ヘッセ行列の対角成分を用いたニュートン型近接勾配法

柳下　翔太郎

近接勾配法は，統計，機械学習，信号処理などの分野において現れる非平滑最適化問題を解くための代表的なアルゴリズムである．近接勾配法は目的関数の滑らかな項の一次近似までしか考慮しないが，近年，二次近似まで考えるニュートン型近接勾配法が注目を浴びている．ニュートン型近接勾配法は二次近似まで考慮することで近接勾配法よりも反復回数を減らすことができる傾向にあるが，一般にはその部分問題を解くための計算コストは近接勾配法のそれよりも高くなる．本稿では，部分問題を解くための計算コストを抑え，特定の状況下で近接勾配法を優越することが理論的に示唆されるニュートン型近接勾配法を紹介する．

キーワード：非平滑最適化，近接勾配法，ニュートン型近接勾配法，重み付き近接写像

1. はじめに

本稿では，以下の最適化問題を解くためのアルゴリズムについて考える：

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad F(x) \coloneqq f(x) + g(x). \tag{1}$$

ここで，$f:\mathbb{R}^n \to \mathbb{R}$ は二階連続的微分可能関数，$g:\mathbb{R}^n \to (-\infty,\infty]$ は下半連続かつ方向微分可能な関数で $\operatorname{dom} g \coloneqq \{x\in\mathbb{R}^n \mid g(x) < \infty\} \neq \emptyset$ であるとし，F が下に有界であるとする．問題 (1) のような最適化問題は統計，機械学習，信号処理などの分野において頻繁に現れる．$\gamma > 0$, $g(x) = \gamma\|x\|_1$ として解にスパース性を誘導する定式化はその典型例であり[1]，多くの場合，g は非平滑な関数である．問題 (1) は大規模であることが多いため，この問題を効率的に解くためのアルゴリズムの研究が盛んに行われている．

近接勾配法[2]は問題 (1) を解くためのアルゴリズムとして，最もよく知られたものの一つである．$L_t > 0$ をステップサイズの逆数として，近接勾配法は

$$\begin{aligned} x^{t+1} &\in \operatorname{prox}_{g/L_t}\Big(x^t - \frac{1}{L_t}\nabla f(x^t)\Big) \\ &= \underset{y\in\mathbb{R}^n}{\operatorname{argmin}}\Big\{f(x^t) + \nabla f(x^t)^\top(y - x^t) \\ &\qquad + \frac{L_t}{2}\|y - x^t\|_2^2 + g(y)\Big\} \end{aligned} \tag{2}$$

によって点列を生成する．ただし，

$$\operatorname{prox}_h(x) \coloneqq \underset{y}{\operatorname{argmin}}\Big\{h(y) + \frac{1}{2}\|y - x\|_2^2\Big\}$$

は近接写像と呼ばれる集合値写像である．すなわち近接勾配法は，滑らかな関数 f を一次近似し二次の正則化項を加えた部分問題を繰り返し解く反復法である．$g(x) = \gamma\|x\|_1$ などの一部の g に関しては，その近接写像を陽に計算できるため，近接勾配法の各反復を効率的に実行することができる．特に，非凸な g であっても変数分離可能であれば，式 (2) は変数ごとの最適化問題を考えればよいため，効率的に各反復を実行できる場合がある．少なくない g に対して近接勾配法の各反復が効率的に実行できる一方で，近接勾配法ではあくまで f の一次の情報までしか用いていないため，悪条件の問題に対してはアルゴリズムが収束するまでに必要な反復回数が多くなってしまうことがしばしばある．

ニュートン型近接勾配法 [2–5] は，そのような問題点に対処しうるアルゴリズムとして近年注目を集めている．ニュートン型近接勾配法は，f の x^t におけるヘッセ行列またはその近似行列である正定値行列を B_t として

$$\begin{aligned} z^t &\in \operatorname{prox}_g^{B_t}\Big(x^t - B_t^{-1}\nabla f(x^t)\Big) \\ &= \underset{y\in\mathbb{R}^n}{\operatorname{argmin}}\Big\{f(x^t) + \nabla f(x^t)^\top(y - x^t) \\ &\qquad + \frac{1}{2}\|y - x^t\|_{B_t}^2 + g(y)\Big\} \end{aligned} \tag{3}$$

