日本オペレーションズ・リサーチ学会
2019年 秋季研究発表会

メモリーレス対称ランクワン法に基づいた ニュートン型近接勾配法の大域的収束性について

05000392　中央大学　　*中山　舜民　NAKAYAMA Shummin
01406032　慶應義塾大学　成島　康史　NARUSHIMA Yasushi

1. はじめに

本稿では，以下の無制約最適化問題

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

に対するニュートン型近接勾配法を考える．ここで，$g:\mathbb{R}^n \to \mathbb{R}$ は連続的微分可能な関数 (凸とは限らない) とし，$h:\mathbb{R}^n \to \mathbb{R}$ は凸関数ではあるが必ずしも微分可能ではないものとする．機械学習においては，g を損失関数，h を正則化項とした問題 (1) がよく現れる．以降，∇g は g の勾配，∂h は h の劣微分，$\|\cdot\|$ は ℓ_2 ノルムとし，正定値対称行列 A による重み付きノルムを $\|x\|_A = \sqrt{x^TAx}$ で表記する．また，凸関数 h に対して正定値対称行列 B を重み行列とした近接写像を

$$\mathrm{Prox}_h^B(\bar{x}) = \operatorname*{argmin}_{x\in\mathbb{R}^n}\left\{h(x) + \frac{1}{2}\|x-\bar{x}\|_B^2\right\} \tag{2}$$

で定義する．本研究で扱うニュートン型近接勾配法の反復式は，任意の初期点 $x_0 \in \mathbb{R}^n$ に対し，以下で与えられる：

$$x_{k+1} = x_k + \eta_k d_k, \tag{3}$$
$$d_k = \mathrm{Prox}_h^{B_k}(x_k - H_k\nabla g(x_k)) - x_k.$$

ここで，$d_k \in \mathbb{R}^n$ は探索方向，$\eta_k \in (0,1]$ はステップ幅と呼ばれる．さらに，B_k は適当な正定値対称行列とし $H_k = B_k^{-1}$ とする．B_k としてヘッセ行列 $\nabla^2 g(x_k)$ そのものや，準ニュートン法の更新公式を用いたニュートン型近接勾配法が提案されている．特に B_k として単位行列 I を選べば，最急降下法に基づいた通常の近接勾配法に対応し，関数 h が ℓ_1 ノルムや標示関数などの単純な正則化関数ならば (2) は閉形式で与えられるため計算が容易になる．一方，B_k が一般的な行列の場合は，近接写像を閉形式で与えることができないため，反復法を用いて近接写像を計算する必要がある．中山ら [3] は近接写像の計算の簡略化のために，メモリーレス Broyden 公式族に基づいて，近接写像を非厳密に求めるニュートン型近接勾配法を提案した．

本講演では，厳密な近接写像を効率的に求めることが可能なメモリーレス対称ランクワン法 [2] に着目した近接勾配法を提案し，その収束性について議論する．

2. メモリーレス対称ランクワン法に基づいたニュートン型近接勾配法

$\gamma_k > 0$ をスケーリングパラメータ，ν_k を正則化パラメータとしたとき，中山ら [3] は行列 B_k に

$$B_k s_{k-1} = \gamma_k y_{k-1} \tag{4}$$

という修正セカント条件に基づいたメモリーレス Broyden 公式族を提案した．ここで，$s_{k-1} = x_k - x_{k-1}$, $y_{k-1} = \nabla g(x_k) - \nabla g(x_{k-1}) + \nu_k s_{k-1}$ とする．メモリーレス Broyden 公式族はランク 2 更新という形で与えられるため，厳密な重み付き近接写像 (2) を与えることが難しい．そのため，中山ら [3] は近接写像を非厳密に求める方法を提案している．一方，$u \in \mathbb{R}^n$ とし，$I \pm uu^T$ という単位行列のランク 1 更新の場合，Becker and Fadili [1, Theorem 7] の定理より，効率的に重み付き近接写像の計算が可能である．

命題 1. 行列 $B = I \pm uu^T$ は正定値対称行列であるとする．このとき，重み付き近接写像は

$$\mathrm{Prox}_h^B(\bar{x}) = \mathrm{Prox}_h^I(\bar{x} \mp \alpha u) \tag{5}$$

で与えられる．ただし，α は

$$u^T\left(\bar{x} - \mathrm{Prox}_h^I(\bar{x} \mp \alpha u)\right) + \alpha = 0 \tag{6}$$

の解である． □

この定理を用いれば，高々1 次元の方程式 (6) を解くことによって，効率的に近接写像が計算できる．さらに，関数 h が ℓ_1 ノルムの場合などといった区分分離可能な関数であれば，厳密な近接写像を求めることが可能である．本研究では，中山ら [2] の提案した行列の正定値性を保持するメモリーレス対称ランクワン公式に着目し，大域的収束性を保証するために，修正セカント条件 (4) に基づいたメモリーレス対称ランクワン公式

$$B_k = I + \frac{(\gamma_k y_{k-1} - s_{k-1})(\gamma_k y_{k-1} - s_{k-1})^T}{s_{k-1}^T(\gamma_k y_{k-1} - s_{k-1})}, \tag{7}$$

$$H_k = I + \frac{(s_{k-1} - \gamma_k y_{k-1})(s_{k-1} - \gamma_k y_{k-1})^T}{\gamma_k y_{k-1}^T(s_{k-1} - \gamma_k y_{k-1})} \tag{8}$$

