

大規模無制約最適化問題に対する 準ニュートン法と近接勾配法

中山 舜民

近接勾配法は微分可能な関数と ℓ_1 ノルムなどの微分不可能な点を含む関数の和を最小化する数値解法である。通常の近接勾配法は最急降下法に基づく方法であるが、近年、ニュートン型近接勾配法と呼ばれる準ニュートン法に基づく近接勾配法の研究が注目を浴びている。準ニュートン法は小・中規模の微分可能な無制約最適化問題に対して有効な数値解法であるが、密行列を必要とすることから大規模な問題に直接適用することが困難である。そのため、大規模問題に対して、メモリーレス準ニュートン法と呼ばれる行列を陽に使用しない準ニュートン法が提案されている。本稿では筆者が提案したメモリーレス準ニュートン法に基づくニュートン型近接勾配法を中心に関連研究を紹介する。

キーワード：無制約最適化問題、準ニュートン法、メモリーレス準ニュートン法、近接勾配法、ニュートン型近接勾配法

1. はじめに

近接勾配法は

$$\min_{x \in \mathbb{R}^n} f(x) = g(x) + h(x) \quad (1)$$

という構造をもつ無制約最適化問題を解く数値解法である。ここで、 $g: \mathbb{R}^n \rightarrow \mathbb{R}$ は連続的・微分可能な関数とし、 $h: \mathbb{R}^n \rightarrow \mathbb{R}$ は微分不可能な点を含む下半連続な関数とする。機械学習の分野では g を損失関数、 h を正則化項として定式化される問題 (1) がよく扱われ、これはスパース最適化と呼ばれる問題の枠組みである。たとえば、 $g(x) = \frac{1}{2} \|Ax - b\|^2$ 、 $h(x) = \lambda \|x\|_1$ とした Least Absolute Shrinkage and Selection Operator (LASSO) などが挙げられる。ここで、 $A \in \mathbb{R}^{m \times n}$ 、 $b \in \mathbb{R}^m$ であり、 $\lambda \geq 0$ はペナルティパラメータである。また、本稿では ℓ_2 ノルムを $\|\cdot\|$ 、 ℓ_1 ノルムを $\|\cdot\|_1$ と表記する。

近接勾配法の詳細を述べる前に、最急降下法と準ニュートン法を紹介する。これらの方法は、問題 (1) において、関数 h がない場合、もしくは関数 h が関数 g と同様に連続的・微分可能な場合の最適化問題を解くための数値解法である。最急降下法は最も基本的な方法であるが、収束が遅いことから実用的な方法ではない。一方、準ニュートン法は小・中規模な問題に対して非常に有効な方法として知られている。しかしながら、次元が大きな問題に直接適用することが困難であ

る。このような大規模問題に対しては行列を陽に使用しない準ニュートン法としてメモリーレス準ニュートン法が有効である。2 節では最急降下法と準ニュートン法について述べ、メモリーレス準ニュートン法を紹介する。3 節では最急降下法に基づく近接勾配法およびニュートン型近接勾配法について解説する。最後に、われわれの提案したメモリーレス準ニュートン法に基づくニュートン型近接勾配法を紹介する。

2. 大規模問題に対する準ニュートン法

本節では、近接勾配法を紹介するための前提知識として最急降下法と準ニュートン法について述べる。また、大規模問題に対する準ニュートン法としてメモリーレス準ニュートン法を紹介する。