とし，適切な $\lambda_t \in (0,1]$ を用いて

$$x^{t+1} = \lambda_t z^t + (1-\lambda_t)x^t$$

とすることで点列の更新を行う．ただし，

$$\operatorname{prox}_h^A(x) \coloneqq \underset{y}{\operatorname{argmin}}\Big\{h(y) + \frac{1}{2}\|y - x\|_A^2\Big\}$$

やぎした　しょうたろう
統計数理研究所リスク解析戦略研究センター
〒 190–8562 東京都立川市緑町 10–3
syagi@ism.ac.jp

[1] $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}}$ は ℓ_p ノルムを表す．
[2] たとえば，文献 [1] を参照されたい．

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

は重み付き近接写像と呼ばれる集合値写像で，$\|x\|_A = \sqrt{x^\top Ax}$ である．ニュートン型近接勾配法は f の二次の情報を用いるため，通常の近接勾配法と比べて反復回数を減らすことができる傾向にある．しかし一般には，g が変数分離可能であっても部分問題 (3) は変数ごとの最適化問題にはならないため，各反復で部分問題を効率よく解くことは難しい． Becker et al. [6] は凸である g と特定の B_t に対して，部分問題 (3) が低次元の方程式を解くことに帰着されることを示したが，g が非凸である場合には適用できない．

このような問題意識から近年，B_t を対角行列に制限したニュートン型近接勾配法が提案された [7, 8]．この方法は，f の二次の情報を通常の近接勾配法よりも取り入れることができる一方で，g が変数分離可能であれば (3) は変数ごとの最適化問題に帰着されるため，各反復での部分問題は通常の近接勾配法と同程度に簡単になるという利点を有する．文献 [7, 8] において，対角行列を用いたニュートン型近接勾配法の通常の近接勾配法に対する優位性が数値実験により示された一方で，その優位性を示唆する理論的な結果は与えられていなかった．

本稿では，著者らが文献 [9] において提案した**近接対角ニュートン法**を紹介する．近接対角ニュートン法は，ヘッセ行列の対角成分のみを用いたニュートン型近接勾配法で，その収束解析の結果から，特定の状況下での近接勾配法に対する優位性が示唆される．紹介する命題の証明や数値実験の結果については紙面の都合上割愛するので，文献 [9] を参照していただきたい．

2. 近接対角ニュートン法

以降では，f のヘッセ行列の対角成分は正であるとする．近接対角ニュートン法は，$x^0 \in \operatorname{dom} g, \eta > 1$, $D_t = \operatorname{diag}(\nabla^2 f(x^t))$ を $\nabla^2 f(x^t)$ の対角成分のみを取り出した行列として

$$x^{t+1} \in \operatorname{prox}_g^{\eta^k D_t}\left(x^t - (\eta^k D_t)^{-1}\nabla f(x^t)\right) \quad (4)$$

によって点列を生成する．ただし，$\beta > 0$ として許容条件

$$f(x^{t+1}) \le f(x^t) + \nabla f(x^t)^\top (x^{t+1} - x^t) + \frac{\beta}{2}\|x^{t+1} - x^t\|^2_{\eta^k D_t} \quad (5)$$

を満たす最小の k を選択するといった直線探索を用いる．近接対角ニュートン法のアルゴリズムを Algorithm 1 にまとめる．

Algorithm 1 問題 (1) に対する近接対角ニュートン法

Input: $x^0 \in \operatorname{dom} g$, $\eta > 1$, $\beta > 0$, $t = 0$.
repeat
  $H_t = \operatorname{diag}(\nabla^2 f(x^t))$ を計算する．
  $x^{t+1} \in \operatorname{prox}_g^{H_t}\left(x^t - H_t^{-1}\nabla f(x^t)\right)$ を求める．
  while $f(x^{t+1}) > f(x^t) + \nabla f(x^t)^\top (x^{t+1} - x^t) + \frac{\beta}{2}\|x^{t+1} - x^t\|^2_{H_t}$ **do**
    $H_t \leftarrow \eta H_t$ と更新する．
    $x^{t+1} \in \operatorname{prox}_g^{H_t}\left(x^t - H_t^{-1}\nabla f(x^t)\right)$ を求める．
  end while
  $t \leftarrow t + 1$ と更新する．
until 適当な停止条件を満たす．

