

無制約最適化問題に対する Broyden family に基づいた 非線形共役勾配法

東京理科大学大学院 中山 舜民 (Shummin Nakayama)
Graduate school, Tokyo University of Science
横浜国立大学 成島 康史 (Yasushi Narushima)
Yokohama National University
東京理科大学 矢部 博 (Hiroshi Yabe)
Tokyo University of Science

1 はじめに

本稿では、以下の無制約最適化問題を考える:

$$\min_{x \in \mathbf{R}^n} f(x)$$

ただし、 $f: \mathbf{R}^n \rightarrow \mathbf{R}$ は連続微分可能な関数とし、その勾配ベクトルを $g(x) = \nabla f(x)$ とする。通常、この問題に対する数値解法として反復法が用いられる。反復法は任意の初期点 $x_0 \in \mathbf{R}^n$ から出発し、反復式

$$x_{k+1} = x_k + \alpha_k d_k \tag{1.1}$$

により点列を更新する。ここで、 x_k を k 回目の近似解、 $\alpha_k > 0$ をステップ幅、 $d_k \in \mathbf{R}^n$ を探索方向と呼ぶ。探索方向 d_k について様々な選び方があり、例えば最急降下法、ニュートン法、準ニュートン法、非線形共役勾配法などがよく知られている。準ニュートン法の探索方向は

$$d_k = -H_k g_k \tag{1.2}$$

で与えられ、行列 H_k は $\nabla^2 f(x_k)^{-1}$ の近似行列である。行列 H_k は各反復ごとにセカント条件

$$H_k y_{k-1} = s_{k-1}$$

を満たすように更新され、 s_{k-1} と y_{k-1} は

$$s_{k-1} = x_{k-1} - x_k, \quad y_{k-1} = g_k - g_{k-1}$$

で与えられる。行列の更新式は BFGS 公式

$$H_k = H_{k-1} - \frac{H_{k-1} y_{k-1} s_{k-1}^T + s_{k-1} y_{k-1}^T H_{k-1}}{s_{k-1}^T y_{k-1}} + \left(1 + \frac{y_{k-1}^T H_{k-1} y_{k-1}}{s_{k-1}^T y_{k-1}} \right) \frac{s_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}}$$

や DFP 公式、対称ランクワン公式などがあり、それらを一つの公式族にまとめた Broyden family

$$H_k = H_{k-1} - \frac{H_{k-1} y_{k-1} y_{k-1}^T H_{k-1}}{y_{k-1}^T H_{k-1} y_{k-1}} + \frac{s_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}} + \theta_{k-1} w_{k-1} w_{k-1}^T,$$

$$w_{k-1} = \sqrt{y_{k-1}^T H_{k-1} y_{k-1}} \left(\frac{s_{k-1}}{s_{k-1}^T y_{k-1}} - \frac{H_{k-1} y_{k-1}}{y_{k-1}^T H_{k-1} y_{k-1}} \right)$$

がある。ここで θ_{k-1} はパラメータであり、 $\theta_{k-1} = 0$ ならば DFP 公式、 $\theta_{k-1} = 1$ ならば BFGS 公式に帰着される。 $\theta_{k-1} \in [0, 1]$ の範囲ならば DFP 公式と BFGS 公式の凸結合を表しており、convex class と呼ばれる。数値実験において、convex class の中では BFGS 公式 ($\theta_{k-1} = 1$) がもっとも優れているとされている。一方、preconvex class と呼ばれる $\theta_{k-1} > 1$ とした範囲を考えたとき、BFGS 公式より優れたパフォーマンスが得られる場合があると報告されている [19]。

一般の準ニュートン法は、行列の保存や、行列の演算を行う必要があるため、大規模な問題に直接適用することが困難である。そこで、Shanno [17] は大規模な問題に対する準ニュートン法として、毎反復で H_{k-1} の代わりに単位行列 I を使うメモリーレス準ニュートン法を提案した。一方、大規模な最適化問題に対する数値解法としては、非線形共役勾配法 (Nonlinear Conjugate Gradient method, 以下 CG 法と呼ぶ) が有名である。CG 法の探索方向は

$$d_k = -g_k + \beta_k d_{k-1} \quad (1.3)$$

で与えられ、 β_k は CG 法を特徴つけるパラメータであり、様々な研究がされている [9, 14]。メモリーレス準ニュートン法は CG 法との関わりが深く、メモリーレス準ニュートン法を基にした CG 法も提案されている [4, 18]。近年、大規模な最適化問題に対する数値解法の需要が増えてきたことから、メモリーレス準ニュートン法が注目されるようになった [11, 12, 13]。

