

制約なし最小化問題に対する勾配法，ニュートン型手法の反復回数の見積もり

山下 信雄
京都大学

上田 健詞
三菱電機株式会社

(受理 2012 年 3 月 8 日; 再受理 2013 年 2 月 25 日)

和文概要 最急降下法やニュートン法など，非線形計画問題に対する解法の多くは反復法である．反復法に関する評価基準としては大域的収束性や局所的な収束率があるが，これらの評価基準では，総合的な計算時間を見積もることはできない．それに対して，初期点から適当な近似解を求めるまでにかかる反復回数の上限の見積もりが，近年盛んに研究されている．本稿は勾配法およびニュートン型手法に関する研究結果をまとめたサーベイ論文である．

キーワード: 非線形計画，勾配法，ニュートン法，大域的な反復回数

1. はじめに

本稿では，特に断らない限り，制約のない非線形計画問題：

$$\begin{aligned} & \text{minimize} && f(x) \\ & \text{subject to} && x \in \mathbb{R}^n \end{aligned} \tag{1.1}$$

を考える．ここで， f は \mathbb{R}^n から \mathbb{R} への目的関数である．

非線形計画問題 (1.1) に対しては，最急降下法，ニュートン法などさまざまな手法が開発されている．これらの手法の多くは，最適解に収束する点列 $\{x^k\}$ を生成する反復法である．

反復法の収束性を議論する際に重要となる基準は，(i) 解を求めることができるか，(ii) どれくらい速く解を求めることができるかの 2 点である．前者としては，大域的収束性という概念で説明されている．これは，任意の初期点 x^0 から始めてもなんらかの解に収束することを保証する性質である．一方，反復法の速さについては，1 次収束，2 次収束，超 1 次収束などの収束率を基準として議論されることが多い¹．収束率は解のそばでの収束の速さを表す．2 次収束する解法は，反復点が解に近づくと，とても速く収束する．一方，1 次収束しかしない解法はなかなか収束しない．そのため，1 次収束しかしない解法は遅いと考えられている．しかし，収束率を用いた議論では，解のそばまで到達する計算時間が考慮されていない．そのため，実際には解のそばまではすぐに到達する手法でも遅い解法とみなされてしまうことがある．

一方，組み合わせ最適化問題においては，解法の評価基準として，(iii) 最悪の場合の計算量を考えることが多い．非線形計画においても，内点法などではそのような計算量が見積もられている．しかし，通常の最急降下法などでは，このような計算量についてはあまり知ら

¹点列 $\{x^k\}$ が解 x^* (あるいは目的関数列 $\{f(x^k)\}$ が最小値 $f(x^*)$) に収束するとき， $\|x^{k+1} - x^*\| \leq C\|x^k - x^*\|^p$ (あるいは $f(x^{k+1}) - f(x^*) \leq C(f(x^k) - f(x^*))^p$) を満たす正の定数 C, p が存在するとき，点列 $\{x^k\}$ (または数列 $\{f(x^k)\}$) は x^* (または $f(x^*)$) に p 次収束するという．

れていない．近年，非線形計画問題に対しても，このような評価基準 (iii) に関連して，大域的反復回数というものが，盛んに研究されている²．大域的反復回数とは，なんらかの近似解を求めるために必要となる反復回数の上限のことである．具体的には， f_{\min} を最小値，つまり $f_{\min} = \inf_x f(x)$ としたときに，

$$f(x^k) - f_{\min} \leq \varepsilon_f \quad (1.2)$$

となるような近似解 x^k を求めるために必要な反復回数 k の上限である．ただし， $\varepsilon_f > 0$ は近似精度を表す正の定数である．なお， f が凸でないようなときには，一般に最小値 f_{\min} を求めることは容易ではない．そのようなときは，停留点，つまり $\nabla f(x^*) = 0$ となる点 x^* を求めることが目的になる．その際の大域的反復回数は

$$\min_{0 \leq i \leq k} \|\nabla f(x^i)\| \leq \varepsilon_g \quad (1.3)$$

を満たす k の上限として与えられる³．ここで，近似精度を表す正の定数 ε_g の “ g ” は gradient の g である．

大域的反復回数は， $O(\varepsilon^{-1})$ とか， $O(\ln \varepsilon^{-1})$ などというように，近似精度を表す定数 ε とランダウ記号 O を用いて表されることが多い．ここで，1回の反復にかかる計算時間 (例えば勾配の計算) を T とすると，大域的反復回数が $O(\varepsilon^{-1})$ である解法の計算量は $O(T\varepsilon^{-1})$ となる．大域的反復回数の大雑把な見積もりとしては，劣1次収束をする場合は $O(\varepsilon^{-1})$ ，(大域的に)1次収束する場合は $O(\ln \varepsilon^{-1})$ と考えるとよい．ここで，劣1次収束とは，ある正の定数 a が存在して，

$$f(x^k) - f_{\min} \leq \frac{a}{k}$$

となることである．このとき， $k \geq \frac{a}{\varepsilon}$ であれば， $f(x^k) - f_{\min} \leq \frac{a}{k} \leq \varepsilon$ が成り立つので，大域的反復回数は $O(\varepsilon^{-1})$ となる．なお，2次収束する場合の反復回数は $O(\ln \ln \varepsilon^{-1})$ となり，著しく少ない回数となる．しかしながら，この反復回数の見積もりは “局所的” なものであることに注意して欲しい．

本稿で扱う大域的反復回数において，特徴的なことは，以下の2点である．

- f の強凸性，2次の十分条件などの，強い正則性の仮定は用いていない．
- 大域的反復回数を与える具体的な式に 問題の規模 n が現れない．

強い正則性とは，大雑把に言えば， f のヘッセ行列 $\nabla^2 f(x)$ が正定値行列になることである．強い正則性を仮定すれば，たいていの解法は1次収束するため，大域的反復回数は $O(\ln \varepsilon^{-1})$ となり，理論的にはあまり面白くない．本稿で紹介する結果は，そのような強い仮定を必要としていないことに注意して欲しい．また，それらの結果は，決定変数の次元 n に依存していない．一方，内点法の反復回数は $O(\sqrt{n})$ というように n に依存して与えられている⁴．

以下では，大域的反復回数は， ε_f を用いたときは目的関数を用いた条件 (1.2)， ε_g を用いたときは勾配を用いた条件 (1.3) を満たすものとする．一般に， $\varepsilon_f = \varepsilon_g$ であれば，(1.3) の方が厳しい条件となる．実際， $f(x) = x^2$ のときには， $f_{\min} = 0$ より $f(x^k) - f_{\min} = (x^k)^2$

²すべてを網羅することは不可能なので，詳しくは [3, 11] やその参考文献を見て欲しい．

³ $\{\|\nabla f(x^k)\|\}$ は単調に減少するとは限らない．この条件は， k 回反復すれば，少なくとも1回は勾配の大きさが ε_g 以下になるという意味である．