f の二階連続的微分可能性やヘッセ行列の対角成分に対する仮定は通常の近接勾配法には課されないものであり，近接対角ニュートン法を適用できる問題の範囲は，近接勾配法のそれよりも限られていることに注意しておく．しかし，近接対角ニュートン法の適用範囲には十分多くの応用上重要な問題が含まれている．たとえば，f が強凸な二階連続的微分可能関数であれば適用可能である．また，線形回帰の文脈でスパース正則化付き最小二乗法の問題を，すなわち $f(x) = \frac{1}{2}\|b - Ax\|_2^2$ のときを考えてみる．ここで，A の j 番目の列ベクトル a_j は j 番目の説明変数ベクトルを表している．この f は一般には強凸関数にはならないが，以下のようにして適用可能であることがわかる．f のヘッセ行列の j 番目の対角成分は $a_j^\top a_j \ge 0$ であるので，$a_j = 0$ となる j が存在する場合に限って仮定が満たされないということがわかる．これはすべての標本に対して 0 という値をとる説明変数が存在するということであり，そのような説明変数は線形回帰においては意味がないので，初めから取り除いておけばよい．つまり，一般性を失わず f のヘッセ行列の対角成分は正であるとしてよい．

近接対角ニュートン法は対角行列を用いるため，各反復で式 (4) の対角重み付き近接写像を効率的に計算することができる．たとえば，g が ℓ_1 ノルム，非負制約の標示関数，単体制約の標示関数などであるときの対角重み付き近接写像は Park et al. [10] で与えられている．また，g が非凸関数であっても変数分離可能であれば，対角重み付き近接写像が陽に求まることも多い．g が変数分離可能でない非凸関数であってもその対角重み付き近接写像が陽に求まることがある．$K \in \{1, \ldots, n-1\}$ として，刈込 ℓ_1 ノルム [11, 12] は

$$T_K(x) := \min_{\substack{\Lambda \subset \{1,\ldots,n\} \\ |\Lambda| = n-K}} \sum_{i \in \Lambda} |x_i|$$

のように定義される変数分離可能でない非凸関数で，

 Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

非ゼロ要素数制約

$$|\{i \in \{1,\dots,n\} \mid x_i \neq 0\}| \leq K$$

の正確なペナルティ関数である [13, 14]．刈込 ℓ_1 ノルムの対角重み付き近接写像は以下のように与えられる．

命題 1. $\gamma > 0$, $x \in \mathbb{R}^n$, D を対角成分が正である対角行列とする．$\mathcal{I}$ を，任意の $i \in I$, $j \notin I$ に対して $\phi(x_i, d_i) \leq \phi(x_j, d_j)$ が成立する $|I| = n - K$ を満たす集合 $I \subset \{1,\dots,n\}$ をすべて集めた集合とする．ただし，

$$\phi(x_i, d_i) = \begin{cases} \frac{d_i}{2} x_i^2 & (|x_i| \leq \frac{\gamma}{d_i}) \\ \gamma |x_i| - \frac{\gamma^2}{2d_i} & (|x_i| > \frac{\gamma}{d_i}) \end{cases}$$

で，d_i は D の i 番目の対角成分を表す．このとき $\mathrm{prox}^D_{\gamma T_K}(x)$ は，ある $I \in \mathcal{I}$ に対して

$$x_i^* = \begin{cases} \mathcal{S}_{\frac{\gamma}{d_i}}(x_i) & (i \in I) \\ x_i & (i \notin I) \end{cases}$$

となる x^* をすべて集めた集合である．ここで，

$$\mathcal{S}_c(\xi) = \begin{cases} \xi - c & (\xi > c) \\ 0 & (-c \leq \xi \leq c) \\ \xi + c & (\xi < -c) \end{cases}$$

である．

命題 1 より，$\phi(x_i, d_i)$ を計算し，その中で K 番目に大きい値を求めれば，刈込 ℓ_1 ノルムの対角重み付き近接写像の一点を計算することができるということがわかる．

3. 近接対角ニュートン法の収束性

以降の収束性の解析においてはさらに，$\mathrm{diag}(\nabla^2 f)$ が一様に正定値であることを仮定する．つまり，ある正の定数 m が存在して，任意の $x \in \mathrm{dom}\, g$ に対して $\mathrm{diag}(\nabla^2 f(x)) - mI$ が半正定値であるとする．ここで，I は単位行列である．近接対角ニュートン法の大域的収束性については以下のように与えられる．