CG 法やメモリーレス準ニュートン法は十分な降下条件を満たすことが重要である。十分な降下条件とは、ある正定数 c が存在して、すべての k に対して

$$g_k^T d_k \leq -c \|g_k\|^2 \quad (1.4)$$

を満たすことである。ここで $\|\cdot\|$ は l_2 ノルムを表す。Nakayama ら [13] では、Spectral-scaling Broyden family [2] に基づいたメモリーレス準ニュートン法を提案し、十分な降下条件を満たすためのパラメータに対する条件を与え、その条件のもとで提案法の大域的収束性を示している。しかし、数値実験で最も優れたパフォーマンスを示したパラメータに対しては十分な降下条件と大域的収束性の保証ができていなかった。一方、Kou and Dai [11] では、Self-scaling BFGS 公式に基づいたメモリーレス準ニュートン法に修正を加えた解法を提案し、十分な降下条件と大域的収束性を示している。本論文では、上述した Nakayama ら [13] の方法の課題を解決するために Nakayama らの方法に Dai and Kou の修正法を取り入れて、Spectral scaling Broyden family に基づいた新しいメモリーレス準ニュートン法を提案する。さらに、その方法に基づいた CG 法を提案する。

2 メモリーレス準ニュートン法

本節では、先行研究の紹介をする。そのための準備として、本稿では以下の反復法のアルゴリズムを定義する。

アルゴリズム 1 (反復法のアルゴリズム)

Step 0: 初期点 x_0 と $0 < \delta < \sigma < 1, \varepsilon > 0$ を与え, $k = 0$ とする.

Step 1: 終了条件 $\|g_k\|_\infty < \varepsilon$ を満たすならばアルゴリズムは停止して x_k を最適解とする.

Step 2: 探索方向 d_k を与える. ただし, $k = 0$ のときは $d_0 = -g_0$ とする.

Step 3: Wolfe 条件

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

を満たすようなステップ幅 $\alpha_k > 0$ を求める.

Step 4: 点 x_{k+1} を (1.1) によって更新する.

Step 5: $k = k + 1$ として Step1 に戻る. \square

また, 以降では目的関数 f に対して以下のような標準的な仮定を設ける.

仮定 2.1 初期点 x_0 における準位集合 $\mathcal{L} = \{x | f(x) \leq f(x_0)\}$ は有界である.

仮定 2.2 目的関数 f は \mathcal{L} の開凸近傍 \mathcal{N} において連続的微分可能とし, 勾配 g は \mathcal{N} 上でリプシッツ連続である.

2.1 Memoryless quasi-Newton method based on Broyden family

Nakayama ら [13] は Spectral-scaling セカント条件 [3] に基づいた Spectral-scaling Broyden family [2]

$$H_k = H_{k-1} - \frac{H_{k-1} y_{k-1} y_{k-1}^T H_{k-1}}{y_{k-1}^T H_{k-1} y_{k-1}} + \frac{1}{\gamma_{k-1}} \frac{s_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}} + \theta_{k-1} w_{k-1} w_{k-1}^T, \quad (2.1)$$

に着目し, (2.1) に基づいたメモリーレス準ニュートン法を提案した. つまり (2.1) で $H_{k-1} = I$ とおき, (1.2) を考え, 探索方向を

$$\begin{aligned} d_k = -g_k + &\left(\theta_{k-1} \frac{y_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} - \left(\hat{\gamma}_{k-1} + \theta_{k-1} \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}} \right) \frac{s_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} \right) d_{k-1} \\ &+ \left(\theta_{k-1} \frac{d_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} + (1 - \theta_{k-1}) \frac{y_{k-1}^T g_k}{y_{k-1}^T y_{k-1}} \right) y_{k-1}, \end{aligned} \quad (2.2)$$

で与えている. ただし, γ_{k-1} はスケーリングパラメータであり, $\hat{\gamma}_{k-1} = 1/\gamma_{k-1}$ とする. 以下の命題は探索方向 (2.2) が十分な降下条件を満たすことを意味している.

命題 2.1 $d_{k-1}^T y_{k-1} > 0$ の下で, 探索方向 (2.2) のパラメータ $\hat{\gamma}_{k-1}, \theta_{k-1}$ が

$$\hat{\gamma}_{k-1} \geq \theta_{k-1} \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}} \quad (2.3)$$

と

$$0 < \theta_{\min} \leq \theta_{k-1} \leq \theta_{\max} < 2 \quad (2.4)$$

を満たすとする。ここで、 θ_{\min} と θ_{\max} は $0 < \theta_{\min} \leq 1 \leq \theta_{\max} < 2$ を満たす定数とする。このとき、探索方向 (2.2) は $c := \min \left\{ \frac{\theta_{\min}}{2}, 1 - \frac{\theta_{\max}}{2} \right\}$ とした十分な降下条件 (1.4) を満たす。

