

無制約最適化問題に対する勾配法について

成島 康史

無制約最適化問題に対する数値解法として、反復法は古くから研究されており、中でも準ニュートン法は小・中規模問題に対して有効な解法として知られている。しかしながら、準ニュートン法は密な近似行列を保存する必要があるため、大規模問題に対しては直接適用できない。そのような理由から、近年行列を陽に使用しない、いわゆる勾配法が注目を集めている。本稿では勾配法の中でも、記憶制限準ニュートン法の流れをくむ方法と、非線形共役勾配法の流れをくむ方法に分類して、それぞれの方法を紹介するとともに、両者の関係性についても解説する。

キーワード：無制約最適化、勾配法、記憶制限準ニュートン法、非線形共役勾配法

1. はじめに

本稿では、無制約最適化問題：

$$\min_{x \in \mathbb{R}^n} f(x), \quad (1)$$

に対する数値解法を考える。ただし、目的関数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ は十分滑らかで、その勾配を $g \equiv \nabla f$ で表すこととする。 n は変数 x の次元であるが、後述するように、 n が大きいときには大規模問題と呼ばれ、扱いが難しくなる。問題 (1) に対する数値解法として、反復法が広く使用されている。反復法は、初期点 $x_0 \in \mathbb{R}^n$ からスタートして、反復式：

$$x_{k+1} = x_k + \alpha_k d_k \quad (2)$$

によって、点列 $\{x_k\}$ を生成する方法で、 $\alpha_k > 0$ はステップ幅、 $d_k \in \mathbb{R}^n$ は探索方向と呼ばれる。通常、反復法では、更新した点での目的関数値 $f(x_{k+1})$ は更新前の目的関数値 $f(x_k)$ よりも小さくなる、つまり $f(x_k + \alpha_k d_k) < f(x_k)$ となるように選ばれる。そのため、探索方向は方向微係数が負、すなわち、

$$g_k^\top d_k < 0 \quad (3)$$

を満たす探索方向であることが必要となる。ただし、 $g_k \equiv g(x_k)$ とする。条件 (3) を降下条件と呼び、降下条件を満たす探索方向は降下方向と呼ばれる。一方、ステップ幅 α_k を決定する手順を直線探索と呼び、何らかの条件を満たすように選択される。最も理想的なステップ幅の選択法は目的関数 f を d_k 方向に 1 次元

の最小化をする、つまり、 $\alpha_k = \operatorname{argmin}_{\alpha > 0} f(x_k + \alpha d_k)$ となる α_k を選択することである。これを正確な直線探索と呼ぶ。目的関数が狭義凸 2 次関数

$$f(x) = \frac{1}{2} x^\top A x - b^\top x \quad (4)$$

である場合には、正確な直線探索を行うことが可能で、正確なステップ幅は

$$\alpha_k = \frac{-g_k^\top d_k}{d_k^\top A d_k} \quad (5)$$

で与えられる。ここで、 $A \in \mathbb{R}^{n \times n}$ は正定値対称行列で、 $b \in \mathbb{R}^n$ とする。しかしながら、目的関数が一般の非線形関数の場合は正確な直線探索を行うことは困難であるため、適当な直線探索条件、たとえば下記のような条件を満たすようなステップ幅が選ばれる：

Wolfe 条件. 与えられた定数 δ, σ ($0 < \delta < \sigma < 1$) に対して、下記を満たす $\alpha > 0$ を選ぶ：

$$f(x_k + \alpha d_k) \leq f(x_k) + \alpha \delta g_k^\top d_k, \quad (6)$$

$$\sigma g(x_k)^\top d_k \leq g(x_k + \alpha d_k)^\top d_k. \quad (7)$$

探索方向 d_k が降下方向であるとする、Wolfe 条件の式 (6) は目的関数値が減少する条件となっている。一方、式 (7) は方向微係数に関する条件となっている。 α を十分小さく選択すれば式 (6) が満たされるのに対し、式 (7) はステップ幅が小さくなりすぎないような条件となっている。ここで、探索方向が降下方向であるときには、Wolfe 条件 (6), (7) を満たすステップ幅が存在することを注意しておく。

反復法は探索方向 d_k の選択法によって大きく性能が異なるため、さまざまな探索方向の選択法が提案

なるしま やすし

慶應義塾大学理工学部管理工学科

〒 223-8522 神奈川県横浜市港北区日吉 3-14-1

narushima@ae.keio.ac.jp

されており、最急降下法やニュートン法、準ニュートン法などがよく知られている。特に準ニュートン法は小・中規模な問題に対して非常に有効な方法として、さまざまなソフトウェアに組み込まれている。しかしながら、準ニュートン法は $n \times n$ (n は変数 x の次元) の密行列を必要とするため、大規模問題に対しては直接適用することは困難となる。そのため、近年では、行列を陽に使用しない、いわゆる勾配法に注目が集まっている。近年注目されている勾配法は大きく分けて 2 通りに分類できる。一つ目が最急降下方向に項を加えることで加速を行う方法であり、非線形共役勾配法がその代表例である。もう一つの方法は、準ニュートン法の近似行列の更新において情報を制限することで、陽に行列を使用しないようにする方法であり、記憶制限準ニュートン法やメモリーレス準ニュートン法などがそれに当たる。本稿では、それらの方法の中でも代表的な方法の紹介を行う。特に、それらの方法は互いに関連性をもつため、それぞれの方法の関連性にも注目することとする。

