

微分不可能な関数を含む非線形方程式系に対する PRP 型平滑化スケーリング共役勾配法の大域的収束性について

成島 康史
横浜国立大学

大谷 亮介
鈴木シンワート株式会社

矢部 博
東京理科大学

(受理 2015 年 5 月 29 日; 再受理 2016 年 7 月 28 日)

和文概要 本論文では微分不可能な関数を含む非線形方程式系の求解問題を取り扱う。このような問題は、例えば、相補性問題や変分不等式問題などから生じることが知られている。平滑化ニュートン法は微分不可能な関数を含む非線形方程式系に対する効果的な反復法としてよく知られているが、行列を保存する必要があるため大規模な問題に対しては必ずしも有効とは限らない。一方、無制約最適化問題に対する共役勾配法は行列を保存しなくてもよいため、大規模問題に対する解法として知られており、中でも自動的に降下方向を生成するようなスケーリング共役勾配法が近年注目されている。本論文では、微分不可能な関数を含む方程式系に対して平滑化手法とスケーリング共役勾配法を組み合わせ、行列を陽に使用しないような数値解法を提案し、その大域的収束性を証明する。

キーワード: 非線形計画, 微分不可能な関数を含む非線形方程式系, 共役勾配法, 大域的収束性

1. はじめに

本論文では、次の非線形方程式系

$$F(x) = 0 \quad (1.1)$$

に対する数値解法を取り扱う。ただし、 $F: \mathbf{R}^n \rightarrow \mathbf{R}^n$ は必ずしも微分可能とは限らない連続関数とし、問題 (1.1) は少なくとも 1 つの解を持つものとする。このような方程式系の求解問題は実社会において生じる基本的な問題の一つで、例えば相補性問題や変分不等式問題を再定式化した際に表れることが知られている [4, 13, 17, 19]。通常、微分可能な非線形方程式系の求解問題に対してはニュートン法などといった関数 F の勾配を利用する方法が広く使用されているが、問題 (1.1) に対してはこのような方法をそのまま適用することはできない。そのため、問題 (1.1) に対する数値解法として、いくつかの方法が提案されているが (例えば、[17–19] などを参照)、その一つに平滑化関数に基づいた平滑化手法がある。ここで、平滑化関数とは次のように定義される。

定義 1.1. 関数 $\tilde{F}: \mathbf{R} \times \mathbf{R}^n \rightarrow \mathbf{R}^n$ が $\mathbf{R}_{++} \times \mathbf{R}^n$ において連続的微分可能で、かつ、すべての $x \in \mathbf{R}^n$ に対して

$$\lim_{t \rightarrow +0} \tilde{F}(t, x) = \tilde{F}(0, x) = F(x)$$

を満たすとき、 \tilde{F} を F の平滑化関数と呼ぶ。ただし、 $\mathbf{R}_{++} = \{t \in \mathbf{R} \mid t > 0\}$ である。

微分不可能な方程式系の具体的な応用例として、例えば非負象限上の非線形相補性問題がある。非線形相補性問題は関数 $G: \mathbf{R}^n \rightarrow \mathbf{R}^n$ に対して

$$x \geq 0, \quad G(x) \geq 0, \quad x^T G(x) = 0 \quad (1.2)$$

を満たす $x \in \mathbf{R}^n$ を見つける問題である。このとき、

$$\phi(a, b) = 0 \iff a \geq 0, b \geq 0, ab = 0 \quad (1.3)$$

を満たすような関数 $\phi: \mathbf{R}^2 \rightarrow \mathbf{R}$ を考えると、問題 (1.2) は次の方程式系に再定式化される:

$$F(x) = \begin{pmatrix} \phi(x_1, G_1(x)) \\ \phi(x_2, G_2(x)) \\ \vdots \\ \phi(x_n, G_n(x)) \end{pmatrix} = 0.$$

ただし、 $x = (x_1, x_2, \dots, x_n)^T$ とし、 $G(x) = (G_1(x), G_2(x), \dots, G_n(x))^T$ とする。ここで、 $\phi_{\min}(a, b) = \min\{a, b\}$ は (1.3) の性質を満たす関数として知られているが、これは微分不可能な方程式系となる。ここで、 $\phi_{\min}(a, b)$ に対する平滑化関数の一つとして

$$\tilde{\phi}_{\min}(t, a, b) = \frac{1}{2} \left(a + b - \sqrt{(a-b)^2 + t^2} \right)$$

などが考えられる。

問題 (1.1) に対して、定義 1.1 で定めた平滑化関数を用いて関数 $H: \mathbf{R} \times \mathbf{R}^n \rightarrow \mathbf{R}^{1+n}$ を

$$H(t, x) = \begin{pmatrix} t \\ \tilde{F}(t, x) \end{pmatrix} \quad (1.4)$$

によって定めると、 $H(t, x) = 0$ ならば $F(x) = 0$ が成立する。ここで、メリット関数 $\Psi: \mathbf{R} \times \mathbf{R}^n \rightarrow \mathbf{R}$ を

$$\Psi(t, x) = \frac{1}{2} \|H(t, x)\|^2 = \frac{1}{2} \{t^2 + \|\tilde{F}(t, x)\|^2\} \quad (1.5)$$

とすれば、最小二乗形式の無制約最適化問題:

$$\min \Psi(t, x) \quad (1.6)$$

の大域的最適解 (の x に対応する成分) は問題 (1.1) の解となる。なお、ベクトルのノルム $\|\cdot\|$ は 2 ノルムを表すこととする。ここで関数 Ψ は $\mathbf{R}_{++} \times \mathbf{R}^n$ 上では連続的微分可能だが、それ以外の領域 ($t \leq 0$ となる場合) では必ずしも連続的微分可能とは限らないことを注意しておく。

方程式系 $H(t, x) = 0$ や最適化問題 $\min \Psi(t, x)$ をニュートン法やそれに類似する反復法で解く方法は既に多くの研究がある (例えば [18] 参照)。しかしながら、ニュートン法は行列を保存する必要があるため、必ずしも次元の大きな、いわゆる大規模問題に対しては有効であるとは限らない。よって本論文では、行列を陽に使用しないような数値解法である共役勾配法に着目する。

一般の(微分可能な)無制約最適化問題:

$$\min f(z), \quad (f: \mathbf{R}^l \rightarrow \mathbf{R}, z \in \mathbf{R}^l)$$

に対する共役勾配法は大規模な無制約最適化問題に対する数値解法として, 近年注目されている. 共役勾配法は, 反復法の一つで, 初期点 $z_0 \in \mathbf{R}^l$ に対して,

$$z_{k+1} = z_k + \alpha_k d_k$$

によって点列 $\{z_k\}$ を生成する. ただし, $\alpha_k > 0$ はステップ幅, d_k は探索方向とよばれる. 共役勾配法の探索方向は

$$d_k = \begin{cases} -\nabla f(z_k), & k = 0, \\ -\nabla f(z_k) + \beta_k d_{k-1}, & k \geq 1 \end{cases}$$

によって与えられる. ここで, β_k は共役勾配法を特徴付けるパラメータである. 共役勾配法は陽に行列を使用しない数値解法であるため, 大規模な問題に対しても直接適用することができる. 共役勾配法は β_k のとりかたによって計算効率が変わるため, 効果的なパラメータ β_k の選び方について多くの研究が行われている(例えば, [2, 5–10, 16] を参照). なかでもよく知られた公式としては Hestenes-Stiefel (HS) [10], Fletcher-Reeves (FR) [5], Polak-Ribière-Polyak (PRP) [16], Dai-Yuan (DY) [2] などがある:

$$\begin{aligned} \beta_k^{HS} &= \frac{\nabla f(z_k)^T \hat{y}_{k-1}}{d_{k-1}^T \hat{y}_{k-1}}, & \beta_k^{FR} &= \frac{\|\nabla f(z_k)\|^2}{\|\nabla f(z_{k-1})\|^2}, \\ \beta_k^{PRP} &= \frac{\nabla f(z_k)^T \hat{y}_{k-1}}{\|\nabla f(z_{k-1})\|^2}, & \beta_k^{DY} &= \frac{\|\nabla f(z_k)\|^2}{d_{k-1}^T \hat{y}_{k-1}}. \end{aligned} \quad (1.7)$$

ここで, $\hat{y}_{k-1} = \nabla f(z_k) - \nabla f(z_{k-1})$ とする. 一方で, 多くの共役勾配法には, 必ずしも十分な降下方向を生成するとは限らないという欠点があり, それを克服するために, 直線探索の方法によらずに十分な降下方向を生成する三項共役勾配法やスケーリング共役勾配法が近年盛んに研究されている(例えば, [1, 11, 14, 15, 20–22] を参照). ここで, 探索方向 d_k が十分な降下条件を満たすとは, ある正の定数 \bar{c} が存在し, 全ての $k \geq 0$ に対し,

$$\nabla f(z_k)^T d_k \leq -\bar{c} \|\nabla f(z_k)\|^2$$

を満たすことをいう.

Narushima et al. [15] はいくつかの三項共役勾配法やスケーリング共役勾配法を含むような族を提案しており, その探索方向は

$$d_k = \begin{cases} -\nabla f(z_k), & k = 0, \\ -\nabla f(z_k) + \beta_k (\nabla f(z_k)^T p_k)^\dagger \{(\nabla f(z_k)^T p_k) d_{k-1} - (\nabla f(z_k)^T d_{k-1}) p_k\}, & k \geq 1 \end{cases} \quad (1.8)$$

で与えられる. ただし, $p_k \in \mathbf{R}^l$ は任意のベクトルで, 記号 \dagger は

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

で定義される. このとき, (1.8) の探索方向 d_k に左から $\nabla f(z_k)^T$ をかけると $\nabla f(z_k)^T d_k = -\|\nabla f(z_k)\|^2$ が得られる. これは探索方向が十分な降下条件を満たすことを意味している. 通常, パラメータベクトル p_k として, $p_k = \nabla f(z_k)$, または, $p_k = \hat{y}_{k-1}$ が選ばれることが多い. 例えば, β_k として $\beta_k = \beta_k^{PRP}$ を選択し, $p_k = \hat{y}_{k-1}$ とすると, 式 (1.8) は $\nabla f(z_k)^T \hat{y}_{k-1} \neq 0$ のとき,

$$d_k = \begin{cases} -\nabla f(z_k), & k = 0, \\ -\nabla f(z_k) + \beta_k^{PRP} d_{k-1} - \frac{\nabla f(z_k)^T d_{k-1}}{\|\nabla f(z_{k-1})\|^2} \hat{y}_{k-1}, & k \geq 1 \end{cases} \quad (1.9)$$

となる. これは, Zhang ら [21] によって提案された PRP 型三項共役勾配法と一致する. 一方, $p_k = \nabla f(z_k)$ とした場合には, (1.8) の探索方向 d_k は

$$d_k = \begin{cases} -\nabla f(z_k), & k = 0, \\ -\left(1 + \beta_k \frac{\nabla f(z_k)^T d_{k-1}}{\|\nabla f(z_k)\|^2}\right) \nabla f(z_k) + \beta_k d_{k-1}, & k \geq 1 \end{cases} \quad (1.10)$$

と表すことができる. ここで $\beta_k = \beta_k^{PRP}$ とすると, Cheng [1] の PRP 型スケーリング共役勾配法と一致する.