また Nakayama らは、大域的収束性を示すために探索方向 (2.2) を次のように修正した:

$$d_k = -g_k + \beta_k^{N1} d_{k-1} + \zeta_k^{N1} y_{k-1} \quad (2.5)$$

$$\beta_k^{N1} = \max \left\{ \theta_{k-1} \frac{y_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} - \left(\hat{\gamma}_{k-1} + \theta_{k-1} \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}} \right) \frac{s_{k-1}^T g_k}{d_{k-1}^T y_{k-1}}, 0 \right\},$$

$$\zeta_k^{N1} = \text{sgn}(\beta_k) \left(\theta_{k-1} \frac{d_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} + (1 - \theta_{k-1}) \frac{y_{k-1}^T g_k}{y_{k-1}^T y_{k-1}} \right).$$

ただし $\text{sgn}(\cdot)$ は

$$\text{sgn}(a) = \begin{cases} 1 & a > 0, \\ 0 & a = 0 \end{cases}$$

とする。さらに彼らは、大域的収束性を示すために以下のような性質を導入した。

Property 1 探索方向 (2.5) を用いたアルゴリズム 1 について考える。正定数 ε が存在してすべての k に対して $\varepsilon \leq \|g_k\|$ が成り立つと仮定する。このとき、すべての k に対して

$$\hat{\gamma}_{k-1} < \bar{c} \|d_{k-1}\|$$

が成り立つならば、アルゴリズム 1 は *Property 1* を持つと言う。

Property 1 を持つ方法の大域的収束性に関する以下の定理を与えた。

定理 2.1 仮定 2.1-2.2 が成立しているとし、(2.5) を用いたアルゴリズム 1 が条件 (2.3) と (2.4) を満たし、さらに *Property 1* を持つとする。このとき、アルゴリズム 1 により生成される点列 $\{x_k\}$ は

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

の意味で大域的に収束する。

具体的なパラメータ $\hat{\gamma}_{k-1}$ として

$$\hat{\gamma}_{k-1} = \theta_{k-1} \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}} \quad (2.6)$$

を選べば、 θ_{k-1} が条件 (2.4) を満たすとき、定理 2.1 より彼らの方法は大域的に収束することが保証される。

さらに Nakayama ら [13] は、数値実験によってパラメータ $\hat{\gamma}_{k-1}$ や θ_{k-1} の選び方が提案法の数値的な効率性にどのような影響があるのかを検証している。その実験において、 $\hat{\gamma}_{k-1}$ の選択法として (2.6) よりも

$$\hat{\gamma}_{k-1} = \theta_{k-1} \frac{s_{k-1}^T y_{k-1}}{s_{k-1}^T s_{k-1}} \quad (2.7)$$

の方が数値的な効率性に優れていることを報告した。ところが、

$$\frac{s_{k-1}^T y_{k-1}}{s_{k-1}^T s_{k-1}} \leq \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}}$$

より、(2.7) は条件 (2.3) を満たすとは限らないため、(2.7) を用いた方法は十分な降下条件や大域的収束性を保証できていない。そのため、(2.7) を使用しつつ、十分な降下条件と大域的収束性を保証することが課題に挙げられている。

また Nakayama ら [13] は、Broyden family のパラメータ θ_{k-1} として

$$\theta_{k-1} = \min \left\{ \theta_{max}, \hat{\theta}_{k-1} \right\},$$

$$\hat{\theta}_{k-1} = 1 + \left| \frac{d_{k-1}^T g_k}{d_{k-1}^T g_{k-1}} \right|, \quad 1 + \frac{|d_{k-1}^T g_k|}{\|d_{k-1}\| \|g_k\|}$$

が優れていることを示した。ここでは (2.4) を満たすように $\theta_{max} = 1.9$ としている。これは、 $\theta_{k-1} = 1$ (BFGS 公式) より優れているパラメータが preconvex class に存在することを意味している。

2.2 Modified memoryless self-scaling BFGS method

Kou and Dai [11] では Self-scaling BFGS 公式に基づいたメモリーレス準ニュートン法にパラメータ $\xi_{k-1} \in [0, 1]$ を加えた以下のような探索方向を提案した:

$$d_k = -g_k + \left(\frac{y_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} - \left(\hat{\gamma}_{k-1} + \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}} \right) \frac{s_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} \right) d_{k-1} + \xi_{k-1} \frac{d_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} y_{k-1}. \quad (2.8)$$

ここで $\hat{\gamma}_{k-1} > 0$ はパラメータである。探索方向 (2.8) は $\xi_{k-1} = 1$ のときは self-scaling BFGS 公式に基づいたメモリーレス準ニュートン法であり、 $\xi_{k-1} = 0$ のときは CG 法になる。彼らは十分な降下条件に関する以下の命題を示している。