2. 最急降下法と準ニュートン法

この節では、本稿の主眼である勾配法の前提知識として、最急降下法と準ニュートン法を紹介する。

2.1 最急降下法

最急降下法は探索方向として、目的関数の 1 次近似：

$$f(x_k + d) \approx \ell(d) = f(x_k) + g_k^\top d$$

を最小にする方向が選択される。 $\ell(d)$ はベクトル d に対して線形であるため、 $\|d\| = \|g_k\|$ であると仮定して $\ell(d)$ を最小にする方向を考えれば、探索方向は勾配ベクトルの逆方向、つまり、

$$d_k = -g_k \quad (8)$$

となる。これを最急降下方向と呼び、最急降下方向を使用した反復法を最急降下法と呼ぶ。最急降下法は反復法の中で最もよく知られた方法であるが、その一方で、実用上はあまり効果的ではないこともよく知られている。たとえば、狭義凸 2 次関数 (4) に対して、正確な直線探索 (5) を用いた最急降下法の収束率は

$$\|x_{k+1} - x^*\|_A \leq \left(\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}} \right) \|x_k - x^*\|_A \quad (9)$$

であることが知られている。ここで、 x^* を最適解、 λ_{\max} と λ_{\min} をそれぞれ行列 A の最大固有値と最小固有値

とし、 $\|x\|_A = \sqrt{x^\top A x}$ を正定値対称行列 A による重み付きノルムとする。上述の関係式から、行列 A の最大固有値と最小固有値の差が非常に大きい (つまり、条件数が大きい) 場合には、 $\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}} \approx 1$ となり、非常に効率が悪くなってしまふ。このような性質は一般の目的関数でも同様であることが知られている (たとえば、文献 [1] などを参照)。

2.2 準ニュートン法

最急降下法は、各反復において目的関数の 1 次近似を最小にする方向を選択するのに対し、ニュートン法は 2 次近似を最小にする方向を選択する方法である。目的関数の 2 次近似：

$$f(x_k + d) \approx q(d) = f(x_k) + g_k^\top d + \frac{1}{2} d^\top \nabla^2 f(x_k) d$$

を最小にする方向は、ヘッセ行列 $\nabla^2 f(x_k)$ が正定値であると仮定すれば、 $\nabla q(d) = 0$ を考えて、

$$d_k = -\nabla^2 f(x_k)^{-1} g_k \quad (10)$$

となる。ニュートン法は局所的に 2 次収束する¹ というよい性質をもっているが、ヘッセ行列 $\nabla^2 f(x_k)$ が正定値ではない場合には降下方向を生成するとは限らないという弱点があり、一般の目的関数において大域的な収束性² を保証することが難しい。そのため、ヘッセ行列を近似行列 $B_k \in \mathbb{R}^{n \times n}$ で置き換えた準ニュートン法が提案されている。準ニュートン法の探索方向は、式 (10) においてヘッセ行列を近似行列で置き換えて、 $d_k = -B_k^{-1} g_k$ または

$$d_k = -H_k g_k \quad (11)$$

¹ 定数 $\eta > 0$ と $p \geq 1$ が存在して、点列 $\{x_k\}$ が

$$\|x_{k+1} - x^*\| \leq \eta \|x_k - x^*\|^p$$

を満たすとき、 $\{x_k\}$ は x^* に p 次収束するという (ただし、 $p = 1$ のときは $\eta \in (0, 1)$ とする)。なお、式 (9) は重み付きノルムを用いた場合の 1 次収束性を表していることを注意しておく。一方、正の数 $\lim_{k \rightarrow \infty} \eta_k = 0$ が存在して、

$$\|x_{k+1} - x^*\| \leq \eta_k \|x_k - x^*\|$$

を満たすとき、 $\{x_k\}$ は x^* に超 1 次収束するという。超 1 次収束は 1 次収束と 2 次収束の中間的な収束速度であると捉えることができる。また、実用的には超 1 次収束性をもつアルゴリズムは十分効率的であると考えてよい。

² 任意の初期点から出発したときに、反復法によって生成された点列 $\{x_k\}$ が目的関数の停留点 (1 次の最適性条件 $g(x) = 0$ を満たす点) に収束することを大域的収束と呼ぶ。目的関数が一般の非線形関数の場合はもう少し緩和して

$$\liminf_{k \rightarrow \infty} \|g(x_k)\| = 0$$

をもって大域的収束と呼ぶことも多い。

で与えられる³. ここで, $H_k = B_k^{-1}$ である. 具体的に近似行列を選択する際には, $B_k \approx \nabla^2 f(x_k)$ であることが望まれる. ここで, $g(x_{k-1})$ の 1 次近似を考えると

$$g(x_{k-1}) \approx g(x_k) - \nabla^2 f(x_k)(x_k - x_{k-1}) \quad (12)$$

という関係式が得られる. よって, 近似行列が満たすべき条件として

$$B_k s_{k-1} = y_{k-1}, \quad \text{または} \quad s_{k-1} = H_k y_{k-1}$$

を考えることができる. これをセカント条件と呼ぶ. ただし,

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

とする. セカント条件を満たす更新公式として DFP 公式, BFGS 公式, SR1 公式などがよく知られている. 逆行列版の (つまり, H_k に対する) BFGS 公式は

$$\begin{aligned} H_k &= V_{k-1}^\top H_{k-1} V_{k-1} + \frac{s_{k-1} s_{k-1}^\top}{s_{k-1}^\top y_{k-1}}, \\ &= H_{k-1} - \frac{H_{k-1} y_{k-1} s_{k-1}^\top + s_{k-1} y_{k-1}^\top H_{k-1}}{s_{k-1}^\top y_{k-1}} \\ &\quad + \left(1 + \frac{y_{k-1}^\top H_{k-1} y_{k-1}}{s_{k-1}^\top y_{k-1}} \right) \frac{s_{k-1} s_{k-1}^\top}{s_{k-1}^\top y_{k-1}}, \quad (13) \end{aligned}$$

