平均値を示した. 太字になっている数値は, その項目において反復回数が一番少なかったものである. また, 初期点は x_0 は $(x_1, ..., x_n)^T = (1, ..., 1)^T$

表 1: パラメータ $\nu_i(k)$ を動かした場合の反復回数

ℓ	1	1	1	1	1	2	2	2	2
ϕ_1	1	1	1	1	1	0.5	0.5	0.5	0.5
ϕ_2	—	—	—	—	—	0.5	0.5	0.5	0.5
$\nu_1(k)$	$k-1$	$k-2$	$k-3$	$k-4$	$k-5$	$k-1$	$k-2$	$k-3$	$k-4$
$\nu_2(k)$	—	—	—	—	—	$k-2$	$k-3$	$k-4$	$k-5$
$\rho_1(k)$	0	0	0	0	0	0	0	0	0
$\rho_2(k)$	—	—	—	—	—	0	0	0	0
$n = 10^2$, $cond = 10^2$	138.3	128.1	132.2	138.1	143.4	157.2	130.9	**127.2**	144.6
$n = 10^2$, $cond = 10^3$	480.8	408.9	330.0	333.0	350.5	616.4	470.0	390.5	**324.7**
$n = 10^2$, $cond = 10^4$	1386.5	1322.7	947.5	746.6	761.0	2254.5	1660.2	1016.8	**739.2**
$n = 10^3$, $cond = 10^2$	141.6	141.4	140.8	144.5	151.2	155.4	**137.9**	150.1	149.7
$n = 10^3$, $cond = 10^3$	488.0	448.2	459.6	**434.5**	478.3	658.8	494.4	448.2	469.8
$n = 10^3$, $cond = 10^4$	1647.6	1567.5	1386.5	1337.8	**1299.1**	2249.2	1732.3	1571.3	1388.4
$n = 10^4$, $cond = 10^2$	153.4	**141.5**	153.6	148.7	152.1	165.3	152.1	156.1	151.7
$n = 10^4$, $cond = 10^3$	496.5	482.4	474.1	489.4	491.1	706.0	553.3	**466.2**	501.6
$n = 10^4$, $cond = 10^4$	1770.5	1586.0	1562.1	**1477.0**	1520.2	2448.5	1683.6	1626.6	1519.3

として, 停止条件は $||g_k||_2 \leq 10^{-8}$ とした. このとき, Barzilai-Borwein 法, 遅延付

き勾配法, 拡張 Barzilai-Borwein 法の反復回数を比較する. はじめに表 1 では, パラメータ $\rho_i(k)$ に関しては $\rho_i(k) = 0$ で固定し, 他のパラメータに関しては $\ell = 1$ の場合, $\nu_1(k) = k-1, k-2, k-3, k-4, k-5$ として, $\ell = 2$ の場合, $\phi_i = 0.5$, $\nu_1(k) = k-1, k-2, k-3, k-4$, $\nu_2(k) = k-2, k-3, k-4, k-5$ として実験を行った. この結果からわかるように, $\ell = 2$ とした拡張 Barzilai-Borwein 法は, Barzilai-Borwein 法や遅延付き勾配法と同等の反復回数で収束している. 次に表 2 では, パラメータ $\nu_i(k)$ に関しては $\nu_i(k) = k-1$ で固定し, 他のパラメータに関しては $\ell = 1$ の場合, $\rho_1(k) = 0, 1, 2, 3$ として, $\ell = 2$ の場合, $\phi_i = 0.5$, $\rho_1(k) = 0, 1, 2, 3$, $\nu_2(k) = 1, 2, 3, 4$ として実験を行った. この結果から $\ell = 2$ とした拡張 Barzilai-Borwein 法は, Barzilai-

