数理解析研究所講究録 1349 巻 2004 年 142-149

Global Convergence of Quasi-Newton Methods Based on Modified Secant Conditions

東京理科大学・数理情報科学科　小笠原 英穂 (Hideho Ogasawara)
東京理科大学・数理情報科学科　矢部 博 (Hiroshi Yabe)
Department of Mathematical Information Science,
Tokyo University of Science
東京理科大学大学院・理学研究科　柏 徹 (Toru Kashiwa)
Graduate School of Science,
Tokyo University of Science

1 はじめに

本稿では, 無制約最小化問題

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

を解くための準ニュートン法を考える. ただし目的関数 $f: \mathbf{R}^n \to \mathbf{R}$ は十分に滑らかであると仮定する.

準ニュートン法は, 初期点 x_1 と初期行列 B_1 を与え, 近似解の点列 $\{x_k\}$ を反復式

$$x_{k+1} = x_k - \alpha_k B_k^{-1} \nabla f(x_k), \qquad k = 1, 2, \ldots$$

によって生成する解法である. ここで α_k はステップ幅, B_k は f のヘッセ行列 $\nabla^2 f(x_k)$ の近似行列である. 特に, $B_k \equiv I$ にとったとき最急降下法, また $B_k \equiv \nabla^2 f(x_k)$ にとったときニュートン法に相当するが, 通常は何らかの更新公式によって更新される. その際, ヘッセ行列を近似するために, B_{k+1} の満たすべき条件として, いわゆるセカント条件というものが課される. この条件はヘッセ行列の曲率情報を近似行列 B_{k+1} に取り込むための条件とみなすことができる. したがって, この条件を満たすように更新することにより, ヘッセ行列に対する B_{k+1} の近似度が上がっていくことが期待できる.

よい近似行列を構成することによって計算効率が大きく向上するので, これまでに, セカント条件を満足するさまざまな更新公式が提案されている. その代表的なものは, BFGS 公式と DFP 公式, そしてそれらを特別な場合として含む Broyden 公式族であろう. とりわけ BFGS 公式の有効性は広く認知されている.

他方, よりよい近似行列を生成するために, 通常のセカント条件自身を拡張する研究も行われている. Ford and Moghrabi [2, 3] は, セカント条件の左辺に現れるベクトル (反復

点の変位) の代わりに過去数回のステップで得られた情報に基づき, 補間多項式を構成する Multi-step 準ニュートン法を提案している. Zhang, Deng and Chen [5], Zhang and Xu [6] は, 通常のセカント条件が目的関数の勾配しか利用していない点を指摘し, 目的関数の値も利用するとともにテンソル項を考慮することによって, 行列 B_{k+1} の精度を高めることを考えた. そのような発想から修正セカント条件を提案し, それに基づいた修正 BFGS 公式, 修正 DFP 公式の局所的超 1 次収束性を示した. 吉野・矢部・小笠原 [4] は Zhang らの研究を拡張し, 修正 BFGS 公式, 修正 DFP 公式を含む修正 Broyden 公式族に対して, その局所的超 1 次収束性を示した.

本稿では修正 Broyden 公式族の大域的収束性について考える. Zhang らの修正セカント条件にパラメタを導入すれば, 適当な仮定の下で, 修正 Broyden 公式族を用いた準ニュートン法は大域的に収束することを示す.

2 修正セカント条件と修正 Broyden 公式族

以下では簡単のために, $g(x) = \nabla f(x)$ と表記する. また必要のない限り, $f_k = f(x_k)$, $g_k = g(x_k)$ などと略記することにする.

Zhang らにより提案された修正セカント条件は次の通りである:

$$B_{k+1}s_k = y_k + \frac{\theta_k}{s_k^\top u}u. \tag{1}$$

ただし, $s_k = x_{k+1} - x_k$, $y_k = g_{k+1} - g_k$, $\theta_k = 6(f_k - f_{k+1}) + 3s_k^\top(g_k + g_{k+1})$ であり, $u \in \mathbf{R}^n$ は $s_k^\top u \neq 0$ となる任意のベクトルである.

はじめ Zhang, Deng and Chen [5] は $u = s_k$ の形のみで提案していたが, 後に Zhang and Xu [6] がこのような一般形で提案し直した.

