修士論文 要旨

非凸近接Newton法の大域計算量および局所収束性の解析

数理情報学専攻 48226208　逸見榛一
指導教員　武田 朗子 教授

1　はじめに

目的関数が，微分可能な関数と凸関数の和で表される非凸非平滑最適化問題には様々な工学的応用が存在する．損失関数と非平滑な正則化項の和の最適化問題という数理モデル化は，least absolute shrinkage and selection operator (LASSO) をはじめとして，機械学習，統計，画像・信号処理などの分野にでも広く用いられている．

このような構造を持つ問題に対する最適化手法として，近接勾配法の枠組で最適な大域計算量を達成する手法や，近接 Newton 法の枠組で超 1 次以上の局所収束性を持つ手法などが提案されてきた．しかしながら，一般的な仮定のもとで，大域的な計算量保証と超 1 次以上の局所収束性の両方を満たすアルゴリズムは知られていなかった．

本論文では，正則化付きの近接 Newton 法を提案し，主に平滑項の勾配の Lipschitz 連続性と局所的な凸性，そして目的関数の metrical subregularity という比較的弱い仮定のもとで，$\tilde{O}\big(\varepsilon^{-2(1+\rho)}\big)$ の反復計算量と局所的な超 1 次収束性を達成できることを示した．ただし，$\rho \in [0,1]$ は大域と局所の収束性のトレードオフを示すアルゴリズムのパラメータである．

また，本手法では，古典的な勾配・Hesse 行列オラクルの評価回数だけでなく，子問題の解法まで含めた反復計算量の保証も与えている．提案手法で用いる子問題の解法では，Lipschitz 定数の自動推定と Hesse 行列・ベクトル積の利用による Hesse 評価の代替を行うため，提案手法は事前情報の少ない問題や，大規模な問題に対して適用できる実用的な手法である．

2　構造を持つ非凸非平滑最適化問題

関数 $f\colon \mathbb{R}^n \to (-\infty,\infty]$ は 2 階連続微分可能，関数 $g\colon \mathbb{R}^n \to (-\infty,\infty]$ は閉真凸関数として，最適化問題

$$\min_{x\in\mathbb{R}^n} F(x) \coloneqq f(x) + g(x) \tag{1}$$

を考える．このような構造を持つ最適化問題は，機械学習・統計のスパース推定や，画像・信号処理におけるノイズの除去等の応用が知られている．また，g を標示関数にとることで，凸最適化問題にも帰着する．

表 1: 提案手法と既存の近接 Newton 法の比較

	計算量保証		オラクル	
近接 Newton 法	大域	局所	$\nabla^2 f\cdot v$	$\nabla^2 f$
[3]	$O(\varepsilon^{-2})$			✓
[2]		>1		✓
[4]		>1		
提案手法	$O(\varepsilon^{-2(1+\rho)})$	$1+\rho$	$\tilde{O}(\varepsilon^{-\frac{\rho}{2}})$	

3　提案手法

前章の背景を踏まえ，本研究では非凸な f に対して，大域計算量及び局所収束性の保証を同時に持つ，新たな近接 Newton 法を提案した．その概略を Algorithm 1に示す．アルゴリズム中の $\mathcal{G}$ は gradient mapping を表し，非平滑な関数の勾配に対応する概念である．

既存の手法に対する大きな改善点として， [5] の近接 Newton 法で用いられていた正則化項を非凸の場合への拡張，大域計算量と局所収束性の両方を同時に保証するための直線探索の条件の設定，そして，子問題の近似の十分条件の設定と，その反復計算量の保証を行った．

なお，最小固有値計算と理論保証で用いる子問題の解法 [1] については，Hesse 行列を直接評価することなく，Hesse 行列・ベクトル積を利用するのみで実装可能であり，大規模な問題に対して適用可能である．既存の近接 Newton 法との比較を表 1に示した．

4　大域計算量

大域計算量の保証のため，以降では ∇f は Lipschitz 連続であることを仮定する．また，

$$\omega(x) \coloneqq \min_{p\in\partial g} \|\nabla f(x) + p\| \le \varepsilon$$

を満たす点を ε-停留点 x と呼び，この x 点を得るために必要な f の関数値，勾配，そして Hesse 行列または Hesse 行列・ベクトル積の評価回数に対する理論保証を行った．まず，提案手法の反復計算量保証を与える．

定理 1. 提案手法はそれぞれ $O\big(\varepsilon^{-2(1+\rho)}\big)$ 回の f の勾配，Hesse 行列評価と，$O\big(\varepsilon^{-2(1+\rho)}\log\varepsilon^{-1}\big)$ 回の関数値評価で ε-停留点を出力する．

Algorithm 1 提案手法（概略）

Input : $x_0 \in \mathbb{R}^n, \overline{\mu}, c > 0, \rho \in [0,1], 0 < \theta, \sigma, \gamma < 1$