命題 2.2 $\xi \in [0, 1]$ は定数とし、 $\xi_{k-1} = \xi$ とする。このとき $d_{k-1}^T y_{k-1} > 0$ の下で、探索方向 (2.8) は十分な降下条件を満たす。

この命題は ξ_{k-1} に条件を課すことで、パラメータ $\hat{\gamma}_{k-1}$ に条件を付加することなく、十分な降下条件を満たすことを意味している。また、(2.5) と同様の補正を加えたとき、大域的な収束性を示している。なお、彼らはパラメータ $\hat{\gamma}_{k-1}$ として、Oren and Luenberger [15, 16] の推奨する

$$\hat{\gamma}_{k-1} \in \left[\frac{s_{k-1}^T y_{k-1}}{s_{k-1}^T s_{k-1}}, \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}} \right]$$

を考え、数値実験においては

$$\hat{\gamma}_{k-1} = \frac{s_{k-1}^T y_{k-1}}{s_{k-1}^T s_{k-1}}$$

を使用している.

3 Broyden family に基づいた非線形共役勾配法

本研究では、Nakayama ら [13] の方法に Kou and Dai [11] と同様の修正をすることを考える. 具体的には、(2.2) 式の 3 項目に Dai and Kou [11] と同様にパラメータ $\xi_{k-1} \in [0, 1)$ を導入した以下の探索方向を考える:

$$\begin{aligned} d_k = & -g_k + \left(\theta_{k-1} \frac{y_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} - \left(\hat{\gamma}_{k-1} + \theta_{k-1} \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}} \right) \frac{s_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} \right) d_{k-1} \\ & + \xi_{k-1} \left(\theta_{k-1} \frac{d_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} + (1 - \theta_{k-1}) \frac{y_{k-1}^T g_k}{y_{k-1}^T y_{k-1}} \right) y_{k-1}. \end{aligned} \quad (3.1)$$

探索方向 (3.1) に関して、以下の命題を得る.

命題 3.1 $d_{k-1}^T y_{k-1} > 0$ の下で、もし次の条件

$$(i) \quad 0 \leq \theta_{k-1} \leq 1, \quad 0 \leq \xi_{k-1} \leq \bar{\xi}.$$

または

$$(ii) \quad 1 < \theta_{k-1} < \bar{c}_1^2, \quad 0 \leq \xi_{k-1} \leq \frac{\bar{c}_1}{\sqrt{\theta_{k-1}}} - 1.$$

が成り立つならば、探索方向 (3.1) は十分な降下条件を満たす. ただし $\bar{\xi}$ を \bar{c}_1 はそれぞれ $0 \leq \bar{\xi} < 1, 1 < \bar{c}_1 < 2$ を満たす定数とする.

証明. 任意の $a, b \in \mathbf{R}^n$ に対して $2a^T b \leq \|a\|^2 + \|b\|^2$ が成立するので

$$\begin{aligned}
g_k^T d_k &= -\|g_k\|^2 + \theta_{k-1}(1 + \xi_{k-1}) \frac{(y_{k-1}^T g_k)(d_{k-1}^T g_k)}{d_{k-1}^T y_{k-1}} \\
&\quad - \left(\hat{\gamma}_{k-1} + \theta_{k-1} \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}} \right) \frac{\alpha_{k-1}(d_{k-1}^T g_k)^2}{d_{k-1}^T y_{k-1}} + \xi_{k-1}(1 - \theta_{k-1}) \frac{(y_{k-1}^T g_k)^2}{y_{k-1}^T y_{k-1}} \\
&= -\|g_k\|^2 + \theta_{k-1} \left(\frac{\sqrt{2} d_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} y_{k-1} \right)^T \left(\frac{(1 + \xi_{k-1})}{\sqrt{2}} g_k \right) \\
&\quad - \left(\hat{\gamma}_{k-1} + \theta_{k-1} \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}} \right) \frac{\alpha_{k-1}(d_{k-1}^T g_k)^2}{d_{k-1}^T y_{k-1}} + \xi_{k-1}(1 - \theta_{k-1}) \frac{(y_{k-1}^T g_k)^2}{y_{k-1}^T y_{k-1}} \\
&\leq -\|g_k\|^2 + \frac{\theta_{k-1}}{2} \left(\left\| \frac{\sqrt{2} d_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} y_{k-1} \right\|^2 + \left\| \frac{(1 + \xi_{k-1})}{\sqrt{2}} g_k \right\|^2 \right) \\
&\quad - \left(\hat{\gamma}_{k-1} + \theta_{k-1} \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}} \right) \frac{\alpha_{k-1}(d_{k-1}^T g_k)^2}{d_{k-1}^T y_{k-1}} + \xi_{k-1}(1 - \theta_{k-1}) \frac{(y_{k-1}^T g_k)^2}{y_{k-1}^T y_{k-1}} \\
&= - \left(1 - \frac{\theta_{k-1}(1 + \xi_{k-1})^2}{4} \right) \|g_k\|^2 - \hat{\gamma}_{k-1} \frac{\alpha_{k-1}(d_{k-1}^T g_k)^2}{d_{k-1}^T y_{k-1}} \\
&\quad + \xi_{k-1}(1 - \theta_{k-1}) \frac{(y_{k-1}^T g_k)^2}{y_{k-1}^T y_{k-1}} \\
&\leq - \left(1 - \frac{\theta_{k-1}(1 + \xi_{k-1})^2}{4} \right) \|g_k\|^2 + \xi_{k-1}(1 - \theta_{k-1}) \frac{(y_{k-1}^T g_k)^2}{y_{k-1}^T y_{k-1}}
\end{aligned}$$

