

無制約最適化問題に対する アルゴリズムの最前線 —非線形共役勾配法を中心に—

成島 康史

無制約最適化問題に対する数値解法は古くから研究されており、ニュートン法や準ニュートン法が有効な方法として知られている。しかし、近年、大規模な無制約最適化問題に対する数値解法として、非線形共役勾配法が盛んに研究されている。本稿ではここ最近の非線形共役勾配法の研究の中でもとりわけ注目の高い数値解法のいくつかを紹介し、その性質や有効性を議論する。

キーワード：無制約最適化問題、非線形共役勾配法、非線形 3 項共役勾配法、CG-DESCENT

1. はじめに

本稿では無制約最適化問題：

$$\text{minimize } f(x)$$

に対する最近の話題を取り上げる。ただし、以降では目的関数 $f: \mathbf{R}^n \rightarrow \mathbf{R}$ は十分滑らかであるとし、目的関数の勾配 ∇f を g で表すこととする。無制約最適化問題に対する数値解法として、反復法が広く用いられている。反復法は任意の初期点 $x_0 \in \mathbf{R}^n$ から出発し、反復式

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

により点列を更新する。ここで、 $\alpha_k > 0$ 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 の選択によりさまざまな方法が提案されており、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。なかでも、ニュートン法や準ニュートン法は変数の規模が小規模、または中規模な無制約最適化問題に対して非常に有効な方法であり、さまざまなソフトウェアに組み込まれている。一方、近年の情報技術の発展に伴い、大規模な問題を解く必要性が増してきているが、問題が大規模な場合はニュートン法や準ニュートン法は目的関数のヘッセ行列やその近似行列を利用する必要があるため、記憶容量や計算量の関係で、問題に対して直接適用することができないことがある。そのため、近年、大規模無制約最適化問題に対する数値解法が盛んに研究されている。

大規模な問題に対する数値解法として大きく 2 通りの方法がある。一つ目は目的関数のヘッセ行列の疎性を利用した方法で、ニュートン法と信頼領域法を組み合わせた方法や、スパース準ニュートン法などがある。どちらも局所的に速い収束性を持つ有効な数値解法であるが、あらかじめヘッセ行列（またはその疎性）を求めておく必要があったり、ニュートン方程式をどのように解くかなどといった問題も残されている。二つ目は行列を使用しない方法で記憶制限準ニュートン法、非線形共役勾配法や Barzilai-Borwein 法 (BB 法) などがある。これらの方法は局所的な速い収束性は保証されていないものの、ヘッセ行列の情報が必要なく、各反復ごとの計算量も非常に少ない。そのような理由から、近年行列を使用しない方法が注目を集めており、なかでも、非線形共役勾配法は非常に盛んに研究されている。本稿では、非線形共役勾配法に着目し、最近の研究を紹介する。

2. 非線形共役勾配法

非線形共役勾配法の歴史は、1952 年に Hestenes and Stiefel [11] によって開発された線形共役勾配法までさかのぼる。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、1964 年に Fletcher and Reeves [5] は

$$\text{Solve } Ax = b \Leftrightarrow \min f(x) = \frac{1}{2}x^T Ax - b^T x$$

なるしま やすし

横浜国立大学

〒240-8501 神奈川県横浜市保土ヶ谷区常盤台 79-4

の関係性をもとに、残差ベクトル $Ax - b$ を勾配ベクトル $g(x)$ で置き換えることで線形共役勾配法を無制約最適化問題に拡張した。これが最初の非線形共役勾配法 (Nonlinear Conjugate Gradient Method, 以下 CG 法と呼ぶ¹) の研究となっている。CG 法のアルゴリズムは以下で与えられる。

アルゴリズム CG.

Step 0. 初期点 x_0 を与え, $k = 0$ として Step 1 へ.

Step 1. 終了判定条件を満たしていたら停止する.

Step 2. 探索方向を

$$d_k = \begin{cases} -g_k, & \text{for } k = 0, \\ -g_k + \beta_k d_{k-1}, & \text{for } k \geq 1, \end{cases} \quad (2)$$

によって計算する.