$l \leftarrow 0$
for $k = 0, 1, \ldots$ **do**
  $\lambda_k \leftarrow |\min\{\lambda_{\min}(\nabla^2 f(x^k)), 0\}|$
  $\mu_k \leftarrow \min\left\{\overline{\mu}, c\left\|\mathcal{G}_L^f(x^k)\right\|^\rho\right\}$
  $\tilde{q}^k(d) := \nabla f(x^k)^\top d + \frac{1}{2}\|d\|^2_{\nabla^2 f(x^k) + (\lambda_k + \mu_k)I}$
  $\min_d\left\{\tilde{q}^k(d) + g(x^k + d)\right\}$ の近似解 d^k を計算
  if $F(x^k + d^k), \|\mathcal{G}_L(x^k + d^k)\|$ が十分減少 **then**
    $t_k \leftarrow 1$, $l_{k+1} \leftarrow k + 1$
  else
    直線探索により t_k を決定
  $x^{k+1} \leftarrow (1 - t_k)x^k + t_k\hat{x}^k$

この計算量保証は，2 次法の反復計算量保証では一般的な方法であるが，子問題の解法の計算量を含まないことに留意する必要がある．本提案手法では，子問題の解法に対しても計算量解析を行なっており，次の全体反復計算量の保証を示した．

定理 2. 提案手法はそれぞれ $O\left(\varepsilon^{-2(1+\rho)-\frac{\rho}{2}} \log \varepsilon^{-1}\right)$ 回の Hesse 行列・ベクトル積の評価，$O\left(\varepsilon^{-2(1+\rho)}\right)$ 回の勾配評価，$O\left(\varepsilon^{-2(1+\rho)} \log \varepsilon^{-1}\right)$ 回の関数値評価で ε-停留点を出力する．

ここで，提案手法中の子問題の解法を用いることで，f の Hesse 行列を直接評価する代わりに Hesse 行列・ベクトル積の評価回数に対して計算量が保証される．また，Algorithm 1 中では省略したが， 同時に Lipschitz 定数が自動推定される．

5 局所収束性

局所収束性の保証のための仮定として，以降では停留点の近傍で，f は凸かつ $\nabla^2 f(x)$ は Lipschitz 連続であるとする．また，さらに $\nabla f + \partial g$ が metrically subregular であるという仮定を置く．この仮定は，非平滑関数に対して強凸性を弱めた概念であり，$g \equiv 0$ の場合は error bound 条件に帰着される．上記の仮定のもとで，提案手法は次の局所収束性を満たす．

定理 3. 提案手法の生成する点列 $\{x^k\}$ に対して，ある $K \in \mathbb{N}, C \in \mathbb{R}^n$ が存在して，任意の $K \geq K$ に対して

$$\operatorname{dist}(x^{k+1}, \mathcal{X}^*) < C \operatorname{dist}(x^k, \mathcal{X}^*)^{1+\rho}$$

が成り立つ．

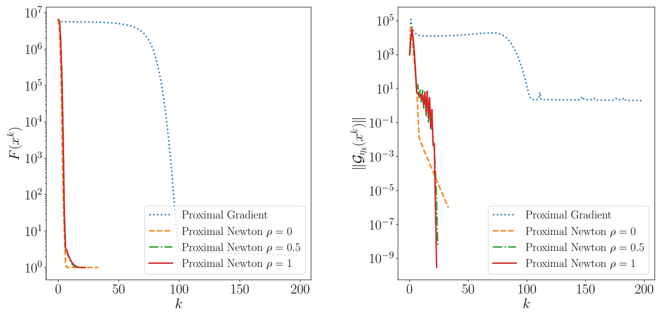

図 1: 横軸に提案手法の反復，縦軸に関数値 $F(x^k)$（左）と gradient mapping のノルム $\|\mathcal{G}_L(x^k)\|$（右）をプロットしたもの．

6 数値実験

平滑項として Rosenbrock 関数：

$$f(x) := \frac{1}{2}\sum_{i=1}^{n-1}\left(100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2\right)$$

非平滑項として ℓ_1 ノルム $g := \lambda\|x\|_1$ を用いて数値実験を行い，提案手法と近接勾配法の比較を行った結果を図 1に示す．なお，$n = 10^6, \lambda = 10^{-6}$ とし，$\rho = 0, 0.5, 1$ の各設定で提案手法を実行した．

提案手法（橙，緑，赤．それぞれ $\rho = 0, 0.5, 1$ に対応）は，近接勾配法（青）より少ない反復計算量でより精度の高い ε-停留点を出力することがわかる．また，ρ に対する挙動の変化として，$\rho = 0$ の場合にはじめは速い収束を見せるものの，停留点の近傍では $\rho > 0$ の設定よりも遅くなり，理論保証通りの振る舞いを確認できる．

参考文献

[1] A. d'Aspremont, D. Scieur, and A. Taylor. Acceleration methods. *Foundations and Trends® in Optimization*, 5(1-2):1–245, 2021.

[2] C. Kanzow and T. Lechner. Globalized inexact proximal Newton-type methods for nonconvex composite functions. *Computational Optimization and Applications*, 78(2):377–410, 2021.

[3] C.-p. Lee and S. J. Wright. Inexact successive quadratic approximation for regularized optimization. *Computational Optimization and Applications*, 72(3): 641–674, 2019.

[4] S. Nakayama, Y. Narushima, and H. Yabe. Inexact proximal DC Newton-type method for nonconvex composite functions. *Computational Optimization and Applications*, 2023.

[5] B. S. Mordukhovich, X. Yuan, S. Zeng, and J. Zhang. A globally convergent proximal Newton-type method in nonsmooth convex optimization. *Mathematical Programming*, 198(1):899–936, 2023.