で与えられる. ただし,

$$V_{k-1} = I - \frac{y_{k-1} s_{k-1}^\top}{s_{k-1}^\top y_{k-1}} \quad (14)$$

とし, $I \in \mathbb{R}^{n \times n}$ を単位行列とする. BFGS 公式では, 更新前の行列 H_{k-1} が正定値対称行列で, $s_{k-1}^\top y_{k-1} > 0$ ならば⁴, 更新後の行列 H_k も正定値対称行列となり, 探索方向は降下方向となる. さらに, BFGS 更新公式を用いた準ニュートン法 (以下では BFGS 法と呼ぶ) は局所的に超 1 次収束することが知られている. このように, BFGS 法はよい性質をもち, 実用上も有効な方法としてさまざまなソフトウェアに組み込まれているが, その一方で, BFGS 法で生成された近似行列は密行列となるため, 大規模問題には直接適用できないという問題点がある.

³ ニュートン法や最急降下法の場合は現在の点 x_k における情報しか必要ないが, それ以外の方法の場合は一つ前の点 x_{k-1} の情報が必要となるため, 通常は, 初期探索方向として最急降下方向 $d_0 = -g_0$ が選択される. 以降, 断りが無い限り, 初期探索方向として最急降下方向を選択することとする.

⁴ たとえば, 直線探索において Wolfe 条件 (6), (7) を満たすようにステップ幅を選択した場合, $s_{k-1}^\top y_{k-1} > 0$ が保証される.

3. 準ニュートン法に基づいた方法

3.1 記憶制限準ニュートン法

記憶制限準ニュートン法は, 1980 年に Nocedal [2] によって提案された方法である. BFGS 更新公式 (13) を用いれば, 生成される H_k と, その m 反復前の H_{k-m} の間の関係は

$$\begin{aligned} H_k &= V_{k-1}^\top \cdots V_{k-m}^\top H_{k-m} V_{k-m} \cdots V_{k-1} \\ &\quad + V_{k-1}^\top \cdots V_{k-m+1}^\top \frac{s_{k-m} s_{k-m}^\top}{s_{k-m}^\top y_{k-m}} \\ &\quad \times V_{k-m+1} \cdots V_{k-1} + \cdots + \\ &\quad V_{k-1}^\top \frac{s_{k-2} s_{k-2}^\top}{s_{k-2}^\top y_{k-2}} V_{k-1} + \frac{s_{k-1} s_{k-1}^\top}{s_{k-1}^\top y_{k-1}} \end{aligned}$$

と表すことができる. ここで, 上式の H_{k-m} をベクトルとの積が容易な正定値対称な初期行列 $H_k^{(0)}$ で置き換えれば, $2m$ 本のベクトル $s_{k-1}, \dots, s_{k-m}, y_{k-1}, \dots, y_{k-m}$ のみで近似行列 H_k を構築できる. 遡る回数 m は記憶数と呼ばれ, $m = k$ の場合には元の BFGS 公式に帰着する. 探索方向の計算 (11) では, 一見, 勾配ベクトル g_k と行列 V_i ($i = k-1, \dots, k-m$) の積が必要に見えるが, V_{k-1} の定義 (14) より, ベクトルの内積だけで計算可能である. さらに, Nocedal [2] は探索方向を求める際に計算量を減らす計算方法も提案しており, たとえば, $H_k^{(0)} = I$ の場合には, $2m$ 回程度の内積計算で探索方向が得られる.

初期行列 $H_k^{(0)}$ の選び方としては, 通常は, 単位行列にスケーリングパラメータ $\gamma_k > 0$ を乗じた対角行列 $\gamma_k I$ が選ばれる. γ_k のよく使用される選択法として,

$$\gamma_k^{(1)} = \frac{s_{k-1}^\top s_{k-1}}{s_{k-1}^\top y_{k-1}}, \quad \gamma_k^{(2)} = \frac{s_{k-1}^\top y_{k-1}}{y_{k-1}^\top y_{k-1}} \quad (15)$$

などが挙げられる. 2 点 x_{k-1}, x_k 間の平均ヘッセ行列を

$$G_k = \int_0^1 \nabla^2 f(x_{k-1} + t s_{k-1}) dt$$

によって定義すると, $y_{k-1} = G_k s_{k-1}$ となるため, G_k を正則行列であると仮定すると (15) はそれぞれ,

$$\gamma_k^{(1)} = \frac{s_{k-1}^\top s_{k-1}}{s_{k-1}^\top G_k s_{k-1}}, \quad \gamma_k^{(2)} = \frac{y_{k-1}^\top G_k^{-1} y_{k-1}}{y_{k-1}^\top y_{k-1}}$$

と表すことができる. したがって, $\gamma_k^{(1)}$ は G_k のレイリー商の逆数となっており, $\gamma_k^{(2)}$ は G_k^{-1} のレイリー商となっていることがわかる. したがって, どちらの選択でも, 粗い近似ではあるものの, $\gamma_k I \approx \nabla^2 f(x_k)^{-1}$