一方, 方程式 (1.1) に対する数値解法として, 平滑化関数を含んだ無制約最適化問題 (1.6) に対する共役勾配法も研究されており, Narushima [12] は Zhang らの三項共役勾配法 (1.9) に倣い平滑化三項共役勾配法を提案した. この方法では探索方向はメリット関数 Ψ の降下方向を常に生成している. さらに Narushima [12] は直線探索において, Armijo 条件の変種を用いたアルゴリズムを構築し, その大域的収束性を証明している.

Narushima [12] は三項共役勾配法 (1.9) を基に平滑化三項共役勾配法を構築しているが, 三項共役勾配法 (1.9) の代わりにスケーリング共役勾配法 (1.10) に基づいた平滑化スケーリング共役勾配法を考えることもできる. 特に, 探索方向 (1.8) において, $p_k = \hat{y}_{k-1}$ とした場合に, 三項共役勾配法 (1.9) と一致するためには, $\nabla f(z_k)^T \hat{y}_{k-1} \neq 0$ という条件が必要だったのに対し, $p_k = \nabla f(z_k)$ とした場合には, こうした条件を課すことなくスケーリング共役勾配法 (1.10) と一致することから, p_k として $\nabla f(z_k)$ を選択した場合のスケーリング共役勾配法 (1.10) の方が自然であるように思われる. そこで, 今回, 我々はスケーリング共役勾配法 (1.10) に基づき, 無制約最適化問題 (1.6) を解くような平滑化スケーリング共役勾配法を提案する.

本論文は以下のように構成される. まず第2節で, スケーリング共役勾配法 (1.10) に倣い, メリット関数 Ψ の降下方向を生成するような平滑化スケーリング共役勾配法を提案する. さらに, 直線探索において Armijo 条件の変種を用いたアルゴリズムを与える. 第3節では提案したアルゴリズムの大域的収束性を示し, 第4節で数値実験の結果を報告する. 最後に第5節では, この論文のまとめと今後の課題を述べる.

ここで, 次節以降に必要な事項を導入しておく. 平滑化関数 \tilde{F} は $\mathbf{R}_{++} \times \mathbf{R}^n$ 上で連続的微分可能なので, $\nabla \tilde{F}$ が存在し,

$$\nabla \tilde{F}(t, x) = \begin{pmatrix} \nabla_t \tilde{F}(t, x) \\ \nabla_x \tilde{F}(t, x) \end{pmatrix}$$

となり, (1.5) から $\nabla \Psi$ は

$$\nabla \Psi(t, x) = \begin{pmatrix} \nabla_t \Psi(t, x) \\ \nabla_x \Psi(t, x) \end{pmatrix} = \begin{pmatrix} t + \nabla_t \tilde{F}(t, x) \tilde{F}(t, x) \\ \nabla_x \tilde{F}(t, x) \tilde{F}(t, x) \end{pmatrix} \in \mathbf{R}^{1+n} \quad (1.11)$$

となる. 以降では, 表記を簡単にするために, $(t, x^T)^T \in \mathbf{R}^{1+n}$ と $(t, x) \in \mathbf{R} \times \mathbf{R}^n$ を同一視することとし, $v = (t, x) \in \mathbf{R}^{1+n}$ と表す. また, 行列 A に対して, $\|A\|$ はベクトルの 2 ノルムから誘導される行列ノルム, つまり $\|A\| = \sqrt{\rho(A^T A)}$ とする. ただし, $\rho(A^T A)$ は $A^T A$ の絶対値最大固有値を表す.

2. 平滑化スケーリング共役勾配法

本節では, スケーリング共役勾配法 (1.10) に倣い, 非線形方程式 (1.1) に対する平滑化スケーリング共役勾配法のアルゴリズムを提案する. 平滑化スケーリング共役勾配法のアルゴリズムでは, k 回目の反復解 $v_k = (t_k, x_k)$ を

$$v_{k+1} = v_k + \alpha_k d_k \quad (2.1)$$

によって更新する. ここで, $d_k \in \mathbf{R}^{1+n}$ は探索方向, $\alpha_k \in (0, 1]$ はステップ幅とする. 以降では, 簡単のために $\tilde{F}(v_k)$ を

$$\tilde{F}_k = \tilde{F}(v_k)$$

と省略し, 他の関数についても同様の省略記号を用いる. 第 1 節でも触れたように Ψ は $\mathbf{R}_{++} \times \mathbf{R}^n$ 以外の範囲では連続的微分可能性が保証されていないので, 全ての k に対して $t_k > 0$ となるようにアルゴリズムを構築する必要がある. そのために次の集合を定義する.

定義 2.1. $\bar{\gamma} \in (0, 1)$, $\bar{t} \in (0, 1]$ となる正の定数 \bar{t} , $\bar{\gamma}$ に対して, 関数 $\gamma: \mathbf{R}^{1+n} \rightarrow \mathbf{R}_+$ と集合 Ω を

$$\gamma(v) = \bar{\gamma} \min\{1, \Psi(v)\}, \quad \Omega = \{v = (t, x) \mid t \geq \bar{t}\gamma(v)\} \quad (2.2)$$

と定義する. ただし, $\mathbf{R}_+ = \{t \in \mathbf{R} \mid t \geq 0\}$ である.

もし $\{v_k\} \subset \Omega$ かつ $\Psi_k > 0$ が成り立つならば, 全ての $k \geq 0$ に対し $t_k > 0$ が成り立つ. また, $\{v_k\} \subset \Omega$ が成り立ち, さらに t_k が 0 に近づく場合には, Ψ_k も 0 に近づくことがいえる.

次に, 探索方向 d_k を

$$d_k = \begin{pmatrix} \bar{t}\gamma_k - t_k \\ \tilde{d}_k \end{pmatrix} \quad (2.3)$$

と定める. ここで, $\bar{t}\gamma_k - t_k \in \mathbf{R}$ は変数 t に関する探索方向, $\tilde{d}_k \in \mathbf{R}^n$ は変数 x に関する探索方向である. また, $\tilde{d}_k \in \mathbf{R}^n$ はスケーリング共役勾配法 (1.10) に基いて以下のように定める:

もし $\nabla_x \Psi_k = 0$ ならば

$$\tilde{d}_k = 0 \quad (2.4)$$

とし, さもなければ

$$\tilde{d}_k = \begin{cases} -\theta_k \nabla_x \Psi_k, & k = 0. \\ -\left(\theta_k + \beta_k \frac{\nabla_x \Psi_k^T \tilde{d}_{k-1}}{\|\nabla_x \Psi_k\|^2}\right) \nabla_x \Psi_k + \beta_k \tilde{d}_{k-1}, & k \geq 1 \end{cases} \quad (2.5)$$

とする。ただし、 $\eta \in (0, 1)$ を定数、 $y_{k-1} = \nabla_x \Psi_k - \nabla_x \Psi_{k-1}$ とし、さらに

$$\theta_k = \begin{cases} 1, & \eta \|\nabla_x \Psi_k\|^2 \geq (\bar{t}\gamma_k - t_k) \nabla_t \tilde{F}_k \tilde{F}_k \text{ のとき,} \\ 1 + \frac{(\bar{t}\gamma_k - t_k) \nabla_t \tilde{F}_k \tilde{F}_k}{\|\nabla_x \Psi_k\|^2}, & \text{その他,} \end{cases} \quad (2.6)$$

$$\beta_k = \frac{\nabla_x \Psi_k^T y_{k-1}}{\|\nabla \Psi_{k-1}\|^2} \quad (2.7)$$

とする。

ここで、Polak-Ribière-Polyak の選択法 (式 (1.7) の β_k^{PRP} を参照) に対応する β_k の分母が $\|\nabla_x \Psi_{k-1}\|^2$ ではなくて $\|\nabla \Psi_{k-1}\|^2$ であることに注意されたい。

上で述べたように、 $\{v_k\} \subset \Omega$ となるようにアルゴリズムを構築する必要があるが、そのためには、すべての $k \geq 0$ に対して $0 < t_{k+1} \leq t_k$ と $\Psi_{k+1} < \Psi_k$ が成り立てば十分である。以下の命題は $0 < t_{k+1} \leq t_k$ が成立することを保証している。

命題 2.1. $v_k \in \Omega$, かつ $t_k > 0$ であるとき、任意の $\alpha \in (0, 1]$ に対して、 $0 < t_k + \alpha(\bar{t}\gamma_k - t_k) \leq t_k$ が成立する。

証明. まず (1.5) と $t_k > 0$ より、 $\gamma_k = \bar{\gamma} \min\{1, \Psi_k\} > 0$ が成立する。したがって、 $t_k + \alpha(\bar{t}\gamma_k - t_k) = (1 - \alpha)t_k + \alpha\bar{t}\gamma_k > 0$ が成り立つ。また、 $v_k \in \Omega$ より、 $\bar{t}\gamma_k - t_k \leq 0$ となるため、 $t_k + \alpha(\bar{t}\gamma_k - t_k) \leq t_k$ が成り立つ。□

補足 2.1. 式 (2.1) と (2.3) より $t_{k+1} = t_k + \alpha_k(\bar{t}\gamma_k - t_k)$ なので、命題 2.1 の α をステップ幅 $\alpha_k \in (0, 1]$ とすれば、 $0 < t_{k+1} \leq t_k$ が成り立つ。

次に $\Psi_{k+1} < \Psi_k$ が成り立つことを示すために、以下の補題を示す。

補題 2.2. 式 (2.5) によって定義された \tilde{d}_k は関係式

$$\nabla_x \Psi_k^T \tilde{d}_k = -\theta_k \|\nabla_x \Psi_k\|^2$$

を満たす。

証明. $k = 0$ のとき、題意は明らかに成り立つ。一方、 $k \geq 1$ のとき、(2.5) の \tilde{d}_k の左から $\nabla_x \Psi_k^T$ を掛けると

$$\nabla_x \Psi_k^T \tilde{d}_k = -\theta_k \|\nabla_x \Psi_k\|^2 - \beta_k \nabla_x \Psi_k^T \tilde{d}_{k-1} + \beta_k \nabla_x \Psi_k^T \tilde{d}_{k-1} = -\theta_k \|\nabla_x \Psi_k\|^2$$

を得る。□

次の命題は、探索方向 d_k がメリット関数 Ψ の降下方向になることを示している。

命題 2.3. $\nabla_x \tilde{F}_k$ が正則のとき、 $v_k \in \Omega$ と $0 < t_k \leq 1$ に対して、

$$\nabla \Psi_k^T d_k \leq -(1 - \eta) \|\nabla_x \Psi_k\|^2 + t_k (\bar{t}\gamma_k - t_k) < 0 \quad (2.8)$$

が成り立つ。

証明. はじめに, d_k の左から $\nabla\Psi_k^T$ をかけると, (1.11) と (2.3) より,

$$\begin{aligned}\nabla\Psi_k^T d_k &= \nabla_x\Psi_k^T \tilde{d}_k + (\bar{t}\gamma_k - t_k)\nabla_t\Psi_k \\ &= \nabla_x\Psi_k^T \tilde{d}_k + t_k(\bar{t}\gamma_k - t_k) + (\bar{t}\gamma_k - t_k)\nabla_t\tilde{F}_k\tilde{F}_k\end{aligned}\quad (2.9)$$

を得る. 次に $\nabla_x\Psi_k = 0$ と $\nabla_x\Psi_k \neq 0$ の 2通りで場合分けをする.