2.1 最急降下法と準ニュートン法

最急降下法と準ニュートン法は

$$\min_{x \in \mathbb{R}^n} f(x)$$

を解くための数値解法である。本節では $f: \mathbb{R}^n \rightarrow \mathbb{R}$ は連続的・微分 (または 2 回連続的・微分) 可能な関数とする。この問題に対して、反復法が広く使用されている。反復法は任意の初期点 $x_0 \in \mathbb{R}^n$ から出発し、反復式

$$x_{k+1} = x_k + \alpha_k d_k \quad (2)$$

により点列を更新する。ここで、 $\alpha_k > 0$ をステップ幅、 $d_k \in \mathbb{R}^n$ を探索方向と呼ぶ。反復法は 1 次の最適性条件を満たす点、すなわち、

$$\nabla f(x^*) = 0 \quad (3)$$

なかやま しゅんみん

中央大学理工学部

〒112-8551 東京都文京区春日 1-13-27

shummin@kc.chuo-u.ac.jp

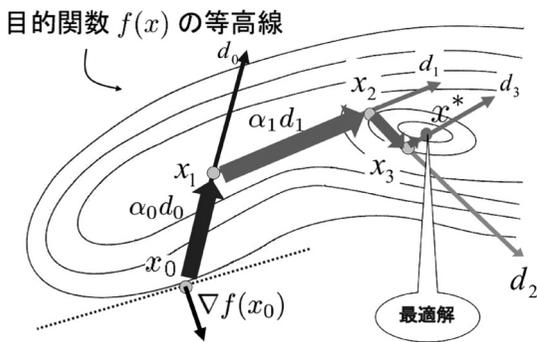


図1 反復法

を満たす点 x^* を求める方法である。ここで、 $\nabla f(x^*)$ は x^* での f の勾配を表し、式 (3) を満たす点を停留点と呼ぶ。図1は2変数の場合を例に、 x_0 から停留点 x^* に収束する反復法の様子を表している。反復法では、目的関数値が減少するように点列を更新する。そのためには、方向微係数が負、すなわち、

$$f'(x_k; d_k) \equiv \lim_{\alpha \rightarrow 0} \frac{f(x_k + \alpha d_k) - f(x_k)}{\alpha} = \nabla f(x_k)^\top d_k < 0 \quad (4)$$

を満たす探索方向である必要がある。条件 (4) を降下条件と呼び、降下条件を満たす探索方向は降下方向と呼ばれる。降下方向に進めば、目的関数値が必ず減少することが保証されている (図1)。降下方向ならば、ステップ幅 α_k を決める直線探索を行うことが可能になる。ステップ幅の選択基準として、Wolfe 条件や Armijo 条件が用いられる [1, 2]。

次に、最急降下法と準ニュートン法の探索方向 d_k の選び方について述べる。

最急降下法の探索方向は目的関数の1次近似

$$f(x_k + d) \approx f(x_k) + \nabla f(x_k)^\top d$$

を最小にする方向 d として

$$d_k = -\nabla f(x_k) \quad (5)$$

が選択される。ただし、 $\|d_k\| = \|\nabla f(x_k)\|$ であると仮定している。 $\nabla f(x_k)^\top d_k = -\|\nabla f(x_k)\|^2 < 0$ であるため、降下条件を満たす。目的関数の勾配 $\nabla f(x)$ がリプシッツ連続¹であるとき、任意の初期点から出発して、最急降下法は停留点に大域的収束することが保証

¹ ある正定数 L が存在して

$\|\nabla f(u) - \nabla f(v)\| \leq L\|u - v\|, \forall u, v \in \mathbb{R}^n$ が成り立つ。

されている。しかしながら、局所的な収束率²は1次収束であることから、実用上はあまり効率的ではない。

次に、準ニュートン法について紹介する。準ニュートン法はニュートン法を改良した方法であり、ニュートン法は2次近似

$$f(x_k + d) \approx f(x_k) + \nabla f(x_k)^\top d + \frac{1}{2} d^\top \nabla^2 f(x_k) d$$