となっている。

3.2 メモリーレス準ニュートン法

メモリーレス準ニュートン法は Shanno [3] によって提案された方法で、その探索方向は、準ニュートン法の更新公式において、一つ前の近似行列を単位行列 I 、もしくはスケーリングパラメータ $\gamma_k^{(1)} > 0$ を乗じた対角行列 $\gamma_k^{(1)} I$ で置き換えることで定義される。たとえば、BFGS 公式に基づくメモリーレス準ニュートン法⁵ を考えると、BFGS 公式 (13) において、 $H_{k-1} = \gamma_k I$ とおき、式 (11) に代入することで、探索方向は以下で与えられる：

$$d_k = -\gamma_k g_k + \left(\frac{\gamma_k g_k^\top y_{k-1}}{s_{k-1}^\top y_{k-1}} - \left(1 + \frac{\gamma_k y_{k-1}^\top y_{k-1}}{s_{k-1}^\top y_{k-1}} \right) \times \frac{g_k^\top s_{k-1}}{s_{k-1}^\top y_{k-1}} \right) s_{k-1} + \frac{\gamma_k g_k^\top s_{k-1}}{s_{k-1}^\top y_{k-1}} y_{k-1}. \quad (16)$$

この探索方向は、上述の記憶制限準ニュートン法において、記憶数 $m = 1$ 、初期行列 $H_k^{(0)} = \gamma_k I$ とした場合と一致する。スケーリングパラメータ γ_k の選択法としては、記憶制限準ニュートン法と同様に式 (15) が用いられることが多いが、式 (15) を用いたメモリーレス準ニュートン法は、一般の目的関数に対する大域的収束性を保証することが困難である。そのため、近年では、大域的な収束性を保証するために修正されたメモリーレス準ニュートン法に注目が集まっている ([4-8] を参照)。

たとえば、Nakayama et al. [7] ではスペクトラルスケーリングセカント（以下、SSセカント）条件 [9] に基づいたメモリーレス準ニュートン法を提案している。SSセカント条件では数値的な安定性を高めるため、近似行列 B_k はスケーリングパラメータを乗じたヘッセ行列 $\gamma_k \nabla^2 f(x_k)$ を近似している。セカント条件の導出に用いた式 (12) の両辺に γ_k を乗じて $B_k \approx \gamma_k \nabla^2 f(x_k)$ とすれば、SSセカント条件：

$$B_k s_{k-1} = \gamma_k y_{k-1} \quad \text{または} \quad s_{k-1} = H_k(\gamma_k y_{k-1})$$

が得られる。SSセカント条件に基づいたメモリーレスBFGS法⁶の探索方向は

$$d_k = -g_k + \left(\frac{g_k^\top y_{k-1}}{s_{k-1}^\top y_{k-1}} - \left(\hat{\gamma}_k + \frac{y_{k-1}^\top y_{k-1}}{s_{k-1}^\top y_{k-1}} \right) \times \frac{g_k^\top s_{k-1}}{s_{k-1}^\top y_{k-1}} \right) s_{k-1} + \frac{g_k^\top s_{k-1}}{s_{k-1}^\top y_{k-1}} y_{k-1} \quad (17)$$

⁵ これをメモリーレスBFGS法と呼び、ほかの方法も同様の呼び方を採用する。

⁶ 実際にはBFGS公式を含んだ公式族であるBroyden公式族に基づいてメモリーレス準ニュートン法を提案している。

で与えられる。ここで、 $\hat{\gamma}_k = 1/\gamma_k$ である。式 (17) は $\hat{\gamma}_k \geq \theta_k/\gamma_k^{(2)}$ ならば、降下条件を満たすことが証明されている。ただし、 $\theta_k \in [\theta_{\min}, \theta_{\max}]$ ($0 < \theta_{\min} \leq 1 \leq \theta_{\max} < 2$) はパラメータである。

3.3 Barzilai-Borwein (BB) 法

Barzilai-Borwein 法（以下、BB法）は、その名のとおり Barzilai and Borwein [10] によって提案された方法である。BB法では準ニュートン法の近似行列を単位行列のスカラー倍、つまり、 $B_k = \lambda_k^{(1)} I$ または $H_k = \lambda_k^{(2)} I$ によって定める。上述のとおり、準ニュートン法では近似行列がセカント条件を満たすように選択されるのが一般的であるが、BB法の場合は選択の自由度の低さ⁷ からセカント条件を満たす $\lambda_k^{(1)}$ （または $\lambda_k^{(2)}$ ）は選択できない。そのため、残差が最も小さくなるようなスカラーを選択する、つまり

$$\lambda_k^{(1)} = \operatorname{argmin}_{\lambda > 0} \{ \|\lambda I s_{k-1} - y_{k-1}\|^2 \}, \quad \text{または} \\ \lambda_k^{(2)} = \operatorname{argmin}_{\lambda > 0} \{ \|s_{k-1} - \lambda I y_{k-1}\|^2 \}$$

により $\lambda_k^{(1)}$ （または $\lambda_k^{(2)}$ ）を決定する。この問題は、簡単に解くことができ、

$$\lambda_k^{(1)} = \frac{s_{k-1}^\top y_{k-1}}{s_{k-1}^\top s_{k-1}}, \quad \lambda_k^{(2)} = \frac{s_{k-1}^\top y_{k-1}}{y_{k-1}^\top y_{k-1}}$$