が成り立つ. 次に (i) と (ii) の場合分けを考える.

(i) $0 \leq \theta_{k-1} \leq 1, 0 \leq \xi_{k-1} < \bar{\xi}$ の場合:

$\xi_{k-1}(1 - \theta_{k-1}) \geq 0$ より

$$\begin{aligned}
g_k^T d_k &\leq - \left(1 - \frac{\theta_{k-1}(1 + \xi_{k-1})^2}{4} - \xi_{k-1}(1 - \theta_{k-1}) \right) \|g_k\|^2 \\
&= - \left((1 - \xi_{k-1}) - \frac{\theta_{k-1}}{4} (1 - 2\xi_{k-1} + \xi_{k-1}^2) \right) \|g_k\|^2 \\
&= - \left((1 - \xi_{k-1}) \left(1 - \frac{\theta_{k-1}}{4} (1 - \xi_{k-1}) \right) \right) \|g_k\|^2 \\
&\leq - \frac{3(1 - \xi_{k-1})}{4} \|g_k\|^2 \\
&\leq - \frac{3(1 - \bar{\xi})}{4} \|g_k\|^2
\end{aligned}$$

となる. $\bar{\xi} < 1$ なので, 十分な降下条件が成り立つ.

(ii) $1 < \theta_{k-1} < \bar{c}_1^2, 0 \leq \xi_{k-1} \leq \frac{\bar{c}_1}{\sqrt{\theta_{k-1}}} - 1$ の場合:

$\xi_{k-1}(1 - \theta_{k-1}) < 0$ より

$$\begin{aligned} g_k^T d_k &\leq - \left(1 - \frac{\theta_{k-1}(1 + \xi_{k-1})^2}{4} \right) \|g_k\|^2 \\ &\leq - \left(1 - \frac{\theta_{k-1}}{4} \frac{c_1^2}{\theta_{k-1}} \right) \|g_k\|^2 \\ &= - \left(1 - \frac{c_1^2}{4} \right) \|g_k\|^2 \end{aligned}$$

となる. $c_1^2 < 4$ なので, 十分な降下条件が成り立つ. \square

Nakayama ら [13] と同様に大域的収束性のために探索方向を次のように修正する:

$$d_k = -g_k + \beta_k^{N2} d_{k-1} + \zeta_k^{N2} y_{k-1} \quad (3.2)$$

$$\beta_k^{N2} = \max \left\{ \theta_{k-1} \frac{y_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} - \left(\hat{\gamma}_{k-1} + \theta_{k-1} \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}} \right) \frac{s_{k-1}^T g_k}{d_{k-1}^T y_{k-1}}, 0 \right\},$$

$$\zeta_k^{N2} = \text{sgn}(\beta_k) \xi_{k-1} \left(\theta_{k-1} \frac{d_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} + (1 - \theta_{k-1}) \frac{y_{k-1}^T g_k}{y_{k-1}^T y_{k-1}} \right).$$

このとき以下の大域的収束性の定理を得る.

定理 3.1 仮定 2.1-2.2 が成立しているとし, スケーリングパラメータを

$$\hat{\gamma}_{k-1} = \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}} \quad \text{または} \quad \hat{\gamma}_{k-1} = \frac{s_{k-1}^T y_{k-1}}{s_{k-1}^T s_{k-1}}$$

とする. また θ_{k-1} と ξ_{k-1} は命題 3.1 の (i) または (ii) を満たすとする. このとき, (3.2) を用いたアルゴリズム 1 により生成される点列 $\{x_k\}$ は

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

の意味で大域的に収束する.

