数理解析研究所講究録
第 1629 巻 2009 年 234-243

制約無し最小化問題に対する Inexact Cubic Regularized Newton 法

京都大学大学院情報学研究科　上田 健詞 (Kenji Ueda)
山下 信雄 (Nobuo Yamashita)
Graduate School of Infomatics,
Kyoto University

1 序論

本稿では，以下の制約無し最小化問題に対するニュートン型手法の計算量を考える．

$$\underset{x\in\mathbf{R}^n}{\text{minimize}}\, f(x) \tag{1.1}$$

ここで，目的関数 $f:\mathbf{R}^n\to\mathbf{R}$ は 2 階連続的微分可能とする．問題 (1.1) に対するアルゴリズムとして，最急降下法やニュートン型の手法など多くの手法が提案され，その性質が理論的に調べられている．これまでのアルゴリズムの研究の多くは，収束率や大域的収束性について議論されてきたが，初期点から停留点を求めるまでに必要な全体の計算量 (または反復回数) はほとんど考慮されていなかった．適切な初期点がわからない問題に対しては，実際に解を見つけるまでに必要な計算時間を収束率によって見積もることができない．一方，大規模な問題など厳密な解を求めることが難しい問題では，計算時間 (反復回数) と解の精度の関係があらかじめわかっていることが望ましい．最近，そのような観点から計算量を見積もることのできるニュートン型のアルゴリズムがいくつか提案され，その計算量からアルゴリズムの優劣が議論されるようになってきている [1, 4, 5].

今，あるアルゴリズムの反復回数を k，そのアルゴリズムによって生成された点列を $\{x_k\}$ とし，$\epsilon>0$ を用いた基準

$$\|\nabla f(x_k)\|\le\epsilon \tag{1.2}$$

を満たす最初の反復を J とする．以下では，J をアルゴリズムの計算量ということにする．適当な仮定の下で，最急降下法の計算量は

$$J=O(\epsilon^{-2})$$

となることが知られている [3]. 最近，ニュートン型の手法として，Nesterov と Polyak [5] は Cubic Regularization of Newton Method (以下，CRN 法) を提案し，その計算量を見積もった．この手法では，k 回目の反復における反復点 x_k において，以下で定義される 3 次モデル関数 $m_k:\mathbf{R}^n\times\mathbf{R}\to\mathbf{R}$ で目的関数を近似する．

$$m_k(d,\sigma)=f(x_k)+\nabla f(x_k)^Ts+\frac{1}{2}s^T\nabla^2 f(x_k)s+\frac{1}{3}\sigma\|s\|^3$$

そして，ある σ_k に対する $m_k(s,\sigma_k)$ の最適解 $s_k^*\in\operatorname{argmin}_{s\in\mathbf{R}^n} m_k(s,\sigma_k)$ を用いて，反復点を $x_{k+1}=x_k+s_k^*$ として更新する．厳密解 s_k^* が求まるとき，CRN 法の計算量は

$$J=O(\epsilon^{-\frac{3}{2}})$$

となり，最急降下法よりもよいということが示されている [5]. ただし，以下で述べるように，s_k^* を求めるためには，それと等価なある非線形方程式を解かなければならない．上記の計算量にはこの非線形方程式を正確に解くために必要な計算量が含まれていない．

s_k^* が 3 次モデル関数 $m_k(s,\sigma_k)$ の最適解であることの必要十分条件は以下の条件を満たす λ_k^* が存在することである [1].

$$\begin{aligned}(\nabla^2 f(x_k)+\lambda_k^*I)s_k^*&=-\nabla f(x_k),\\ \lambda_k^*&=\sigma_k\|s_k^*\|,\\ \nabla^2 f(x_k)+\lambda_k^*I&\succeq 0.\end{aligned} \tag{$*$}$$

これより，$\lambda > \min(0, -\lambda_{\min}(\nabla^2 f(x_k)))$ である領域において，λ に対する一変数の方程式

$$\lambda - \sigma_k \| - (\nabla^2 f(x_k) + \lambda I)^{-1} \nabla f(x_k) \| = 0 \tag{1.3}$$

を解くことによって，(*) を満たす解 λ_k^* と $s_k^* = -(\nabla^2 f(x_k) + \lambda_k^* I)^{-1} \nabla f(x_k)$ を求めることができる．ただし，一変数の方程式であっても，ニュートン法で解くときには，その各反復で n 変数の線形方程式を解かなければならない．

Cartis, Gould, Toint [1] は CRN 法を拡張した Adaptive Cubic Overestimation method (以下，ACO 法) を提案した．ACO 法では，3 次正則化パラメータ σ_k の更新に信頼領域法のアイデアを取り入れ，さらに，3 次モデル関数の最適解 s_k^* の近似解 $\hat{s}_k$ がある適当な条件を満たせば，それを採用しても計算量のオーダーが変わらないことを示した．その条件の 1 つは，近似解がある部分空間上で 3 次モデル関数の最適解であることであり，そのため，近似解を求めるときに (1.3) と同様の非線形方程式を解かなければならない．