修正セカント条件 (1) は従来のセカント条件に, 任意ベクトル u を含む修正項が付加された形になっている. この項は局所的には補正項として働くであろうが, 時に大域的には補正を悪化させる攪乱項として作用してしまうことも起こり得る. すなわち, 従来のセカント条件の方が望ましい場合もある. 例えば, B_{k+1} の正定値性を保持するためには, $s_k^\top\left(y_k + \frac{\theta_k}{s_k^\top u}u\right) = s_k^\top y_k + \theta_k > 0$ である必要があるが, θ_k は負の値もとり得るので, $s_k^\top y_k + \theta_k > 0$ となる保証はない. そのような場合には従来のセカント条件で更新する方がよいだろう. しかし修正セカント条件は, そのままの形では従来のセカント条件に一致することはない.

そこで本稿では, 従来のセカント条件と修正セカント条件の間を自然に移行できるようにするために, (1) にパラメタ $\rho \geq 0$ を導入した次のセカント条件を考える:

$$B_{k+1}s_k = z_k, \qquad z_k = y_k + \frac{\rho\theta_k}{s_k^\top u}u. \tag{2}$$

ここで, $\rho = 0$ とおけば従来のセカント条件になり, $\rho = 1$ とおけば Zhang らの修正セカント条件 (1) になることに注意する.

パラメタ付修正セカント条件 (2) に基づいた次の Broyden 公式族を修正 Broyden 公式族と呼ぶことにする.

$$\begin{aligned} B_{k+1} &= B_k - \frac{B_k s_k s_k^\top B_k}{s_k^\top B_k s_k} + \frac{z_k z_k^\top}{s_k^\top z_k} + \phi_k (s_k^\top B_k s_k) v_k v_k^\top, \\ v_k &= \frac{z_k}{s_k^\top z_k} - \frac{B_k s_k}{s_k^\top B_k s_k}. \end{aligned} \tag{3}$$

ここで, ϕ_k は Broyden パラメタである. 特に $\rho = 1$ で $\phi_k = 0$, $\phi_k = 1$ とおいた場合が, それぞれ Zhang ら [5, 6] によって扱われた修正 BFGS 公式, 修正 DFP 公式である. やはり $\rho = 1$ であるが ϕ_k を一様有界に拡張した場合が, 吉野・矢部・小笠原 [4] によって扱われた修正 Broyden 公式族である.

3 アルゴリズム

前節で提案した修正 Broyden 公式族を用いた準ニュートン法のアルゴリズムは, 以下のようになる.

アルゴリズム [QN]

Step 0. 初期点 $x_1 \in \mathbf{R}^n$ と初期行列 $B_1 \in \mathbf{R}^{n\times n}$ を与える. $k := 1$ とおく.

Step 1. 連立 1 次方程式 $B_k d_k = -g_k$ を解く.

Step 2. 直線探索によりステップ幅 α_k を求め, $x_{k+1} := x_k + \alpha_k d_k$ とおく.

Step 3. 停止条件を満たしていれば終了し, 満たしていなければ Step 4 へ進む.

Step 4. B_k を修正 Broyden 公式族 (3) により更新する.

Step 5. $k := k + 1$ とおき, Step 1 へ戻る.

Step 2 の直線探索で, ステップ幅 α_k は次の Wolfe の基準を満たすように選ぶものとする:

$$f(x_k + \alpha_k d_k) \le f_k + \sigma_1 \alpha_k g_k^\top d_k, \tag{4}$$

$$g(x_k + \alpha_k d_k)^\top d_k \ge \sigma_2 g_k^\top d_k. \tag{5}$$

ここに, σ_1, σ_2 は $0 < \sigma_1 < \sigma_2 < 1$ なる定数である.

修正セカント条件を用いた準ニュートン法の特徴は, 式 (2) に含まれるベクトル u を自由に選べることにある. これによって準ニュートン法の計算効率が高められる可能性がある. ベクトル u の選び方はいろいろと考えられるが, 新たに計算するよりも既存のベクトルを利用する方が実用的であろう. Zhang, Deng and Chen [5] は $u = s_k$ の場合を, Zhang and Xu [6] は $u = g_k$, g_{k+1}, y_k の場合を数値実験しており, $u = y_k$ にとった場合が他に比べて良好な結果を与えたと報告している.

4 仮定

修正 Broyden 公式族を用いた準ニュートン法の大域的収束性を示すために, 次のような仮定を設ける.

仮定

A1. 目的関数 f は 2 回連続的微分可能である.