具体的なパラメータ θ_{k-1}, ξ_{k-1} の選択法として, 命題 3.1 の条件を満たす定数としては

- $\theta_{k-1} = 0.5, \quad \xi_{k-1} = 0.5$
- $\theta_{k-1} = 1, \quad \xi_{k-1} = 0.8$
- $\theta_{k-1} = 2, \quad \xi_{k-1} = 0.4$

などが挙げられる. また, 定数ではなく, 可変的な θ_{k-1} を考えることもできる. 例えば, $\xi_{k-1} = 0.8$ としたときにパラメータ θ_{k-1} を

$$\theta_{k-1} = 1 + \mu_{k-1}, \quad \mu_{k-1} = \max \{ -0.2, \min \{ \bar{\mu}_{k-1}, 0.2 \} \}$$

と選ぶことができる. ここで [13] で挙げられている $\bar{\mu}_{k-1}$ の選択法として

$$\bar{\mu}_{k-1} = \frac{d_{k-1}^T g_k}{-g_{k-1}^T d_{k-1}}, \quad \left| \frac{d_{k-1}^T g_k}{-g_{k-1}^T d_{k-1}} \right|, \quad \frac{|d_{k-1}^T g_k|}{\|d_{k-1}\| \|g_k\|}, \quad \frac{d_{k-1}^T g_k}{\|d_{k-1}\| \|g_k\|}$$

などを考えることもできる。これらのパラメータを使用し、かつ、正確な直線探索を行った場合、探索方向 (3.1) は Hestenes-Stiefel 公式 ([9] 参照) を用いた CG 法 (1.3) と一致するため、CG 法の一つとみなすことができる。このことから、上記のパラメータを使用した提案法を新しい CG 法として提案する。

4 数値実験

本節では数値実験において、パラメータ θ_{k-1} , $\hat{\gamma}_{k-1}$, ξ_{k-1} の選択が提案法に与える影響を調査する。下記の方法の数値実験結果を報告する:

Name	θ_{k-1}	$\hat{\gamma}_{k-1}$	ξ_{k-1}
ML1(BFGS)	1	$\frac{s_{k-1}^T y_{k-1}}{s_{k-1}^T s_{k-1}}$	1
ML2	$1 + \min \left\{ \frac{ g_k^T d_{k-1} }{\ g_k\ \ d_{k-1}\ }, 0.9 \right\}$	$\theta_{k-1} \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}}$	1
ML3	$1 + \min \left\{ \frac{ g_k^T d_{k-1} }{\ g_k\ \ d_{k-1}\ }, 0.9 \right\}$	$\theta_{k-1} \frac{s_{k-1}^T y_{k-1}}{s_{k-1}^T s_{k-1}}$	1
BCG1 (BFGS)	1	$\frac{s_{k-1}^T y_{k-1}}{s_{k-1}^T s_{k-1}}$	0.8
BCG2	$1 + \min \left\{ \frac{ g_k^T d_{k-1} }{ g_{k-1}^T d_{k-1} }, 0.2 \right\}$	$\frac{s_{k-1}^T y_{k-1}}{s_{k-1}^T s_{k-1}}$	0.8
CGD	CG_DESCENT Ver5.3 [7, 10]		

ML1-ML3 は [13] の数値実験で使用されたメモリーレス準ニュートン法である。その中で ML3 は大域的収束性が保証されていないが最も優れていた解法であり、ML2 は大域的収束性が保証している中で最も優れていた解法である。また、ML1 は Self-scaling BFGS 公式に基づいたメモリーレス準ニュートン法である。BCG1-BCG2 は今回の提案法であり、 $\xi_{k-1} = 0.8$ とおいた。CGD は CG-DESCENT [7, 10] と呼ばれる CG 法であり、ベンチマークとしてよく使われる手法である。

今回の数値実験では Hager and Zhang が提案した非線形共役勾配 [8] に基づいたソフトウェアである CG-DESCENT Ver 5.3 [7, 10] の探索方向を修正して実験を行った。直線探索などの設定は CG-DESCENT Ver 5.3 に倣っている。また、収束判定条件として

$$\|g_k\|_\infty \leq 10^{-6}$$

を使用した。テスト問題として CUTEr 問題集 [1, 6] から 138 問 (付録参照) を選択して実験を行った。

Dolan and Moré [5] の提案したパフォーマンスプロファイルを基に、各方法の計算時間による比較を行う。各方法のパフォーマンスプロファイル $P(\tau)$ の $\tau = \bar{\tau}$ のときの値は、その解法が全て問題の中で最も早く解くことができた方法の求解時間の τ 倍以内に解くことのできた問題の割合を表している。 $\tau = 1$ のときの値は、その方法がすべての方法の中で最も早く解けた問題数の割合を表し、 τ が十分大きいときは、その方法の解くことがで

きた問題数の割合を意味する。どの τ においても、 $P(\tau)$ が1に近いほうが好ましく、複数の数値解法のパフォーマンスプロファイルのグラフを並べたときにはグラフが上に位置するほど効率が良い数値解法と考えられる。