で与えられる⁸。このとき、探索方向はそれぞれ

$$d_k = -\frac{1}{\lambda_k^{(1)}} g_k, \quad d_k = -\lambda_k^{(2)} g_k$$

となる。探索方向からわかるとおり、BB法はスケーリング付きの最急降下法とみなすこともできる⁹。特に、直線探索を行わない場合のBB法は、ステップ幅として $\alpha_k = 1/\lambda_k^{(1)}$ または $\alpha_k = \lambda_k^{(2)}$ とした最急降下法と考えることもできる。さらに（メモリーレス準ニュートン法が記憶数 $m = 1$ の記憶制限準ニュートン法だったのに対し）、BB法は記憶数 $m = 0$ とし、スケーリングパラメータ (15) を用いた場合の記憶制限準ニュートン法に一致することを注意しておく。

⁷ セカント条件を B_k に対する方程式と考えよう。 B_k が通常の対称行列の場合、変数の数は $n(n+1)/2$ 。方程式の本数が n 本であり、優決定問題となる。一方、BB法の場合は変数の数が一つだけとなるため、劣決定問題となり、一般的には解をもたないこととなる。

⁸ BB法においては B_k のセカント条件を考えるか、 H_k のセカント条件を考えるかで2通りの異なる方法が導出される。

⁹ BB法は最急降下法の一つとして扱われることもあれば、準ニュートン法の一つとして扱われることもある。ここでは、準ニュートン法の一つとして扱うこととする。

目的関数が狭義凸 2 次関数 (4) の場合には直線探索を使用しない BB 法は、正確な直線探索 (5) を用いた最急降下法 (8) と対比して考えることができるため、多くの研究者によって狭義凸 2 次関数に対する、直線探索を使用しない BB 法の収束性が研究されている [11–15]. 特に、文献 [13] では、非常に強い仮定の下ではあるが、局所的な超 1 次収束性を証明しており、BB 法の有用性を裏づけている。また、一般の目的関数に対して直線探索を用いた BB 法についても盛んに研究が行われており、非単調直線探索を用いた BB 法 [16] や、BB 法の変種 [14, 17] などが提案されている。

4. 非線形共役勾配法

4.1 一般的な非線形共役勾配法

共役勾配法の歴史は、1952 年に Hestenes and Stiefel [18] によって開発された線形共役勾配法まで遡る。線形共役勾配法は正定値対称行列を係数行列に持つ連立一次方程式系 $Ax = b$ を解くための反復法であり、現在ではその変種も含めて連立一次方程式系に対する主流な数値解法の一つとなっている。一方、Fletcher and Reeves [19] は線形共役勾配法を狭義凸 2 次関数 (4) の最小化問題に対する反復法と捉えて、線形共役勾配法における残差ベクトル $Ax - b$ を勾配ベクトル $g(x)$ で置き換えることにより非線形共役勾配法を提案している。

一般的に、非線形共役勾配法の探索方向は

$$d_k = -g_k + \beta_k d_{k-1} \quad (18)$$

で与えられる。ここで β_k は非線形共役勾配法を特徴づけるパラメータであり、通常 β_k は、目的関数 f が狭義凸 2 次関数 (4) で、正確な直線探索 (5) が用いられた場合は、線形共役勾配法に一致するように選ばれる。線形共役勾配法では β_k は一意に決定されるが、非線形共役勾配法では様々な選択が可能であり、その選択法によって数値的な効率性が大きく異なる。そのため、 β_k の選択法の研究が盛んに行われており、Fletcher–Reeves (FR), Hestenes–Stiefel (HS), Polak–Ribière (PR), Dai–Yuan (DY) 公式などがよく知られている (たとえば、文献 [1, 20, 21] などを参照) :

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \quad \beta_k^{HS} = \frac{g_k^\top y_{k-1}}{d_{k-1}^\top y_{k-1}},$$

$$\beta_k^{PR} = \frac{g_k^\top y_{k-1}}{\|g_{k-1}\|^2}, \quad \beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^\top y_{k-1}}.$$

上記四つの方法は分子 2 種類、分母 2 種類の 4 種

類で考えることができるが、分子の種類で分類するのが妥当である。実際、正確な直線探索の場合には $\alpha_{k-1} g_k^\top d_{k-1} = g_k^\top s_{k-1} = 0$ となるため、すべての k に対して $\|g_{k-1}\|^2 = -g_{k-1}^\top d_{k-1} = d_{k-1}^\top y_{k-1}$ が成立する。したがって、 $\beta_k^{FR} = \beta_k^{DY}$ と $\beta_k^{HS} = \beta_k^{PR}$ が成り立つ。また、FR 法と DY 法は直線探索に条件を課すことで降下条件が保証されるのに対し、HS 法と PR 法は必ずしも降下条件を満たすとは限らないという欠点がある。さらに、大域的な収束性を証明する方法も FR 法と DY 法、HS 法と PR 法でそれぞれ分類可能である。なお、上記 4 種類の方法の大域的収束性に関しては文献 [20] が詳しい。非線形共役勾配法は β_k の選択によって数値的な効率性が異なるが、上記 4 種類の中では HS 法と PR 法が効果的であることが知られている。上記 4 種類のほかにもさまざまな β_k の選択法が提案されているが、中でも β_k^{HS} の修正法として捉えられるものが多い。たとえば、Hager and Zhang [22] や Dai and Liao [23] の方法はそれぞれ

$$\beta_k^{HZ} = \beta_k^{HS} - \lambda \frac{\|y_{k-1}\|^2}{(d_{k-1}^\top y_{k-1})^2} g_k^\top d_{k-1},$$

$$\beta_k^{DL} = \beta_k^{HS} - t \frac{g_k^\top s_{k-1}}{d_{k-1}^\top y_{k-1}}$$