Step 3. 直線探索によりステップ幅 α_k を計算し, (1)

により点列 $\{x_k\}$ を更新する.

Step 4. $k := k + 1$ として Step 1 へ戻る.

ここで, Step 2 において $g_k \equiv g(x_k)$ であり, β_k は CG 法を特徴づけるパラメータである. 通常, パラメータ β_k は目的関数が狭義凸 2 次関数で, かつ正確な直線探索の場合には線形共役勾配法に一致するように選ばれる. 一方, 目的関数が一般の非線形関数の場合, β_k の選択法によってアルゴリズムの数値的な振る舞いが大きく異なる. そのため, 有効な β_k の選択法に対して多くの研究が行われており, よく知られた公式としては Fletcher-Reeves (FR), Hestenes-Stiefel (HS), Polak-Ribière (PR), Dai-Yuan (DY) などがある:

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

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

ただし, $y_{k-1} = g_k - g_{k-1}$ である. これらの方法の大域的収束性についても多くの研究があり, それらは Hager and Zhang のサーベイ論文 [8] が詳しい. ここで, 簡単のために β_k^{FR} を用いた CG 法を FR 法などと呼ぶこととし, ほかも同様の表記を用いることとする. 上記 4 つの選択法以外にも多くの β_k の選択法が提案されており, 例えば, Dai and Liao [3] は準ニュートン法の考えであるセカント条件に基づいた β_k の選択法を提案している. まず, 彼らは準ニュートン法の探索方

向の式 $B_k d_k = -g_k$ とセカント条件 $B_k s_{k-1} = y_{k-1}$ を利用して関係式

$$d_k^T y_{k-1} = d_k^T (B_k s_{k-1}) = (B_k d_k)^T s_{k-1} = -g_k^T s_{k-1}$$

を導いた. ただし, $s_{k-1} = x_k - x_{k-1}$ であり, B_k は $\nabla^2 f(x_k)$ の対称な近似行列とする. 彼らは, この条件に非負パラメータ t を導入した条件 $d_k^T y_{k-1} = -t g_k^T s_{k-1}$ の探索方向 d_k に (2) を代入し逆算することで

$$\beta_k^{DL} = \frac{g_k^T (y_{k-1} - t s_{k-1})}{d_{k-1}^T y_{k-1}}$$

を導いた. DL 法は上記の 4 つの方法よりも実用上有効であることが知られている. しかしながら, 必ずしも降下方向を生成するとは限らないという弱点がある. ここで, β_k^{DL} は β_k^{HS} に修正項 $-t g_k^T s_{k-1} / d_{k-1}^T y_{k-1}$ を加えたものであることを注意しておく. β_k^{DL} のほかにも多くの β_k の選択法が提案されているが, それらは, 上記 4 つのうちのどれかに関連していることが多い.

ここで, 再度, 上記 4 つの方法に注目する. このとき, β_k は分子 2 種類, 分母 2 種類の 4 通りであるが, 分子の種類によって分類するのが妥当であろう. 実際, 正確な直線探索の場合には $g_k^T d_{k-1} = 0$ となるため, $d_{k-1}^T y_{k-1} = \|g_{k-1}\|^2$ 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. さらに, HS 法と PR 法はほかの 2 つよりも数値的な性能が良いことが知られているが, 必ずしも降下方向を生成するとは限らないという弱点がある. 一方, FR 法と DY 法は直線探索において適当な条件を課すことで降下方向を生成するが, 実用上, HS 法や PR 法ほど有用ではない. そのため, 最近, 直線探索によらず十分な降下条件を満たすような非線形共役勾配法が盛んに研究されている. なお, 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 で

$$g_k^T d_k \leq -c \|g_k\|^2$$

を満たすことを意味する.