を最小化する方向 d を選択する。ここで、ヘッセ行列 $\nabla^2 f(x_k)$ が正定値であれば

$$d_k = -\nabla^2 f(x_k)^{-1} \nabla f(x_k) \quad (6)$$

となる。ニュートン法は局所的に2次収束をするという利点があるが、 $\nabla^2 f(x_k)$ が正定値である保証がないため、探索方向が降下方向であるとは限らないという弱点がある。そのため、 $\nabla^2 f(x_k)$ を正定値対称な近似行列 B_k で置き換えた準ニュートン法が提案されている。その探索方向 d_k は

$$d_k = -B_k^{-1} \nabla f(x_k)$$

与えられる。具体的に近似行列を選択する際には、 $B_k \approx \nabla^2 f(x_k)$ であることが望まれ、正定値対称になるように一つ前の近似行列 B_{k-1} を更新して B_k を計算する。ここで、 $\nabla f(x_{k-1})$ の1次近似を考えると

$$\nabla f(x_{k-1}) \approx \nabla f(x_k) - \nabla^2 f(x_k)(x_k - x_{k-1}) \quad (7)$$

という関係式が得られることから、近似行列が満たすべき条件としてセカント条件：

$$B_k s_{k-1} = y_{k-1} \text{ または } s_{k-1} = H_k y_{k-1}$$

を考えることができる。ただし、 $H_k = B_k^{-1}$ とし、

$$s_{k-1} = x_k - x_{k-1}, \quad y_{k-1} = \nabla f(x_k) - \nabla f(x_{k-1})$$

とする。セカント条件を満たす B_k は無数に存在するため、いろいろな更新公式が提案されている。なかでもBFGS (Broyden-Fletcher-Goldfarb-Shanno) 公式：

$$B_k = B_{k-1} - \frac{B_{k-1} s_{k-1} s_{k-1}^\top B_{k-1} + y_{k-1} y_{k-1}^\top}{s_{k-1}^\top B_{k-1} s_{k-1} + y_{k-1}^\top y_{k-1}},$$

$$H_k = H_{k-1} - \frac{H_{k-1} y_{k-1} s_{k-1}^\top + s_{k-1} y_{k-1}^\top H_{k-1}}{s_{k-1}^\top y_{k-1}} + \left(1 + \frac{y_{k-1}^\top H_{k-1} y_{k-1}}{s_{k-1}^\top y_{k-1}} \right) \frac{s_{k-1} s_{k-1}^\top}{s_{k-1}^\top y_{k-1}} \quad (8)$$

が最も有名かつ有効な更新公式として知られてい

² 収束率は最適解に収束する速度を表す。2次収束、超1次収束、1次収束の順番で速く、少ない反復回数で最適解に到達する。

る³。BFGS 公式のほかにも、DFP(Davidon-Fletcher-Powell) 公式や対称ランクワン公式などが有名である。それぞれの更新公式の詳細については文献 [1, 2] などを参照されたい。さらに、パラメータ θ_k を導入して、BFGS 公式を含むような公式族として Broyden 公式族：

$$H_k = H_{k-1} - \frac{H_{k-1}y_{k-1}y_{k-1}^\top H_{k-1}}{y_{k-1}^\top H_{k-1}y_{k-1}} + \frac{s_{k-1}s_{k-1}^\top}{s_{k-1}^\top y_{k-1}} + \theta_k (y_{k-1}^\top H_{k-1}y_{k-1}) w_{k-1} w_{k-1}^\top \quad (9)$$

が知られている。ただし、

$$w_{k-1} = \frac{s_{k-1}}{s_{k-1}^\top y_{k-1}} - \frac{H_{k-1}y_{k-1}}{y_{k-1}^\top H_{k-1}y_{k-1}}$$

である。Broyden 公式族は $\theta_k = 0$ のときには DFP 公式に一致し、 $\theta_k = 1$ のときには BFGS 公式に一致する。準ニュートン法は適当な仮定のもとで、局所的に超 1 次収束をすることが知られている。また、 $\{B_k\}$ (または $\{H_k\}$) に対してある正の定数 c_1 と c_2 が存在して

$$c_1 \|u\|^2 \leq u^\top B_k u \leq c_2 \|u\|^2, \quad \forall u \in \mathbb{R}^n \quad (10)$$