• $\nabla_x\Psi_k = 0$ の場合:

$\nabla_x\tilde{F}_k$ が正則であることと, $\nabla_x\Psi_k = \nabla_x\tilde{F}_k\tilde{F}_k = 0$ より, $\tilde{F}_k = 0$ となる. また, (2.4), (2.9) と $\nabla_x\Psi_k = 0$, $\tilde{F}_k = 0$, $v_k \in \Omega$ より,

$$\nabla\Psi_k^T d_k = t_k(\bar{t}\gamma_k - t_k) \leq 0 \quad (2.10)$$

を得る. 次に, $\bar{t}\gamma_k - t_k < 0$ を示すために, $\bar{t}\gamma_k - t_k = 0$ と仮定すると, $\tilde{F}_k = 0$ より, $t_k = \bar{t}\gamma_k = \bar{t}\bar{\gamma} \min\{1, \Psi_k\} < \Psi_k = \frac{1}{2}t_k^2$ が成り立つ. このとき $t_k > 0$ より $t_k > 2$ となるが, これは $0 < t_k \leq 1$ に反する. したがって, $\bar{t}\gamma_k - t_k < 0$ が成立する. ここで (2.10) と $\bar{t}\gamma_k - t_k < 0$, $\nabla_x\Psi_k = 0$ より (2.8) が示された.

• $\nabla_x\Psi_k \neq 0$ の場合:

補題 2.2 と (2.9) より

$$\nabla\Psi_k^T d_k = -\theta_k\|\nabla_x\Psi_k\|^2 + t_k(\bar{t}\gamma_k - t_k) + (\bar{t}\gamma_k - t_k)\nabla_t\tilde{F}_k\tilde{F}_k \quad (2.11)$$

が成立する. まず, $\eta\|\nabla_x\Psi_k\|^2 \geq (\bar{t}\gamma_k - t_k)\nabla_t\tilde{F}_k\tilde{F}_k$ の場合, (2.6) と (2.11) より,

$$\begin{aligned}\nabla\Psi_k^T d_k &= -\|\nabla_x\Psi_k\|^2 + t_k(\bar{t}\gamma_k - t_k) + (\bar{t}\gamma_k - t_k)\nabla_t\tilde{F}_k\tilde{F}_k \\ &\leq -(1-\eta)\|\nabla_x\Psi_k\|^2 + t_k(\bar{t}\gamma_k - t_k)\end{aligned}$$

が成立する. 一方, $\eta\|\nabla_x\Psi_k\|^2 < (\bar{t}\gamma_k - t_k)\nabla_t\tilde{F}_k\tilde{F}_k$ の場合, 同じく (2.6) と (2.11) より,

$$\begin{aligned}\nabla\Psi_k^T d_k &= -\left(1 + \frac{(\bar{t}\gamma_k - t_k)\nabla_t\tilde{F}_k\tilde{F}_k}{\|\nabla_x\Psi_k\|^2}\right)\|\nabla_x\Psi_k\|^2 + t_k(\bar{t}\gamma_k - t_k) + (\bar{t}\gamma_k - t_k)\nabla_t\tilde{F}_k\tilde{F}_k \\ &= -\|\nabla_x\Psi_k\|^2 + t_k(\bar{t}\gamma_k - t_k) \\ &\leq -(1-\eta)\|\nabla_x\Psi_k\|^2 + t_k(\bar{t}\gamma_k - t_k)\end{aligned}$$

が成立する. したがって, $v_k \in \Omega$ より $\bar{t}\gamma_k - t_k \leq 0$ が成り立つことと $\nabla_x\Psi_k \neq 0$ であることから, (2.8) が示された. \square

ここで, $0 < t_{k+1} \leq t_k$ と $\Psi_{k+1} < \Psi_k$ を満たす平滑化スケージング共役勾配法 (smoothing and scaling conjugate gradient method: SSCG) のアルゴリズムを与える.

アルゴリズム SSCG.

Step 0. $\bar{t} \in (0, 1]$, $\bar{\gamma} \in (0, 1)$, $\eta \in (0, 1)$, $\sigma \in (0, 1)$, $\delta \in (0, 1)$ を与える. $t_0 = \bar{t}$ とし, 初期点 $v_0 = (t_0, x_0) \in \Omega$ を与え, $k = 0$ とする.

Step 1. $\|F(x_k)\| = 0$ ならば反復終了. x_k を解とする.

Step 2. 探索方向 d_k を (2.3)–(2.7) により求める.

Step 3. 次の式を満たす最も小さな非負の整数 l を求める:

$$\Psi(v_k + \sigma^l d_k) \leq \Psi(v_k) - \delta \|\sigma^l d_k\|^2. \quad (2.12)$$

そして, ステップ幅を $\alpha_k = \sigma^l$ とする.

Step 4. v_{k+1} を (2.1) により更新する.

Step 5. $k := k + 1$ とし, Step 1 へ戻る

Step 0 では初期点を $v_0 \in \Omega$ となるように選ばなければならないが, (2.2) の γ_k の定義より, 任意の定数 $\bar{t} \in (0, 1]$ に対して $\bar{t}\gamma(v_0) = \bar{t}\gamma \min\{1, \Psi(v_0)\} \leq \bar{t}$ が成り立つので, $t_0 = \bar{t}$ とすれば, $t_0 \geq \bar{t}\gamma_0$ が成立する. したがって, $t_0 = \bar{t}$ とおけば $v_0 \in \Omega$ が成立する.

アルゴリズム SSCG の Step 3 において分割 (バックトラック) 法を用いているが, 次のように 2 次補間を用いることもできる:

$0 < \sigma_{\min} \leq \sigma_{\max} < 1$ となる正定数 $\sigma_{\min}, \sigma_{\max}$ に対して,

Step 3'. $l = 0$ とし, $\alpha^{(l)} = 1$ と定める.

Step 3.1. もし,

$$\Psi(v_k + \alpha^{(l)} d_k) \leq \Psi(v_k) - \delta \|\alpha^{(l)} d_k\|^2$$

が成り立つならば, $\alpha_k = \alpha^{(l)}$ とし, Step 4 へ進む.

Step 3.2. $\sigma^{(l)} \in [\sigma_{\min}, \sigma_{\max}]$ を選び, $\alpha^{(l+1)} = \sigma^{(l)} \alpha^{(l)}$ とする.

Step 3.3. $l := l + 1$ とし, Step 3.1 に戻る

ここで, $\sigma_{\min} = \sigma_{\max} = \sigma$ ととると, Step 3' は Step 3 と一致する. 一方で, $\sigma^{(l)}$ を

$$\sigma^{(l)} = \max \left\{ \sigma_{\min}, \min \left\{ \sigma_{\max}, \frac{0.5 \alpha^{(l)} \nabla \Psi_k^T d_k}{\Psi_k + \alpha^{(l)} \nabla \Psi_k^T d_k - \Psi(v_k + \alpha^{(l)} d_k)} \right\} \right\}$$

と定めると, Step 3' は 2 次補間の直線探索となる. 2 次補間を用いた直線探索についての詳細は, 例えば, [16] などを参照のこと. 今回, 我々は Step 3 を用いたアルゴリズム SSCG の大域的収束性のみを証明するが, Step 3' を用いたアルゴリズム SSCG の大域的収束性も同様の方法で証明できる.

次にアルゴリズム SSCG の実行可能性と生成される点列 $\{v_k\}$ が Ω に含まれることを示す. そのために, まずは \tilde{F} に関して以下の条件を課す.

仮定 2.1.

A1. メリット関数 Ψ の初期点 v_0 において定義される集合:

$$L = \{v \in \mathbf{R}^{1+n} \mid \Psi(v) \leq \Psi(v_0), t \in [0, 1]\}$$

は有界である.

A2. 任意の $v \in (\mathbf{R}_{++} \times \mathbf{R}^n) \cap \Omega$ に対して, $\nabla_x \tilde{F}(v)$ は正則である.

いま, F が強圧的である, つまり $\lim_{\|x\| \rightarrow \infty} \|F(x)\| = \infty$ を満たす場合を考えよう. このとき, 平滑化の方法をうまく選べば, 任意の非負の実数 t に対して

$$\lim_{\|x\| \rightarrow \infty} \|\tilde{F}(t, x)\| = \infty \quad (2.13)$$

が成立すると仮定することは自然である. このとき, (2.13) から仮定 A1 を導けるので, F が強圧的である場合には仮定 A1 は自然な仮定であるといえる.

以下では, 仮定 2.1 の下でアルゴリズム SSCG の実行可能性を証明する. 以降では, 全ての $k \geq 0$ に対して $F_k \neq 0$ を仮定しても一般性を失わないものとする. さもなければ, (1.1) の解が得られたこととなり, アルゴリズム SSCG は有限回で終了する.

命題 2.4. 仮定 2.1 が成り立つとする. このとき, すべての k に対して探索方向 d_k は定義可能で, 内部反復 (Step 3 の直線探索) は有限回で終了する. さらに, アルゴリズム SSCG によって生成される点列 $\{v_k\}$ に対して, $\{v_k\} \subset \Omega$ かつ $\{t_k\} \subset (0, 1]$ が成立する.

証明. まず, $v_0 \in \Omega$ と $t_0 \in (0, 1]$ は明らかに成り立つ. また, (2.3)–(2.7) より, d_0 は定義可能である. さらに, 仮定 A2 と命題 2.3 より d_0 は降下方向なので, (2.12) を満たす整数 l が存在する. したがって, v_1 は定義可能である. また, 補足 2.1 より $t_1 \in (0, 1]$ が成立する. $k = 1$ の場合, 下記の証明と同様の手順で $v_1 \in \Omega$ を示せるため証明は省略する.

ここで, v_1, \dots, v_k は定義可能で, $v_1, \dots, v_k \in \Omega$, かつ $t_0, \dots, t_k \in (0, 1]$ であると仮定する. いま, $\nabla \Psi_{k-1} = 0$ であるとする, (1.11), $v_{k-1} \in \Omega$, $t_{k-1} \in (0, 1]$ と仮定 A2 より, $\tilde{F}_{k-1} = 0$ となり, 再度 (1.11) より, $t_{k-1} = 0$ が成立する. よって, 矛盾するので, $\nabla \Psi_{k-1} \neq 0$ を得る. したがって, (2.3)–(2.7) より, d_k は定義可能である. さらに, 仮定 A2 と命題 2.3 より d_k は降下方向なので, (2.12) を満たす整数 l が存在する. よって, v_{k+1} は定義可能となる. さらに, 補足 2.1 より $t_{k+1} \in (0, 1]$ が成立する. ここで, $v_k \in \Omega$ から $t_k \geq \bar{t}\gamma_k$ が成り立ち, また, $\alpha_k \in (0, 1]$ なので

$$t_{k+1} = t_k + \alpha_k(\bar{t}\gamma_k - t_k) = (1 - \alpha_k)t_k + \alpha_k\bar{t}\gamma_k \geq (1 - \alpha_k)\bar{t}\gamma_k + \alpha_k\bar{t}\gamma_k = \bar{t}\gamma_k$$

となる. また, (2.12) より, $\Psi_k \geq \Psi_{k+1}$ なので, $\gamma_k \geq \gamma_{k+1}$ が成り立つ. したがって, $t_{k+1} \geq \bar{t}\gamma_{k+1}$ が成立し, $v_{k+1} \in \Omega$ となる. 以上, 帰納法により命題は証明された. \square

補足 2.2. (2.12) より $0 \leq \delta \|\alpha_k d_k\|^2 \leq \Psi_k - \Psi_{k+1}$ が成り立ち, $\{\Psi_k\}$ が下に有界かつ単調減少列であることから, $\lim_{k \rightarrow \infty} (\Psi_k - \Psi_{k+1}) = 0$ となる. よって,

$$\lim_{k \rightarrow \infty} \|\alpha_k d_k\| = 0 \quad (2.14)$$

が成り立つ. また, 命題 2.4 より $\{t_k\} \subset (0, 1]$ となるので, $\{\Psi_k\}$ の単調減少性より, $\{v_k\} \subset L$ となる. さらに, 仮定 A1 から L は有界であるので, $\{v_k\}$ は有界である.