A2. レベル集合 $\mathcal{D} = \{x \in \mathbf{R}^n \mid f(x) \leq f(x_1)\}$ は凸集合である. また, ある正の定数 m, M が存在して,

$$m\|p\|^2 \leq p^\top \nabla^2 f(x) p \leq M\|p\|^2, \qquad \forall p \in \mathbf{R}^n, \quad \forall x \in \mathcal{D}.$$

A3. (2) のベクトル u は次式を満たす: ある定数 $\mu \in (0, 1]$ が存在して,

$$|s_k^\top u| \geq \mu\|s_k\|\|u\|, \qquad \forall k \geq 1.$$

A4. (2) のパラメタ ρ は $0 \leq \rho < \dfrac{1}{3(1 - m/M)}$ を満たす.
($m = M$ のときは $0 \leq \rho < \infty$ とみなす.)

仮定 A2 より, f は $\mathcal{D}$ 上で一様凸関数となるから, 唯一の大域的最小解 $x^* \in \mathcal{D}$ が存在する.

5 大域的収束性

最急降下方向 $-g_k$ と s_k とのなす角を ψ_k とする. すなわち,

$$\cos\psi_k \triangleq \frac{-g_k^\top s_k}{\|g_k\|\|s_k\|}.$$

大域的収束性の証明は, Byrd, Nocedal and Yuan [1] が通常のセカント条件の下で証明した方法に倣う. 証明のポイントは, $\cos\psi_k > 0$ を部分列の意味で, 正の定数によって下から押えることにある.

そのために, 仮定 A1–4 のもとで, 以下の補助定理を準備する.

補助定理 1 ある正の定数 c_1, c_2 が存在して, 次の不等式が成り立つ:

$$\text{(i)} \qquad c_1\|g_k\|\cos\psi_k \leq \|s_k\| \leq c_2\|g_k\|\cos\psi_k,$$

$$\text{(ii)} \qquad 0 < f_{k+1} - f_* \leq [1 - \sigma_1 m c_1 \cos^2\psi_k](f_k - f_*).$$

ここで $f_* = f(x^*)$ である.

2番目の不等式は数列 $\{f_{k+1} - f_*\}$ が $1 - \sigma_1 m c_1 \cos^2 \psi_k$ の比率で単調減少することを示している.

補助定理 2 ステップ幅 α_k は次の不等式を満たす:

$$c_1 \frac{s_k^\top B_k s_k}{\|s_k\|^2} \le \alpha_k \le c_2 \frac{s_k^\top B_k s_k}{\|s_k\|^2}.$$

$\frac{\|B_k s_k\|}{\|s_k\|} \le \|B_k\| < \mathrm{Tr}(B_k)$ と上の右側の不等式から, $\cos\psi_k$ が次のように下から評価されることに注意する:

$$\begin{aligned}\cos\psi_k &= \frac{\|s_k\|}{\|B_k s_k\|} \frac{s_k^\top B_k s_k}{\|s_k\|^2} \qquad (B_k s_k = -\alpha_k g_k \text{より}) \\ &\ge \frac{\alpha_k}{c_2 \mathrm{Tr}(B_k)}. \end{aligned} \tag{6}$$

したがって 直観的には α_k を下から, また $\mathrm{Tr}(B_k)$ を上から評価してやればよい. そのために, 次の評価式を用いる.

補助定理 3 ある正の定数 c_3, m_1, M_1, M_2 が存在して, 次の不等式が成り立つ:

$$\begin{aligned}\frac{\|z_k\|^2}{s_k^\top z_k} &\le M_1, \\ \frac{s_k^\top B_k s_k}{s_k^\top z_k} &\le \frac{c_3 \alpha_k}{1 - \sigma_2}, \\ \frac{\|B_k s_k\|^2}{s_k^\top B_k s_k} &\ge \frac{\alpha_k}{c_2 \cos^2 \psi_k}, \\ \frac{|z_k^\top B_k s_k|}{s_k^\top z_k} &\le \frac{M_2 \alpha_k}{c_1 m_1 \cos\psi_k}. \end{aligned}$$

これらを用いて $\mathrm{Tr}(B_{k+1})$ を評価する. まず, B_{k+1} の定義式 (3) から

$$\begin{aligned}\mathrm{Tr}(B_{k+1}) &= \mathrm{Tr}(B_k) + \frac{\|z_k\|^2}{s_k^\top z_k} - \frac{\|B_k s_k\|^2}{s_k^\top B_k s_k} + \phi_k (s_k^\top B_k s_k) \|v_k\|^2, \\ \|v_k\|^2 &= \frac{\|z_k\|^2}{(s_k^\top z_k)^2} - 2 \frac{z_k^\top B_k s_k}{(s_k^\top z_k)(s_k^\top B_k s_k)} + \frac{\|B_k s_k\|^2}{(s_k^\top B_k s_k)^2} \end{aligned}$$