表 2: パラメータ $\rho_i(k)$ を動かした場合の反復回数

ℓ	1	1	1	1	1	2	2	2	2
ϕ_1	1	1	1	1	1	0.5	0.5	0.5	0.5
ϕ_2	—	—	—	—	—	0.5	0.5	0.5	0.5
$\nu_1(k)$	$k-1$	$k-1$	$k-1$	$k-1$	$k-1$	$k-1$	$k-1$	$k-1$	$k-1$
$\nu_2(k)$	—	—	—	—	—	$k-1$	$k-1$	$k-1$	$k-1$
$\rho_1(k)$	0	1	2	3	4	0	1	2	3
$\rho_2(k)$	—	—	—	—	—	1	2	3	4
$n = 10^2$ $cond = 10^2$	**129.6**	136.8	141.2	142.3	140.6	163.6	165.2	169.6	183.8
$n = 10^2$ $cond = 10^3$	**393.1**	475.1	446.5	482.7	469.2	672.0	678.3	715.6	722.8
$n = 10^2$ $cond = 10^4$	**1314.8**	1512.2	1348.5	1395.9	1460.8	2514.7	2404.7	2487.9	2678.5
$n = 10^3$ $cond = 10^2$	144.5	**144.3**	145.9	151.8	150.8	172.7	179.2	190.4	179.8
$n = 10^3$ $cond = 10^3$	**433.3**	504.8	464.9	487.2	496.8	671.8	630.1	729.6	711.8
$n = 10^3$ $cond = 10^4$	**1472.1**	1683.8	1725.0	1541.5	1610.4	2525.9	2517.4	2689.9	2690.7
$n = 10^4$ $cond = 10^2$	147.5	**146.4**	147.3	153.4	155.8	186.3	178.0	194.6	194.8
$n = 10^4$ $cond = 10^3$	**496.2**	513.1	502.5	506.4	563.8	711.1	734.8	718.9	689.0
$n = 10^4$ $cond = 10^4$	1777.4	1797.1	1753.9	1899.0	**1733.1**	2597.9	2697.7	2765.5	2798.7

Borwein 法や遅延付き勾配法より反復回数が増加している. よってパラメータ $\rho_i(k)$ は, なるべく変動させないで固定した方が良いことがわかる. 最後に表 3 では, パラ

メータ $\nu_i(k)$, $\rho_i(k)$ に関しては $\nu_1(k)=k-4$, $\nu_2(k)=k-5$, $\rho_i(k)=1$ で固定し, 他のパラメータに関しては $\ell=1$ の場合, $\phi_1=1$ として, $\ell=2$ の場合, $\phi_1=0.75, 0.5, 0.25$, $\phi_2=0.25, 0.5, 0.75$ として実験を行った. この結果から $\ell=2$ とした拡張 Barzilai-

表 3: パラメータ ϕ_i を動かした場合の反復回数

ℓ	1	2	2	2	1
ϕ_1	1	0.75	0.5	0.25	1
ϕ_2	—	0.25	0.5	0.75	—
$\nu_1(k)$	$k-4$	$k-4$	$k-4$	$k-4$	$k-5$
$\nu_2(k)$	—	$k-5$	$k-5$	$k-5$	—
$\rho_1(k)$	1	1	1	1	1
$\rho_2(k)$	—	1	1	1	—
$n=10^2$ $cond=10^2$	139.6	136.4	**135.5**	148.7	149.4
$n=10^2$ $cond=10^3$	**311.8**	364.8	348.4	370.0	320.5
$n=10^2$ $cond=10^4$	712.7	654.2	874.6	769.2	**608.4**
$n=10^3$ $cond=10^2$	146.7	**143.4**	150.0	144.8	145.6
$n=10^3$ $cond=10^3$	455.3	470.2	475.1	**449.3**	470.4
$n=10^3$ $cond=10^4$	1304.9	1305.0	1327.8	1357.4	**1187.5**
$n=10^4$ $cond=10^2$	150.8	150.8	152.4	**147.7**	154.5
$n=10^4$ $cond=10^3$	495.6	509.8	499.8	**481.0**	488.5
$n=10^4$ $cond=10^4$	**1456.0**	1471.9	1611.6	1516.1	1493.7

Borwein 法は, パラメータ ϕ_i の選び方によって, Barzilai-Borwein 法や遅延付き勾配法より反復回数を減少させることが出来ることがわかる. しかし, 今回の実験では顕著な傾向は存在せず, どのようなパラメータを選べば反復回数が減少するかは今のところ明らかではない.

5 おわりに

本論文では, 拡張 Barzilai-Borwein 法を提案し, その大域的収束性を示した. また, 数値実験により有効性を検証した. 今後の課題は, 提案した方法に含まれるパラメータの選び方, Q–超 1 次収束性の証明, および一般関数への適用などがあげられる.

参考文献

[1] J. Barzilai and J. M. Borwein, Two point step size gradient methods, *IMA J. Numer. Anal.*, 8 (1988), pp. 141-148.

[2] Y. H. Dai and L. Z. Liao, R-linear convergence of the Barzilai and Borwein gradient method, *IMA J. Numer. Anal.*, 22 (2002), pp. 1-10.

[3] A. Friedlander, J. M. Martinez, B. Molina, and M. Raydan, Gradient method with retards and generalizations, *SIAM J. Numer. Anal.*, 36 (1999), pp. 275-289.

[4] M. Raydan, On the Barzilai and Borwein choice of steplength for the gradient method, *IMA J. Numer. Anal.*, 13 (1993), pp. 321-326.