定理 1. ∇f がリプシッツ連続であることを仮定し，$0 < \beta < 1$, $\{x^t\}_{t=0}^{\infty}$ を Algorithm 1 から生成された点列とする．このとき，$\{x^t\}_{t=0}^{\infty}$ の任意の集積点は問題 (1) の方向停留点[3]である． さらに，もし $\{x^t\}_{t=0}^{\infty}$ が有界であれば，$\lim_{t\to\infty} \mathrm{dist}(x^t, \mathcal{X}^*) = 0$ が成立する[4]．ここで，$\mathcal{X}^*$ は問題 (1) の方向停留点をすべて集めた集合である．もし g が凸であれば，上記の結果は $1 \leq \beta < 2$ に対しても成立する．

大域的収束性以外にも，g に対する凸性などの適当な仮定のもとで劣一次収束性などを示すことができる．しかしながら，それらは通常の近接勾配法に対しても示すことができるため，近接対角ニュートン法特有のものではない．以下で述べる二次収束性は，近接勾配法にはない性質である．

定理 2. $1 < \beta < 2$, $\{x^t\}_{t=0}^{\infty}$ を Algorithm 1 から生成された点列とし $f(x) = \sum_{i=1}^n f_i(x_i)$ とする．ただし，$f_i : \mathbb{R} \to \mathbb{R}$ である．f が m-強凸，∇f と $\nabla^2 f$ がリプシッツ連続，g が凸であることを仮定する．このとき，$\{x^t\}_{t=0}^{\infty}$ は問題 (1) の一意な大域的最適解 x^* に収束し，その収束は二次収束である．つまり，ある $C > 0$ が存在して

$$\|x^{t+1} - x^*\|_2 \leq C \|x^t - x^*\|_2^2$$

が成立する．

Li et al. [15] が平滑な無制約最適化問題（つまり，問題 (1) で $g = 0$ の場合）に対して提案した対角 BFGS 法は，定理 2 と同様の f の変数分離可能性の仮定のもとで，局所超一次収束性を有する．一方，近接対角ニュートン法は $g \neq 0$ であってもよく，加えて大域的な二次収束性を備えている．大雑把に定理 2 を解釈すれば，f のヘッセ行列が対角行列に近ければ，近接対角ニュートン法は有効であるように思える．しかしながら，定理 2 は f の完全な変数分離可能性を仮定しているため，それは拡大解釈のようにも思える．狭義凸二次関数に対する一次収束性について考えることで，上記の解釈が的外れでなかったことがわかる．

定理 3. $0 < \beta \leq 1$, $\{x^t\}_{t=0}^{\infty}$ を Algorithm 1 から生成された点列，$f(x) = \frac{1}{2} x^\top A x + b^\top x$, $D = \mathrm{diag}(A)$ とする．ただし，A は正定値行列である．g が凸関数であると仮定し，x^* を問題 (1) の一意な大域的最適解とする．このとき，

$$F(x^t) - F(x^*) \leq \frac{\overline{\eta}\tau}{2} \left(1 - \frac{\sigma}{\overline{\eta}\tau}\right)^t \|x^0 - x^*\|_D^2$$

[3] $F'(x^*; d) := \lim_{\eta \searrow 0} \frac{F(x^* + \eta d) - F(x^*)}{\eta} \geq 0$ が任意の方向 d に対して成立するとき，$x^* \in \mathrm{dom}\, g$ を問題 (1) の方向停留点であるという．方向停留点であることは局所最適解であることの必要条件である．

[4] $\mathrm{dist}(x, \mathcal{X}) = \inf_{y \in \mathcal{X}} \|y - x\|_2$ は点 x と集合 $\mathcal{X}$ の距離を表す．

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

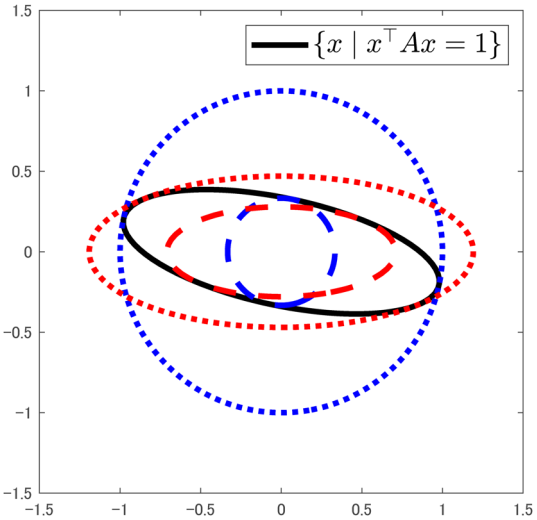

図 1　内接または外接する円と楕円

が成立する．ここで，

$$\sigma = \inf_{v \neq 0} \frac{v^\top Av}{v^\top Dv}, \quad \tau = \sup_{v \neq 0} \frac{v^\top Av}{v^\top Dv},$$