で与えられる。ただし、 $\lambda > 1/4$ と $t \geq 0$ はパラメータである。どちらの方法も正確な直線探索を用いた場合には HS 法と一致することを注意しておく。

ここで、記憶数 $m = 1$ とし、 $H_k^{(0)} = I$ とした記憶制限準ニュートン法、つまりメモリーレス準ニュートン法 (16) を考えてみよう。正確な直線探索が用いられた場合、 $g_k^\top s_{k-1} = 0$ より、式 (16) は

$$d_k = -g_k + \frac{g_k^\top y_{k-1}}{d_{k-1}^\top y_{k-1}} d_{k-1}$$

となる。これは β_k^{HS} を用いた非線形共役勾配法に他ならない。さらに、正確な直線探索を用いない場合には、探索方向 (16) はパラメータ β_k や ζ_k を適当に定義すれば

$$d_k = -g_k + \beta_k d_{k-1} + \zeta_k y_{k-1}$$

と表すことができる。この場合には非線形 3 項共役勾配法として捉えることも可能である。メモリーレス準ニュートン法は非線形 (3 項) 共役勾配法と非常に関係性が強く、メモリーレス準ニュートン法を基として、非線形 (3 項) 共役勾配法を導出・提案している論文も数多く存在する (たとえば、文献 [4, 24, 25] などを

参照).

4.2 非線形 3 項共役勾配法

前節でもメモリーレス準ニュートン法との関係に基づいて、非線形 3 項共役勾配法について言及したが、本節では Narushima et al. [24] の方法を紹介することとする。彼らは、次の非線形 3 項共役勾配法の族：

$$d_k = \begin{cases} -g_k, & g_k^\top p_k = 0, \\ -g_k + \beta_k d_{k-1} + \zeta_k p_k, & g_k^\top p_k \neq 0 \end{cases} \quad (19)$$

を提案している。ただし、 $p_k \in \mathbb{R}^n$ をパラメータベクトルとし、

$$\zeta_k = -\beta_k \frac{g_k^\top d_{k-1}}{g_k^\top p_k} p_k$$

とする。式 (19) は、正確な直線探索を用い、 $g_k^\top p_k \neq 0$ の場合には元の非線形共役勾配法 (18) に帰着されることを注意しておく。ここで、式 (19) の左側から g_k^\top をかけると、

$$g_k^\top d_k = -\|g_k\|^2 \quad (20)$$

となることがわかる。したがって、非線形 3 項共役勾配法 (19) はパラメータ β_k の選択にかかわらず式 (20) の意味で降下条件を満たす。通常の共役勾配法の場合、数値的な効率のよいパラメータである β_k^{HS} や β_k^{PR} は降下条件を満たすとは限らないが、非線形 3 項共役勾配法 (19) の場合は β_k^{HS} や β_k^{PR} を用いても降下条件が保証される。一方、探索方向 (19) は $g_k^\top p_k \neq 0$ の場合には

$$d_k = -g_k + \beta_k \left(I - \frac{p_k g_k^\top}{g_k^\top p_k} \right) d_{k-1}$$

と書き換えることできる。これは、通常の共役勾配法 (18) の第二項を射影行列 $I - p_k g_k^\top / g_k^\top p_k$ で射影していることを意味する。ここで、 $I - p_k g_k^\top / g_k^\top p_k$ は p_k に沿った $\text{Span}\{g_k\}$ の直交補空間への射影行列であり、特に、 $p_k = g_k$ とした場合には正射影行列となる。なお、パラメータベクトル p_k の選択法としては、 g_k や y_{k-1} が用いられることが多く、 β_k や p_k の選択によって、式 (19) は文献 [25–28] によって提案された方法に帰着する。

前節でも述べたが、一般的に非線形 3 項共役勾配法とメモリーレス準ニュートン法は非常に近い関係の方法である。両者に特にはっきりした境界はないが、あえて挙げるとするならば、メモリーレス準ニュートン法は探索方向を $d_k = -H_k g_k$ として表したときに、 H_k

に対して対称性やセカント条件を意識していることが多いが、非線形 3 項共役勾配法ではそうではないことが多い。また、非線形 3 項共役勾配法の場合は (2 項の) 非線形共役勾配法との関係性を意識していることが多いというのも特徴であるといえるだろう。

5. 数値実験

本節では前節までに紹介してきた方法のうち、下記の四つの方法の数値実験結果を報告する¹⁰：

BB : Barzilai–Borwein 法 ($\lambda_k = \lambda_k^{(2)}$),

HS : 非線形共役勾配法 (18)

$$(\beta_k = \max\{\beta_k^{HS}, 0\}),$$

3PR : 非線形 3 項共役勾配法 (19)

$$(\beta_k = \max\{\beta_k^{PR}, 0\}, p_k = g_k),$$

mless : メモリーレス準ニュートン法 (17)

$$(\hat{\gamma}_k = 1/\lambda_k^{(2)}),$$

上記の方法の実装においては非線形共役勾配法のソフトウェアである CG-DESCENT [29] を修正してコードを作成し、直線探索などの設定は CG-DESCENT の設定にならって Wolfe 条件 (6), (7) を満たすようなステップ幅が選択されている。ただし、HS は必ずしも降下方向を生成するとは限らないため、降下方向を生成しない場合は最急降下方向 ($d_k = -g_k$) にスイッチしている。収束判定条件は