となる. 次に $\|v_k\|$ を消去して整理し, 第 2 項目以降に各項の評価式を代入すると次の評価式を得る (ここで $\phi_k \in [0,1]$ と仮定する):

$$\begin{aligned}\mathrm{Tr}(B_{k+1}) &= \mathrm{Tr}(B_k) + \frac{\|z_k\|^2}{s_k^\top z_k} + \phi_k \frac{\|z_k\|^2}{s_k^\top z_k} \frac{z_k^\top B_k s_k}{s_k^\top z_k} - (1 - \phi_k) \frac{\|B_k s_k\|^2}{s_k^\top B_k s_k} - 2\phi_k \frac{z_k^\top B_k s_k}{s_k^\top z_k} \\ &\le \mathrm{Tr}(B_k) + M_1 + \eta_k \alpha_k. \end{aligned} \tag{7}$$

ただし,

$$\eta_k \triangleq \frac{\phi_k c_3 M_1}{1-\sigma_2} - \frac{1-\phi_k}{c_2 \cos^2 \psi_k} + \frac{2\phi_k M_2}{c_1 m_1 \cos \psi_k}. \tag{8}$$

ところで, (6) より $\alpha_k \le c_2 \mathrm{Tr}(B_k) \cos \psi_k \le c_2 \mathrm{Tr}(B_k)$ だから, (7) の右辺第3項目は

$$\eta_k \alpha_k \le \left(\frac{c_3 M_1}{1-\sigma_2} + \frac{2M_2}{c_1 m_1} \right) c_2 \mathrm{Tr}(B_k)$$

と押えられ, 結局 (7) は次のように評価される:

$$\mathrm{Tr}(B_{k+1}) \le M_1 + c_4 \mathrm{Tr}(B_k).$$

ただし,

$$c_4 = 1 + \frac{c_3 M_1}{1-\sigma_2} + \frac{2M_2}{c_1 m_1}.$$

これより, 最終的に, ある正の定数 c_5 が存在して, $\mathrm{Tr}(B_{k+1}) \le c_5^k$ となることが導かれる.

一方, $\det(B_{k+1})$ を評価すると,

$$\begin{aligned}
\det(B_{k+1}) &\ge \det\left(B_k - \frac{B_k s_k s_k^\top B_k}{s_k^\top B_k s_k} + \frac{z_k z_k^\top}{s_k^\top z_k} \right) \\
&= \det(B_k) \frac{s_k^\top z_k}{s_k^\top B_k s_k} \\
&\ge \det(B_k) \frac{1-\sigma_2}{c_3 \alpha_k} \qquad \text{(補助定理 3 より)} \\
&\ge \det(B_1) \prod_{j=1}^{k} \frac{1-\sigma_2}{c_3 \alpha_j}
\end{aligned}$$

を得, $\det(B_{k+1}) \le \left(\frac{\mathrm{Tr}(B_{k+1})}{n} \right)^n$ (相加・相乗平均の関係) とから, 次の補助定理を得る. この結果はステップ幅 α_k が, (幾何平均の意味で) 平均的には下に押えられる, ということを意味している.

補助定理 4 $\phi_k \in [0,1]$ とする. このとき, ある正の定数 c_6 が存在して, 次の不等式が成り立つ:

$$\prod_{j=1}^{k} \alpha_j \ge c_6^k, \qquad \forall k \ge 1.$$

以上の補助定理 1–4 を利用すれば, 修正 Broyden 公式族を用いた準ニュートン法の大域的収束性が示される.

定理 1 (大域的収束) 点列 $\{x_k\}$ はアルゴリズム [QN] によって生成されるものとする. ただし, $\phi_k \in [0,\delta]$ $(0 \le \delta < 1)$ とし, α_k は Wolfe の基準 (4)–(5) を満たすように選ぶ. このとき, 任意の正定値対称な初期行列 B_1 に対して, 点列 $\{x_k\}$ は最小解 x^* に収束する.

(証明) 概略のみ示す. $\cos\psi_k$ が真に正となるような $\{x_k\}$ の部分列が存在すれば, 数列 $\{f_k - f_*\}$ の単調減少性 (補助定理 1) から $x_k \to x^*$ を示すことができる. そこで, 背理法によって $\cos\psi_k \to 0$ と仮定し, 矛盾を導く.