⁴実際の数値計算では，内点法の反復回数も次元の影響をうけないことが知られている．また，勾配法でも1回の反復の計算 (勾配の計算など) には n に依存した計算時間が必要であることに注意して欲しい．

であり， $|\nabla f(x^k)| = 2|x^k|$ となることからわかる．この例では， $O(\varepsilon_g^{-2})$ と $O(\varepsilon_f^{-1})$ は同じようなものとみなせる．一方， $f(x) = e^{-x}$ のときは， $f(x^k) - f_{\min} = e^{-x} = |\nabla f(x^k)|$ となるから， $\varepsilon_f = \varepsilon_g$ であれば，条件 (1.2) と条件 (1.3) は等価になる．よって，大域的反復回数が $O(\varepsilon_g^{-2})$ となる解法と $O(\varepsilon_f^{-1})$ となる解法があれば， $O(\varepsilon_f^{-1})$ となる解法の方が優れていることになる．

大域的反復回数の解析は，数理計画だけでなく，機械学習，信号処理などの数理計画の応用分野でも盛んに研究されている．特に，目的関数 f が特別な形をした凸関数であるような問題に対しては，その応用分野の方が進んだ結果を出していることがある．一方， f が凸でないときの大域的反復回数は，ここ数年の間に数理計画の研究者によって，様々な研究成果が出されている．これまでに大域的収束性や局所的収束率についての結果は，様々なサーベイ論文や教科書にまとめられている．しかし，大域的反復回数について，凸および非凸の場合を俯瞰的に紹介しているものはない．そこで，本稿では，非線形計画問題に対する勾配法やニュートン型手法の大域的反復回数に関する結果を紹介する．

これらの結果を俯瞰的に見る上で重要となる概念がリプシッツ連続性である．このリプシッツ連続性については次節でその詳細について説明する．3 節では，目的関数 f が凸な場合において，リプシッツ連続性を用いて，勾配法の大域的反復回数を与える．一方，4 節では f が凸でない場合の解法，特にニュートン法に基づく手法の結果を紹介する．最後に，5 節において本稿のまとめを与える．

2. リプシッツ連続性について

ある関数 $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ が集合 $S \subseteq \mathbb{R}^n$ 上でリプシッツ連続であるとは，

$$\|F(x) - F(y)\| \leq L\|x - y\|, \quad \forall x, y \in S$$

が成り立つ定数 L が存在することである．この定数 L のことをリプシッツ定数とよぶ．ここで，右辺と左辺に現れる $\|\cdot\|$ はそれぞれの適当なノルムとする．以下では，特に断らない限り， \mathbb{R}^n 上のベクトルに対してはユークリッドノルム， $\mathbb{R}^{n \times n}$ 上の行列に対してはスペクトルノルム⁵ を考えるものとし， S は全空間 \mathbb{R}^n とする．

本稿では，目的関数 f の勾配 ∇f とヘッセ行列 $\nabla^2 f$ のリプシッツ連続性に基づいて大域的な反復回数を見積もる．以下では，それぞれのリプシッツ定数を L_g と L_H と書くことにする．リプシッツ定数の大きさをイメージするために，リプシッツ関数 f が 2 次関数，つまり $f(x) = \frac{1}{2}x^T Ax + b^T x$ と表されている場合を考えよう． $\nabla f(x) = Ax + b$ となるから， A の最大の特異値を σ_{\max} とすると， ∇f は定数 $L_g = \sigma_{\max}$ のリプシッツ連続であることがいえる．また，ヘッセ行列は定数行列 $\nabla^2 f(x) = A$ となるから， $L_H = 0$ となる．ここで，決定変数の次元 n が大きいからといって，リプシッツ定数も大きくなるとは限らないことに注意しよう．実際，多くの応用問題においては，リプシッツ定数は，問題の規模の影響を受けず，ある一定の値となる．また，一般には，勾配のリプシッツ連続性は，ヘッセ行列のリプシッツ連続性よりも厳しい条件といえる．実際， $f(x) = x^3$ を考えると，この関数の勾配はリプシッツ連続とはならないが，ヘッセ行列は $L_H = 6$ のリプシッツ連続となることがわかる⁶．

次に，リプシッツ連続性によって成り立つ性質を紹介する．その性質から，最急降下法や 3 次正則化ニュートン法を導くことができる．

⁵ $n \times n$ 行列 A に対して $\|A\| = \max_{v \in \mathbb{R}^n, \|v\|=1} \|Av\|$.

⁶有界な集合上では，リプシッツ連続となる場合が多い．そのため，生成される点列 $\{x^k\}$ が有界であると仮定し，その点列を含む有界集合上でリプシッツ連続性を仮定することはよく行われている．

目的関数の勾配 ∇f がリプシッツ連続であるとき，任意の $x, d \in \mathbb{R}^n$ に対して，

$$\begin{aligned} f(x+d) - f(x) &= \int_0^1 \nabla f(x + \tau d)^T d \, d\tau \\ &= \nabla f(x)^T d + \int_0^1 (\nabla f(x + \tau d) - \nabla f(x))^T d \, d\tau \\ &\leq \nabla f(x)^T d + \int_0^1 \|\nabla f(x + \tau d) - \nabla f(x)\| \|d\| \, d\tau \\ &\leq \nabla f(x)^T d + \frac{L_g}{2} \|d\|^2, \end{aligned}$$

つまり，

$$f(x+d) \leq f(x) + \nabla f(x)^T d + \frac{L_g}{2} \|d\|^2$$

が成り立つ．ここで，

$$m_1(x, d) \equiv f(x) + \nabla f(x)^T d + \frac{L_g}{2} \|d\|^2$$

とすると，この不等式は

$$f(x+d) \leq m_1(x, d) \quad (2.1)$$

と書ける．関数 m_1 は f の 1 次の近似式に 2 次の正則化項をつけたものとみなせる．

不等式 (2.1) を用いれば，次のようにして“最急降下法”が導ける．関数 $m_1(x^k, d)$ は d に関して凸 2 次関数であるから， $m_1(x^k, \cdot)$ を最小とする \bar{d} は，最適性の 1 次の必要条件 $0 = \nabla_d m_1(x^k, \bar{d}) = \nabla f(x^k) + L_g \bar{d}$ より， $\bar{d} = -\frac{1}{L_g} \nabla f(x^k)$ である．不等式 (2.1) に $x = x^k$ ， $d = \bar{d}$ を代入すると，

$$f\left(x^k - \frac{1}{L_g} \nabla f(x^k)\right) \leq m_1\left(x^k, -\frac{1}{L_g} \nabla f(x^k)\right) \leq m_1(x^k, 0) = f(x^k) \quad (2.2)$$