節の最後に直線探索について簡単に言及しておく. 通常, ニュートン法や準ニュートン法などの方法は, $\alpha_k = 1$ が自然であるため, 大域的収束性を妨げない程度の直線探索を行うことが望ましい. 一方, CG 法の場合, もととの線形共役勾配法が正確な直線探索を行う方法であるため, 同じ CG 法であっても直線探索によって大きく効率が変わる. 特に, 直線探索に手間をかけたほうが, 全体として効率的であることが多

¹ 通常, CG 法は線形共役勾配法を指すことが多いが, 本稿では主題となっている非線形共役勾配法を CG 法と呼ぶこととする.

い。そのため、さまざまな直線探索条件が考案されており、なかでも、最も一般的なのは下記の Wolfe 条件である：

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k, \quad (3)$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k. \quad (4)$$

ここで、 δ と σ は $0 < \delta < \sigma < 1$ を満たす定数とする。また、Wolfe 条件を強めた強い Wolfe 条件：(3) と

$$|g(x_k + \alpha_k d_k)^T d_k| \leq \sigma |g_k^T d_k|, \quad (5)$$

を課す場合も多い。さらに、Wolfe 条件の 1 式目である (3) のみ (Armijo 条件と呼ぶ) を用いることもある。通常、直線探索条件を満たす α_k を見つけるために、2 分法などを利用するのが一般的であるが、CG 法の場合には通常の 2 分法はあまり有効でないことが多く、2 次補間や挟み込みなどを利用することが好ましい。そのため、CG 法とそれに適した直線探索を合わせて研究することも行われている (例えば、[2, 7] など)。

3. 3 項 CG 法・スケーリング CG 法

近年、直線探索に依存せず常に十分な降下方向を生成する CG 法が注目されている。十分な降下方向を保証するための方法はいくつかあるが、その一つとして CG 法の探索方向 (2) を修正するという考えがある。例えば、Zhang et al. [15] は FR 法の第一項を修正してスケーリング FR 法：

$$d_k = -\frac{d_{k-1}^T y_{k-1}}{\|g_{k-1}\|^2} g_k + \beta_k^{FR} d_{k-1} \quad (6)$$

を提案している²。ここで、帰納法を用いると直線探索によらず、 $g_k^T d_k = -\|g_k\|^2$ を満たすことが簡単に確認できる。これは $c = 1$ とした十分な降下方向にほかならない。さらに、Zhang et al. [16, 17] は下記の 3 項 PR 法と 3 項 HS 法

$$d_k = -g_k + \beta_k^{PR} d_{k-1} - \frac{g_k^T d_{k-1}}{\|g_{k-1}\|^2} y_{k-1}, \quad (7)$$

$$d_k = -g_k + \beta_k^{HS} d_{k-1} - \frac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}} y_{k-1} \quad (8)$$

を提案しており、Cheng [1] は修正 PR 法：

$$d_k = -g_k + \beta_k^{PR} \left(I - \frac{g_k g_k^T}{\|g_{k-1}\|^2} \right) d_{k-1} \quad (9)$$

を提案している。これらの方法もスケーリング FR 法と同様に $g_k^T d_k = -\|g_k\|^2$ を満たしている。

一方、Narushima et al. [12] は (6)–(9) を含むよう

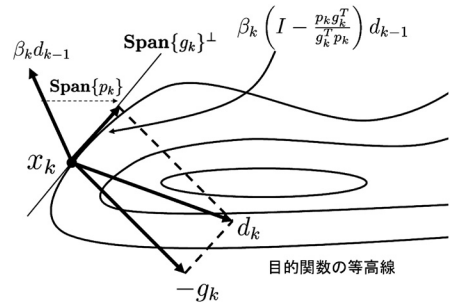


図 1 探索方向のイメージ図

な 3 項 CG 法の族

$$d_k = -g_k + \beta_k (g_k^T p_k)^\dagger (g_k^T p_k) d_{k-1} - \beta_k (g_k^T p_k)^\dagger (g_k^T d_{k-1}) p_k \quad (10)$$