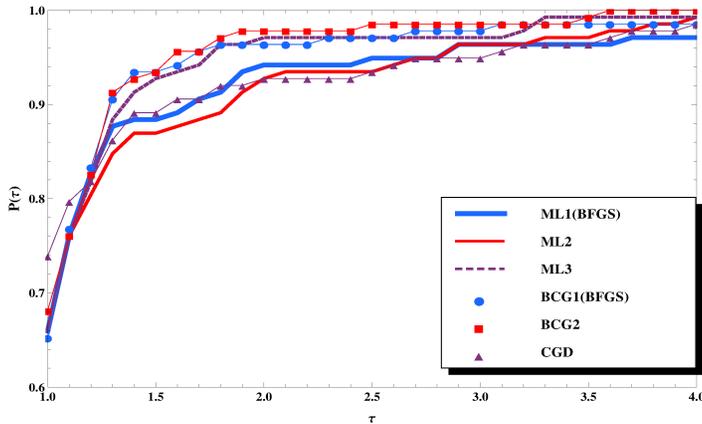


図 1: 計算時間による比較

図 1 では実験を行った方法のパフォーマンスが与えられている。この図より 6 つの方法の効率性は

$$\text{BCG2} \geq \text{BCG1} > \text{ML3} > \text{ML1} \approx \text{CGD} > \text{ML2}$$

となっていることがわかる。ML2 と ML3 のパフォーマンスを比べるとスケーリングパラメータが与える影響が大きいことがわかる。ML1–ML3 と BCG1–BCG2 を比べると、 ξ_{k-1} を加えることで、今回の提案法は大域的収束性の保証に加え、数値的性能も向上していることがわかる。微かながら BCG2 が BCG1 より良いパフォーマンスが得られたことから、[13] と同様に BFGS 公式より preconvex class を考えた Broyden family が良いことがわかる。また、ML2 を除いたどの方法もベンチマークである CGD より優れている。

5 終わりに

Nakayama ら [13] の提案法は、数値実験において効果的なスケーリングパラメータを用いた方法の大域的収束性が保証できていないという課題があった。今回我々は、Nakayama ら [13] の方法に Kou and Dai [11] と同様の修正を加えた方法を提案した。提案法にスケーリングパラメータの条件を付加することなく、十分な降下条件を満たし、大域的に収束することを示した。さらに数値実験において、提案法の有用性を検証した。提案法のパラメータが従来法に比べて多いため、より適切なパラメータの選択法が今後の課題に挙げられる。

謝辞

本研究の一部は JSPS 科研費 JP17K00039 の助成を受けて行われている。

参考文献

- [1] I. Bongartz, A.R. Conn, N.I.M Gould and P.L. Toint, CUTE: constrained and unconstrained testing environment, *ACM Transactions on Mathematical Software*, **21** (1995), 123–160.
- [2] Z. Chen and W. Cheng, Spectral-scaling quasi-Newton methods with updates from the one parameter of the Broyden family, *Journal of Computational and Applied Mathematics*, **248** (2013), 88–98.
- [3] W.Y. Cheng and D.H. Li, Spectral scaling BFGS method, *Journal of Optimization Theory and Applications*, **146** (2010), 305–319.
- [4] Y.H. Dai and C.X. Kou, A nonlinear conjugate gradient algorithm with an optimal property and an improved wolfe line search, *SIAM Journal on Optimization*, **23**, (2013), 296–320.
- [5] E.D. Dolan and J.J Moré, Benchmarking optimization software with performance profiles, *Mathematical Programming*, **91** (2002), 201–213.
- [6] N.I.M Gould, D. Orban, and P.L. Toint, CUTER and SifDec: A constrained and unconstrained testing environment, revisited, *ACM Transactions on Mathematical Software*, **29** (2003), 373–394.
- [7] W.W. Hager, Hager’s web page : <http://people.clas.ufl.edu/hager/>.
- [8] W.W. Hager and H. Zhang, A new conjugate gradient method with guaranteed descent and an efficient line search, *SIAM Journal on Optimization*, **16** (2005), 170–192.
- [9] W.W. Hager and H. Zhang, A survey of nonlinear conjugate gradient method, *Pacific Journal of Optimization*, **2** (2006), 35–58.
- [10] W.W. Hager and H. Zhang, Algorithm 851: CG_DESCENT, a conjugate gradient method with guaranteed descent, *ACM Transactions on Mathematical Software*, **32** (2006), 113–137.
- [11] C.X. Kou and Y.H. Dai, A modified self-scaling memoryless Broyden-Fletcher-Goldfarb-Shanno method for unconstrained optimization, *Journal of Optimization Theory and Applications*, **165** (2015), 209–224.