が成り立つとき、準ニュートン法は大域的収束する。準ニュートン法は、探索方向の計算に行列とベクトルの積を計算する必要があるため、大規模問題に直接適用することが困難であるという弱点をもつ。次節では大規模な問題に適用できるように工夫した準ニュートン法を紹介する。

2.2 メモリーレス準ニュートン法

大規模な問題に対する準ニュートン法として、BB (Barzilai-Borwein) 法 [3] や記憶制限準ニュートン法 [4]、メモリーレス準ニュートン法 [5] が知られている。ここではメモリーレス準ニュートン法を紹介する。準ニュートン法の更新公式において、一つ前の近似行列を単位行列 I に置き換えることで定義される。たとえば $H_{k-1} = I$ とした BFGS 公式 (8) を用いれば、探索方向は

$$d_k = -\nabla f(x_k) + \left(\frac{\nabla f(x_k)^\top y_{k-1}}{s_{k-1}^\top y_{k-1}} - \left(1 + \frac{y_{k-1}^\top y_{k-1}}{s_{k-1}^\top y_{k-1}} \right) \times \frac{\nabla f(x_k)^\top s_{k-1}}{s_{k-1}^\top y_{k-1}} \right) s_{k-1} + \frac{\nabla f(x_k)^\top s_{k-1}}{s_{k-1}^\top y_{k-1}} y_{k-1}$$

と与えられる。上記の探索方向はベクトルの演算のみで計算が可能である。そのため、メモリーレス準ニュートン法は行列を陽に使用しないことから、近年、大規模な最適化問題に対する数値解法として活発に研究が行われている。メモリーレス準ニュートン法に関する研究はわれわれのサーベイ論文 [6] を参照されたい。

3. 近接勾配法

本節では、最急降下法に基づく近接勾配法およびニュートン型近接勾配法について述べる。最後にわれわれが提案したメモリーレス準ニュートン法に基づくニュートン型近接勾配法を紹介する。

3.1 最急降下法に基づく近接勾配法

関数 g が L -平滑、すなわち、勾配 $\nabla g(x)$ がリプシッツ連続である問題 (1) に対して、近接勾配法を用いることができる。近接勾配法は反復式

$$x_{k+1} = \operatorname{argmin}_{x \in \mathbb{R}^n} \left\{ g(x_k) + \nabla g(x_k)^\top (x - x_k) + \frac{1}{2t_k} \|x - x_k\|^2 + h(x) \right\} \quad (11)$$

によって点列を更新する反復法である。ただし、 $t_k > 0$ はパラメータである。上式は目的関数 f において、微分可能な関数 g のみを x_k で 1 次近似したもの (関数 h は近似しない) と近接項 $\frac{1}{2t_k} \|x - x_k\|^2$ の和を最小にすることを意味している。また、式 (11) を変形することで

$$x_{k+1} = \operatorname{argmin}_{x \in \mathbb{R}^n} \left\{ h(x) + \frac{1}{2t_k} \|x - (x_k - t_k \nabla g(x_k))\|^2 \right\} \quad (12)$$

が得られる。ここで、関数 \tilde{h} に対する近接写像：

$$\operatorname{Prox}_{\tilde{h}}(u) \equiv \operatorname{argmin}_{x \in \mathbb{R}^n} \left(\tilde{h}(x) + \frac{1}{2} \|x - u\|^2 \right) \quad (13)$$

を定義することで式 (12) を

$$x_{k+1} = \operatorname{Prox}_{t_k \tilde{h}}(x_k - t_k \nabla g(x_k)) \quad (14)$$

と表すことができる。式からわかるとおり、近接勾配法はステップ幅を t_k とした最急降下法の探索方向 (5) と近接写像 (13) を組み合わせた方法である。関数 h が凸関数である場合に部分問題 (12) が強凸関数の最小化になっており、 x_{k+1} は一意に定まる。さらに、 ℓ_1 ノルムや ℓ_2 ノルムなど多くの場合には近接写像 (14) の解析解が知られており、部分問題 (12) を数値計算で解く必要がない。たとえば、 $h(x) = \lambda \|x\|_1$ とした場合の