$\overline{\eta} = \max\{\frac{1}{\tau}, \frac{\eta}{\beta}\} \geq 1$ である．

$\frac{\sigma}{\overline{\eta}\tau}$ が大きければ大きいほど，収束が速いということに注意しておく．$\overline{\eta}$ は直線探索を行っている都合で出てきてしまう定数であるので，$\frac{\sigma}{\tau}$ について考える．明らかに $\frac{\sigma}{\tau} \leq 1$ であり，A が対角行列であることと $\frac{\sigma}{\tau} = 1$ であることが同値であるため，$\frac{\sigma}{\tau}$ は A が対角行列に近いかどうかを表す指標と解釈することができる．つまり，A が対角行列に近ければ近いほど速い収束が期待できるということを定理 3 は示唆する．

一方で近接勾配法の一次収束性は，$\lambda_{\max}$，$\lambda_{\min}$ をそれぞれ A の最大固有値，最小固有値として，

$$F(x^t) - F(x^*) \leq \frac{\hat{\eta}\lambda_{\max}}{2}\left(1 - \frac{\lambda_{\min}}{\hat{\eta}\lambda_{\max}}\right)^t \|x^0 - x^*\|_2^2$$

のように与えられる [16](Theorem 10.29)．ただし，$\hat{\eta} \geq 1$ である．$\hat{\eta}$ も直線探索などの都合で出てきてしまう定数であるので無視して，$\frac{\lambda_{\min}}{\lambda_{\max}}$ を $\frac{\sigma}{\tau}$ と比べてみる．その逆数 $\frac{\lambda_{\max}}{\lambda_{\min}}$ は条件数と呼ばれ，条件数が大きければ大きいほど最適化問題が悪条件であるとされる．$\frac{\sigma}{\tau}$ のときと同様に考えてみると，$\frac{\lambda_{\min}}{\lambda_{\max}}$ は A が単位行列の定数倍から遠ざかれば遠ざかるほど小さくなる量であるといえる（このとき，その逆数である条件数は大きくなる）．

図 1 の点線の円は $\{x \mid \lambda_{\min} x^\top x = 1\}$，破線の円は $\{x \mid \lambda_{\max} x^\top x = 1\}$，点線の楕円は $\{x \mid x^\top(\sigma D)x = 1\}$，破線の楕円は $\{x \mid x^\top(\tau D)x = 1\}$ をそれぞれ表している．実は，$\frac{\sigma}{\tau}$ は $\{x \mid x^\top Ax = 1\}$ に内接する楕円と外接する楕円の長半径（ないしは短半径）の比の二乗となっている．同様に，$\frac{\lambda_{\min}}{\lambda_{\max}}$ は内接する円と外接する円の半径の比の二乗である．すなわち，A が対角行列に近い場合は，図 1 のように $\{x \mid x^\top Ax = 1\}$ の軸が座標軸に対して垂直に近くなり，楕円の長半径の比が円の半径の比よりも小さくなる．その結果，条件数が大きい状況でも A が対角行列に近ければ，通常の近接勾配法よりも近接対角ニュートン法が速く収束することを期待できるということになる．

4.　非単調な直線探索

非単調な直線探索 [17, 18] は，生成する点列の目的関数値が必ずしも単調減少しないような緩い許容条件を用いることで，関数値や勾配などの評価回数を抑えることを狙った直線探索法であり，近年では近接勾配法への応用も提案されている [19, 20]．許容条件 (5) は適切な β のもとで，近接対角ニュートン法が生成する点列の目的関数値は単調減少させる．そこで，近接対角ニュートン法に対しても，同様の狙いで非単調な直線探索を組み合わせることを考える．具体的には許容条件 (5) の代わりに，

$$F(x^{t+1}) \leq \max_{\max\{0,t-M+1\} \leq s \leq t} F(x^s) - \frac{\alpha}{2}\|x^{t+1} - x^t\|_{\eta^k D_t}^2$$

を許容条件として用いる．ただし，$\eta > 1$，$0 < \alpha < 1$，$M \geq 1$ で，$D_t = \mathrm{diag}(\nabla^2 f(x^t))$ である．非単調な直線探索を用いた近接対角ニュートン法のアルゴリズムを Algorithm 2 にまとめる．