を得る．ここで，左辺にあらわれる $x^k - (1/L_g)\nabla f(x^k)$ を次の反復点 x^{k+1} とする反復法を考えよう．このとき，不等式 (2.2) から， $f(x^{k+1}) \leq f(x^k)$ が成り立つ．つまり，この反復法によって生成される点列は，目的関数値が増加せず（実際には，以下の (2.3) で示すように減少する），解に収束することが期待できる．この反復法は，ステップ幅を $1/L_g$ とした最急降下法にほかならない．

一方，(2.1) において $x = x^k$ ， $d = -t_k \nabla f(x^k)$ ， $x^{k+1} = x^k - t_k \nabla f(x^k)$ を代入すると，

$$f(x^{k+1}) \leq f(x^k) - t_k \left(1 - \frac{L_g t_k}{2}\right) \|\nabla f(x^k)\|^2 \quad (2.3)$$

を得る．つまり，最急降下法 $x^{k+1} = x^k - t_k \nabla f(x^k)$ では，ステップ幅 t_k を $2/L_g$ よりも小さく選べば， $f(x^{k+1}) < f(x^k)$ とできる．4 節では，この不等式を利用して，最急降下法の大域的な反復回数を見積もる．

上記の議論では，勾配のリプシッツ連続性に基づいた不等式 (2.1) から最急降下法が導かれること，さらにそのステップ幅 t_k をリプシッツ定数に基づいて定めれば，目的関数を減少させることができることがわかった．そこで次にヘッセ行列 $\nabla^2 f$ がリプシッツ連続である場合を考えよう．ヘッセ行列がリプシッツ連続であるとき，(2.1) と同様にして，次の 2 つの不等式を示すことができる [9]．

$$\|\nabla f(x) + \nabla^2 f(x)d - \nabla f(x+d)\| \leq \frac{L_H}{2} \|d\|^2 \quad (2.4)$$

$$f(x+d) \leq f(x) + \nabla f(x)^T d + \frac{1}{2} d^T \nabla^2 f(x) d + \frac{L_H}{6} \|d\|^3$$

2 つめの不等式の右辺を $m_2(x, d)$ ，つまり

$$m_2(x, d) \equiv f(x) + \nabla f(x)^T d + \frac{1}{2} d^T \nabla^2 f(x) d + \frac{L_H}{6} \|d\|^3$$

とおくと，この不等式は

$$f(x+d) \leq m_2(x, d) \tag{2.5}$$

と書ける．関数 $m_2(x, d)$ は f の 2 次の近似式に 3 次の正則化項をつけたものである．

最急降下法を導いたときと同じように， $m_2(x^k, d)$ を d に関して最小化する最小解 d^k が求まったとしよう⁷．このとき， $m_2(x^k, 0) = f(x^k)$ であることから，

$$f(x^k + d^k) \leq m_2(x^k, d^k) \leq m_2(x^k, 0) = f(x^k)$$

が成り立つ．よって， $x^{k+1} = x^k + d^k$ とした反復法では， $f(x^{k+1}) \leq f(x^k)$ となることが保証できる．このことは，2 次の近似式に 3 次の正則化項を加えた m_2 を最小化することによって，新たな手法を導くことができることを示唆している．この手法が Nesterov と Polyak によって提案された 3 次正則化ニュートン法である [9]．3 次正則化ニュートン法では， $\nabla^2 f(x)$ が正定値行列でなくても，目的関数値を減らすことができる．また， d^k が十分小さいときは， d^k は $f(x) + \nabla f(x)^T d + \frac{1}{2} d^T \nabla^2 f(x) d$ の最小解の近似になるため，ニュートン方向とほぼ同じになり，強い正則性の仮定の下ではニュートン法と同様の高速な収束が期待できる．

3. 凸計画に対する解法の大域的反復回数

この節では凸計画問題に対する勾配法の大域的反復回数について紹介する．

3.1. 近接勾配法

凸計画問題に対して，適当な正則化項を目的関数に加えて性質を良くした問題を考え，その正則化された問題を逐次的に解くことによって最適解を求める手法がある⁸．近接点法はそのような手法のひとつであり，以下のように点列 $\{x^k\}$ を生成する [10]．

$$x^{k+1} = \operatorname{argmin}_{x \in \mathbb{R}^n} \left\{ t_k f(x) + \frac{1}{2} \|x - x^k\|^2 \right\}$$

ただし， t_k は正のパラメータであり， argmin は最小化問題

$$\begin{aligned} & \text{minimize} && t_k f(x) + \frac{1}{2} \|x - x^k\|^2 \\ & \text{subject to} && x \in \mathbb{R}^n \end{aligned}$$

の解 (集合) を表している⁹．以下では，次の反復点 x^{k+1} を与える最小化問題を部分問題とよぶことにする．近接点法は， t_k を上手に調整すれば，理論的に優れた収束性を持っている

⁷3 次関数の最小化問題を解かなければならないが，その特別な構造を利用して，信頼領域法の部分問題と同様にして解けることが示されている [9]．

⁸次節で紹介する 3 次正則化ニュートン法，正則化ニュートン法もそのような手法であるが， f が凸関数でなくても適用できる．

⁹この問題の最適解は唯一であるため " $x^{k+1} =$ " としているが，最適解が複数存在する場合は " $x^{k+1} \in$ " と書かなければならない．

る [6, 10, 16] . しかし, 部分問題は非線形計画問題であり, 厳密な解を求めることは容易ではない. そこで, 目的関数を線形近似した関数で置き換えた次の部分問題を逐次的に解くことを考える.

$$x^{k+1} = \operatorname{argmin}_{x \in \mathbb{R}^n} \left\{ t_k f(x^k) + t_k \nabla f(x^k)^T (x - x^k) + \frac{1}{2} \|x - x^k\|^2 \right\}$$

部分問題は, 制約のない凸 2 次計画問題であるため, 最適性の 1 次の条件を考えれば,

$$x^{k+1} = x^k - t_k \nabla f(x^k)$$

となる. これは, 最急降下法そのものである.