補足 2.3. 命題 2.4 から $\{v_k\} \subset (\mathbf{R}_{++} \times \mathbf{R}^n) \cap \Omega$ となる. さらに仮定 A2 から, $\nabla_x \tilde{F}_k$ は任意の k に対して正則となる.

3. 大域的収束性

この節ではアルゴリズム SSCG の大域的収束性を示す. もし $\{v_k\} \subset \Omega$ が $\lim_{k \rightarrow \infty} t_k = 0$ を満たすならば, 任意の k に対して

$$t_k \geq \bar{t}\gamma \min\{1, \Psi_k\} > 0$$

が成り立つので, $\lim_{k \rightarrow \infty} \Psi_k = 0$ を得る. これは第 1 節で述べたように, (1.1) の解が得られたこととなる. 後述する定理でアルゴリズムの大域的収束性を背理法によって証明するために, $\lim_{k \rightarrow \infty} t_k \neq 0$ の仮定の下でいくつかの性質を証明する.

補題 3.1. 仮定 2.1 を満たすとき, もし $\lim_{k \rightarrow \infty} t_k \neq 0$ ならば,

$$\liminf_{k \rightarrow \infty} \|\nabla \Psi(v_k)\| \neq 0 \quad (3.1)$$

が成り立つ.

証明. 補足 2.2 から $\{v_k\}$ は有界であるので, 少なくとも一つの集積点を持つ. また, $\{t_k\}$ は下に有界な非増加列であり, $\lim_{k \rightarrow \infty} t_k \neq 0$ であることから, $\{t_k\}$ は正の極限 $\hat{t} \in (0, 1]$ を持つ. 次に, (3.1) を背理法で証明するために,

$$\lim_{k \in K_1, k \rightarrow \infty} v_k = \hat{v}, \quad \|\nabla \Psi(\hat{v})\| = 0 \quad (3.2)$$

を満たす $\{v_k\}$ の集積点 \hat{v} と部分列 $K_1 \subset \mathbf{N}$ が存在すると仮定する. ただし, \mathbf{N} は自然数全体の集合であるとする. まず, 仮定 A1 と L の定義より, L は有界閉集合となる. また, Ω も閉集合であるため, $([\hat{t}, 1] \times \mathbf{R}^n) \cap \Omega \cap L$ は有界閉集合となる. また, 命題 2.4 と補足 2.2 より, $\{v_k\} \subset ([\hat{t}, 1] \times \mathbf{R}^n) \cap \Omega \cap L$ となるので, $\hat{v} \in ([\hat{t}, 1] \times \mathbf{R}^n) \cap \Omega \cap L$ が成立する. したがって, 仮定 A2 より $\nabla_x \tilde{F}(\hat{v})$ は正則となる. また, (1.11), (3.2) より $\nabla_x \Psi(\hat{v}) = \nabla_x \tilde{F}(\hat{v}) \tilde{F}(\hat{v}) = 0$ が成り立つため, $\nabla_x \tilde{F}(\hat{v})$ の正則性より, $\tilde{F}(\hat{v}) = 0$ を得る. したがって, (1.11) から $\|\nabla \Psi(\hat{v})\| = \hat{t} \neq 0$ となるが, これは $\|\nabla \Psi(\hat{v})\| = 0$ に矛盾する. このことから, (3.1) が成り立つことが示された. \square

補題 3.2. 仮定 2.1 と $\lim_{k \rightarrow \infty} t_k \neq 0$ が成り立つとする. このとき, アルゴリズム SSCG で生成される探索方向の列 $\{d_k\}$ は有界となる.

証明. まず, 補題 3.1 より, 全ての k に対し,

$$\|\nabla \Psi_k\| \geq \varepsilon \quad (3.3)$$

を満たす正の定数 $\varepsilon > 0$ が存在する. また, $\gamma_k < 1$ と $t_k \leq 1$ から

$$|\bar{t}\gamma_k - t_k| \leq \bar{t}\gamma_k + t_k < \bar{t} + 1 \quad (3.4)$$

となるので, $\bar{t}\gamma_k - t_k$ は有界となる. したがって, $\|d_k\|^2 = \|\tilde{d}_k\|^2 + (\bar{t}\gamma_k - t_k)^2$ の有界性を示すには, $\{\tilde{d}_k\}$ の有界性を示せばよい. $\nabla_x \Psi_k \neq 0$ の場合, (2.5), (2.7), (3.3) より

$$\begin{aligned} \|\tilde{d}_k\| &\leq \|\theta_k \nabla_x \Psi_k\| + |\beta_k| \frac{\|\nabla_x \Psi_k\| \|\tilde{d}_{k-1}\|}{\|\nabla_x \Psi_k\|^2} \|\nabla_x \Psi_k\| + |\beta_k| \|\tilde{d}_{k-1}\| \\ &\leq \|\theta_k \nabla_x \Psi_k\| + 2 \frac{\|\nabla_x \Psi_k\| \|y_{k-1}\|}{\|\nabla \Psi_{k-1}\|^2} \|\tilde{d}_{k-1}\| \\ &\leq \|\theta_k \nabla_x \Psi_k\| + 2 \frac{\|\nabla_x \Psi_k\| \|y_{k-1}\|}{\varepsilon^2} \|\tilde{d}_{k-1}\| \end{aligned} \quad (3.5)$$

が成り立つ．次に， $|\theta_k|$ を評価する．式 (1.11) と (2.6) より，

$$|\theta_k| \leq 1 + \frac{|(\bar{t}\gamma_k - t_k)\nabla_t \tilde{F}_k \tilde{F}_k|}{\|\nabla_x \tilde{F}_k \tilde{F}_k\|} \quad (3.6)$$

が成立するので，右辺の第2項を評価すればよい．補題3.1の証明の中で述べたように， $\{t_k\}$ の極限值 $\hat{t} > 0$ に対し， $\{v_k\} \subset ([\hat{t}, 1] \times \mathbf{R}^n) \cap \Omega \cap L$ が成り立ち，さらに $([\hat{t}, 1] \times \mathbf{R}^n) \cap \Omega \cap L$ は有界閉集合となる．また， \tilde{F} の $([\hat{t}, 1] \times \mathbf{R}^n) \cap \Omega \cap L$ 上での連続的微分可能性より， $\nabla_x \tilde{F}$ は $([\hat{t}, 1] \times \mathbf{R}^n) \cap \Omega \cap L$ 上で連続である．したがって， $\nabla_x \tilde{F}(v)$ の特異値は $v \in ([\hat{t}, 1] \times \mathbf{R}^n) \cap \Omega \cap L$ で連続となるため， $\nabla_x \tilde{F}(v)$ の特異値は $([\hat{t}, 1] \times \mathbf{R}^n) \cap \Omega \cap L$ 上で最小値を持つ．さらに，補足2.3から任意の $k \geq 0$ に対し $\nabla_x \tilde{F}_k$ は正則なので， $\nabla_x \tilde{F}_k$ の最小特異値は正となる．以上のことから， $\nabla_x \tilde{F}_k$ の最小特異値 ρ_k に対し，全ての $k \geq 0$ において， $c_1 \leq \rho_k$ が成り立つような正の定数 c_1 が存在する．したがって，すべての $k \geq 0$ に対して

$$\|(\nabla_x \tilde{F}_k)^{-1}\| \leq \frac{1}{c_1}$$

が成り立つ．また，(3.4)より

$$\begin{aligned} \frac{|(\bar{t}\gamma_k - t_k)\nabla_t \tilde{F}_k \tilde{F}_k|}{\|\nabla_x \tilde{F}_k \tilde{F}_k\|} &= \frac{\|(\nabla_x \tilde{F}_k)^{-1}\| |(\bar{t}\gamma_k - t_k)\nabla_t \tilde{F}_k \tilde{F}_k|}{\|(\nabla_x \tilde{F}_k)^{-1}\| \|\nabla_x \tilde{F}_k \tilde{F}_k\|} \\ &\leq \frac{\|(\nabla_x \tilde{F}_k)^{-1}\| \|\nabla_t \tilde{F}_k\| \|\tilde{F}_k\| |\bar{t}\gamma_k - t_k|}{\|(\nabla_x \tilde{F}_k)^{-1}\| \|\nabla_x \tilde{F}_k \tilde{F}_k\|} \\ &< \frac{\bar{t} + 1}{c_1} \|\nabla_t \tilde{F}_k\| \end{aligned} \quad (3.7)$$

が成立する．ここで， $\{v_k\} \subset ([\hat{t}, 1] \times \mathbf{R}^n) \cap \Omega \cap L$ かつ $([\hat{t}, 1] \times \mathbf{R}^n) \cap \Omega \cap L$ が有界閉集合であることと， $\nabla_t \tilde{F}$ は $([\hat{t}, 1] \times \mathbf{R}^n) \cap \Omega \cap L$ 上で連続であることから， $\{\|\nabla_t \tilde{F}_k\|\}$ は上に有界である．したがって，(3.6)と(3.7)より

$$|\theta_k| \leq 1 + \frac{|(\bar{t}\gamma_k - t_k)\nabla_t \tilde{F}_k \tilde{F}_k|}{\|\nabla_x \tilde{F}_k \tilde{F}_k\|} \leq c_2 \quad (3.8)$$

を満たす正の定数 c_2 が存在する．

次に $\|\nabla_x \Psi_k\|$ と $\|y_{k-1}\|$ を評価する． \tilde{F} は $([\hat{t}, 1] \times \mathbf{R}^n) \cap \Omega \cap L$ 上で連続的微分可能であることから， $\nabla_x \Psi$ は $([\hat{t}, 1] \times \mathbf{R}^n) \cap \Omega \cap L$ 上で連続となる．さらに， $([\hat{t}, 1] \times \mathbf{R}^n) \cap \Omega \cap L$ は有界閉集合なので，

$$\|\nabla_x \Psi_k\| \leq c_3 \quad (3.9)$$

を満たす正の定数 c_3 が存在する．また，(2.14)より $\lim_{k \rightarrow \infty} \|v_k - v_{k-1}\| = 0$ が成り立つこと，および，有界閉集合 $([\hat{t}, 1] \times \mathbf{R}^n) \cap \Omega \cap L$ 上で $\nabla_x \Psi$ が連続であることより，全ての $k \geq k_0$ に対して，

$$\|y_{k-1}\| = \|\nabla_x \Psi_k - \nabla_x \Psi_{k-1}\| \leq \frac{b\varepsilon^2}{2c_3} \quad (3.10)$$

が成り立つような正の整数 k_0 と正の定数 $b \in (0, 1)$ が存在する. したがって, 全ての $k \geq k_0$ に対して, (3.5), (3.8), (3.9), (3.10) より

$$\|\tilde{d}_k\| \leq c_2 c_3 + 2c_3 \frac{b\varepsilon^2}{2c_3 \varepsilon^2} \|\tilde{d}_{k-1}\| = c_4 + b\|\tilde{d}_{k-1}\|$$

が成り立つ. ただし, $c_4 = c_2 c_3$ である. 一方, $\nabla_x \Psi_k = 0$ の場合には, (2.4) より $\|\tilde{d}_k\| = 0$ であるので, $\|\tilde{d}_k\| \leq c_4 + b\|\tilde{d}_{k-1}\|$ が成立する. したがって,

$$\|\tilde{d}_k\| \leq c_4(1 + b + b^2 + \cdots + b^{k-k_0-1}) + b^{k-k_0}\|\tilde{d}_{k_0}\| \leq \frac{c_4}{1-b} + \|\tilde{d}_{k_0}\|$$

となり, $\{\tilde{d}_k\}$ が有界であることが示された. \square

以上の準備のもとで, アルゴリズム SSCG の大域的収束性に関する次の定理を得る.