目的関数が凸関数である問題に対しては，その性質を利用することによって，より効率のよいアルゴリズムを構築することができる．Nesterov [4] は f の凸性を利用した CRN 法の高速化手法を提案し，その計算量が

$$J = O(\epsilon^{-\frac{1}{3}})$$

となることを示した．また，Polyak [6] はある特殊なステップ幅を用いた 2 次の正則化ニュートン法を提案し，その計算量が

$$J = O(\epsilon^{-4})$$

となることを示した．この計算量は CRN 法の計算量よりも劣る．しかし，2 次の正則化ニュートン法は CRN 法と違い，線形方程式を 1 回解くだけで次の反復点が求まるという長所がある．なお，Polyak の手法のステップ幅は，$\nabla f(x)$ の Lipschitz 定数を用いて計算しているため，その定数が未知の場合には上記の結果は成り立たない．

本稿では，まず，3 次モデル関数の近似解で反復点を更新する枠組みとして inexact CRN 法 (以下，ICRN 法) を提案する．次に，ICRN 法の具体例として inexact ACO 法 (以下，IACO 法) を提案する．さらに，ニュートン法と IACO 法を組み合わせたハイブリッド法を提案する．このハイブリッド法では，2 次の十分条件が成り立つとき，解の十分近くでニュートン法が用いられることを示す．

本稿では以下の表記を用いる．ベクトル $x \in \mathbb{R}^n$ に対して，x のノルム $\|x\|$ をユークリッドノルム $\|x\| := \sqrt{x^T x}$ で定義する．対称行列 $M \in \mathbb{R}^{n \times n}$ に対する最大固有値と最小固有値をそれぞれ，$\lambda_{\max}(M), \lambda_{\min}(M)$ と表す．また，$A \in \mathbb{R}^{n \times n}$ のノルム $\|A\|$ を $\|A\| := \sqrt{\lambda_{\max}(A^T A)}$ で定義する．

2 Inexact Cubic Regularized Newton 法

本節では，ICRN 法を提案し，その収束性について述べる．k 回目の反復における反復点を x_k とし，目的関数 f の 3 次モデル関数 $m_k : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ を以下のように定義する．

$$m_k(s, \sigma) = f(x_k) + g_k^T s + \frac{1}{2} s^T B_k s + \frac{1}{3} \sigma \|s\|^3$$

ここで，$g_k = \nabla f(x_k)$，B_k は目的関数 f のヘッセ行列 $\nabla^2 f(x_k)$ の近似行列である．提案手法では，ある σ_k に対する $m_k(s, \sigma_k)$ の最小化問題

$$\underset{s \in \mathbb{R}^n}{\text{minimize}}\, m_k(s, \sigma_k)$$

の近似解 s_k を用いて，$x_{k+1} = x_k + s_k$ として点列を生成する．なお，[5] では s_k として $m_k(s, \sigma_k)$ の最適解を用い，[1] では，ある部分空間 $\mathrm{L}(\subseteq \mathbb{R}^n)$ 上の最適解，つまり $s_k \in \operatorname{argmin}_{s \in \mathrm{L}} m_k(s, \sigma_k)$ を用いている．ICRN 法のアルゴリズムは以下の通りである．

Inexact cubic regularized Newton 法

Step1. 初期点 x_0 を与える. $k := 0$ とする.
Step2. 適当な終了条件を満たしていれば停止. そうでなければ Step3 へ.
Step3. ある σ_k に対する 3 次モデル関数最小化問題

$$\underset{s\in\mathbb{R}^n}{\mathrm{minimize}}\, m_k(s,\sigma_k)$$

の近似解 s_k を求める.
Step4. $x_{k+1} = x_k + s_k$ とする.
Step5. $k := k+1$ として Step2 へ.

このアルゴリズムの重要なところは Step3 における正則化パラメータ σ_k の決め方と探索方向 s_k の求め方である. その具体的な内容については次節で与える. また, Step3 で $\sigma_k = 0$ とすると, ICRN 法は inexact ニュートン法となるため, ICRN 法は inexact ニュートン法を拡張したものと考えることができる.

ICRN 法の収束性について述べる. そのために, 目的関数 f に対して次のことを仮定する.

仮定 1. 目的関数 f は下に有界である.

提案アルゴリズムで求めたステップ s_k が次の条件を満たすとする.

近似条件 1. 次の条件を満たすある $p>0, q>0$ が存在する.

$$f(x_k) - f(x_{k+1}) \geq p \min_{0\leq l\leq k+1} \|g_l\|^q, \quad \forall k \geq 0.$$

[5] では, パラメータ σ_k を十分大きい値に固定し, s_k が $m_k(s,\sigma_k)$ の厳密な最適解であれば, $q = 3/2$ とある定数 p に対して近似条件 1 を満たすことが示されている.