- [12] S. Nakayama, Y. Narushima and H. Yabe, A memoryless symmetric rank-one method with sufficient descent property for unconstrained optimization, *Journal of the Operations Research Society of Japan*, to appear.
- [13] S. Nakayama, Y. Narushima and H. Yabe, Global convergence of memoryless quasi-Newton methods based on Broyden family for unconstrained optimization, submitted.
- [14] Y. Narushima and H. Yabe, A survey of sufficient descent conjugate gradient methods for unconstrained optimization, *SUT Journal of Mathematics*, **50** (2014), 167–203.
- [15] S.S. Oren, Self-scaling variable metric (SSVM) algorithms, Part II: implementation and experiments, *Management Science*, **20** (1974), 863–874.
- [16] S.S. Oren and D.G. Luenberger, Self-scaling variable metric (SSVM) algorithms, Part I: Criteria and sufficient conditions for scaling a class of algorithms, *Management Science*, **20** (1974), 845–862.
- [17] D.F. Shanno, Conjugate gradient methods with inexact searches, *Mathematics of Operations Research*, **3** (1978), 244–256.
- [18] L. Zhang, W. Zhou, and D.H. Li, A descent modified Polak-Ribière-Polyak conjugate gradient method and its global convergence, *IMA Journal of Numerical Analysis*, **26** (2006), 629–640.
- [19] Y. Zhang and R.P. Tewarson, Quasi-Newton Algorithms with updates from the preconvex part of Broyden’s family, *IMA Journal of Numerical Analysis*, **8** (1988), 487–509.

付録

表 1: テスト問題集 (問題名と次元数)

問題名	次元数	問題名	次元数	問題名	次元数	問題名	次元数
AKIVA	2	DIXMAANC	3000	HEART8LS	8	PENALTY1	1000
ALLINITU	4	DIXMAAND	3000	HELIX	3	PENALTY2	200
ARGLINA	200	DIXMAANE	3000	HIELOW	3	PENALTY3	200
ARGLINB	200	DIXMAANF	3000	HILBERTA	2	POWELLSG	5000
ARWHEAD	5000	DIXMAANG	3000	HILBERTB	10	POWER	10000
BARD	3	DIXMAANH	3000	HIMMELBB	2	QUARTC	5000
BDQRTIC	5000	DIXMAANI	3000	HIMMELBF	4	ROSENBR	2
BEALE	2	DIXMAANJ	3000	HIMMELBG	2	S308	2
BIGGS6	6	DIXMAANK	3000	HIMMELBH	2	SCHMVETT	5000
BOX3	3	DIXMAANL	3000	HUMPS	2	SENSORS	100
BOX	10000	DIXON3DQ	10000	JENSMP	2	SINEVAL	2
BRKMCC	2	DJTL	2	KOWOSB	4	SINQUAD	5000
BROWNAL	200	DQDRTIC	5000	LIARWHD	5000	SISSER	2
BROWNBS	2	DQRTIC	5000	LOGHAIRY	2	SNAIL	2
BROWNDEN	4	EDENSCH	2000	MANCINO	100	SPARSINE	5000
BROYDN7D	5000	EG2	1000	MARATOSB	2	SPARSQUR	10000
BRYBND	5000	ENGVAL1	5000	MEXHAT	2	SPMSRTLS	4999
CHAINWOO	4000	ENGVAL2	3	MOREBV	5000	SROSENBR	5000
CHNROSNB	50	ERRINROS	50	MSQRTALS	1024	STRATEC	10
CLIFF	2	EXPFIT	2	MSQRTBLS	1024	TESTQUAD	5000
COSINE	10000	EXTROSNB	1000	NONCVXU2	5000	TOINTGOR	50
CRAGGLVY	5000	FLETGBV2	5000	NONDIA	5000	TOINTGSS	5000
CUBE	2	FLETCHCR	1000	NONDQUAR	5000	TOINTPSP	50
CURLY10	10000	FMINSRF2	5625	OSBORNEA	5	TOINTQOR	50
CURLY20	10000	FMINSURF	5625	OSBORNEB	11	TQUARTIC	5000
CURLY30	10000	FREUROTH	5000	OSCIPATH	10	TRIDIA	5000
DECONVU	63	GENHUMPS	5000	PALMER1C	8	VARDIM	200
DENSCHNA	2	GENROSE	500	PALMER1D	7	VAREIGVL	50
DENSCHNB	2	GROWTHLS	3	PALMER2C	8	VIBRBEAM	8
DENSCHNC	2	GULF	3	PALMER3C	8	WATSON	12
DENSCHND	3	HAIRY	2	PALMER4C	8	WOODS	4000
DENSCHNE	3	HATFLDD	3	PALMER5C	6	YFITU	3
DENSCHNF	2	HATFLDE	3	PALMER6C	8	ZANGWIL2	2
DIXMAANA	3000	HATFLDFL	3	PALMER7C	8		
DIXMAANB	3000	HEART6LS	6	PALMER8C	8		