定理 3.3. 仮定 2.1 が成り立つならば, アルゴリズム SSCG によって生成される点列 $\{v_k\}$ は少なくとも一つの集積点を持ち, $\lim_{k \rightarrow \infty} t_k = 0$ が成り立つ. さらに, 任意の集積点 v^* において $H(v^*) = 0$ が満たされる. したがって, $v^* = (0, x^*)$ となる x^* は (1.1) の解である.

証明. 補足 2.2 より $\{v_k\}$ は少なくとも一つの集積点を持つ. また, $\{t_k\}$ は下に有界かつ非増加列なので, $\lim_{k \rightarrow \infty} t_k = \hat{t} \in [0, 1]$ となる極限が存在する. ここで, $\hat{t} = 0$ を背理法で証明するために, $\hat{t} > 0$ であると仮定する. まず, (2.14) より次の (A), (B):

$$(A) : \liminf_{k \rightarrow \infty} \alpha_k = 0, \quad (B) : \liminf_{k \rightarrow \infty} \|d_k\| = 0$$

のうち, 少なくとも一方は成り立つ.

はじめに, (A) が成り立つ場合を考えると, 補足 2.2 と補題 3.2 より, $\{v_k\}$ と $\{d_k\}$ がそれぞれ有界であるため,

$$\lim_{k \in K_2, k \rightarrow \infty} \alpha_k = 0, \quad \lim_{k \in K_2, k \rightarrow \infty} d_k = \hat{d}_{(A)}, \quad \lim_{k \in K_2, k \rightarrow \infty} v_k = \hat{v}_{(A)}$$

を満たす $\{d_k\}$ の集積点 $\hat{d}_{(A)} \in \mathbf{R}^{1+n}$ と $\{v_k\}$ の集積点 $\hat{v}_{(A)} \in ([\hat{t}, 1] \times \mathbf{R}^n) \cap \Omega \cap L$, および部分列 $K_2 \subset \mathbf{N}$ が存在する. また, $\lim_{k \in K_2, k \rightarrow \infty} \alpha_k = 0$ より, $k \in K_2$ が十分大きい場合には, アルゴリズムの Step 3 の直線探索 (2.12) において $l = 0$ とはならないため,

$$\frac{\Psi(v_k + (\alpha_k/\sigma)d_k) - \Psi_k}{(\alpha_k/\sigma)} > -\frac{\delta\alpha_k}{\sigma}\|d_k\|^2$$

が成り立つ. ここで, $k \rightarrow \infty$ ($k \in K_2$) として極限を取ると, $\{d_k\}$ が有界であることと, Ψ が $([\hat{t}, 1] \times \mathbf{R}^n) \cap \Omega \cap L$ 上で連続的微分可能であることから, 不等式 $\nabla\Psi(\hat{v}_{(A)})^T \hat{d}_{(A)} \geq 0$ を得る. 一方で, 命題 2.3 より, 不等式 $\nabla\Psi(\hat{v}_{(A)})^T \hat{d}_{(A)} \leq 0$ が成立するため,

$$\nabla\Psi(\hat{v}_{(A)})^T \hat{d}_{(A)} = 0$$

が成り立つ.

次に (B) が成り立つ場合を考えると, $\{v_k\}$ の有界性より,

$$\lim_{k \in K_3, k \rightarrow \infty} d_k = \hat{d}_{(B)} = 0, \quad \lim_{k \in K_3, k \rightarrow \infty} v_k = \hat{v}_{(B)}$$

を満たす $\{d_k\}$ の集積点 $\widehat{d}_{(B)} \in \mathbf{R}^{1+n}$ と $\{v_k\}$ の集積点 $\widehat{v}_{(B)} \in ([\widehat{t}, 1] \times \mathbf{R}^n) \cap \Omega \cap L$, および部分列 $K_3 \subset \mathbf{N}$ が存在する. したがって, $\nabla \Psi(\widehat{v}_{(B)})^T \widehat{d}_{(B)} = 0$ が成り立つ.

以上のことから,

$$\lim_{k \in K_4, k \rightarrow \infty} \nabla \Psi(v_k)^T d_k = 0, \quad \lim_{k \in K_4, k \rightarrow \infty} d_k = \widehat{d}, \quad \lim_{k \in K_4, k \rightarrow \infty} v_k = \widehat{v}$$

を満たす $\{d_k\}$ の集積点 $\widehat{d} \in \mathbf{R}^{1+n}$ と $\{v_k\}$ の集積点 $\widehat{v} \in ([\widehat{t}, 1] \times \mathbf{R}^n) \cap \Omega \cap L$, および部分列 $K_4 \subset \mathbf{N}$ が存在することが示された. ここで, γ の連続性と Ψ の連続的微分可能性より, (2.8) 式において $k \rightarrow \infty$ ($k \in K_4$) と極限を取ると,

$$0 \leq -(1 - \eta) \|\nabla_x \Psi(\widehat{v})\|^2 + \widehat{t}(\bar{t}\gamma(\widehat{v}) - \widehat{t}) \leq 0$$

を得る. このことと $\widehat{t} > 0$, $\bar{t}\gamma(\widehat{v}) - \widehat{t} \leq 0$ より,

$$\nabla_x \Psi(\widehat{v}) = \nabla_x \tilde{F}(\widehat{v}) \tilde{F}(\widehat{v}) = 0, \quad (3.11)$$

$$\bar{t}\gamma(\widehat{v}) - \widehat{t} = 0 \quad (3.12)$$

が成り立つ. したがって, (3.11) と $\widehat{v} \in ([\widehat{t}, 1] \times \mathbf{R}^n) \cap \Omega \cap L$, および仮定 A2 より, $\tilde{F}(\widehat{v}) = 0$ となる. さらに, (1.5), (2.2), (3.12), $\bar{t}\gamma < 1$ より

$$\widehat{t} = \bar{t}\gamma(\widehat{v}) = \bar{t}\gamma \min\{1, \Psi(\widehat{v})\} < \Psi(\widehat{v}) = \frac{1}{2}\widehat{t}^2$$

を得るが, これは $\widehat{t} \in (0, 1]$ に矛盾するため,

$$\lim_{k \rightarrow \infty} t_k = 0$$

が成り立つ.

さらに, $v^* = (0, x^*)$ を $\{v_k\}$ の任意の集積点とすると, $([0, 1] \times \mathbf{R}^n) \cap \Omega \cap L$ が有界閉集合なので, $v^* \in ([0, 1] \times \mathbf{R}^n) \cap \Omega \cap L$ となり

$$0 \geq \bar{t}\gamma \min\{1, \Psi(v^*)\}$$

が成り立つ. よって $\Psi(v^*) = 0$, すなわち $H(v^*) = 0$ が成り立つ. したがって, (1.4) より $v^* = (0, x^*)$ となる x^* が $F(x^*) = 0$ を満たすことがわかる. \square

4. 数値実験

この節ではアルゴリズム SSCG を用いた数値実験の結果について記述する. 今回, 我々は Matlab R2012b でコーディングを行い, 以下のスペックの計算機:

計算機 : Dell Precision T1650 (64bit),

CPU : Intel(R) Core(TM) i7-3770 @3.40GHz 3.40GHz,

メモリ : 16.0 GB RAM,

を用いて数値実験を行った．また，関数 $F(x) = [F_1(x), \dots, F_n(x)]^T$ の成分 $F_i(x)$ をそれぞれ

$$\begin{aligned}
 P1 : F_i(x) &= \begin{cases} e^{\sqrt{x_i^2 + x_{i+1}^2}} - 1, & i \text{ は奇数} \\ x_{i-1} - x_i, & i \text{ は偶数} \end{cases} \\
 P2 : F_i(x) &= \begin{cases} e^{\sqrt{x_i^2 + x_{i+1}^2}} - 1, & i \text{ は奇数} \\ \min\{x_{i-1}, x_i\}, & i \text{ は偶数} \end{cases} \\
 P3 : F_i(x) &= \begin{cases} \max\{0, x_i + x_{i+1}^2 + 2\} - 2, & i \text{ は奇数} \\ \sqrt{x_{i-1}^2 + x_i^2}, & i \text{ は偶数} \end{cases} \\
 P4 : F_i(x) &= \begin{cases} e^{\sqrt{x_i^2 + x_{i+1}^2}} - 1, & i \text{ は奇数} \\ \max\{x_{i-1}, x_i\}, & i \text{ は偶数} \end{cases} \\
 P5 : F_i(x) &= \begin{cases} e^{|\max\{x_i, x_{i+1}\}|} - 1, & i \text{ は奇数} \\ \min\{x_{i-1}, x_i\}, & i \text{ は偶数} \end{cases} \\
 P6 : F_i(x) &= n - 1 + e^{|x_i|} - \sum_{j=1}^n \cos x_j,
 \end{aligned}$$

と定め，非線形方程式 (1.1) の解を求める問題 P1–P6 を扱った．ここで，問題 P1–P6 の真の解はいずれも $x^* = [0, \dots, 0]^T$ である．また，問題 P1–P6 の関数 $F(x)$ は微分不可能な点を持つ関数を含んでいる．ここで，次のような微分不可能な点を持つ関数 f に対し，平滑化関数 \tilde{f} をそれぞれ

$$\begin{aligned}
 f(\alpha) = \sqrt{\alpha} &\longrightarrow \tilde{f}(t, \alpha) = \sqrt{\alpha + t^2} \\
 f(\alpha) = |\alpha| &\longrightarrow \tilde{f}(t, \alpha) = \sqrt{\alpha^2 + t^2} \\
 f(\alpha) = \max\{\alpha, \beta\} &\longrightarrow \tilde{f}(t, \alpha, \beta) = \frac{1}{2} \left(\alpha + \beta + \sqrt{(\alpha - \beta)^2 + t^2} \right) \\
 f(\alpha) = \min\{\alpha, \beta\} &\longrightarrow \tilde{f}(t, \alpha, \beta) = \frac{1}{2} \left(\alpha + \beta - \sqrt{(\alpha - \beta)^2 + t^2} \right)
 \end{aligned}$$

と定義した．これらの平滑化関数 \tilde{f} を用いて次の平滑化アルゴリズム

SSCG	: アルゴリズム SSCG (2 分割法)
SSCG_q	: アルゴリズム SSCG (2 次補間)
STCG	: 平滑化三項共役勾配法 [12] (2 分割法)
STCG_q	: 平滑化三項共役勾配法 [12] (2 次補間)
SNewton	: 平滑化ニュートン法 [18, 19] (2 分割法)
SNewton_q	: 平滑化ニュートン法 [18, 19] (2 次補間)

の数値実験を行った．ここで，アルゴリズム SSCG における Step 3 (つまりバックトラック法) において， $\sigma = 0.5$ とした方法を 2 分割法と呼んでいる．STCG および STCG_q は提案法と同じ直線探索法を用いている．一方，SNewton および SNewton_q は非線形方程式 $H(v) = 0$ を解くためにニュートン方程式:

$$H(v_k) + \nabla H(v_k)^T d_k = \gamma(v_k) \bar{v}$$

を解いて探索方向 d_k を決定している. ここで, $\bar{v} = (\bar{t}, 0, \dots, 0) \in \mathbf{R}^{1+n}$ である. SNewton および SNewton_q では直線探索において, Armijo 条件を満たすようにステップ幅を決定している.