リプシッツ連続性を用いると, 最急降下法の最大反復回数は簡単に見積もることができる. 以下では, 簡単のため, $t_k = 1/L_g$ とする¹⁰. ただし, L_g は勾配 ∇f のリプシッツ定数である. 最急降下法の更新式より,

$$\|x^{k+1} - x^*\|^2 = \left\| x^k - \frac{1}{L_g} \nabla f(x^k) - x^* \right\|^2 = \|x^k - x^*\|^2 + \frac{2}{L_g} \nabla f(x^k)^T (x^* - x^k) + \frac{1}{L_g^2} \|\nabla f(x^k)\|^2$$

を得る. ただし, x^* は任意の最小解である. ここで, f の凸性¹¹ を利用すると,

$$\|x^{k+1} - x^*\|^2 - \|x^k - x^*\|^2 \leq \frac{2}{L_g} (f(x^*) - f(x^k)) + \frac{1}{L_g^2} \|\nabla f(x^k)\|^2$$

を得る. また, (2.3) より,

$$f(x^k) - f(x^{k+1}) \geq \frac{1}{2L_g} \|\nabla f(x^k)\|^2 \quad (3.1)$$

であるから,

$$\|x^{k+1} - x^*\|^2 - \|x^k - x^*\|^2 \leq \frac{2}{L_g} (f(x^*) - f(x^{k+1}))$$

が成り立つ. 第 0 反復から第 k 反復までのこの不等式を足し合わせると,

$$\frac{2}{L_g} \sum_{i=0}^k (f(x^{i+1}) - f(x^*)) \leq \|x^0 - x^*\|^2 - \|x^{k+1} - x^*\|^2 \leq \|x^0 - x^*\|^2$$

となる. (2.3) より, 任意の $i \leq k$ に対して, $f(x^{i+1}) \geq f(x^{k+1})$ だから,

$$f(x^{k+1}) - f(x^*) \leq \frac{L_g \|x^0 - x^*\|^2}{2(k+1)}$$

を得る. つまり, 最急降下法の大域的な反復回数は $O(\varepsilon_f^{-1})$ となる.

目的関数 f が微分不可能なときは, $\nabla f(x^k)$ のかわりに劣勾配 $\partial f(x^k)$ を用いた劣勾配法が提案されている. 劣勾配法では, 勾配のリプシッツ連続性を用いた解析はできない. また, 最小解 x^* においても, 劣勾配 $\partial f(x^*)$ は 0 でないベクトルを含むことがある. そのときは, ステップ幅 t_k をある定数に固定すると解に収束しないことがある. 実際, $f(x) = |x|$ を考え

¹⁰ 実用的にはステップ幅 t_k は適宜調整したほうがよい.

¹¹ $\nabla f(x^k)^T (x^* - x^k) \leq f(x^*) - f(x^k)$.

ると， $x^k \neq 0$ における勾配は -1 または 1 となるため， $|x^{k+1} - x^k| = t_k$ となり， t_k が定数のときは $\{x^k\}$ は収束しない．そのため， $t_k = 1/k$ とするなどの工夫が必要となる．しかし，このような減少するステップ幅を採用すると収束は遅くなり，大域的反復回数は $O(\varepsilon_f^{-2})$ となることが知られている [7]．なお，この解析においては，(勾配やヘッセ行列ではなく) 目的関数のリプシッツ連続性が用いられている．

最近，目的関数 f が微分可能な凸関数 h と簡単な構造を持つ凸関数 P の和で表された問題とその解法が注目を浴びている [1, 11]．そのような問題として， P が L_1 正則化項 $\sum_{i=1}^n |x_i|$ となる問題や，表示関数

$$P(x) = \begin{cases} 0 & \text{if } x \in X \\ \infty & \text{if } x \notin X \end{cases}$$

で与えられる問題が挙げられる． P が表示関数で与えられたときは，制約付きの最適化問題

$$\begin{aligned} & \text{minimize} && h(x) \\ & \text{subject to} && x \in X \end{aligned}$$

と等価である．

P が微分不可能であるときは，最急降下法は適用できない．一方，劣勾配法では速い収束が望めない．そこで， h に対応する部分には最急降下法を， P に対応する部分には近接点法を使う次の近接勾配法が提案されている [11]．

$$x^{k+1} = \operatorname{argmin}_{x \in \mathbb{R}^n} \left\{ t_k h(x^k) + t_k \nabla h(x^k)^T (x - x^k) + t_k P(x) + \frac{1}{2} \|x - x^k\|^2 \right\}$$

なお， $P(x) \equiv 0$ のときは，最急降下法となる． $P(x) \not\equiv 0$ の場合でも，最急降下法と同様の解析をすることによって， ∇h がリプシッツ連続であるとき大域的反復回数が $O(\varepsilon_f^{-1})$ となることを示せる [11]．

3.2. 近接勾配法の加速手法

近接勾配法は，現在の点 x^k における勾配の情報しか用いていない．これまでに得られた情報を用いて高速化する工夫が提案されている [1, 7]．そのうちのひとつに，点列を

$$\begin{aligned} y^k &= x^k + \alpha_k (x^k - x^{k-1}) \\ x^{k+1} &= \operatorname{argmin}_{x \in \mathbb{R}^n} \left\{ \langle t_k \nabla h(y^k), x \rangle + t_k P(x) + \frac{1}{2} \|x - y^k\|^2 \right\} \end{aligned}$$

と生成する手法がある．ここで， $\alpha_k = 0$ とすれば，通常の近接勾配法である．

もちろん， α_k は何でもよいわけではない． α_k として次のものを用いるものが FISTA とよばれる高速化法である [1]．

$$\begin{aligned} \alpha_k &= \frac{\theta_k}{\theta_{k-1}} - \theta_k \\ \theta_{k+1} &= \frac{\sqrt{\theta_k^4 + 4\theta_k^2} - \theta_k}{2} \end{aligned}$$

この手法の大域的反復回数は， ∇h がリプシッツ連続のとき， $O(\varepsilon_f^{-\frac{1}{2}})$ となる [1]．

他にもさまざまな高速化法が提案されているが，それらの大域的反復回数は高々 $O(\varepsilon_f^{-\frac{1}{2}})$ である．Nesterov [7] は，勾配のみに基づく手法は，大域的反復回数が $O(\varepsilon_f^{-\frac{1}{2}})$ より良くならないことを特殊例を用いて示している¹²．そのため，[7] で提案した $O(\varepsilon_f^{-\frac{1}{2}})$ の手法を彼は

¹²反復回数が $n/2$ 回よりも少ない場合のみである．

“optimal method” と名づけている．

4. 非凸な問題に対する解法の大域的反復回数

この節では，“非凸”な制約なし最小化問題に対する勾配法やニュートン型手法の大域的反復回数について紹介する．以下では， $\lambda_{\min}(\nabla^2 f(x^k))$ は $\nabla^2 f(x^k)$ の最小固有値を表すものとする．

4.1. 最急降下法

3.1 節において， f が凸であるとき，最急降下法の大域的反復回数が $O(\varepsilon_f^{-1})$ となることを示した．ここでは， f が凸でないときでも，(1.3) を満たす大域的反復回数が $O(\varepsilon_g^{-2})$ となることを示す．

以下では，3.1 節と同様に， $t_k = 1/L_g$ の場合を考える．第 0 反復から第 k 反復までの不等式 (3.1) を辺々足し合わせると，

$$f(x^0) - f(x^{k+1}) \geq \sum_{i=0}^k \frac{1}{2L_g} \|\nabla f(x^i)\|^2 \geq \frac{k+1}{2L_g} \min_{0 \leq i \leq k} \|\nabla f(x^i)\|^2$$