探索方向 s_k が近似条件 1 を満たすとき, ICRN 法の計算量は以下の定理で与えられる.

定理 2.1. 仮定 1 が成り立つとする. また, $\{x_k\}$ を ICRN 法で生成された点列とし, 探索方向 s_k は近似条件 1 を満たすとする. このとき, $\|g_k\| \leq \epsilon$ を満たす最初の反復 J は

$$J = O(\epsilon^{-q})$$

となる.

証明. f は下に有界であるので,

$$f(x) \geq f_{\min}, \quad \forall x \in \mathbb{R}^n$$

を満たす $f_{\min}$ が存在する. 条件 1 より, ある $p>0, q>0$ が存在して,

$$f(x_0) - f_{\min} \geq f(x_0) - f(x_k) = \sum_{l=0}^{k-1}(f(x_l) - f(x_{l+1})) \geq pk \min_{0\leq l\leq k} \|g_l\|^q$$

となる. したがって,

$$\min_{0\leq l\leq k} \|g_l\| \leq \left(\frac{f(x_0) - f_{\min}}{pk}\right)^{\frac{1}{q}}$$

が得られる. これより,

$$k \geq \left(\frac{f(x_0) - f_{\min}}{p}\right)\epsilon^{-q} \Rightarrow \min_{0\leq l\leq k} \|g_l\| \leq \epsilon$$

が成り立つので，$\|g_k\| \leq \epsilon$ を満たす最初の反復 J は

$$J = O(\epsilon^{-q})$$

となる． □

ICRN 法の Step3 で近似条件 1 を満たす s_k を求めることができれば，3 次モデル関数 $m_k(s, \sigma_k)$ の最小化問題を厳密に解かなくても上の定理で得られる収束性を持つことがわかる．ただし，上の定理では，ICRN 法の Step3 の内部反復でかかる計算量については何も述べていない．

3 正則化パラメータ σ_k と探索方向 s_k の選び方

本節では，近似条件 1 を満たす正則化パラメータ σ_k と探索方向 s_k の具体的な選び方について議論する．以下では，目的関数のヘッセ行列 $\nabla^2 f$ とヘッセ行列の近似行列 B_k に対して次のことを仮定する．

仮定 2.

(a) 目的関数 f のヘッセ行列 $\nabla^2 f(x)$ は有界である; すなわち，以下の関係を満たす $U_H > 0$ が存在する．

$$\|\nabla^2 f(x)\| \leq U_H, \quad \forall x \in \mathbb{R}^n$$

(b) $\nabla^2 f(x)$ は Lipschitz 連続である; すなわち，以下の関係を満たす $L_H > 0$ が存在する．

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L_H \|x - y\|, \quad \forall x, y \in \mathbb{R}^n$$

(c) $\nabla^2 f(x_k)$, B_k, s_k に対して以下の関係を満たす $C_H \geq 0$ が存在する．

$$\|(\nabla^2 f(x_k) - B_k)s_k\| \leq C_H \|s_k\|^2, \quad \forall k \geq 0$$

仮定 2 (c) は，$B_k = \nabla^2 f(x_k)$ や $B_k = \nabla^2 f(x_k) + C_H\|s\|I$ とすれば成り立つ．

次に，近似条件 1 を調べるために重要な役割を果たす命題を述べておく．この命題は $\|s_k\|$ と $\|g_{k+1}\|$ の関係を与えている．

命題 3.1. ([1], Lemma 4.6 , Lemma 6.4). 仮定 2 が成り立つとする．ある $0 \leq c_1 < 1, c_2 \geq 0$ が存在して，

$$\|\nabla_s m_k(s_k, \sigma_k)\| \leq \|g_k\| \min(c_1, c_2\|s_k\|), \quad \forall k \geq 0 \tag{3.1}$$

が成り立つとする．さらに，正則化パラメータ σ_k が上に有界，すなわち，ある $\sigma_{\max} > 0$ が存在して，

$$\sigma_k \leq \sigma_{\max}, \quad \forall k \geq 0.$$

が成り立つとする．このとき，

$$\|s_k\|^2 \geq \kappa_1 \|g_{k+1}\|, \quad \kappa_1 := \frac{1 - c_1}{(\frac{1}{2}L_H + C_H + c_2 U_H + \sigma_{\max})} \tag{3.2}$$

が成り立つ．

探索方向 s_k が 3 次モデル関数 $m_k(s, \sigma_k)$ の最適解であるときは，$\nabla_s m_k(s_k, \sigma_k) = 0$ となるため，この命題の仮定 (3.1) が成り立つ．