³ ここでは B_k と H_k の両方を紹介しているが、実際には H_k を更新するだけでよい。

近接写像 (14) は

$$(\text{Prox}_{t_k h}(u))_i = \begin{cases} (u)_i - t_k \lambda & ((u)_i \geq t_k \lambda) \\ 0 & (-t_k \lambda < (u)_i < t_k \lambda) \\ (u)_i + t_k \lambda & ((u)_i \leq -t_k \lambda) \end{cases}$$

で与えられる. ここで, $(u)_i$ はベクトル u の第 i 成分を表す. また, 関数 h が凸関数であれば $t_k = \frac{1}{L}$ とした近接勾配法 (14) は停留点 x^* に収束することが知られている. この場合の停留点は

$$0 \in \nabla g(x^*) + \partial h(x^*) \quad (15)$$

を満たす点 x^* のことである⁴. ただし, $\partial h(x^*)$ は h の劣微分を表す. 通常, リプシッツ乗数 L は未知のため,

$$g(x_{k+1}) \leq g(x_k) + \nabla g(x_k)^\top (x_{k+1} - x_k) + \frac{1}{2t_k} \|x_{k+1} - x_k\|^2 \quad (16)$$

を満たす t_k を採用することが多い. 関数 g が L -平滑であれば任意の $u, v \in \mathbb{R}^n$ に対して

$$g(u) \leq g(v) + \nabla g(v)^\top (u - v) + \frac{L}{2} \|u - v\|^2$$

が成り立つことから, 式 (16) を満たすまで t_k を小さくしていくことで, t_k を求めることが可能である. たとえば, 正のパラメータ $\eta < 1$ と適当な初期ステップ幅 \bar{t} を用いて, $t_k = \bar{t}\eta^j$ が式 (16) を満たす非負整数 j を求めることでステップ幅 t_k を定めることができる. 初期ステップ幅として BB 法を採用する方法や, 関数値が単調に減少しないことを許す非単調な直線探索により t_k を選択する近接勾配法 [7] や, Nesterov の加速 [8] を加えた近接勾配法など, さまざまな改良がされている.

近年, 機械学習の分野では関数 h として Smoothly Clipped Absolute Deviation (SCAD) [9] や Minimax Concave Penalty (MCP) [10] など非凸な正則化項を使用する研究が注目を集めている. 非凸最適化問題に対する近接勾配法 (14) は $t_k < 1/L$ である場合に停留点 (15) に収束する [11]. 関数が凸の場合と非凸の場合には劣微分の定義が異なるため, h が凸関数である場合の停留点とは少しか概念が異なることに注意する. 近年, われわれ [12] は近接勾配法の停留点への収束の議論を再考することで, 新たな仮定を設けることなく, 近接勾配法が方向停留点, すなわち,

$$f'(x^*; d) \geq 0, \quad \forall d \in \mathbb{R}^n$$

を満たす点 x^* に収束することを示した. 方向停留点は停留点よりも最適性条件として強い概念であることが知られている [13].

3.2 ニュートン型近接勾配法

準ニュートン法に基づく近接勾配法であるニュートン型近接勾配法 [14] を紹介する. まず, 微分可能な関数 g を x_k で 2 次近似したものと微分不可能な関数 h の和の最小化

$$x_k^+ = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ g(x_k) + \nabla g(x_k)^\top (x - x_k) + \frac{1}{2} (x - x_k)^\top B_k (x - x_k) + h(x) \right\} \quad (17)$$

を考える. ただし, B_k は $\nabla^2 g(x_k)$ の正定値対称な近似行列とする. ここで $\|x\|_{B_k} = \sqrt{x^\top B_k x}$, $H_k = B_k^{-1}$ とすれば, 式 (12) と同様に式 (17) を

$$x_k^+ = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ h(x) + \frac{1}{2} \|x - (x_k - H_k \nabla g(x_k))\|_{B_k}^2 \right\} \quad (18)$$