を提案している。ただし、 p_k は任意の n 次ベクトルであり、 a^\dagger は

$$a^\dagger = \begin{cases} \frac{1}{a} & a \neq 0, \\ 0 & a = 0 \end{cases}$$

であるような一般化逆数とする。ここで、(10) は、直線探索やパラメータ β_k の選択によらず、 $g_k^T d_k = -\|g_k\|^2$ の意味で十分な降下方向を満たしている。(10) は $\beta_k = \beta_k^{FR}$ かつ $p_k = g_k$ のときには (6) に、 $\beta_k = \beta_k^{PR}$ かつ $p_k = g_k$ のときには (9) に帰着される。また、 $g_k^T y_{k-1} \neq 0$ と仮定した場合には、 $\beta_k = \beta_k^{PR}$ かつ $p_k = y_{k-1}$ のときには (7) に、 $\beta_k = \beta_k^{HS}$ かつ $p_k = y_{k-1}$ のときには (8) にそれぞれ帰着される。一方、探索方向 (10) は $g_k^T p_k \neq 0$ の場合には

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

と書きかえることができる。これは、CG 法の探索方向 (2) の第二項を射影行列 $I - p_k g_k^T / g_k^T p_k$ で射影していることを意味する (図 1 参照)。ここで、 $I - p_k g_k^T / g_k^T p_k$ 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。

また、Narushima et al. は

$$\|d_k\|^2 \leq \psi_k^2 \|d_{k-1}\|^2 + \|g_k\|^2, \\ \psi_k = \beta_k \|g_k\| \|p_k\| (g_k^T p_k)^\dagger$$

と表せることを利用し、3 項 CG 法に対する性質を定義し、さらに、その大域的収束性を与えている。

Property A. 3 項 CG 法 (1), (10) を考える。さら

² 以降では断りがない限り $d_0 = -g_0$ とする。

に、正の定数 $\gamma, \bar{\gamma}$ が存在して、すべての k に対して $0 < \gamma \leq \|g_k\| \leq \bar{\gamma}$ が成立していると仮定する. このとき、すべての k に対し、 $|\psi_k| \leq b$ を満たし、さらに $\|s_{k-1}\| \leq \eta$ ならば $|\psi_k| \leq 1/b$ を満たす定数 $b > 1$ と $\eta > 0$ が存在するとき、3項CG法はProperty Aを持つという.

このProperty Aは $0 < \gamma \leq \|g_k\|$ のとき、つまり、収束しない場合には $|\psi_k|$ が有界であり、ステップ $s_{k-1} = x_k - x_{k-1}$ が小さいときには $|\psi_k|$ も十分小さくなるという性質を表しており、この性質を持つ3項CG法に対して以下の大域的収束性を得ることができる.

定理. 初期点 x_0 における単位集合 $\mathcal{L} = \{x \mid f(x) \leq f(x_0)\}$ は有界で、その開凸近傍 \mathcal{N} において目的関数 f は一回連続微分可能、かつ、その勾配 g はLipschitz連続であるとする. $\beta_k \geq 0$ かつProperty Aを満たす3項CG法(1), (10)を考える. さらに、直線探索においてステップ幅 α_k は強いWolfe条件(3), (5)を満たすように選択されるものとする. このとき、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.

さらに、Narushima et al. はこの定理の系として、 $\beta_k = \max\{0, \beta_k^{HS}\}$ かつ $p_k = y_k$ (または $p_k = g_k$) と $\beta_k = \max\{0, \beta_k^{PR}\}$ かつ $p_k = y_k$ (または $p_k = g_k$) とした3項CG法の大域的収束性を与えている.

また、3項CG法(10)に関連する研究もいくつか行われており、例えば、Sugiki et al. [14] はDL法の弱点を補うために、(10)において、 $\beta_k = \beta_k^{DL}$ 、 $p_k = y_{k-1} - ts_{k-1}$ とした3項CG法を提案している.