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

を使用しており、実行時間が 600 (秒) を超えた場合もアルゴリズムを停止している。テスト問題は CUTEr 問題集 [30] から 135 問を選んで実験を行った。

本稿において、われわれは、各方法の比較を行うために、Dolan and Moré [31] で提案されたパフォーマンスプロファイルを用いた。各方法のパフォーマンスプロファイル $P(\tau)$ の $\tau = \tau$ のときの値は、各問題に対する最速の方法の τ 倍以内に、その方法が求解できた問題の割合を表している。 $\tau = 1$ のときの値は、その方法がすべて方法の中で、最も速く解くことができた問題の割合を表しており、一方、 τ が十分大きいときは、解くことのできた問題の割合を表すこととなる。どの τ においても、1 に近いほうが好ましく、複数の数値解法を比較する場合、パフォーマンスプロファイルが上

¹⁰非線形 (3 項) 共役勾配法では、大域的収束性を保証するために、パラメータ β_k を修正することが多い。本稿における実験では、修正されたパラメータを使用している。

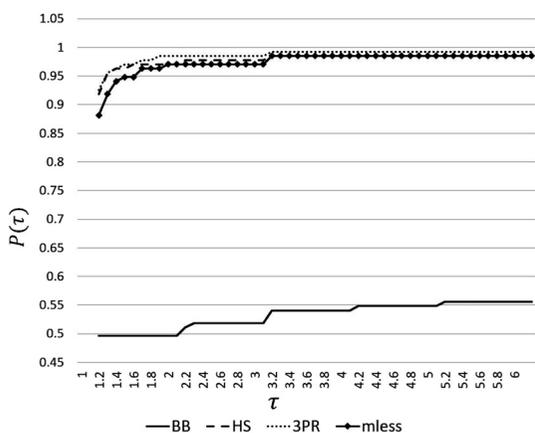


図 1 パフォーマンスプロファイル その 1

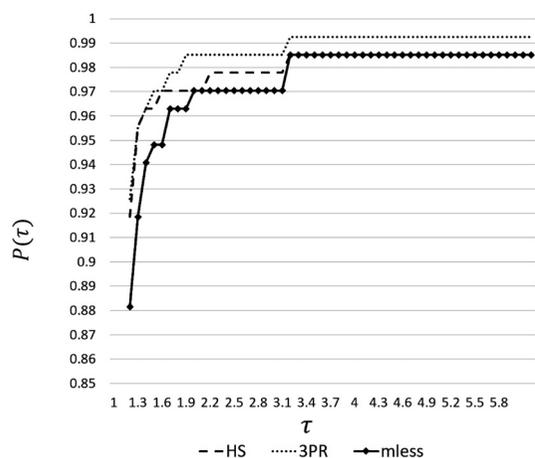


図 2 パフォーマンスプロファイル その 2

に位置する方法ほど効率がよいと考えることができる。

図 1 では、実験した四つの方法のパフォーマンスプロファイルが描かれている。図から明らかのように BB の効率はほかの三つの方法と比べて劣っていることがわかる。次に、BB 以外の方法を比較するために、BB を除いた三つの方法でパフォーマンスプロファイルを作成し、図 2 に掲載した。図 2 から、計算効率は $3PR > HS > mless$ となっていることがわかる。

6. おわりに

本稿では、近年注目されている勾配法について、記憶制限準ニュートン法と非線形共役勾配法の 2 通りに分類して紹介してきた。前者は準ニュートン法の立場から情報を減らしていく方法であるのに対し、後者は最急降下法に項を加えて情報を加えていく方法であると考えられる。考え方の違う両者であるが、メモリーレス準ニュートン法と非線形共役勾配法（特に

HS 法) の関係を見ればわかるとおり、非常に関係性の深い方法であり、記憶制限準ニュートン法は準ニュートン法と非線形共役勾配法をつなぐ、中間的な方法であると考えられる。その一方、その取扱いに関しては異なる部分も存在する。記憶制限準ニュートン法は記憶数が大きくなれば、準ニュートン法に近づくため、ステップ幅 $\alpha_k = 1$ が自然であり、直線探索は大域的な収束性を保証するための手段と考えられる。その一方、非線形共役勾配法は、元となった線形共役勾配法が正確な直線探索を用いているため、直線探索ありきの方法であると考えられる。実際に、数値実験などでは準ニュートン法や記憶制限準ニュートン法で記憶数を大きくとった場合には、直線探索はそれほど厳しくなく¹¹実行したほうがよい結果が得られることが多いのに対し、メモリーレス準ニュートン法や非線形共役勾配法では直線探索に手間をかけたほうがよい結果が得られることが多い。

非線形共役勾配法は元々古くからある方法であるが、大規模な問題に対して有効な方法として、2000 年ごろから特に活発に研究されており、非線形共役勾配法に適した直線探索法の発展も相まって、CG-DESCENT のような高性能なソフトウェアも開発されてきている。その一方、メモリーレス準ニュートン法は、ごく最近まであまり注目を集めてきてはいなかったが、非線形共役勾配法と同様に、1 回の反復に非常に少ない計算量しか必要としないうえに、近似行列を考えた場合には、その対称性や正定値性なども活用することが可能である。このような性質は、勾配法を制約付き最適化問題や微分不可能な関数を含む最適化問題に拡張する際に非常に重要となる。そのような観点からも、メモリーレス準ニュートン法は今後の発展が期待される方法であるといえるだろう。

謝辞 本稿執筆の機会を与えていただいた本特集オーガナイザーの後藤順哉先生と「オペレーションズ・リサーチ」編集委員の皆様へ感謝いたします。本研究の一部は JSPS 科研費基盤研究 (C)17K00039 の助成を受けて実施されている。

参考文献

- [1] J. Nocedal and S. J. Wright, *Numerical Optimization, Springer Series in Operations Research*, 2nd edition, Springer, 2006.