3.1 Inexact Adaptive Cubic Overestimation 法

序論で述べたように，3 次モデル関数 $m_k(s, \sigma_k)$ の最適解 s_k^* を求めるためには $(*)$ を満たす λ_k^*, s_k^* を求めなければならない．この節では，まず，CRN 法や ACO 法と同じ計算量が維持できる $(*)$ の近似解の条件を与える．そして，この条件に基づいた IACO 法を提案する．

次の補題は 3 次モデル関数の減少量の下界値を与えるものである．

補題 3.2. c_3 を $\frac{2}{3} < c_3 \le 1$ を満たす定数とする．$(\bar\lambda_k, \bar s_k)$ が以下の 3 つの条件を満たすとする．

$$\bar\lambda_k \ge c_3\sigma_k\|\bar s_k\|, \tag{3.3}$$
$$(B_k + \bar\lambda_k I)\bar s_k = -g_k, \tag{3.4}$$
$$B_k + \bar\lambda_k I \succeq 0 \tag{3.5}$$

このとき，

$$f(x_k) - m_k(\bar s_k, \sigma_k) \ge \frac{1}{2}\left(c_3 - \frac{2}{3}\right)\sigma_k\|\bar s_k\|^3$$

が成り立つ．

証明．

$$\begin{aligned}
f(x_k) - m_k(\bar s_k, \sigma_k) &= -g_k^T\bar s_k - \frac{1}{2}\bar s_k^T B_k\bar s_k - \frac{1}{3}\sigma_k\|\bar s_k\|^3 \\
&= \bar s_k^T(B_k + \bar\lambda_k I)\bar s_k - \frac{1}{2}\bar s_k^T B_k\bar s_k - \frac{1}{3}\sigma_k\|\bar s_k\|^3 && (3.6)\\
&= \frac{1}{2}\bar s_k^T(B_k + \bar\lambda_k I)\bar s_k + \frac{1}{2}(\bar\lambda_k - \frac{2}{3}\sigma_k\|\bar s_k\|)\|\bar s_k\|^2 \\
&\ge \frac{1}{2}\left(c_3 - \frac{2}{3}\right)\sigma_k\|\bar s_k\|^3 && (3.7)
\end{aligned}$$

ここで，(3.6) 式の導出には (3.4) 式を，(3.7) 式の不等式の関係は (3.3) 式と (3.5) 式を用いた． □

CRN 法では，$\bar s_k$ として 3 次モデル関数 $m_k(s, \sigma_k)$ の最適解を用いるため，$\bar\lambda_k = \sigma_k\|\bar s_k\|$ が成り立つ．したがって，3 次モデル関数の減少量の下界値は

$$f(x_k) - m_k(\bar s_k, \sigma_k) \ge \frac{1}{6}\sigma_k\|\bar s_k\|^3 \tag{3.8}$$

で与えられる．また，ACO 法では，$\bar s_k$ として 3 次モデル関数 $m_k(s, \sigma_k)$ のある部分空間上の最適解を用いるが，この場合も (3.8) が成り立つことが示されている．

次の補題は命題 3.1 の仮定 (3.1) の十分条件を与えるものである．

補題 3.3. c_1 と c_2 を $0 < c_1 < 1, c_2 > 0$ を満たす定数とする．$(\bar\lambda_k, \bar s_k)$ が以下の 2 つの条件を満たすとする．

$$|\bar\lambda_k - \sigma_k\|\bar s_k\|| \cdot \|\bar s_k\| \le \|g_k\|\min(c_1, c_2\|\bar s_k\|), \tag{3.9}$$
$$(B_k + \bar\lambda_k I)\bar s_k = -g_k \tag{3.10}$$

このとき，

$$\|\nabla_s m_k(\bar s_k, \sigma_k)\| \le \|g_k\|\min(c_1, c_2\|\bar s_k\|)$$

が成り立つ．

証明．

$$\begin{aligned}
\|\nabla_s m_k(\bar s_k, \sigma_k)\| &= \|g_k + (B_k + \sigma_k\|\bar s_k\|I)\bar s_k(\sigma_k)\| \\
&= \|(-\bar\lambda_k + \sigma_k\|\bar s_k\|)\bar s_k\| && (3.11)\\
&= |\bar\lambda_k - \sigma_k\|\bar s_k\|| \cdot \|\bar s_k\| \\
&\le \|g_k\|\min(c_1, c_2\|\bar s_k\|) && (3.12)
\end{aligned}$$

ここで，(3.11) 式の導出には (3.10) 式を，(3.12) 式の不等式の関係は (3.9) 式を用いた． □

補題 3.2, 3.3 の仮定をまとめたものを近似条件として提案する.

近似条件 2. c_1, c_2, c_3 を $0 < c_1 < 1, c_2 > 0, \frac{2}{3} < c_3 \le 1$ を満たす定数とする. $(\bar{\lambda}_k, \bar{s}_k)$ は以下の条件を満たす.