4. Hager-Zhang 法 (CG-DESCENT)

前節では探索方向を修正することで降下方向を保証する方法を紹介したが、探索方向ではなく、パラメータ β_k を修正することで降下方向を保証する方法も提案されている. Hager and Zhang [7] はパラメータ β_k を修正することで常に十分な降下方向を生成するCG法を提案した:

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

ここで、 λ はパラメータであり、 $\lambda > 1/4$ かつ $d_{k-1}^T y_{k-1} \neq 0$ ならば、HZ法は $g_k^T d_k \leq -(1 - 1/(4\lambda))\|g_k\|^2$ を満たす. さらに、Hager and Zhang は β_k を

$$\beta_k^{HZ+} = \max\{\eta_k, \beta_k^{HZ}\},$$

$$\eta_k = \frac{-1}{\|d_{k-1}\| \min\{\eta, \|g_{k-1}\|\}} \quad (\eta > 0)$$

と修正し、Wolfe条件の下でHZ法の大域的収束性を証明している. さらにHZ法のソフトウェアであるCG_DESCENT [9]を開発し、Web上 (<http://www.math.ufl.edu/~hager/>) で公開している³. Hager and Zhang は、数値誤差の影響で、解の近傍においてArmijo条件を満たすステップ幅 α_k を見つけるのが困難となり、結果として直線探索が失敗してしまうことを指摘し、そのような場合にはWolfe条件(3)-(4)から近似Wolfe条件:

$$-(1 - 2\delta)g_k^T d_k \geq g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k$$

へ直線探索条件を変更することで、CG_DESCENTにおける直線探索の効率性を高めている⁴. このCG_DESCENTは非常に有効なソフトウェアというだけでなく、近年では、新しく提案したCG法の有効性を検証するための比較対象としても広く認知されている.

また、HZ法に関連した研究も盛んに行われている. パラメータ β_k^{HZ} は β_k^{HS} の修正法とみなすことができるため、 β_k^{HS} 以外のパラメータを修正することで、HZ法に倣った方法を提案することができる. 実際、 $\beta_k = g_k^T v_k / u_k$ ($v_k \in \mathbf{R}^n$, $u_k \in \mathbf{R} - \{0\}$) の形式をした β_k の場合、

$$\beta_k = \frac{g_k^T v_k}{u_k} - \lambda \frac{\|v_k\|^2 g_k^T d_{k-1}}{u_k^2}$$

と修正を施すことで、 $g_k^T d_k \leq -(1 - 1/(4\lambda))\|g_k\|^2$ の意味で十分な降下方向を生成することが保証される. このことを利用して、 β_k^{HS} 以外のパラメータを修正した方法が数多く提案されている. 詳しくは [13]などを参照されたい.

Dai and Kou [2] はBFGS準ニュートン法とHZ法の関係性を指摘し、 β_k^{HZ} に含まれるパラメータ λ の選択法を研究している. BFGS準ニュートン法では、 x_{k-1} における目的関数のヘッセ行列 $\nabla^2 f(x_{k-1})$ の逆行列の近似行列を H_{k-1} としたとき、探索方向は

³ HZ法が実装されているのはVersion 5.3まで. Version 6.0以降では後述するHZ法の改良法が実装されている.

⁴ Hager and Zhangの論文では近似Wolfe条件への変更を考慮した場合のアルゴリズムの大域的収束性は保証していない.

$$d_k = -H_k g_k$$

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

で表せる. ここで, $H_{k-1} = \frac{1}{\tau_k} I$ とおくと τ_k 倍された探索方向 $d'_k (= \tau_k d_k)$ は

$$d'_k = -g_k + \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} d_{k-1} - \left(\tau_k + \frac{\|y_{k-1}\|^2}{s_{k-1}^T y_{k-1}}\right) \times \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}} d_{k-1} + \frac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}} y_{k-1}$$

となる. さらに, 探索方向を $d_k = \arg \min_d \{\|d - d'_k\| \mid d = -g_k + \beta d_{k-1}, \beta \in \mathbf{R}\}$ によって定めると,

$$d_k = -g_k + \beta_k^{DK} d_{k-1} \\ \beta_k^{DK} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} - \left(\tau_k + \frac{\|y_{k-1}\|^2}{s_{k-1}^T y_{k-1}} - \frac{s_{k-1}^T y_{k-1}}{\|s_{k-1}\|^2}\right) \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}}$$