¹¹たとえば、Wolfe 条件 (6), (7) を使用する場合、パラメータ δ や σ の値を変えることで直線探索の厳しさを調節することができる。

- [2] J. Nocedal, "Updating quasi-Newton matrices with limited storage," *Mathematics of Computation*, **35**, pp. 773–782, 1980.
- [3] D. F. Shanno, "Conjugate gradient methods with inexact searches," *Mathematics of Operations Research*, **3**, pp. 244–256, 1978.
- [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**, pp. 296–320, 2013.
- [5] 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**, pp. 209–224, 2015.
- [6] 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*, **61**, pp. 53–70, 2018.
- [7] S. Nakayama, Y. Narushima and H. Yabe, "Memoryless quasi-Newton methods based on spectral-scaling Broyden family for unconstrained optimization," *Journal of Industrial and Management Optimization*, to appear, DOI: 10.3934/jimo.2018122.
- [8] S. Nakayama, "A hybrid method of three-term conjugate gradient method and memoryless quasi-Newton method for unconstrained optimization," *SUT Journal of Mathematics*, **54**, pp. 79–98, 2018.
- [9] W. Y. Cheng and D. H. Li, "Spectral scaling BFGS method," *Journal of Optimization Theory and Applications*, **146**, pp. 305–319, 2010.
- [10] J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," *IMA Journal of Numerical Analysis*, **8**, pp. 141–148, 1988.
- [11] Y. H. Dai and L. Z. Liao, "R-linear convergence of the Barzilai and Borwein gradient method," *IMA Journal of Numerical Analysis*, **22**, pp. 1–10, 2002.
- [12] Y. H. Dai and Y. Yuan, "Analysis of monotone gradient methods," *Journal of Industrial and Management Optimization*, **1**, pp. 181–192, 2005.
- [13] A. Friedlander, J. M. Martinez, B. Molina and M. Raydan, "Gradient method with retards and generalizations," *SIAM Journal on Numerical Analysis*, **36**, pp. 275–289, 1999.
- [14] Y. Narushima, T. Wakamatsu and H. Yabe, "Extended Barzilai–Borwein method for unconstrained minimization problems," *Pacific Journal of Optimization*, **6**, pp. 591–613, 2010.
- [15] M. Raydan, "On the Barzilai and Borwein choice of steplength for the gradient method," *IMA Journal of Numerical Analysis*, **13**, pp. 321–326, 1993.
- [16] M. Raydan, "The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem," *SIAM Journal on Optimization*, **7**, pp. 26–33, 1997.
- [17] Y. H. Dai, W. W. Hager, K. Schittkowski and H. Zhang, "The cyclic Barzilai–Borwein method for unconstrained optimization," *IMA Journal of Numerical Analysis*, **26**, pp. 604–627, 2006.
- [18] M. R. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," *Journal of Research of the National Bureau of Standards*, **49**, pp. 409–436, 1952.
- [19] R. Fletcher and C. M. Reeves, "Function minimization by conjugate gradients," *The Computer Journal*, **7**, pp. 149–154, 1964.
- [20] W. W. Hager and H. Zhang, "A survey of nonlinear conjugate gradient method," *Pacific Journal of Optimization*, **2**, pp. 35–58, 2006.
- [21] Y. Narushima and H. Yabe, "A survey of sufficient descent conjugate gradient methods for unconstrained optimization," *SUT Journal of Mathematics*, **50**, pp. 167–203, 2014.
- [22] W. W. Hager and H. Zhang, "A new conjugate gradient method with guaranteed descent and an efficient line search," *SIAM Journal on Optimization*, **16**, pp. 170–192, 2005.
- [23] Y. H. Dai and L. Z. Liao, "New conjugacy conditions and related nonlinear conjugate gradient methods," *Applied Mathematics and Optimization*, **43**, pp. 87–101, 2001.
- [24] Y. Narushima, H. Yabe and J. A. Ford, "A three-term conjugate gradient method with sufficient descent property for unconstrained optimization," *SIAM Journal on Optimization*, **21**, pp. 212–230, 2011.
- [25] 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**, pp. 629–640, 2006.
- [26] W. Cheng, "A two-term PRP-based descent method," *Numerical Functional Analysis and Optimization*, **28**, pp. 1217–1230, 2007.
- [27] L. Zhang, W. Zhou and D. H. Li, "Global convergence of a modified Fletcher–Reeves conjugate gradient method with Armijo-type line search," *Numerische Mathematik*, **104**, pp. 561–572, 2006.
- [28] L. Zhang, W. Zhou and D. H. Li, "Some descent three-term conjugate gradient methods and their global convergence," *Optimization Methods and Software*, **22**, pp. 697–711, 2007.
- [29] W. W. Hager and H. Zhang, "Algorithm 851: CG_DESCENT, a conjugate gradient method with guaranteed descent," *ACM Transactions on Mathematical Software*, **32**, pp. 113–137, 2006.
- [30] 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**, pp. 373–394, 2003.
- [31] E. D. Dolan and J. J. Moré, "Benchmarking optimization software with performance profiles," *Mathematical Programming*, **91**, pp. 201–213, 2002.