(a) $|\bar{\lambda}_k - \sigma_k\|\bar{s}_k\|| \cdot \|\bar{s}_k\| \le \|g_k\| \min(c_1, c_2\|\bar{s}_k\|)$
(b) $\bar{\lambda}_k \ge c_3\sigma_k\|\bar{s}_k\|$
(c) $(B_k + \bar{\lambda}_k I)\bar{s}_k = -g_k$
(d) $B_k + \bar{\lambda}_k I \succeq 0$

近似条件 2 を満たす探索方向 $\bar{s}_k$ を用いて反復点を更新するアルゴリズム IACO 法を以下に示す.

Inexact adaptive cubic overestimation 法

Step1. 初期点 x_0 を与える. パラメータ $\sigma_0, \sigma_{\min}, c_1, c_2, c_3, \gamma_1, \gamma_2, \eta_1, \eta_2$ を以下のように与える.

$$\sigma_0 \ge \sigma_{\min} > 0,\ 0 \le c_1 < 1,\ c_2 \ge 0,\ \frac{2}{3} < c_3 \le 1,$$
$$1 < \gamma_1 \le \gamma_2,\ 0 < \eta_1 \le \eta_2 \le 1$$

$k := 0$ とする.

Step2. 適当な終了条件を満たしていれば停止. そうでなければ Step3 へ.

Step3. Step3.1. 近似条件 2 を満たす $(\bar{\lambda}_k, \bar{s}_k)$ を求める.

Step3.2.

$$\rho_k(s, \sigma) := \frac{f(x_k) - f(x_k + s)}{f(x_k) - m_k(s, \sigma)}$$

とする. $\rho_k(\bar{s}_k, \sigma_k) < \eta_1$ ならば, $\sigma_k \in [\gamma_1\sigma_k, \gamma_2\sigma_k]$ として Step3.1 へ戻る.
$\eta_2 > \rho_k(\bar{s}_k, \sigma_k) \ge \eta_1$ ならば, $\sigma_k \in [\sigma_k, \gamma_1\sigma_k]$ として Step4 へ.
$\rho_k(\bar{s}_k, \sigma_k) \ge \eta_2$ ならば, $\sigma_k \in [\sigma_{\min}, \sigma_k]$ として Step4 へ.

Step4. $x_{k+1} = x_k + \bar{s}_k$ とする. $k := k + 1$ として Step2 へ.

CRN 法は, Step3.1 で 3 次モデル関数の最適解を $\bar{s}_k$ とし, Step3.2 で σ_k を

$$\rho_k(\bar{s}_k, \sigma_k) < 1 \Rightarrow \sigma_k := 2\sigma_k,$$
$$\rho_k(\bar{s}_k, \sigma_k) \ge 1 \Rightarrow \sigma_{k+1} = \sigma_k$$

と更新した IACO 法と考えることができる. また, Step3.1 で 3 次モデル関数のある部分空間上の最適解で命題 3.1 の仮定 (3.1) を満たすような解を $\bar{s}_k$ としたものが ACO 法に対応している.

以下では, IACO 法の性質について調べる. まず, Step3 の内部反復回数が k に依存しない定数で抑えられることを示す. そのために次の補題が必要となる.

補題 3.4. ([1], Lemma 5.2). 仮定 2 が成り立つとする. このとき, 任意の s_k に対して,

$$\sigma_k \ge \frac{3(L_H + C_H)}{2} \Rightarrow f(x_{k+1}) \le m_k(s_k, \sigma_k) \Rightarrow \rho_k(s_k, \sigma_k) \ge 1$$

が成り立つ. さらに,

$$\sigma_{\min} \le \sigma_k \le \sigma_{\max}, \quad \sigma_{\max} := \frac{3\gamma_2(L_H + C_H)}{2}$$

が成り立つ.

この補題より, IACO 法の Step3 の内部反復回数, すなわち, $\bar{s}_k$ を計算する回数の上限を与えることができる.

定理 3.5. 仮定 2 が成り立つとする．このとき，各 k に対して，IACO 法の Step3 で $\bar{s}_k$ を計算する回数は，高々

$$\left\lceil \log_{\gamma_1}\left(\frac{\sigma_{\max}}{\sigma_{\min}}\right)+1\right\rceil$$

回である．

証明． $\bar{s}_k$ が採用されないとき，IACO 法における正則化パラメータ σ_k の更新方法より，σ_k は少なくとも γ_1 倍される．したがって，定理の主張が成り立つ． □

次に，IACO 法で生成された点列 $\{x_k\}$ が近似条件 1 を $q=3/2$ で満たすことを示す．

定理 3.6. 仮定 2 が成り立つとする．$\bar{s}_k$ を IACO 法の Step3.1 で求めた探索方向とする．このとき，$\rho_k(\bar{s}_k,\sigma_k)\geq\eta_1$ が成り立つならば，

$$f(x_k)-f(x_{k+1})\geq p_1\|g_{k+1}\|^{\frac{3}{2}}\geq p_1\min_{0\leq l\leq k+1}\|g_l\|^{\frac{3}{2}}$$