今回の実験では, SSCG, SSCG_q, STCG, STCG_q ではパラメータを $\sigma = 0.5$, $\bar{\gamma} = 0.9$, $\bar{t} = \min\{0.1, 1/\sqrt{n}\}$, $\delta = 0.1$, $\eta = 0.1$, $\sigma_{\min} = 0.1$, $\sigma_{\max} = 0.9$ とした. 一方, SNewton および SNewton_q では $\bar{t} = \min\{0.1, 1/n\}$ として実験を行っている. 問題 P1–P5 において $\nabla H(v)$ は疎行列となるため, SNewton および SNewton_q で探索方向 d_k を定める際に行列の疎性を利用していることを注意しておく. また, 問題 P6 は $\nabla H(v)$ が密行列となっている.

今回の実験では, 各問題の次元を $n = 1000, 2000, 4000$ の 3 通りに設定し, さらに, それぞれの場合に対し 100 個の初期点を生成して実験を行った. つまり, 各解法ごとに 1800 通りの実験を行ったこととなる. 初期点に関しては P1–P5 に関しては $[-5, 5]^n$ から, P6 に関しては $[-1, 1]^n$ からランダムに点を生成した. また, 各アルゴリズムの停止条件を

$$\|F(x_k)\| \leq 10^{-5}$$

と定め数値実験を行った. また, 反復回数が 1000 回を超えた場合や目的関数値の計算においてオーバーフローを起こした場合もアルゴリズムを終了している.

今回, 各手法間の性能を比較するためにパフォーマンスプロファイル [3] を使用する. n_s 種類の解法と n_p 種類の問題に対して, 解法 s のパフォーマンスプロファイル $P_s: \mathbf{R} \rightarrow [0, 1]$ は以下のように定義される:

\mathcal{P} と \mathcal{S} をそれぞれ, 問題と解法の集合とする. 各問題 $p \in \mathcal{P}$ と各解法 $s \in \mathcal{S}$ に対して, $t_{p,s}$ を

$t_{p,s}$ = 解法 s が問題 p を解くのに必要な CPU 時間 (または反復回数や関数評価回数など)

によって定義する. さらに, 比率 $r_{p,s}$ を

$$r_{p,s} = \frac{t_{p,s}}{\min_{s' \in \mathcal{S}} t_{p,s'}}$$

とする. このとき, 解法 s のパフォーマンスプロファイルは

$$P_s(\tau) = \frac{1}{n_p} |\{p \in \mathcal{P} \mid r_{p,s} \leq \tau\}|$$

によって定められる. ただし, 有限集合 A に対して, $|A|$ はその要素数を表すこととする. したがって, パフォーマンスプロファイル $P_s(\tau)$ は最も早く求解できた解法の CPU 時間の τ 倍以内に解けた問題の割合を表している. よって, いくつかの解法のパフォーマンスプロファイルのグラフを並べた場合に, 最も上に位置するパフォーマンスプロファイルを持つ解法が最も効率的であると解釈できる. また, $\tau = 1$ のときのパフォーマンスプロファイルの値は最も早く求解できた問題の割合を表し, τ が十分大きいときのパフォーマンスプロファイルの値は解くことのできた問題の割合を表している.

図 1–3 は, それぞれ, CPU 時間, 反復回数, 関数評価回数を評価基準とした場合のパフォーマンスプロファイルである. また, 図 4–9 では, 各問題ごとの CPU 時間を評価基準としたパフォーマンスプロファイルを与えている. まず, 図 1 を見ると, SSCG, STCG, SNewton_q

は $n = 8$ においてもパフォーマンスプロファイルの値が 1 に届いていない。これは、それらの解法では解けなかった問題が存在することを示している。実際、図 4-9 から分かるように、SSCG, STCG は問題 P1, P2, P4, P6 においては解けていない問題があり、また、SNewton_q は問題 P3 に対しては、初期点や次元に関係なく解けていない。一方、SSCG_q, STCG_q を比較すると、ほんのわずかだけ STCG_q が優勢ながら、SSCG_q と STCG_q はほぼ同等の性能であることが分かる。さらに、その 2 つの解法と SNewton を比較すると、SNewton よりも SSCG_q, STCG_q の方が効果的であることが分かる。最後に図 2-3 を見ると、程度の差はあるものの反復回数や関数評価回数を評価基準とした場合も CPU 時間を評価基準とした場合と同様の傾向があることが分かる。

以上のことをまとめると、今回の実験においては、提案法と平滑化三項共役勾配法 [12] は 2 分割法よりも 2 次補間の方が相性がよく、平滑化ニュートン法 [18, 19] は 2 次補間よりも 2 分割法の方が相性が良かった。また、SSCG_q と STCG_q はほぼ同等であり、どちらも SNewton より効果的であることが分かった。

最後に表 1 では各問題を解くのにかった実行時間を掲載している。各問題は各次元ごとに 100 個の初期点を与えており、表 1 の数値は平均実行時間 (秒) を表している。ただし、上述した通り、初期点によっては解けなかったことがあったため、その場合は平均を算出する際に除外している。また、“-” は、初期点によらず、その問題は解けなかったことを意味している。問題 P1-5 に関してみると、SSCG, SSCG_q, STCG, STCG_q は、どの問題も高々 4 秒以内に解けていることが分かる。一方、SNewton と SNewton_q は次元が大きくなるにつれて、他の方法よりも求解に時間がかかるようになっていくのが分かる。この傾向は、行列 $\nabla H(v)$ が密行列であるような問題 P6 に対しては、より顕著であり、 $n = 4000$ のとき、SSCG_q や STCG_q と比べて、SNewton や SNewton_q は非常に時間がかかっているのが分かる。このことから、SSCG_q は STCG_q と同様に大規模な問題に対して有効であることが分かった。

5. おわりに

本論文では、微分不可能な関数を含む非線形方程式系に対する行列を用いない数値解法として、平滑化スケーリング共役勾配法を提案した。提案法の探索方向は、常にメリット関数 Ψ の降下方向となる。さらに、直線探索において Armijo 条件の変種を用いたアルゴリズムを構築し、その大域的収束性を証明した。また、数値実験によって、提案法と平滑化三項共役勾配法 [12]、平滑化ニュートン法 [18, 19] を比較して、有効性を検証した。検証の結果、今回の数値実験結果では、提案法は平滑化ニュートン法よりも効率的であり、平滑化三項共役勾配法とほぼ同等の性能であることが分かった。

本論文では、平滑化三項共役勾配法 [12] の改良を目的としていたが、提案法の性能は平滑化三項共役勾配法と同程度であった。この理由の一つは探索方向の設定であることが推測できる。探索方向 (2.5) では分母に $\|\nabla_x \Psi_k\|^2$ が存在する。 $\nabla_x \Psi_k = 0$ ならば $\tilde{d}_k = 0$ となるが、 $\nabla_x \Psi_k \neq 0$ 、かつ $\nabla_x \Psi_k \rightarrow 0$ となった場合には、数値的に不安定な状況に陥る可能性がある。したがって、今後の課題としては、分母に $\|\nabla_x \Psi_k\|$ の項がないような平滑化スケーリング共役勾配法の構築があげられる。