を得る． $f(x^k) \geq f_{\min}$ であるから，

$$\min_{0 \leq i \leq k} \|\nabla f(x^i)\| \leq \sqrt{\frac{2L_g(f(x^0) - f_{\min})}{k+1}}$$

となる．この不等式の右辺を ε_g 以下に抑えるために必要な k を見積もると，大域的反復回数 $O(\varepsilon_g^{-2})$ が示せる．

ところで，この見積もり $O(\varepsilon_g^{-2})$ は，反復回数の上界値であるため，実際にはもっと良い見積もりが解析できるのではないかと期待されるかもしれない．残念ながら，Cartis ら [2] は，エルミート補間を利用した特殊な目的関数を構成することで，最急降下法の反復回数が $O(\varepsilon_g^{-2+\tau})$ (τ は任意の正の定数) となる例を与えている．したがって，勾配のリプシッツ連続性に加えて，さらなる仮定をおかない限り， $O(\varepsilon_g^{-2})$ よりも良い見積もりを得ることはできない．

4.2. ニュートン法

ニュートン法は，目的関数の 2 次近似式

$$m_3(x^k, d) = f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T \nabla^2 f(x^k) d$$

をモデル関数とし， $m_3(x^k, \cdot)$ の最小解 d^k を探索方向とし， $x^{k+1} = x^k + d^k$ として点列を生成する反復法である．ただし， m_3 は d に関して凸関数になるとは限らない．そこで，通常は，最小解 d^k が存在し，なおかつ大域的収束性を保証するために信頼領域法 [4] と組み合わせて用いられる．信頼領域法では，探索方向 d^k は以下の制約付きの部分問題の解として与えられる．

$$\begin{aligned} & \text{minimize} && m_3(x^k, d) \\ & \text{subject to} && \|d\| \leq \Delta_k. \end{aligned} \tag{4.1}$$

ただし， Δ_k は各反復で以下のように調整される信頼半径とよばれるパラメータである．モデル関数の減少量に対する目的関数の減少量の比を

$$\rho(x^k, d) = \frac{f(x^k) - f(x^k + d)}{f(x^k) - m_3(x^k, d)}$$

で表す． $\rho(x^k, d^k)$ が大きいとき，モデル関数が目的関数をうまく近似できているため信頼半径を大きくし， $x^{k+1} = x^k + d^k$ として反復点を更新する．一方， $\rho(x^k, d^k)$ が小さいとき，モデル関数が目的関数をうまく近似できていないため，反復点は更新せず $x^{k+1} = x^k$ とし，信頼半径を小さくして探索方向 d^{k+1} を計算する．このように反復点を更新することで，目的関数値が減少する点列を生成でき，適当な仮定のもとで大域的収束性を保証できる．

部分問題 (4.1) は凸計画問題ではないが， d^k が大域的最適解となるための必要十分条件として次のものが知られている [4]．ただし $\lambda \in \mathbb{R}$ はラグランジュ乗数である．

- (i) $\nabla f(x^k) + \nabla^2 f(x^k)d^k + \lambda d^k = 0$
- (ii) $\|d^k\| \leq \Delta_k$, $\lambda \geq 0$, $(\|d^k\| - \Delta_k)\lambda = 0$
- (iii) $\nabla^2 f(x^k) + \lambda I$ は半正定値行列

はじめの 2 項目は，部分問題 (4.1) の Karush-Kuhn-Tucker 条件そのものである．最後の項目は最適性の 2 次の必要条件をやや厳しくした条件である．部分問題を正確に解く際には，条件 (ii) より $\|d^k\| = \Delta_k$ の場合と $\|d^k\| < \Delta_k$ の場合にわけて考える [4]¹³．前者のときは，条件 (i) と $\|d^k\| = \Delta_k$ を同時に満たす λ の 1 次元方程式

$$\left\| (\nabla^2 f(x^k) + \lambda I)^{-1} \nabla f(x^k) \right\| = \Delta_k \quad (4.2)$$

を (iii) が満たされるように

$$\lambda \geq \max \{0, -\lambda_{\min}(\nabla^2 f(x^k))\}$$

の範囲¹⁴で解き，条件 (i) を用いて $d^k = -(\nabla^2 f(x^k) + \lambda I)^{-1} \nabla f(x^k)$ とする．後者の場合は，条件 (ii) より $\lambda = 0$ となるので，線形方程式 $\nabla^2 f(x^k)d = -\nabla f(x^k)$ の解が d^k となる．

ヘッセ行列のリブシッツ連続性を用いることで，このニュートン法は最悪の場合でも最急降下法と同じ振る舞いをすることが保証できる．そのため，最急降下法と同等の大域的反復回数 $O(\varepsilon_g^{-2})$ を示すことができる [5]．ただし，この反復回数には部分問題 (4.1) (あるいは方程式 (4.2)) を解く計算量は考慮されていない．

最急降下法の場合と同様に，ニュートン法の大域的反復回数が $O(\varepsilon_g^{-2+\tau})$ (τ は任意の正の定数) となる例が存在する [2]．したがって，ニュートン法でも，さらなる仮定をおかない限り， $O(\varepsilon_g^{-2})$ よりも良い結果を得ることができない．

4.3. 3 次正則化ニュートン法

3 次正則化ニュートン法は (2.5) で定義された関数 $m_2(x^k, d)$ の制約なし部分問題

$$\begin{aligned} & \text{minimize} && m_2(x^k, d) \\ & \text{subject to} && d \in \mathbb{R}^n \end{aligned} \quad (4.3)$$

の最小解 d^k を用いて， $x^{k+1} = x^k + d^k$ と更新する反復法である [9]¹⁵．なお， m_2 は目的関数の 2 次近似関数 m_3 に 3 次の正則化項を加えた関数である．

¹³ 実用的には部分問題を近似的に解けば十分である．

¹⁴ 部分問題 (4.1) は必ず解を持つため，(i)-(iii) を満たす λ は必ず存在する．(iii) を満たす必要条件がこの不等式であるから， λ をこの範囲に限定しても問題ない．

¹⁵ ヘッセ行列のリブシッツ定数 L_H が既知でない場合に対応した実用的な 3 次正則化ニュートン法も提案されている [3]．

部分問題 (4.3) は，制約なしの問題ではあるが，目的関数が非凸な 3 次関数であるため，信頼領域法の部分問題 (4.1) と同様，簡単に解くことはできない．部分問題 (4.3) の最適性の 1 次の必要条件は，

$$\nabla f(x^k) + \left(\nabla^2 f(x^k) + \frac{1}{2} L_H \|d^k\| I \right) d^k = 0 \quad (4.4)$$

となる．Nesterov と Polyak [9] は，ある種の双対性を考慮することによって，(4.4) を加味した次の条件 (a) と (b) が最適性の必要十分条件となることを示した．