と書き直せる. ニュートン型近接勾配法は, x_k^+ と x_k の凸結合を用いて

$$x_{k+1} = \alpha_k x_k^+ + (1 - \alpha_k) x_k, \quad \alpha_k \in (0, 1]$$

により点列を更新する. この反復式は, 探索方向を

$$d_k = x_k^+ - x_k \quad (19)$$

とした反復式 (2) と一致する. 通常, 近接勾配法 (14) とは異なり, 重み付き近接写像 (18) は数値計算によって求める必要がある. そのため, 近接写像を一度計算して, 探索方向 (19) を定めてから直線探索を行う. ニュートン型近接勾配法は

$$\begin{aligned} & f(x_k + \alpha_k d_k) \\ & \leq f(x_k) + \delta \alpha_k (\nabla g(x_k)^\top d_k + h(x_k^+) - h(x_k)) \end{aligned}$$

を満たすように直線探索を行う. ただし, $\delta \in (0, 1)$ とする. さらに, 重み付き近接写像の計算を軽くするために, 近接写像を非厳密に計算する非厳密ニュートン型近接勾配法 [15] が提案されている. 部分問題 (18) を厳密に解いた場合, 最適性条件

$$0 \in \nabla g(x_k) + B_k (x_k^+ - x_k) + \partial h(x_k^+)$$

が成り立つ. この関係を用いて

⁴ h が微分可能であれば $\nabla f(x^*) = \nabla g(x^*) + \nabla h(x^*)$ であるため, 式 (15) は式 (3) と同値であることに注意する.

$$r_k \in \nabla g(x_k) + B_k(x_k^+ - x_k) + \partial h(x_k^+) \quad (20)$$

となるような勾配残差 $r_k \in \mathbb{R}^n$ が

$$\|r_k\|_{B_k} \leq (1 - \sigma_k) \|x_k^+ - x_k\|_{H_k} \quad (21)$$

を満たす x_k^+ を非厳密な重み付き近接写像として採用する。ただし、 $\bar{\sigma} \in (0, 1]$ を定数とし、 $\sigma_k \in [\bar{\sigma}, 1]$ とする。この条件のもとでは

$$\nabla g(x_k)^\top d_k + h(x_k^+) - h(x_k) \leq -\bar{\sigma} \|d_k\|_{B_k}^2$$

が成り立つことから、 B_k が正定値であれば

$$f'(x_k; d_k) = \lim_{\alpha \rightarrow 0} \frac{f(x_k + \alpha d_k) - f(x_k)}{\alpha} \leq -\bar{\sigma} \|d_k\|_{B_k}^2 < 0$$

となり、降下方向になっていることがわかる。よって、直線探索を行うことが可能である。

関数 g と h が凸関数である場合に、Lee et al. [14] は近接写像を厳密に計算し、 B_k が式 (10) を満たすという仮定のもとで、ニュートン型近接勾配法が大域的収束することを示した。また、近接写像を非厳密に解いた場合でも、適当な仮定のもとで超1次収束することを示した。

3.3 メモリーレス準ニュートン法に基づく非厳密ニュートン型近接勾配法

最後に、われわれが提案したメモリーレス準ニュートン法に基づく非厳密ニュートン型近接勾配法 [16] を紹介する。この方法はわれわれが提案したメモリーレス準ニュートン法 [17] と Li and Fukushima [18] の修正セカント条件を組み合わせた非厳密ニュートン型近接勾配法である。はじめに、式 (7) と同様に $\nabla g(x_{k-1})$ の1次近似と、スケールパラメータ $\gamma_k > 0$ と近似行列の正定値性を保証するための補正パラメータ $\nu_k \geq 0$ を導入して

$$\begin{aligned} & \gamma_k (\nabla g(x_{k-1}) + \nu_k s_{k-1}) \\ & \approx \gamma_k (\nabla g(x_k) - (\nabla^2 g(x_k) + \nu_k I) s_{k-1}) \end{aligned}$$