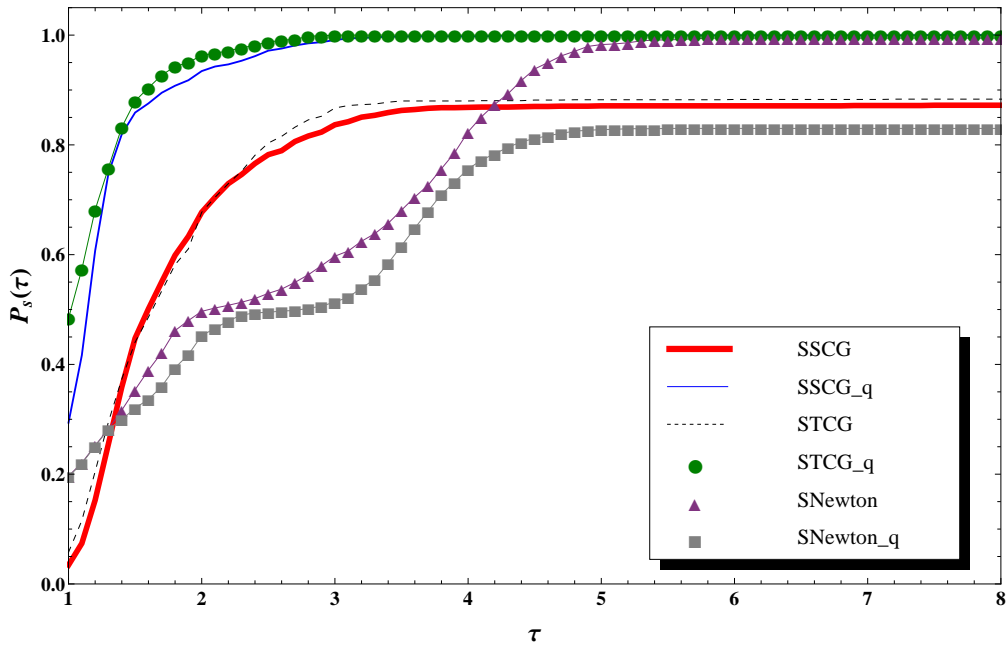


図 1: CPU 時間を評価基準としたパフォーマンスプロファイル

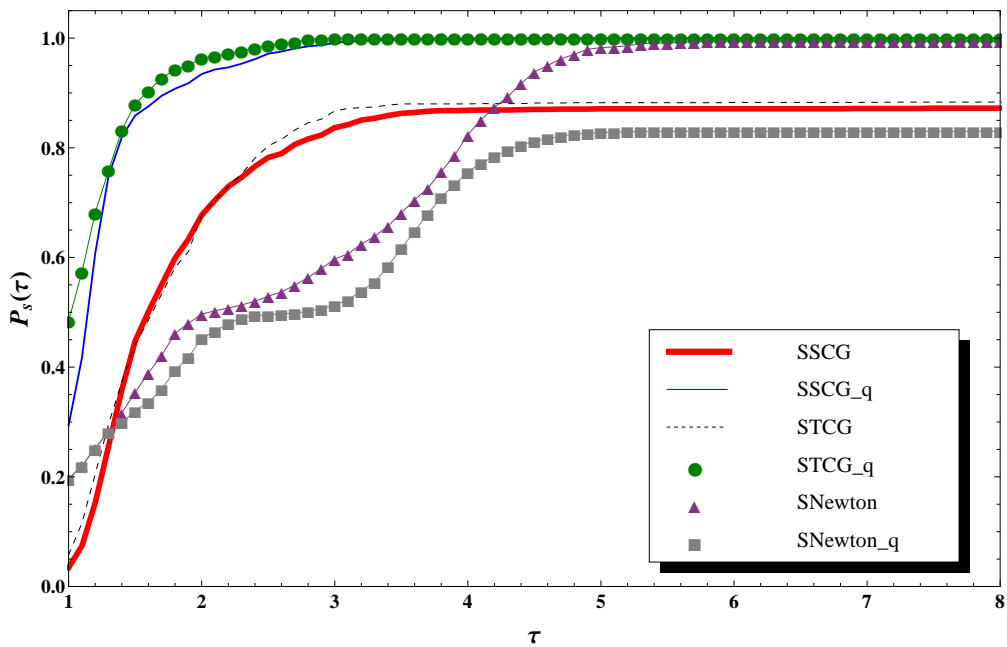


図 2: 反復回数を評価基準としたパフォーマンスプロファイル

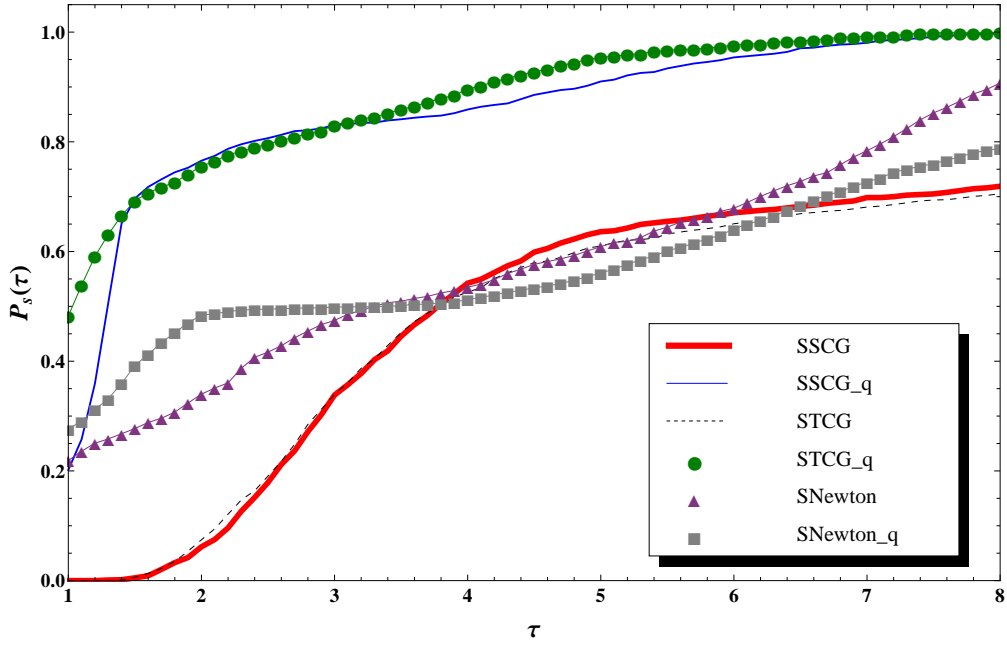


図 3: 関数評価回数を評価基準としたパフォーマンスプロファイル

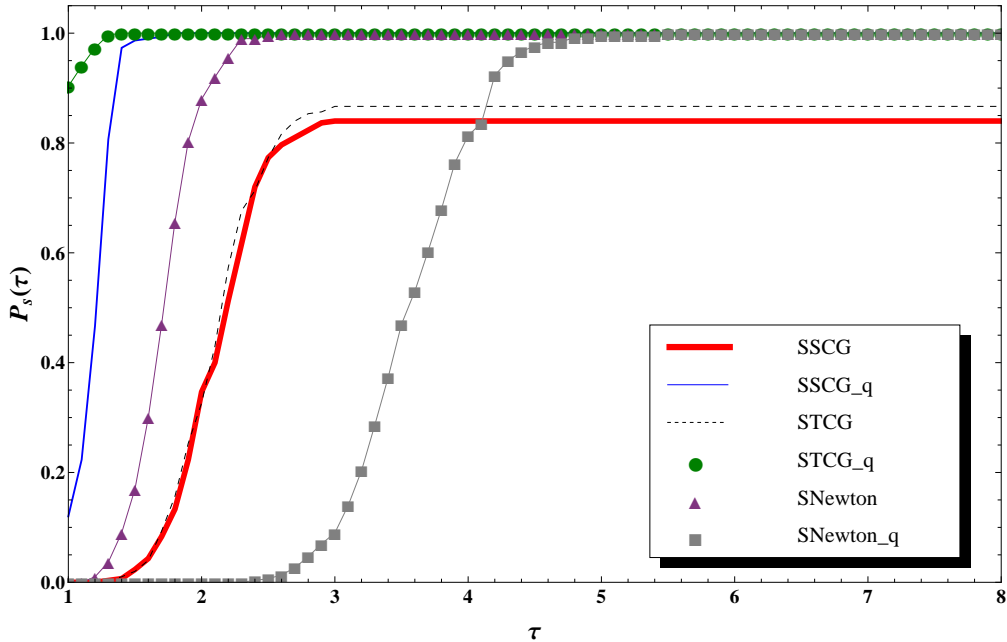


図 4: 問題 P1 のパフォーマンスプロファイル (CPU 時間)

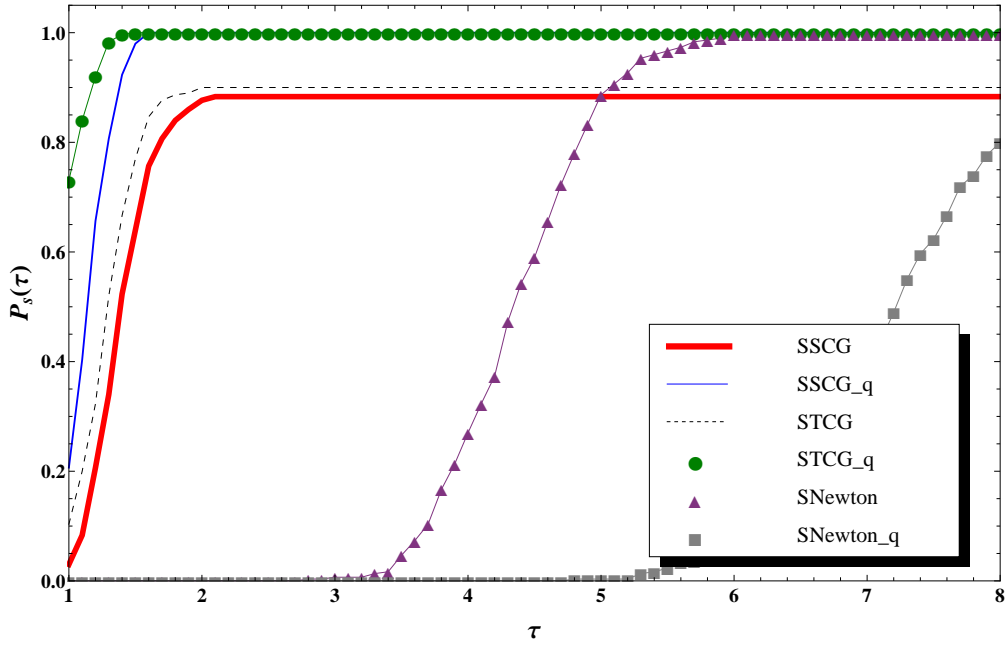


図 5: 問題 P2 のパフォーマンスプロファイル (CPU 時間)

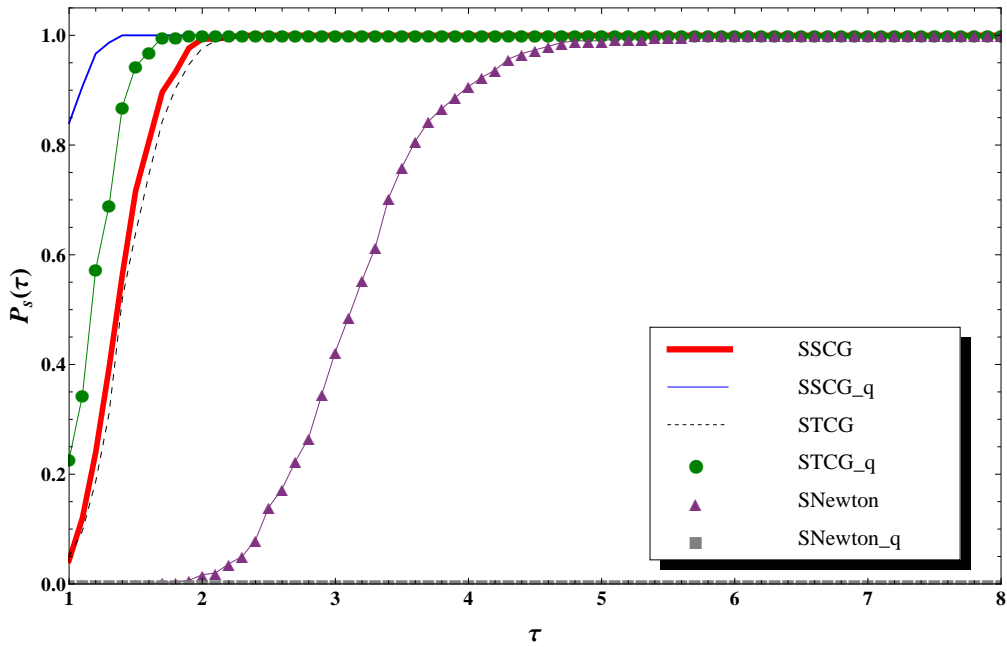


図 6: 問題 P3 のパフォーマンスプロファイル (CPU 時間)