Algorithm 2 非単調な近接対角ニュートン法

Input: $x^0 \in \mathrm{dom}\, g$, $\eta > 1$, $0 < \alpha < 1$, $M \geq 1$, $t = 0$.
repeat
　$H_t = \mathrm{diag}(\nabla^2 f(x^t))$ を計算する．
　$x^{t+1} \in \mathrm{prox}_g^{H_t}\left(x^t - H_t^{-1}\nabla f(x^t)\right)$ を求める．
　while $F(x^{t+1}) > \max_{\max\{0,t-M+1\} \leq s \leq t} F(x^s) - \frac{\alpha}{2}\|x^{t+1} - x^t\|_{H_t}^2$ **do**
　　$H_t \leftarrow \eta H_t$ と更新する．
　　$x^{t+1} \in \mathrm{prox}_g^{H_t}\left(x^t - H_t^{-1}\nabla f(x^t)\right)$ を求める．
　end while
　$t \leftarrow t + 1$ と更新する．
until 適当な停止条件を満たす．

非単調な近接対角ニュートン法に対しても，定理 1 や 2 と同様に大域的収束性や f が変数分離可能である場合の二次収束性を示すことができる．

 Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

定理 4. ∇f がリプシッツ連続であることを仮定し，$\{x^t\}_{t=0}^{\infty}$ を Algorithm 2 から生成された点列とする．さらに，F が $\mathcal{D}$ 上で一様連続になるような領域 $\mathcal{D} \supset \{x^t\}_{t=0}^{\infty}$ の存在を仮定する．このとき，$\{x^t\}_{t=0}^{\infty}$ の任意の集積点は問題 (1) の方向停留点である．さらに，もし $\{x^t\}_{t=0}^{\infty}$ が有界であれば，$\lim_{t\to\infty} \mathrm{dist}(x^t, \mathcal{X}^*) = 0$ が成立する．ここで，$\mathcal{X}^*$ は問題 (1) の方向停留点をすべて集めた集合である．

定理 5. $\{x^t\}_{t=0}^{\infty}$ を Algorithm 2 から生成された点列とし $f(x) = \sum_{i=1}^{n} f_i(x_i)$ とする．ただし，$f_i : \mathbb{R} \to \mathbb{R}$ である．f が m-強凸，∇f と $\nabla^2 f$ がリプシッツ連続，g が凸であること，および F が $\mathcal{D}$ 上で一様連続になるような領域 $\mathcal{D} \supset \{x^t\}_{t=0}^{\infty}$ の存在を仮定する．このとき，$\{x^t\}_{t=0}^{\infty}$ は問題 (1) の一意な大域的最適解 x^* に収束し，その収束は二次収束である．

5. おわりに

本稿では，著者らが文献 [9] において提案した近接対角ニュートン法を紹介した．その収束性について述べることで，近接対角ニュートン法が，通常の近接勾配法対して優位であるような状況が理論的に示唆されるという，これまでの対角行列を用いたニュートン型近接勾配法にはない特徴をもっていることを確認した．今回割愛した数値実験では，狭義凸二次関数 $f(x) = \frac{1}{2}x^\top Ax + b^\top x$ に対して，g が凸および非凸のどちらの場合でも，A が対角行列に近いほど近接対角ニュートン法の性能が良くなることが確認された．g が凸関数である場合は理論の範疇である一方で，非凸の場合はその優位性を示唆する理論的な結果は得られていない．非凸な g に対しては Becker et al. [6] のような部分問題 (3) を低次元の方程式を解くことに帰着させる公式がないため，近接対角ニュートン法はとりわけ g が非凸である場合に有効なアルゴリズムであると考えられる．そのため，g が非凸関数である場合での近接対角ニュートン法の優位性を理論的に示すことは，最も重要な課題である．

また，近接対角ニュートン法のヘッセ行列の対角成分を用いるというアイデアは非常にシンプルなものであり，さまざまな近接型アルゴリズムと組む合わさる可能性を秘めている．たとえば Pock and Chambolle [21] や Giselsson and Boyd [22] はそれぞれ，ある近接型アルゴリズムに対して対角重み付き近接写像を用いることを提案しており，それに対して対角ニュートンのアイデアを組み合わせれば，何かご利益があるかもしれない．本稿が悪条件な問題に悩まされる誰かの助けとなれば幸いである．

謝辞 本稿を執筆する機会をくださったオーガナイザの関谷和之先生と機関誌編集委員のみなさまに深く御礼申し上げます．また，本稿で紹介した研究における共同研究者である中山舜民先生に，この場を借りて感謝いたします．

参考文献

[1] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends® in Optimization*, **1**, pp. 127–239, 2014.