という近似式を考える。ここで、 $z_{k-1} = \nabla g(x_k) - \nabla g(x_{k-1}) + \nu_k s_{k-1}$ とし、 $\gamma_k (\nabla^2 g(x_k) + \nu_k I)$ の近似行列を B_k とすれば

$$B_k s_{k-1} = \gamma_k z_{k-1} \quad (22)$$

という修正 Spectral Scaling セカント条件を考えることができる。 ν_k は $s_{k-1}^\top z_{k-1} \geq \bar{\nu} \|s_{k-1}\|^2$ を満たすように選び、 $\underline{\gamma} \leq \gamma_k \leq \bar{\gamma}$ とする。ただし $\bar{\nu}, \underline{\gamma}, \bar{\gamma}$ は正の定数とする。われわれ [16] は、式 (22) を満たすメモリーレス Broyden 公式族：

$$B_k = I - \frac{s_{k-1} s_{k-1}^\top}{s_{k-1}^\top s_{k-1}} + \gamma_k \frac{z_{k-1} z_{k-1}^\top}{s_{k-1}^\top z_{k-1}} + \phi_k u_{k-1} u_{k-1}^\top \quad (23)$$

を提案した。ただし

$$u_{k-1} = \sqrt{s_{k-1}^\top s_{k-1}} \left(\frac{z_{k-1}}{s_{k-1}^\top z_{k-1}} - \frac{s_{k-1}}{s_{k-1}^\top s_{k-1}} \right)$$

であり、定数 $\phi_1 \in [0, 1)$ 、 $\phi_2 > 0$ に対して Broyden 公式族のパラメータ ϕ_k を $\phi_1 \phi_k^* \leq \phi_k \leq \phi_2$ の範囲で選べば、行列 (23) は式 (10) を満たすことを示した。ここで、

$$\phi_k^* = -\frac{(s_{k-1}^\top z_{k-1})^2}{(s_{k-1}^\top s_{k-1})(z_{k-1}^\top z_{k-1}) - (s_{k-1}^\top z_{k-1})^2} < 0$$

である。逆行列 H_k は更新式 (9) の y_{k-1} を $\gamma_k z_{k-1}$ で置き換え、 $H_{k-1} = I$ とし、 $\theta_k = \frac{\phi_k^*(1-\phi_k)}{\phi_k - \phi_k^*}$ としたもので与えられる⁵。この行列を用いた非厳密ニュートン型近接勾配法は大域的収束する。

定理 1. 点列 $\{x_k\}$ は行列 (23) を用いた非厳密ニュートン型近接勾配法により生成されるとし、関数 g が L -平滑であり、関数 h が下半連続な凸関数とする。目的関数 f が下に有界であれば $\lim_{k \rightarrow \infty} \|d_k\| = 0$ を満たす。さらに、点列 $\{x_k\}$ が有界であれば任意の集積点は停留点 (15) である。

4. おわりに

本稿では、微分可能な無制約最適化問題に対する数値解法として最急降下法と準ニュートン法について述べた。さらに、大規模な問題に対する準ニュートン法としてメモリーレス準ニュートン法を紹介した。次に、微分不可能な構造をもつ問題 (1) に対しては最急降下法に基づく近接勾配法やニュートン型近接勾配法について述べた。最後に、われわれが提案したメモリーレス準ニュートン法に基づく非厳密ニュートン型近接勾配法を紹介した。微分可能な無制約最適化問題に対しては、最急降下法に比べて準ニュートン法の方が優れた数値解法であることが知られているが、問題 (1) のような微分不可能な点を含む関数に対しては、最急降下法に基づく近接勾配法よりニュートン型近接勾配法の方が良いとは限らない。それは、近接写像 (12) に比べて重み付き近接写像 (18) の方が近接写像を計算す