$\cos\psi_k \to 0$ ならば, η_k の定義式 (8) より $\eta_j \to -\infty$ となる. すなわち, ある番号 K_0 が存在して, 次の不等式が成り立つ:

$$\eta_j < -2M_1/c_6, \qquad \forall j \ge K_0.$$

したがって (7) より, すべての $k \ge K_0$ に対して,

$$\begin{aligned} 0 < \mathrm{Tr}(B_{k+1}) &\le \mathrm{Tr}(B_{K_0}) + M_1(k+1-K_0) + \sum_{j=K_0}^{k} \eta_j\alpha_j \\ &< \mathrm{Tr}(B_{K_0}) + M_1(k+1-K_0) - \frac{2M_1}{c_6}\sum_{j=K_0}^{k} \alpha_j. \end{aligned} \tag{9}$$

ところで, 補助定理 4 の不等式に相加・相乗平均の不等式を使えば,

$$\begin{aligned} \Bigl(\sum_{j=1}^{k}\alpha_j\Bigr)/k &\ge \Bigl(\prod_{j=1}^{k}\alpha_j\Bigr)^{1/k} && \text{(相加・相乗平均)} \\ &\ge c_6 && \text{(補助定理 4 より)} \end{aligned}$$

を得るから, $\sum_{j=1}^{k}\alpha_j \ge c_6 k$ すなわち $\sum_{j=K_0}^{k}\alpha_j \ge c_6 k - \sum_{j=1}^{K_0-1}\alpha_j$. これを (9) に代入すると,

$$\begin{aligned} 0 &< \mathrm{Tr}(B_{K_0}) + M_1(k+1-K_0) - \frac{2M_1}{c_6}c_6 k + \frac{2M_1}{c_6}\sum_{j=1}^{K_0-1}\alpha_j \\ &= \mathrm{Tr}(B_{K_0}) + M_1(1-K_0-k) + \frac{2M_1}{c_6}\sum_{j=1}^{K_0-1}\alpha_j. \end{aligned}$$

ところが上式で $k \to \infty$ にとれば, 右辺 $\to -\infty$ となり, これは矛盾である. □

6 最後に

本稿では, 修正セカント条件にパラメタ ρ を導入し, それに基づく新しい修正 Broyden 公式族を考えた. またその公式族を用いた準ニュートン法を提案し, 大域的収束性を示した.

しかしいくつかの課題も残されている. $\rho = 1$ の場合がパラメタ導入前の修正 Broyden 公式族に相当するが, この値は大域的収束性のために仮定したパラメタ ρ の範囲に必ずしも

入っていない (仮定 A4). したがって, このままではパラメタのない修正 Broyden 公式族のもつ優れた局所的収束特性が, パラメタ付修正 Broyden 公式族に引き継がれない.

そこで, 1 つは大域的収束性を保証できるパラメタの範囲をもっと広げられないか検討する必要がある. もう 1 つは, アルゴリズムの上で大域的収束性と局所的収束性を整合させるための工夫を組み入れるべきであろう. 例えば解の近傍では ρ を 1 にとる (あるいは近づけていく) ということが考えられる. このようなことも含め, 数多くの数値実験による検証もまた今後の課題である.

参考文献

[1] R. Byrd, J. Nocedal and Y. Yuan, "Global convergence of a class of quasi-Newton methods on convex problems," *SIAM Journal on Numerical Analysis* **24** (1987), 1171–1190.

[2] J. A. Ford and I. A. Moghrabi, "Multi-step quasi-Newton methods for optimization," *Journal of Computational and Applied Mathematics* **50** (1994), 305–323.

[3] J. A. Ford and I. A. Moghrabi, "Using function-values multi-step quasi-Newton methods," *Journal of Computational and Applied Mathematics* **66** (1996), 201–211.

[4] 吉野 雅之, 矢部 博, 小笠原 英穂, "修正セカント条件に基づいた準ニュートン法の局所的超 1 次収束性について," 最適化：モデリングとアルゴリズム *16*, 統計数理研究所共同研究レポート **161**, pp. 9–19, 2003.

[5] J. Z. Zhang, N. Y. Deng and L. H. Chen, "New quasi-Newton equation and related methods for unconstrained optimization," *Journal of Optimization Theory and Applications* **102** (1999), 147–167.

[6] J. Zhang and C. Xu, "Properties and numerical performance of quasi-Newton methods with modified quasi-Newton equations," *Journal of Computational and Applied Mathematics* **137** (2001), 269–278.