を考える．この更新式は

$$\gamma_k > \frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}} \text{ or } \gamma_k < \frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}}$$

を満たすように選べば，B_k が正定値対称行列となる．また，

$$u = \frac{\gamma_k y_{k-1} - s_{k-1}}{\sqrt{|s_{k-1}^T(\gamma_k y_{k-1} - s_{k-1})|}}$$

とすれば，(7) の分母 $s_{k-1}^T(\gamma_k y_{k-1} - s_{k-1})$ の符号が正であるときは $B_k = I + uu^T$，符号が負であるときは $B_k = I - uu^T$ と表すことができる．よって，命題 1 を適用することで，重み付き近接写像を効率的に求めることが可能である．

メモリーレス対称ランクワン法に基づいたニュートン型近接勾配法のアルゴリズムは以下の通りである．

アルゴリズム 1

Step 0. 初期点 x_0 と $0 < \delta < 1, 0 < \beta < 1, \epsilon > 0$ を与え，$k := 0$ とする．

Step 1. (7), (8) により B_k, H_k を与え，d_k を計算する．ただし，$k = 0$ のときは $B_k = H_k = I$ とする．

Step 2. $\|d_k\| < \epsilon$ を満たすならば，アルゴリズムを停止して x_k を解とする．

Step 3. バックトラッキングにより Armijo 条件:

$$f(x_k + \beta^i d_k) \le f(x_k) + \delta\beta^i\lambda_k,$$
$$\lambda_k = \nabla g(x_k)^T d_k + h(x_k + d_k) - h(x_k)$$

を満たす最小の非負整数 i を求め，$\eta_k = \beta^i$ とする．

Step 4. (3) によって x_{k+1} を計算する．

Step 5. $k := k + 1$ として Step1 に戻る．

ステップ 1 では行列 B_k や H_k を与えているが，実際には行列を生成するのではなく，$x_k, \nabla g(x_k), s_{k-1}, y_{k-1}$ のベクトル積のみで探索方向 d_k の計算が可能である．

3. 提案手法の収束性

本節ではメモリーレス対称ランワン法に基づいたニュートン型近接勾配法の収束性について議論する．以下の命題は，適当な仮定の下で B_k が一様正定値かつ有界な行列になることを示している．

命題 2. ∇g は Lipschitz 連続とする．ν_k は有界かつある正定数 $\bar{\nu}$ が存在して $s_{k-1}^T y_{k-1} \ge \bar{\nu}\|s_{k-1}\|^2$ を満たすように選ばれ，スケーリングパラメータ γ_k は $\underline{\gamma} \le \gamma_k \le \overline{\gamma}$ を満たし，さらに

$$\gamma_k \ge (1+\rho_1)\frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}} \text{ or } \gamma_k \le (1-\rho_2)\frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}}$$

を満たすように選ぶ．ここで，$\underline{\gamma}, \overline{\gamma}, \rho_1, 0 < \rho_2 < 1$ は正定数とする．このとき，正の定数 m と M が存在して，任意の $v \in \mathbb{R}^n$ に対して以下を満たす:

$$m\|v\|^2 \le v^T B_k v \le M\|v\|^2.$$

□

次に，アルゴリズムの収束性に関する命題を与える．

命題 3. 命題 2 と同じ仮定が成り立つとし，点列 $\{x_k\}$ はアルゴリズム 1 により生成されるとする．このとき，x_k が停留点 (つまり，$0 \in \nabla g(x_k) + \partial h(x_k)$ を満たす点) である必要十分条件は $d_k = 0$ である． □

命題 3 から，終了判定条件として「$\|d_k\| < \epsilon$ を満たせばアルゴリズムを終了する」を考えることができる．さらに，上記の性質を用いることで，下記の大域的収束性の定理を得る．

定理 1. 命題 2 と同じ仮定が成り立つとし，点列 $\{x_k\}$ はアルゴリズム 1 により生成されるとする．このとき，目的関数 f が下に有界であれば

$$\lim_{k\to\infty} \|d_k\| = 0$$

が成立する．さらに，生成される $\{x_k\}$ が有界であれば，その任意の集積点は停留点である． □

さらに，g が強凸関数の場合には，以下の収束速度に関する定理を示すことができる．

定理 2. 命題 2 と同じ仮定が成り立つとし，g が μ-強凸関数であり，h は凸関数とする．アルゴリズム 1 により生成される点列 $\{x_k\}$ は (1) の唯一解 x^* に以下の意味で R-1 次収束する．

$$\|x_k - x^*\| \le \rho^k \sqrt{\frac{2}{\mu}(f(x_0) - f(x^*))}.$$

ここで $\rho \in (0, 1)$ は定数である． □

参考文献

[1] S. Becker and J. Fadili, A quasi-Newton proximal splitting method, In *Advances in Neural Information Processing Systems (NIPS)*, **25** (2012), 2618–2626.

[2] S. Nakayama, Y. Narushima and H. Yabe, A memoryless symmetric rank-one method with sufficient descent property for unconstrained optimization, *Journal of the Operations Research Society of Japan*, **61** (2018), 53–70.

[3] 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近接勾配法”，日本オペレーションズ・リサーチ学会 2019 年春季研究発表会アブストラクト集，224–225 (2019)．