⁵ 一見、行列を使用しているように見えるが、単位行列とベクトルのみで行列が構成されているため、通常のメモリーレス準ニュートン法と同様に大規模問題へ適用が可能である。

る手間がはるかに大きいためである。近年, Becker et al. [19] は重み付き近接写像 (18) において B_k が特殊な構造をもつ場合に, 低次元の半平滑な方程式を解くことで近接写像を計算する方法を提案している。 B_k として行列 (23) を選んだ場合にもこの計算方法を適用することが可能であるため, より効果的なニュートン型近接勾配法の開発が期待される。

謝辞 本研究の一部は JSPS 科研費若手研究 20K14986, 基盤研究 (C)20K11698 の助成を受けて実施されている。本稿を執筆する機会を下さった理化学研究所の奥野貴之先生, 筑波大学の高野祐一先生にこの場を借りて御礼申し上げます。

参考文献

- [1] J. Nocedal and S. J. Wright, *Numerical Optimization, Springer Series in Operations Research*, 2nd edition, Springer, 2006.
- [2] 矢部博, 『工学基礎 最適化とその応用』, 数理工学社, 2006.
- [3] J. Barzilai and J. M. Borwein, “Two-point step size gradient methods,” *IMA Journal of Numerical Analysis*, **8**, pp. 141–148, 1988.
- [4] J. Nocedal, “Updating quasi-Newton matrices with limited storage,” *Mathematics of Computation*, **35**, pp. 773–782, 1980.
- [5] D. F. Shanno, “Conjugate gradient methods with inexact searches,” *Mathematics of Operations Research*, **3**, pp. 244–256, 1978.
- [6] 成島康史, 中山舜民, 矢部博, “無制約最適化問題に対するメモリーレス準ニュートン法について,” *応用数理*, **29** (4), pp. 8–17, 2020.
- [7] 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 *Proceedings of the 30th International Conference on Machine Learning*, pp. 37–45, 2013.
- [8] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, Springer, 2003.
- [9] J. Fan and R. Li, “Variable selection via nonconcave penalized likelihood and its oracle properties,” *Journal of the American Statistical Association*, **96**, pp. 1348–1360, 2001.
- [10] C. H. Zhang, “Nearly unbiased variable selection under minimax concave penalty,” *The Annals of Statistics*, **38**, pp. 894–842, 2010.
- [11] H. Attouch, J. Bolte and B. F. Svaiter, “Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward-backward splitting, and regularized Gauss–Seidel methods,” *Mathematical Programming*, **137**, pp. 91–129, 2013.
- [12] S. Nakayama and J. Gotoh, “On the superiority of PGMs to PDCAs in nonsmooth nonconvex sparse regression,” *arXiv preprint*, arXiv:2007.01169, 2020.
- [13] J. S. Pang, M. Razaviyayn and A. Alvarado, “Computing B-stationary points of nonsmooth DC programs,” *Mathematics of Operations Research*, **42**, pp. 95–118, 2016.
- [14] 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.
- [15] J. Li, M. S. Andersen and L. Vandenbergh, “Inexact proximal Newton methods for self-concordant functions,” *Mathematical Methods of Operations Research*, **85**, pp. 19–41, 2017.
- [16] 中山舜民, 成島康史, 矢部博, “メモリーレス Broyden 公式族に基づいた非厳密 Newton 型近接勾配法,” *日本オペレーションズ・リサーチ学会春期研究発表会アブストラクト集*, pp. 224–225, 2019.
- [17] S. Nakayama, Y. Narushima and H. Yabe, “Memoryless quasi-Newton methods based on spectral-scaling Broyden family for unconstrained optimization,” *Journal of Industrial and Management Optimization*, **15**, pp. 1773–1793, 2019.
- [18] D. H. Li and M. Fukushima, “A modified BFGS method and its global convergence in nonconvex minimization,” *Journal of Computational and Applied Mathematics*, **129**, pp. 15–35, 2001.
- [19] 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, 2019.