が成り立つ．ただし，$p_1:=(3c_3-2)\eta_1\sigma_{\min}\kappa_1^{\frac{3}{2}}/6$ である．

証明．

$$\begin{aligned}
f(x_k)-f(x_{k+1}) &\geq \eta_1(f(x_k)-m_k(\bar{s}_k,\sigma_k)) \\
&\geq \frac{\eta_1}{2}\left(c_3-\frac{2}{3}\right)\sigma_k\|\bar{s}_k\|^3 && (3.13)\\
&\geq \frac{\eta_1\sigma_{\min}}{2}\left(c_3-\frac{2}{3}\right)\|\bar{s}_k\|^3 && (3.14)\\
&\geq \frac{\eta_1\sigma_{\min}}{2}\left(c_3-\frac{2}{3}\right)(\kappa_1\|g_{k+1}\|)^{\frac{3}{2}} && (3.15)\\
&= \frac{\eta_1\sigma_{\min}\kappa_1^{\frac{3}{2}}}{2}\left(c_3-\frac{2}{3}\right)\|g_{k+1}\|^{\frac{3}{2}} && (3.16)
\end{aligned}$$

ここで，(3.13) 式の導出には補題 3.2 を，(3.14) 式の導出には補題 3.4 を，(3.15) 式の導出には補題 3.3 と命題 3.1 を用いた． □

この結果と定理 2.1 より，ICRN 法の計算量は以下の定理で与えられる．

定理 3.7. 仮定 1, 2 が成り立つとする．$\{x_k\}$ を IACO 法で生成された点列とする．このとき，$\|g_k\|\leq\epsilon$ を満たす最初の反復 J は

$$J=O\left(\epsilon^{-\frac{3}{2}}\right)$$

となる．

上の定理より，IACO 法の計算量は CRN 法や ACO 法の計算量と同じになることがわかる．IACO 法では，3 次モデル関数 $m_k(s,\sigma_k)$ の最適解 s_k^* を反復法で求めるとき，厳密な最適解が得られるまで反復を行う必要がなく，近似条件 2 を満たした時点で反復を終了できる．そのため，探索方向 $\bar{s}_k$ を求める計算量の削減が期待できる．

3.2 ハイブリッド法

3.1 節で提案した IACO 法は，Step3 で近似条件 2 を満たす探索方向を求めなければならない．そのため，各反復で，通常は n 次元の線形方程式を何回も解く必要がある．一方，ニュートン法によって目的関数値を十

分減少させることができるとき，ニュートン法を用いた方が計算時間を削減することができる．そこで，この節では，3.1 節で提案した IACO 法とニュートン法とのハイブリッド法を提案する．

提案アルゴリズムは以下の通りである．

ハイブリッド法

Step1. 初期点 x_0 を与える．パラメータ $\sigma_0, \sigma_{\min}, c_1, c_2, c_3, c_4, \gamma_1, \gamma_2, \eta_1, \eta_2$ を以下のように与える．

$$\sigma_0 \geq \sigma_{\min} > 0,\ 0 \leq c_1 < 1,\ c_2 \geq 0,\ \frac{2}{3} < c_3 \leq 1,\ c_4 > 0,$$
$$1 < \gamma_1 \leq \gamma_2,\ 0 < \eta_1 \leq \eta_2 \leq 1$$

$k := 0$ とする．

Step2. 適当な終了条件を満たしていれば停止．そうでなければ Step3 へ．

Step3. Step3.1. Step3.1.1. 線形方程式 $B_k s + g_k = 0$ を解く．解ければその解を $\bar{s}_k$ として Step3.1.2 へ．解けなければ Step3.2 へ．

Step3.1.2.

$$f(x_k) - f(x_k + \bar{s}_k) \geq c_4 \|\bar{s}_k\|^3 \tag{3.17}$$

を満たせば $\sigma_{k+1} = \sigma_k$ として Step4 へ．そうでなければ Step3.2 へ．

Step3.2. Step3.2.1. 近似条件 2 を満たす $(\bar{\lambda}_k, \bar{s}_k)$ を求める．

Step3.2.2.

$$\rho_k(s, \sigma) := \frac{f(x_k) - f(x_k + s)}{f(x_k) - m_k(s, \sigma)}$$

とする．

$\rho_k(\bar{s}_k, \sigma_k) < \eta_1$ ならば，$\sigma_k \in [\gamma_1 \sigma_k, \gamma_2 \sigma_k]$ として Step3.2.1 へ戻る．

$\eta_2 > \rho_k(\bar{s}_k, \sigma_k) \geq \eta_1$ ならば，$\sigma_k \in [\sigma_k, \gamma_1 \sigma_k]$ として Step4 へ．