(a) d^k は最適性の 1 次の必要条件 (4.4) を満たす．

(b) $\nabla^2 f(x^k) + \frac{1}{2} L_H \|d^k\| I$ は半正定値．

この条件 (a)，(b) は信頼領域法の部分問題の必要十分条件 (i)-(iii) に類似している．

条件 (a)，(b) は， $r = \|d^k\|$ とすると，

$$\left\| \left(\nabla^2 f(x^k) + \frac{L_H}{2} r \right)^{-1} \nabla f(x^k) \right\| = r$$

$$\frac{1}{2} L_H r \geq \max \{0, -\lambda_{\min}(\nabla^2 f(x^k))\}$$

と表せる．つまり，条件 (a)，(b) は r に関する 1 次元の方程式となる．この 1 次元の方程式の解を r^k とすると，部分問題 (4.3) の解は $d^k = -(\nabla^2 f(x^k) + \frac{L_H}{2} r^k)^{-1} \nabla f(x^k)$ と与えられる．この 1 次元の方程式は，信頼領域法の部分問題と等価な方程式 (4.2) と同じ形をしているため，方程式 (4.2) を解くために開発された手法を用いて解くことができる．

次に，この最適解の必要十分条件 (a)，(b) を用いて大域的反復回数を解析してみよう．最適性の 1 次の必要条件 (4.4) に d^k を掛けると，

$$-\nabla f(x^k)^T d^k = (d^k)^T \left(\nabla^2 f(x^k) + \frac{1}{2} L_H \|d^k\| I \right) d^k \geq 0 \quad (4.5)$$

を得る．最後の不等式は条件 (b) による．さらに，(4.5) の等式を式変形すると，

$$-\frac{1}{2} (d^k)^T \nabla^2 f(x^k) d^k = \frac{1}{2} \nabla f(x^k)^T d^k + \frac{1}{4} L_H \|d^k\|^3$$

を得る．この式と不等式 (4.5) を用いると，

$$\begin{aligned} f(x^k) - m_2(x^k, d^k) &= -\nabla f(x^k)^T d^k - \frac{1}{2} (d^k)^T \nabla^2 f(x^k) d^k - \frac{L_H}{6} \|d^k\|^3 \\ &= -\frac{1}{2} \nabla f(x^k)^T d^k + \frac{L_H}{12} \|d^k\|^3 \\ &\geq \frac{L_H}{12} \|d^k\|^3 \end{aligned}$$

が導ける．よって，(2.5) より，

$$f(x^k) - f(x^{k+1}) = f(x^k) - f(x^k + d^k) \geq f(x^k) - m_2(x^k, d^k) \geq \frac{L_H}{12} \|d^k\|^3$$

が成り立つから，この不等式を足し合わせることで，

$$f(x^0) - f_{\min} \geq \sum_{i=0}^k \{f(x^i) - f(x^{i+1})\} \geq \frac{L_H}{12} \sum_{i=0}^k \|d^i\|^3 \quad (4.6)$$

を得る．

一方， $\nabla^2 f$ のリプシッツ連続性から導かれる性質 (2.4) と (4.4) とより，

$$\|\nabla f(x^k) + \nabla^2 f(x^k)d^k - \nabla f(x^k + d^k)\| \leq \frac{L_H}{2}\|d^k\|^2$$

$$\|\nabla f(x^k) + \nabla^2 f(x^k)d^k\| = \frac{L_H}{2}\|d^k\|^2$$

が成り立つ．これらの不等式より，

$$\|\nabla f(x^{k+1})\| = \|\nabla f(x^k + d^k)\| \leq L_H\|d^k\|^2$$

を得る．この不等式と (4.6) を合わせると，

$$f(x^0) - f_{\min} \geq \frac{k+1}{12\sqrt{L_H}} \min_{0 \leq i \leq k} \|\nabla f(x^{i+1})\|^{\frac{3}{2}}$$

となることから，

$$\left(\frac{12\sqrt{L_H}(f(x^0) - f_{\min})}{k+1} \right)^{\frac{2}{3}} \geq \min_{0 \leq i \leq k} \|\nabla f(x^{i+1})\|$$

を得る．この不等式の左辺を ε_g 以下に抑えるために必要な k を見積もると，大域的反復回数は $O(\varepsilon_g^{-\frac{3}{2}})$ となることがわかる．論文 [9] ではさらに最適性の 2 次の必要条件も含めた条件：

$$\min_{0 \leq i \leq k} \max \left\{ \|\nabla f(x^i)\|, \max \{0, -\lambda_{\min}(\nabla^2 f(x^i))\}^2 \right\} \leq \varepsilon_g$$

を満たす反復回数が $O(\varepsilon_g^{-\frac{3}{2}})$ となることも示している．

なお，(信頼領域法を用いた) ニュートン法の場合と同様に，この見積もり $O(\varepsilon_g^{-\frac{3}{2}})$ には部分問題 (4.3) を解くための計算量は考慮されていない．また，3 次正則化ニュートン法に対しては，大域的反復回数が $O(\varepsilon_g^{-\frac{3}{2}+\tau})$ (τ は任意の正の定数) となる問題例が存在する [2]．そのため， $O(\varepsilon_g^{-\frac{3}{2}})$ はタイトな上限である．

3 次正則化ニュートン法は，探索方向の計算に時間がかかるため，凸計画問題のような性質のよい問題に対しては，特に効率のよい解法というわけではない．そこで，凸計画問題に対しては，結果だけを簡単に紹介しよう．まず， $\nabla^2 f$ がリプシッツ連続のとき，通常の 3 次正則化ニュートン法の大域的反復回数は $O(\varepsilon_f^{-\frac{1}{2}})$ となる [9]．さらに，3.2 節で紹介したような加速手法を組み合わせることによって， $O(\varepsilon_f^{-\frac{1}{3}})$ とすることができる [8]．

4.4. 正則化ニュートン法

3 次正則化ニュートン法では，各反復で，計算量が未知な方程式：

$$\nabla f(x^k) + \left(\nabla^2 f(x^k) + \frac{1}{2}L_H\|d\|I \right) d = 0$$

を解かなければならない．この方程式が難しいのは，行列 $\frac{1}{2}L_H\|d\|I$ が含まれているからである．一方，解のそばでは，3 次正則化ニュートン法はニュートン法と同様の振る舞いをする．さらに，ニュートン法の探索方向 d^k は 0 に収束し，なおかつ，強い正則性¹⁶とリプシッ

¹⁶ 解 x^* における $\nabla^2 f(x^*)$ が正定値行列になること．なお，本稿では強い正則性を仮定しない場合の大域的反復回数を紹介しているが，ここでは正則化ニュートン法のアイデアを説明するためにあえて仮定している．