と表すことができる. ここで, $\tau_k I \approx \nabla^2 f(x_{k-1})$ を考慮して, $\tau_k = s_{k-1}^T y_{k-1} / \|s_{k-1}\|^2$ とおくと β_k^{DK} は $\lambda = 1$ とした β_k^{HZ} に一致する. Dai and Kou は数値実験で HZ 法の λ を変えた場合や, DK 法の τ_k を変えた場合の各方法の比較を行っており, $\lambda = 1$ とした HZ 法 (つまり, $\tau_k = s_{k-1}^T y_{k-1} / \|s_{k-1}\|^2$ とした DK 法) が最も効果的であることを確認している. さらに, Dai and Kou は近似 Wolfe 条件を用いた場合の HZ 法の大域的収束性が保証されていないことを指摘し, 直線探索において修正 Wolfe 条件: (4) と

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \min\{\epsilon \|g_k^T d_k\|, \delta \alpha_k g_k^T d_k + \eta_k\}$$

を用いた DK 法の大域的収束性を証明している. ただし, ϵ は正の定数で, $\{\eta_k\}$ は $\sum_{k=0}^{\infty} \eta_k < +\infty$ を満たす正項級数とする.

Hager and Zhang も自身の HZ 法の改良を試みている [10]. 彼らは数値実験において CG-DESCENT が収束しない問題では g_k が過去数本の探索方向によって張られる空間 $\mathcal{S}_k = \text{Span}\{d_{k-1}, \dots, d_{k-m}\}$ に含まれる, すなわち $g_k \in \mathcal{S}_k$ に近い状況となっていることを指摘している. この場合, k 回目以降の探索方向は \mathcal{S}_k に含まれることとなり, 点列が停滞する原因となる. Hager and Zhang は部分空間最小化

$$\min_{z \in \mathcal{S}_k} f(x_k + z) \quad (11)$$

によってこのような状況の回避法を構築している. 部分空間最小化問題 (11) の解を z_k とした場合, (11) の

一次の最適性条件は

$$g(x_k + z_k)^T v = 0 \quad \text{for all } v \in \mathcal{S}_k$$

となる. したがって, 部分空間最小化を行うことで点列の停滞の原因である $g_k \in \mathcal{S}_k$ という状況の回避が可能となる. さらに, 彼らは部分空間最小化に加えて前処理付 HZ 法:

$$d_k = -P_k g_k + \beta_k^+ d_k \quad (12)$$

$$\beta_k = \frac{P_k g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} - \lambda \frac{y_{k-1}^T P_k y_{k-1}}{(d_{k-1}^T y_{k-1})^2} g_k^T d_{k-1}$$

$$\beta_k^+ = \max \left\{ \beta_k, \eta \frac{g_{k-1}^T d_{k-1}}{d_{k-1}^T P_{k-1} d_{k-1}} \right\} \quad (\eta > 0)$$

を組み込んだアルゴリズムを提案している. ここで, P_k は前処理行列であり, $P_k = I$ の場合は通常の HZ 法に帰着される. 以下は Hager and Zhang による HZ 法の改良法の概要である. (ただし $0 < \theta_1 < \theta_2 < 1$ とする)

Step 1. $\text{dist}\{\mathcal{S}_k, g_k\} > \theta_1 \|g_k\|$ ならば $P_k = I$ とした前処理付 HZ 法 (つまり通常の HZ 法) を実行する⁵. もし, $\text{dist}\{\mathcal{S}_k, g_k\} \leq \theta_1 \|g_k\|$ となった場合は Step 2 へ.

Step 2. 部分空間最小化問題 (11) に対し, $P_k = Z \hat{H}_k Z^T$ とした前処理付 HZ 法を実行する. ただし, Z は \mathcal{S}_k の直交基底を列成分に持つ行列とし, \hat{H}_k は部分空間 \mathcal{S}_k における準ニュートン法の近似行列 (逆行列版) とする. $\text{dist}\{\mathcal{S}_k, g_k\} \geq \theta_2 \|g_k\|$ が満たされたら Step 3 へ.