$\rho_k(\bar{s}_k, \sigma_k) \geq \eta_2$ ならば，$\sigma_k \in [\sigma_{\min}, \sigma_k]$ として Step4 へ．

Step4. $x_{k+1} = x_k + \bar{s}_k$ とする．$k := k+1$ として Step2 へ．

ハイブリッド法でも，IACO 法と同様に，補題 3.4, 定理 3.5 が成り立つ．

ハイブリッド法では，ニュートン法によって得られる探索方向を採用する条件として (3.17) を用いている．ある解 x^* において最適性の 2 次の十分条件が成り立つとき，x^* の十分近くではニュートン法による探索方向が採用されることを示す．

定理 3.8. ハイブリッド法で生成される点列 $\{x_k\}$ は x^* に収束すると仮定する．さらに，

$$\lambda_{\min}(\nabla^2 f(x^*)) > 0 \tag{3.18}$$

が成り立つとする．また，B_k, s_k に対して

$$\lim_{k \to \infty} \frac{\|(\nabla^2 f(x_k) - B_k) s_k\|}{\|s_k\|} = 0, \quad \forall k \geq 0 \tag{3.19}$$

が成り立つと仮定する．このとき，任意の $c_4 > 0$ に対して，$k \geq \bar{k}$ であれば線形方程式 $B_k s_k + g_k = 0$ の解 s_k に対して

$$f(x_k) - f(x_k + s_k) \geq c_4 \|s_k\|^3$$

を満たす $\bar{k}$ が存在する．

証明. Taylor の定理より，ある $\tau \in (0, 1)$ が存在して，

$$f(x_k + s_k) = f(x_k) + g_k^T s_k + \frac{1}{2} s_k^T \nabla^2 f(x_k + \tau s_k) s_k$$

が成り立つ．これより，

$$
\begin{aligned}
f(x_k) - f(x_k + s_k) &= -g_k^T s_k - \frac{1}{2} s_k^T \nabla^2 f(x_k + \tau s_k) s_k \\
&= s_k^T B_k s_k - \frac{1}{2} s_k^T \nabla^2 f(x_k + \tau s_k) s_k \qquad (3.20) \\
&= \frac{1}{2} s_k^T \nabla f(x_k) s_k + s_k^T (B_k - \nabla^2 f(x_k)) s_k + \frac{1}{2} s_k^T (\nabla^2 f(x_k) - \nabla^2 f(x_k + \tau s_k)) s_k \\
&\geq \frac{1}{2} \lambda_{\min}(\nabla^2 f(x_k)) \|s_k\|^2 - \frac{\|(B_k - \nabla^2 f(x_k)) s_k\|}{\|s_k\|} \cdot \|s_k\|^2 \\
&\quad - \frac{1}{2} \|\nabla^2 f(x_k) - \nabla^2 f(x_k + \tau s_k)\| \cdot \|s_k\|^2 \\
&= \frac{1}{2} \left(\lambda_{\min}(\nabla^2 f(x_k)) - \frac{2\|(B_k - \nabla^2 f(x_k)) s_k\|}{\|s_k\|} - \|\nabla^2 f(x_k) - \nabla^2 f(x_k + \tau s_k)\| \right) \|s_k\|^2 \qquad (3.21)
\end{aligned}
$$

となる．ここで，(3.20) 式の導出には $B_k s_k = -g_k$ を用いた．(3.18) 式より，

$$
\lim_{k \to \infty} \|s_k\| = 0 \qquad (3.22)
$$

を得る．したがって，任意の $c_4 > 0$ に対して，$k \geq k_1$ であれば

$$
\frac{4c_4}{\lambda_{\min}(\nabla^2 f(x^*))} \|s_k\| \leq 1 \qquad (3.23)
$$

を満たす k_1 が存在する．よって，$k \geq k_1$ のとき，(3.23) 式を (3.21) 式に適用すると

$$
\begin{aligned}
&f(x_k) - f(x_k + s_k) \geq \\
&\frac{2c_4}{\lambda_{\min}(\nabla^2 f(x^*))} \left(\lambda_{\min}(\nabla^2 f(x_k)) - \frac{2\|(B_k - \nabla^2 f(x_k)) s_k\|}{\|s_k\|} - \|\nabla^2 f(x_k) - \nabla^2 f(x_k + \tau s_k)\| \right) \|s_k\|^3
\end{aligned} \qquad (3.24)
$$

を得る．$x_k \to x^*$ であるから，(3.18)，(3.19)，(3.22) 式より，

$$
\lim_{k \to \infty} \left(\lambda_{\min}(\nabla^2 f(x_k)) - \frac{2\|(B_k - \nabla^2 f(x_k)) s_k\|}{\|s_k\|} - \|\nabla^2 f(x_k) - \nabla^2 f(x_k + \tau s_k)\| \right) = \lambda_{\min}(\nabla^2 f(x^*))
$$