ツ連続性のもとでは $\|d^k\| = O(\|\nabla f(x^k)\|)$ となることが知られている．そこで，上田と山下 [12] は，行列 $\frac{1}{2}L_H\|d\|I$ のかわりに， $c\|\nabla f(x^k)\|^\delta I$ で置き換えた正則化ニュートン法を提案した¹⁷．その正則化ニュートン法では，制約なしの部分問題

$$\begin{aligned} & \text{minimize} && m_4(x^k, d) \\ & \text{subject to} && d \in \mathbb{R}^n \end{aligned} \quad (4.7)$$

の最小解 d^k を用いて， $x^{k+1} = x^k + d^k$ と反復点を更新する．ただし， $m_4(x^k, d)$ は次式で定義される関数である．

$$m_4(x^k, d) = f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T (\nabla^2 f(x^k) + \mu_k I) d$$

ここで， μ_k は，

$$\mu_k = c \max \{0, -\lambda_{\min}(\nabla^2 f(x^k))\} + \nu_k \|\nabla f(x^k)\|^\delta \quad (4.8)$$

であり， c, δ, ν_k は $c > 1, \delta > 0, \nu_k > 0$ を満たすパラメータである．

正則化ニュートン法では，3次正則化ニュートン法における行列 $\frac{1}{2}L_H\|d\|I$ を $\mu_k I$ で置き換えたため，たとえヘッセ行列がリプシッツ連続であっても， μ_k を調整しない限り，大域的収束性を保証できない．そこで， μ_k に含まれるパラメータ ν_k を，信頼領域法における信頼半径 Δ_k のように各反復で調整することにより，大域的収束性を持たせる方法が考案されている [12]．

正則化ニュートン法の部分問題 (4.7) は強凸2次関数の最小化問題となるため，最適性の1次の必要条件を表した線形方程式

$$(\nabla^2 f(x^k) + \mu_k I) d = -\nabla f(x^k)$$

を解くだけで探索方向が得られる．したがって，3次正則化ニュートン法の部分問題 (4.3) よりも，はるかに容易に解ける．

ヘッセ行列のリプシッツ連続性を用いることで，正則化ニュートン法の大域的反復回数は $O(\varepsilon_g^{-2})$ となることが示されている [12]．これは最急降下法やニュートン法と同じオーダーとなる．ただし，この反復回数には $\nabla^2 f(x^k)$ の最小固有値の計算時間 (反復回数) は考慮されていない．

ヘッセ行列のリプシッツ連続性に加えて，次の条件のもとでは，正則化ニュートン法の大域的反復回数はさらに良くなる [12]．

仮定 4.1. $\delta \leq \min\{\frac{1}{2}, \bar{\delta}\}$ かつ $\max\{0, -\lambda_{\min}(\nabla^2 f(x^k))\} \leq \kappa \|\nabla f(x^k)\|^\delta, \forall k \geq 0$ を満たす正の定数 $\delta, \bar{\delta}, \kappa$ が存在する．

目的関数 f が凸関数であれば，仮定 4.1 は満たされる．以下の例が示すように，非凸な場合であっても，仮定 4.1 が満たされることがある．

例 4.1. 目的関数 f を以下で定義される非凸な1変数関数とする．

$$f(x) = |x|^3 - 10x^2 + 50x$$

¹⁷この正則化ニュートン法は正則化パラメータ μ_k を (4.8) の形に限定したものである．一般の (2次) 正則化ニュートン法は3次正則化ニュートン法よりも古い歴史をもつ．

このとき，ヘッセ行列はリプシッツ連続となり，

$$\begin{aligned}\nabla f(x) &= 3x|x| - 20x + 50, \\ \nabla^2 f(x) &= 6|x| - 20, \\ \max\{0, -\lambda_{\min}(\nabla^2 f(x))\} &\leq 3\|\nabla f(x)\|^{\frac{1}{2}}, \quad \forall x \in \mathbb{R}\end{aligned}$$

が成り立つ．そのため， $\bar{\delta} = \frac{1}{2}$ ， $\kappa = 3$ とした場合の仮定 4.1 を満たす．

仮定 4.1 のもとでは，正則化ニュートン法の大域的反復回数は $O(\varepsilon_g^{-\frac{2+\delta}{1+\delta}})$ となることが示されており，特に， $\delta = \frac{1}{2}$ と選ぶことができるとき， $O(\varepsilon_g^{-\frac{5}{3}})$ となる [12]．また， f が凸関数であるときは， $O(\varepsilon_f^{-\frac{2}{3}})$ となる [12]．

正則化ニュートン法については，上記の大域的反復回数 $O(\varepsilon_g^{-2})$ や $O(\varepsilon_g^{-\frac{2+\delta}{1+\delta}})$ がこれ以上良くなれないかどうかについては，まだわかっていない．

4.5. Levenberg-Marquardt 法

この節では，問題 (1.1) の特別な場合である非線形最小 2 乗問題を考える．つまり，目的関数 f は

$$f(x) = \frac{1}{2}\|F(x)\|^2$$

与えられているとする．ここで， F は \mathbb{R}^n から \mathbb{R}^m への連続的微分可能な写像とする．

非線形最小 2 乗問題に対する効率的な解法の一つとして，Levenberg-Marquardt 法 (以下，LM 法) がある．LM 法では，次のモデル関数

$$m_5(x^k, d) = \frac{1}{2}\|F(x^k) + \nabla F(x^k)d\|^2 + c\nu_k\|\nabla f(x^k)\|^\delta\|d\|^2$$

の制約なしの最小解 d^k を用いて， $x^{k+1} = x^k + d^k$ と反復点を更新する．ここで， c, δ, ν_k は正のパラメータである．[14] で提案されている LM 法は，前副節の正則化ニュートン法 [12] と同様に，各反復でパラメータ ν_k をうまく調整することにより，大域的収束性を保証している．

探索方向 d^k を求めるための部分問題は強凸 2 次関数の最小化問題となるため，正則化ニュートン法の場合と同様に，線形方程式

$$(\nabla F(x^k)^T \nabla F(x^k) + c\nu_k\|\nabla f(x^k)\|^\delta I) d = -\nabla F(x^k)^T F(x^k)$$

を解くだけで探索方向が得られる．また，正則化ニュートン法と違い，ヘッセ行列などの固有値を計算する必要がない．

F のヤコビ行列 $\nabla F(x)^T$ のリプシッツ連続性を用いることで，LM 法の大域的反復回数が $O(\varepsilon_g^{-2})$ となることが示されている [14]．さらに，上田と山下 [15] は， F が微分可能でないときも， ∇f がリプシッツ連続であれば， $O(\varepsilon_g^{-2})$ となることを示している．なお，これらの大域的反復回数は，4.1 節や 4.2 節で紹介した最急降下法やニュートン法の大域的反復回数と同じである．ただし，強い正則性を仮定すれば，LM 法によって生成された点列は超 1 次収束するため，最急降下法よりも局所的な振る舞いは良い．また，ニュートン法などの f のヘッセ行列を用いる手法では， F の 2 階微分が必要であるのに対して，LM 法は F の 1 階微分しか必要としない．