Step 3. 全空間での最小化に戻る際, 初回の探索方向として $P_k = Z \hat{H}_k Z^T + \sigma_k (I - Z Z^T)$ とした前処理付 HZ 法 (12) を実行し, Step 1 へ戻る. ただし,

$$\sigma_k = \max \left\{ \sigma_{\min}, \min \left\{ \sigma_{\max}, \frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}} \right\} \right\}$$

とする.

この方法は CG-DESCENT の Version 6.0 以降で実装されており, Version 5.3 以前のものと比較し, 非常に高性能であることが報告されている. また, 上で述べたアルゴリズムは概略だけであるが, [10] では Step 2 と 3 において計算量を減らす工夫や部分空間における準ニュートン法の近似行列 (逆行列版) の生成法などが詳しく述べられている. さらに, Step 2 と 3 では前処理付 HZ 法 (12) の代わりに前処理付最急降下法,

⁵ ベクトル x と集合 \mathcal{S} に対して $\text{dist}\{\mathcal{S}, x\} = \inf\{\|y - x\| \mid y \in \mathcal{S}\}$ とする.

すなわち $d_k = -P_k g_k$ を用いることもできる。この場合は、ある種の準ニュートン法を適用していると解釈できる。なお、部分空間最小化や前処理付最急降下法の考え方やサブルーチンは HZ 法特有のものではなく、ほかの CG 法に対しても加速手法として用いることが可能であることを注意しておく。

5. 数値実験

本節では前節までに紹介してきた CG 法のうち、下記の 6 つの方法の数値実験結果を報告する：

- HS : HS 法
- DL : DL 法 ($t = 1$)
- CGD5 : CG-DESCENT 5.3 [9]
- CGD6 : CG-DESCENT 6.6 [10]
- NYF1 : Narushima-Yabe-Ford 法 [12]
($\beta_k = \max\{0, \beta_k^{HS}\}$, $p_k = g_k$)
- NYF2 : NYF1 に [10] の加速手法を導入

CGD5, CGD6 ではソフトウェアのデフォルトの設定を用いた。また、HS, DL, および NYF1 は CG-DESCENT 5.3 を修正してコードを作成した。直線探索等の設定は CG-DESCENT 5.3 の設定に倣っている。ただし、HS と DL 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。NYF2 は CG-DESCENT 6.6 を修正してコードを作成した。なお、前節の最後で述べたように、Step 2 と 3 では前処理付 HZ 法の代わりに前処理付最急降下法を使用している。また、そのほかの直線探索などの設定は CG-DESCENT 6.6 の設定に倣っている。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 500 (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 [6] から 135 問を選んで実験を行った。

今回、各方法間の比較を行うために、Dolan and Moré [4] の提案し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、その解法がすべての問題の中で、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、その方法がすべて方法の中で、最も早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

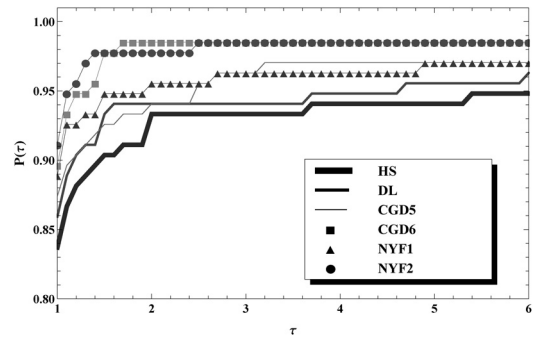


図 2 パフォーマンスプロファイル

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に位置する方法ほど効率が良いと考えることができる。

図 2 では実験を行った 6 つの方法のパフォーマンスプロファイルが与えられている。図 2 より、6 つの方法の効率性は

$$HS < DL < CGD5 \leq NYF1 < CGD6 \approx NYF2$$

となっていることがわかる。HS や DL は最近開発されたほかの 4 つの方法に比べると効率性が低いことがわかる。また、CGD5 と NYF1 は CGD6 と NYF2 ほどではないが、十分効果的であるといえるだろう。一方、CGD6 と NYF2 は非常に効果的であることがわかる。特に、CGD6 と CGD5, および NYF1 と NYF2 の違いは [10] の加速手法であるので、この結果は、いかにこの加速手法が有用であるかを物語っている。なお、上で並べた効率性の順が年代順と一致しているのは、当然という見方もできるが、個人的には興味深い結果であった。