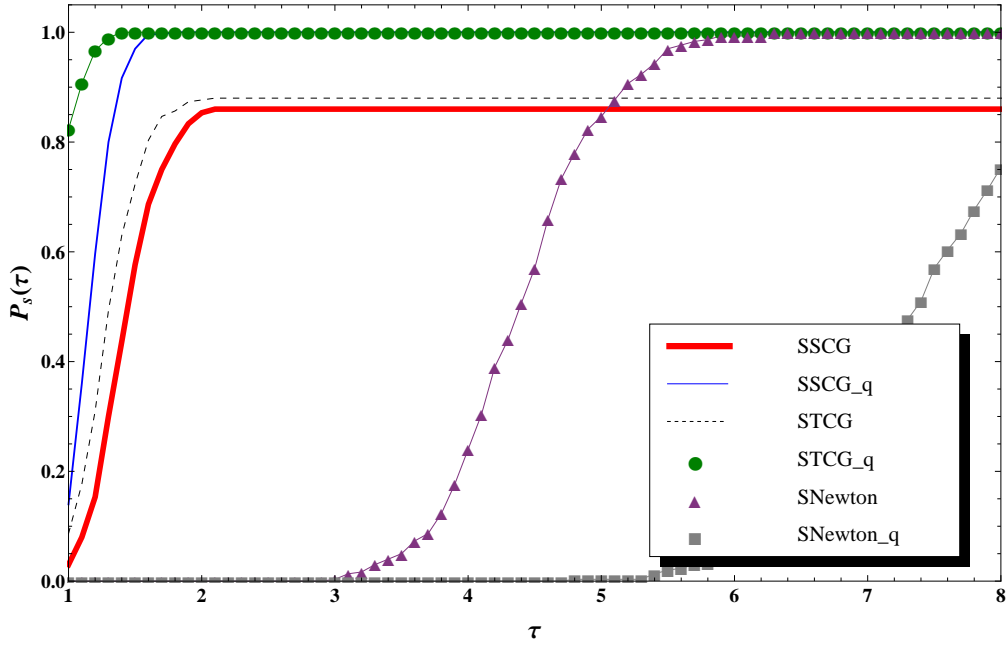


図 7: 問題 P4 のパフォーマンスプロファイル (CPU 時間)

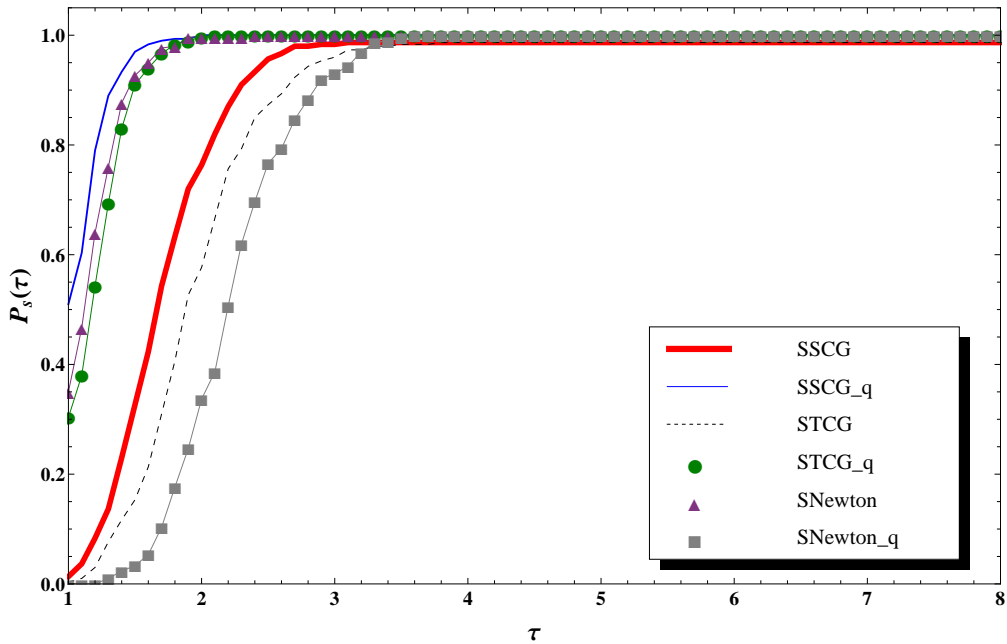


図 8: 問題 P5 のパフォーマンスプロファイル (CPU 時間)

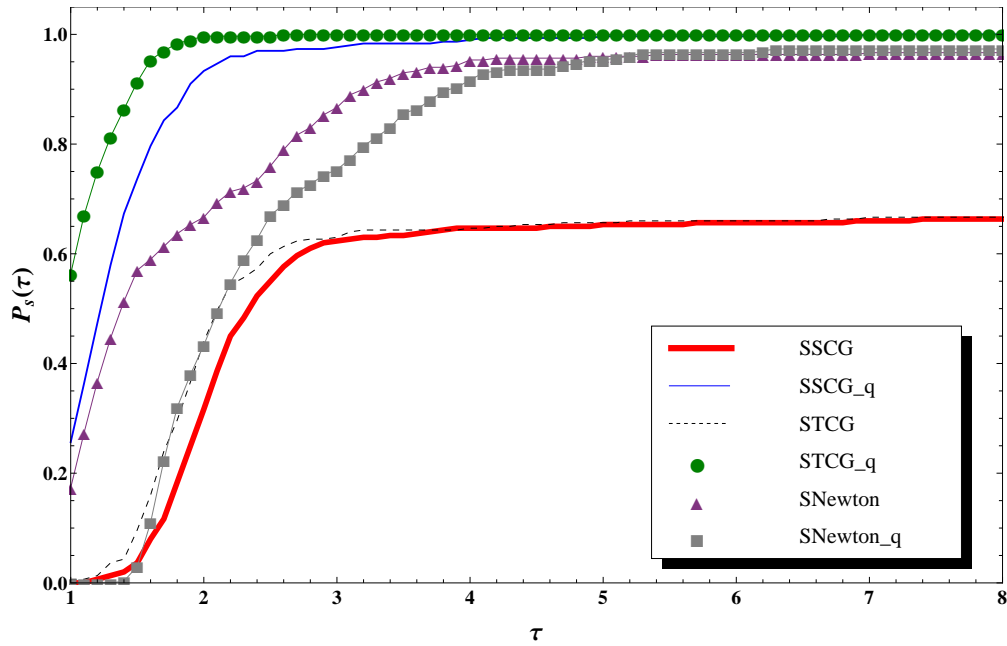


図 9: 問題 P6 のパフォーマンスプロファイル (CPU 時間)

表 1: 平均実行時間 (秒)

P	次元 n	SSCG	SSCG_q	STCG	STCG_q	Snewton	Snewton_q
1	1000	0.25	0.168	0.253	0.134	0.248	0.489
	2000	0.818	0.544	0.835	0.474	0.785	1.649
	4000	2.923	2.008	3.067	1.795	2.794	6.032
2	1000	0.269	0.252	0.25	0.218	0.957	1.482
	2000	0.829	0.811	0.789	0.723	3.312	5.084
	4000	2.999	2.845	2.895	2.711	10.745	18.875
3	1000	0.212	0.164	0.22	0.184	0.454	—
	2000	0.723	0.527	0.713	0.612	1.599	—
	4000	2.554	1.879	2.769	2.267	6.364	—
4	1000	0.254	0.254	0.253	0.209	0.941	1.466
	2000	0.796	0.771	0.736	0.686	2.944	5.095
	4000	3.018	2.885	2.758	2.635	10.544	18.716
5	1000	0.239	0.158	0.264	0.17	0.167	0.304
	2000	0.737	0.497	0.848	0.56	0.555	1.03
	4000	2.822	1.829	3.254	1.963	1.87	3.675
6	1000	4.747	3.097	4.211	2.633	2.418	11.374
	2000	15.857	10.16	14.854	8.753	9.796	42.681
	4000	—	23.891	—	21.058	46.796	143.946

謝辞

本論文の執筆にあたり、査読者より貴重なご意見をいただいた。この場を借りて感謝したい。また、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)および若手研究(B)(25870239)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 Y. Yuan: A nonlinear conjugate gradient method with a strong global convergence property. *SIAM Journal on Optimization*, **10** (1999), 177–182.
- [3] E.D. Dolan and J.J. Moré: Benchmarking optimization software with performance profiles. *Mathematical Programming*, **91** (2002), 201–213.
- [4] F. Facchinei and J.S. Pang: *Finite-Dimensional Variational Inequalities and Complementarity Problems I, II* (Springer-Verlag, New York, 2003).
- [5] R. Fletcher and C.M. Reeves: Function minimization by conjugate gradients. *The Computer Journal*, **7** (1964) 149–154.
- [6] J.A. Ford, Y. Narushima, and H. Yabe: Multi-step nonlinear conjugate gradient methods for unconstrained minimization. *Computational Optimization and Applications*, **40** (2008), 191–216.
- [7] J.C. Gilbert and J. Nocedal: Global convergence properties of conjugate gradient methods for optimization. *SIAM Journal on Optimization*, **2** (1992), 21–42.
- [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 methods. *Pacific Journal of Optimization*, **2** (2006), 35–58.
- [10] 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.
- [11] W. Nakamura, Y. Narushima, and H. Yabe: Nonlinear conjugate gradient methods with sufficient descent properties for unconstrained optimization. *Journal of Industrial and Management Optimization*, **9** (2013), 595–619.
- [12] Y. Narushima: A smoothing conjugate gradient method for solving systems of nonsmooth equations. *Applied Mathematics and Computation*, **219** (2013), 8646–8655.
- [13] Y. Narushima, N. Sagara, and H. Ogasawara: A smoothing Newton method with Fischer–Burmeister function for second-order cone complementarity problems. *Journal of Optimization Theory and Applications*, **149** (2011), 79–101.
- [14] Y. Narushima and H. Yabe: A survey of sufficient descent conjugate gradient methods for unconstrained optimization. *SUT Journal of Mathematics*, **50** (2015), 167–203.
- [15] 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.

- [16] J. Nocedal and S.J. Wright: *Numerical Optimization, Second Edition, Springer Series in Operations Research* (Springer Verlag, New York, 2006).
- [17] L. Qi and J. Sun: A nonsmooth version of Newton's method. *Mathematical Programming*, **58** (1993), 353–367.
- [18] L. Qi and D. Sun: A survey of some nonsmooth equations and smoothing Newton methods. In A Eberhard, R. Hill, D. Ralph, and B.M. Glover (eds.): *Progress in Optimization* (Springer, 1999), 121–146.
- [19] L. Qi, D. Sun, and G. Zhou: A new look at smoothing Newton methods for nonlinear complementarity problems and box constrained variational inequalities. *Mathematical Programming*, **87** (2000), 1–35.
- [20] 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.
- [21] 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.
- [22] 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.

成島康史
横浜国立大学
国際社会科学研究院
〒240-8501 神奈川県横浜市
保土ヶ谷区常盤台79-4
E-mail: narushima@ynu.ac.jp

ABSTRACT

GLOBAL CONVERGENCE OF SMOOTHING AND SCALING PRP TYPE
CONJUGATE GRADIENT METHOD FOR SYSTEM OF
NONSMOOTH EQUATIONS

Yasushi Narushima Ryouyusuke Ootani Hiroshi Yabe
Yokohama National University Suzuyo Shinwart Corp. Tokyo University of Science

This paper treats numerical methods to solve systems of nonsmooth equations. Such problems arise in solving variational inequality problems, complementarity problems and so forth. Although smoothing Newton methods are known as efficient methods for solving systems of nonsmooth equations, these cannot be applied directly to large-scale problems because of the storage of memories for matrices. On the other hand, particular attention is paid to conjugate gradient methods for solving large-scale unconstrained optimization problems, because they do not require the use of matrices.

In this paper, combining the smoothing technique and the PRP type scaling conjugate gradient method, we propose a smoothing and scaling conjugate gradient method which does not use any matrices. Moreover, we show its global convergence. Finally, some numerical results are given.