正則化ニュートン法の場合と同様に，LM 法の大域的反復回数が $O(\varepsilon_g^{-2})$ より良くならない例があるどうかは，まだわかっていない．

5. おわりに

本稿では、非線形計画問題に対するいくつかの解法に対して、初期点から近似解を求めるまでに必要とする大域的反復回数の結果を紹介した。凸な問題の解法に対して、 $f(x^k) - f(x^*) \leq \varepsilon_f$ を満たす近似解を求めるために必要な大域的反復回数の結果を表1にまとめる。なお、表中の“リプシッツ連続”とは解析においてリプシッツ連続性を仮定した関数である。

表 1: 凸な問題の解法に対する大域的反復回数

解法	大域的反復回数	リプシッツ連続
劣勾配法	$O(\varepsilon_f^{-2})$	f
近接勾配法 (3.1 節)	$O(\varepsilon_f^{-1})$	∇h
加速-近接勾配法 (3.2 節)	$O(\varepsilon_f^{-\frac{1}{2}})$	∇h
正則化ニュートン法 (4.4 節)	$O(\varepsilon_f^{-\frac{2}{3}})$	$\nabla^2 f$
3次正則化ニュートン法 (4.3 節)	$O(\varepsilon_f^{-\frac{1}{2}})$	$\nabla^2 f$
加速-3次正則化ニュートン法 (4.3 節)	$O(\varepsilon_f^{-\frac{1}{3}})$	$\nabla^2 f$

また、非凸な問題の解法に対しては、 $\|\nabla f(x^k)\| \leq \varepsilon_g$ を満たす近似解を求めるために必要な大域的反復回数の結果を表2にまとめる。

表 2: 非凸な問題の解法に対する大域的反復回数

解法	大域的反復回数	リプシッツ連続
最急降下法 (4.1 節)	$O(\varepsilon_g^{-2})$	∇f
ニュートン法 (4.2 節)	$O(\varepsilon_g^{-2})$	$\nabla^2 f$
正則化ニュートン法 (4.4 節)	$O(\varepsilon_g^{-2})$	$\nabla^2 f$
3次正則化ニュートン法 (4.3 節)	$O(\varepsilon_g^{-\frac{3}{2}})$	$\nabla^2 f$
Levenberg-Marquardt 法 (4.5 節)	$O(\varepsilon_g^{-2})$	∇F

本稿で紹介した大域的反復回数の見積もりの結果は、(正則性などを仮定しない)非常に広い関数のクラスに対して得られた、ある意味大雑把な結果である。実用面を考えれば、実際に取り扱う問題のクラスに限定したより厳密な大域的反復回数を知る必要がある。そのため、今後は、実問題とその解法に応じて、リプシッツ定数も含めた、厳密かつ具体的な大域的反復回数を見積もる研究が重要となるであろう。

参考文献

- [1] A. Beck and M. Teboulle: A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM Journal on Imaging Sciences*, **2** (2009), 183–202.
- [2] C. Cartis, N.I.M. Gould, and P.L. Toint: On the complexity of steepest descent, Newton’s and regularized Newton’s methods for nonconvex unconstrained optimization problems, *SIAM Journal on Optimization*, **20** (2010), 2833–2852.
- [3] C. Cartis, N.I.M. Gould, and P.L. Toint: Adaptive cubic regularization methods for unconstrained optimization. PartII: worst-case function- and derivative-evaluation complexity, *Mathematical Programming*, **130** (2011), 295–319.

- [4] A.R. Conn, N.I.M. Gould, and P.L. Toint: *Trust-Region Methods* (SIAM, Philadelphia, 2000).
- [5] S. Gratton, A. Sartenaer, and P.L. Toint: Recursive trust-region methods for multiscale nonlinear optimization, *SIAM Journal on Optimization*, **19** (2008), 414–444.
- [6] F.J. Luque: Asymptotic convergence analysis of the proximal point algorithm, *SIAM Journal on Control and Optimization*, **22** (1984), 277–293.
- [7] Yu. Nesterov: *Introductory Lectures on Convex Optimization* (Kluwer Academic, Dordrecht, 2004).
- [8] Yu. Nesterov: Accelerating the cubic regularization of Newton’s method on convex problems, *Mathematical Programming*, **112** (2008), 159–181.
- [9] Yu. Nesterov and B.T. Polyak: Cubic regularization of Newton method and its global performance, *Mathematical Programming*, **108** (2006), 177–205.
- [10] R.T. Rockafellar: Monotone operators and the proximal point algorithm, *SIAM Journal on Control and Optimization*, **14** (1976), 877–898.
- [11] P. Tseng: Approximation accuracy, gradient methods, and error bound for structured convex optimization, *Mathematical Programming*, **125** (2010), 263–295.
- [12] K. Ueda and N. Yamashita: A regularized Newton method without line search for unconstrained optimization, Technical Report 2009-007, Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University (2009).
- [13] K. Ueda and N. Yamashita: Convergence properties of the regularized Newton method for the unconstrained nonconvex optimization, *Applied Mathematics and Optimization*, **62** (2010), 27–46.
- [14] K. Ueda and N. Yamashita: On a global complexity bound of the Levenberg-Marquardt Method, *Journal of Optimization Theory and Applications*, **147** (2010), 443–453.
- [15] K. Ueda and N. Yamashita: Global complexity bound analysis of the Levenberg-Marquardt method for nonsmooth equations and its application to the nonlinear complementarity problem, *Journal of Optimization Theory and Applications*, **152** (2012), 450–467.
- [16] N. Yamashita and M. Fukushima: The proximal point algorithm with genuine superlinear convergence for the monotone complementarity problem, *SIAM Journal on Optimization*, **11** (2001), 364–379.

山下信雄

京都大学大学院情報学研究科

〒 606-8501 京都市左京区吉田本町

E-mail: nobuo@i.kyoto-u.ac.jp

ABSTRACT

THE GLOBAL COMPLEXITY BOUNDS OF
THE GRADIENT METHODS AND THE NEWTON-TYPE METHODS FOR
THE UNCONSTRAINED MINIMIZATION PROBLEM

Nobuo Yamashita
Kyoto University

Kenji Ueda
Mitsubishi Electric Corporation

This is a survey of the recent researches on the global complexity bounds of the gradient methods and the Newton-type methods for the nonlinear programming problem. The classical benchmarks of these iterative methods, such as global convergence and rates of convergence, do not represent a total computational time. The global complexity bound is the worst number of iterations to find an appropriate solution, and hence it is proportional to the worst computational time. In recent years the global complexity bounds of the gradient methods and the Newton-type methods have been vigorously investigated. In the survey we report some of these results.