[2] G. H. Chen and R. T. Rockafellar, "Convergence rates in forward–backward splitting," *SIAM Journal on Optimization*, **7**, pp. 421–444, 1997.

[3] S. Becker and J. Fadili, "A quasi-newton proximal splitting method," In *Advances in Neural Information Processing Systems 25*, 2012.

[4] J. D. Lee, Y. Sun and M. Saunders, "Proximal Newton-type methods for minimizing composite functions," *SIAM Journal on Optimization*, **24**, pp. 1420–1443, 2014.

[5] R. H. Byrd, J. Nocedal and F. Oztoprak, "An inexact successive quadratic approximation method for L-1 regularized optimization," *Mathematical Programming*, **157**, pp. 375–396, 2016.

[6] S. Becker, J. Fadili and P. Ochs, "On quasi-newton forward-backward splitting: Proximal calculus and convergence," *SIAM Journal on Optimization*, **29**, pp. 2445–2481, 2014.

[7] Y. Park, S. Dhar, S. Boyd and M. Shah, "Variable metric proximal gradient method with diagonal Barzilai–Borwein stepsize," In *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3597–3601, 2020.

[8] S. Bonettini, I. Loris, F. Porta and M. Prato, "Variable metric inexact line-search-based methods for nonsmooth optimization," *SIAM Journal on Optimization*, **26**, pp. 891–921, 2016.

[9] S. Yagishita and S. Nakayama, "Proximal diagonal Newton methods for composite optimization problems," *arXiv*, arXiv:2310.06789, 2023.

[10] Y. Park, S. Dhar, S. Boyd and M. Shah, "Variable metric proximal gradient method with diagonal Barzilai–Borwein stepsize," *arXiv*, arXiv:1910.07056, 2019.

[11] Z. Luo, Y. Wang and X. Zhang, "New improved penalty methods for sparse reconstruction based on difference of two norms," *Technical Report*, DOI:10.13140/RG.2.1.3256.3369, 2013.

[12] X. Huang, Y. Liu, L. Shi, S. V. Huffel and J. A. Suykens, "Two-level ℓ_1 minimization for compressed sensing," *Signal Processing*, **108**, pp. 459–475, 2015.

[13] J. Gotoh, A. Takeda and K. Tono, "DC formulations and algorithms for sparse optimization problems," *Mathematical Programming*, **169**, pp. 141–176, 2018.

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

[14] S. Yagishita and J. Gotoh, "Exact penalization at d-stationary points of cardinality-or rank-constrained problem," *arXiv*, arXiv:2209.02315, 2022.

[15] D. Li, X. Wang and J. Huang, "Diagonal BFGS updates and applications to the limited memory BFGS method," *Computational Optimization and Applications*, **81**, pp. 829–856, 2022.

[16] A. Beck, *First-Order Methods in Optimization*, Society for Industrial and Applied Mathematics, 2017.

[17] L. Grippo, F. Lampariello and S. Lucidi, "A nonmonotone line search technique for Newton's method," *SIAM Journal on Numerical Analysis*, **23**, pp. 707–716, 1986.

[18] L. Grippo and M. Sciandrone, "Nonmonotone globalization techniques for the Barzilai-Borwein gradient method," *Computational Optimization and Applications*, **23**, pp. 143–169, 2002.

[19] S. J. Wright, R. D. Nowak and M. A. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Transactions on Signal Processing*, **57**, pp. 2479–2493, 2009.

[20] P. Gong, C, Zhang, Z. Lu, J. Huang and J. Ye, "A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems," In *International Conference on Machine Learning*, pp. 37–45, 2013.

[21] T. Pock and A. Chambolle, "Diagonal preconditioning for first order primal-dual algorithms in convex optimization," In *2011 International Conference on Computer Vision*, pp. 1762–1769, 2011.

[22] P. Giselsson and S. Boyd, "Diagonal scaling in Douglas-Rachford splitting and ADMM," In *53rd IEEE Conference on Decision and Control*, pp. 5033–5039, 2014.

 Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.