が成り立つ．したがって，$k \geq k_2$ であれば

$$
\lambda_{\min}(\nabla^2 f(x_k)) - \frac{2\|(B_k - \nabla^2 f(x_k)) s_k\|}{\|s_k\|} - \|\nabla^2 f(x_k) - \nabla^2 f(x_k + \tau s_k)\| \geq \frac{\lambda_{\min}(\nabla^2 f(x^*))}{2} \qquad (3.25)
$$

を満たす k_2 が存在する．(3.24)，(3.25) 式より，$\bar{k} = \max(k_1, k_2)$ とすると，$k \geq \bar{k}$ を満たす任意の k に対して，

$$
f(x_k) - f(x_k + s_k) \geq c_4 \|s_k\|^3
$$

が成り立つ． □

この定理の仮定 (3.19) は，例えば，準ニュートン法で用いられる BFGS 公式で B_k を更新すれば成り立つ．特に，仮定 2 (c) が成り立つとき，(3.19) 式を満たす．

次に，ハイブリッド法でニュートン法による探索方向が採用された場合も，IACO 法と同様に，近似条件 1 を $q = 3/2$ で満たすことを示す．

定理 3.9. 仮定 2 が成り立つとする．$\bar{s}_k$ は $B_k \bar{s}_k + g_k = 0$ かつ (3.17) を満たすとする．このとき，

$$
f(x_k) - f(x_{k+1}) \geq p_2 \|g_{k+1}\|^{\frac{3}{2}} \geq p_2 \min_{0 \leq l \leq k+1} \|g_l\|^{\frac{3}{2}}
$$

が成り立つ．ただし，$p_2 := c_4 \kappa_1^{\frac{3}{2}}$ である．

証明. ニュートン法によって得られる探索方向は，正則化パラメータ σ_k を一時的に 0 としたときの 3 次モデル関数 $m_k(s,0)$ の最適解であるので，$\nabla_s m_k(\bar{s}_k,0)=0$ となり，命題 3.1 が成り立つ．したがって，

$$\begin{aligned} f(x_k) - f(x_k+\bar{s}_k) &\geq c_4\|\bar{s}_k\|^3 \\ &\geq c_4(\kappa_1\|g_{k+1}\|)^{\frac{3}{2}} \\ &= c_4\kappa_1^{\frac{3}{2}}\|g_{k+1}\|^{\frac{3}{2}} \end{aligned}$$

が成り立つ． □

この結果と定理 2.1 より，ハイブリッド法の計算量は IACO 法の計算量と同じになる．

定理 3.10. 仮定 1, 2 が成り立つとする．$\{x_k\}$ をハイブリッド法で生成された点列とする．このとき，$\|g_k\| \leq \epsilon$ を満たす最初の反復 J は

$$J = O\left(\epsilon^{-\frac{3}{2}}\right)$$

となる．

4 Concluding Remarks

本稿では，3 次モデル関数の近似解で反復点を更新する枠組みとして ICRN 法を提案した．次に，その具体的なアルゴリズムとして IACO 法を提案し，その手法が CRN 法や ACO 法と同様の反復回数のオーダーを持つための近似条件を与えた．さらに，ニュートン法と IACO 法を組み合わせたハイブリッド法を提案した．最適性の 2 次の十分条件が成り立つとき，解の十分近くではニュートン法による探索方向が採用され，それを用いても反復回数のオーダーが IACO 法と変わらないことを示した．ニュートン法により得られる探索方向は線形方程式を 1 回解くだけで求まり，非線形方程式 (1.3) を解く必要がないため，計算時間の削減が期待できる．

今後の課題としては，

- CRN 法，ACO 法，IACO 法，ハイブリッド法を比較するための数値実験を行う
- IACO 法で用いる近似条件 2 を満たす探索方向を求めるために必要となる線形方程式を解く回数を見積もる

などが挙げられる．

参考文献

[1] C. Cartis, N. I. M. Gould and P. L. Toint, *Adaptive cubic overestimation methods for unconstrained optimization*, Technical Report, University of Oxford, 2007.

[2] A. R. Conn, N. I. M. Gould and P. L. Toint: *Trust-Region Methods*, SIAM, Philadelphia, USA, 2000.

[3] Y. Nesterov: *Introductory Lectures on Convex Optimization*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004.

[4] ——, *Accelerating the cubic regularization of Newton's method on convex problems*, Math. Program., Ser.B, **112** (2008), pp. 159–181.

[5] Y. Nesterov and B. T. Polyak, *Cubic regularization of Newton method and its global performance*, Math. Program., Ser.A, **108** (2006), pp. 177–205.

[6] R. A. Polyak, *Regularized newton method for unconstrained convex optimization*, Math. Program., Ser.B.