6. 終わりに

本稿では無制約最適化問題に対するアルゴリズムを扱ったが、無制約最適化問題に限らず、最適化問題の大規模化というのはあらゆる分野で起こっており、それらに対するアルゴリズムの整備は急務であるといえる。従来より、無制約最適化問題に対してはニュートン法のような行列を使用する方法が主流であったが、大規模になればなるほど CG 法のような手軽な (実行しやすい) 方法に注目が集まるのは自然な流れのようにも感じられる。CG 法のように行列を使用しない方法は局所的に速い収束性が保証されないという弱点があるが、数値実験上は非常に高性能なアルゴリズムも開発されてきており、今後の CG 法の発展を予感させる。もちろん、CG 法がすべてではない。本稿では紹介で

きなかったが、ほかにも有用な数値解法はたくさんあり、問題に合わせて最適なアルゴリズムを選択することが重要である。CG法の研究者としては「このような場合にはCG法が一番良い」といったことを研究し、啓蒙していく必要があると感じている。

最後に、5節の数値実験の結果が年代順に高性能になっているのを思い出し、今後のCG法の発展を期待して締めくくりたいと思う。

謝辞 本稿の執筆の機会を与えていただいた「オペレーションズ・リサーチ」編集委員の皆様には感謝いたします。本研究の一部は日本学術振興会科学研究費補助金基盤研究(C)(25330030)からの支援を受けて行われている。

参考文献

- [1] W. Cheng, A two-term PRP-based descent method, *Numerical Functional Analysis and Optimization*, **28** (2007), 1217–1230.
- [2] 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.
- [3] Y.-H. Dai and L.Z. Liao, New conjugacy conditions and related nonlinear conjugate gradient methods, *Applied Mathematics and Optimization*, **43** (2001), 87–101.
- [4] E. D. Dolan and J. J. Moré, Benchmarking optimization software with performance profiles, *Mathematical Programming*, **91** (2002), 201–213.
- [5] R. Fletcher and C. M. Reeves, Function minimization by conjugate gradients, *The Computer Journal*, **7** (1964), 149–154.
- [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 and H. Zhang, A new conjugate gradient method with guaranteed descent and an efficient line search, *SIAM Journal on Optimization*, **16** (2005), 170–192.
- [8] W. W. Hager and H. Zhang, A survey of nonlinear conjugate gradient methods, *Pacific Journal of Optimization*, **2** (2006), 35–58.
- [9] 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.
- [10] W. W. Hager and H. Zhang, The limited memory conjugate gradient method, *SIAM Journal on Optimization*, **23** (2013), 2150–2168.
- [11] M. R. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems, *Journal of Research of the National Bureau of Standards*, **49** (1952), 409–436.
- [12] 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** (2011), 212–230.
- [13] 成島康史, 大規模無制約最適化問題に対する非線形共役勾配法の最近の研究動向, *応用数理*, **22** (2012), 27–39.
- [14] K. Sugiki, Y. Narushima, and H. Yabe, Globally convergent three-term conjugate gradient methods that use secant conditions and generate descent search directions for unconstrained optimization, *Journal of Optimization Theory and Applications*, **153** (2012), 733–757.
- [15] 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** (2006), 561–572.
- [16] 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.
- [17] L. Zhang, W. Zhou, and D. H. Li, Some descent three-term conjugate gradient methods and their global convergence, *Optimization Methods and Software*, **22** (2007), 697–711.