メモリーレスBFGS公式に基づく非厳密ニュートン型近接勾配法における内部反復の改良について

05000392 中央大学 *中山 舜民 NAKAYAMA Shummin
01406032 慶應義塾大学 成島 康史 NARUSHIMA Yasushi
01702330 東京理科大学 矢部 博 YABE Hiroshi

1. はじめに

本稿では，次の最適化問題を扱う：

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

ただし，g は (凸関数とは限らない) 十分滑らかな関数とし，h は凸関数ではあるが必ずしも微分可能ではないとする．以降，∇g は g の勾配，∂h は h の劣微分，$\|\cdot\|$ は ℓ_2 ノルムとする．また，凸関数 h に対して正定値対称行列 A を重み行列とした近接写像を

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

で定義する．ただし，$\|x\|_A \equiv \sqrt{x^TAx}$ は重み付きノルムである．ここで，$A = I$ の時は Prox_h^A は通常の近接写像に一致する．よって，$A = I$ の場合は Prox_h と表記する．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：

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

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 が一般の正定値対称行列の場合は，重み付き近接写像は陽には計算できないため，(2) を部分問題として反復法を用いて計算する必要がある．(2) の右辺は元の問題 (1) と同じ次元の最適化問題であるため，例えば問題が大規模な場合などには部分問題を解くのに非常に手間がかかる．そのため，Nakayama et al. [2] は重み付き近接写像の計算の簡略化のために，重み付き近接写像を非厳密に求めるメモリーレス Broyden 公式族に基づいたニュートン型近接勾配法を提案している．

一方，Becker et al. [1] は (2) において，正定値対称行列 A が $A = D \pm \sum_{i=1}^r u_iu_i^T$ ($D \in \mathbb{R}^{n\times n}$ は対角行列，$u_i \in \mathbb{R}^n$ $(i = 1, \ldots, r)$ とする) で表されるときに，r 次の（微分不可能な）非線形方程式の解を用いて，Prox_h^A の閉形式を与えている．

本発表では，Nakayama et al. のニュートン型近接勾配法において，Prox_h^A を部分問題として解くために，Becker et al. の結果を取り入れたアルゴリズムを考える．

2. メモリーレス準ニュートン法に基づいた非厳密近接勾配法

Nakayama et al. [2] は修正スペクトラルスケーリングセカント条件

$$B_k s_{k-1} = \gamma_k z_{k-1},\ z_{k-1} = y_{k-1} + \nu_k s_{k-1} \tag{4}$$

に基づいたメモリーレス Broyden 公式族を用いたニュートン型近接勾配法を提案している．ここで，$s_k = x_k - x_{k-1}$, $y_{k-1} = \nabla g(x_k) - \nabla g(x_{k-1})$ であり，$\gamma_k > 0$, $\nu_k \geq 0$ はパラメータである．今回は，メモリーレス Broyden 公式族の中でも以下のメモリーレス BFGS 公式を考える[1]：

$$B_k = I - \frac{s_{k-1}s_{k-1}^T}{s_{k-1}^Ts_{k-1}} + \gamma_k\frac{z_{k-1}z_{k-1}^T}{s_{k-1}^Tz_{k-1}}. \tag{5}$$

パラメータ γ_k と ν_k を適切に選択することで，(5) の B_k は一様に正定値，かつ有界であることが保証される．したがって，$\mathrm{Prox}_h^{B_k}$ は定義可能であり，Nakayama et al. [2] はそれに基づいた以下の非厳密ニュートン型近接勾配法を提案している．

[1] $H_k = B_k^{-1}$ も B_k^{-1} を計算するのではなく，更新公式から計算可能だが，紙面の都合上，その更新公式は省略する．

アルゴリズム 1.

Step 0: 初期点 $x_0 \in \mathbb{R}^n$ を与え，$\bar{\theta} \in (0,1]$, $k=0$ とする．

Step 1: 終了条件を満たすならばアルゴリズムを停止して x_k を最適解とする．

Step 2: (5) により B_k(と H_k) を求める．

Step 3: $\theta_k \in [\bar{\theta}, 1]$ を与え，

$$\|r_k\|_{H_k} \leq (1-\theta_k)\|x_k^+ - x_k\|_{B_k}, \quad (6)$$

$$r_k \in \nabla g(x_k) + B_k(x_k^+ - x_k) + \partial h(x_k^+) \quad (7)$$

を満たす $x_k^+ \approx \mathrm{Prox}_h^{B_k}(x_k - H_k \nabla g(x_k))$ を求め，探索方向を $d_k = x_k^+ - x_k$ とする．

Step 4: 直線探索によりステップ幅 $\eta_k > 0$ を求め，(3) によって点 x_{k+1} を更新する．

Step 5: $k = k+1$ として Step1 に戻る． □

ここで，(6), (7) の r_k は残差ベクトルであり，$r_k = 0$ の時は (7) は (2) の最適性条件に一致する．したがって，(6) より，$\theta_k = 1$ の時は重み付き近接写像を厳密に計算していることになる．

3. 部分問題に対する一般化ニュートン法

この節ではアルゴリズム 1 の k 回目の反復における Step 3. を考え，以降では $\bar{x} = x_k - H_k \nabla g(x_k)$,

$$u_1 = \sqrt{\frac{\gamma_k}{s_{k-1}^T z_{k-1}}} z_{k-1}, \quad u_2 = \frac{s_{k-1}}{\|s_{k-1}\|}$$

と置くことにする[2]．このとき，[1, 定理 3.4] を用いることで，直ちに以下の定理が得られる．

定理 1. $\alpha = (\alpha_1, \alpha_2)^T \in \mathbb{R}^2$ に対して，$\xi(\alpha) = \bar{x} + \alpha_1 u_1 - \alpha_2 (I + u_1 u_1^T)^{-1} u_2$ とし，α^* を $\mathcal{L}(\alpha) = 0$ の唯一解であるとする[3]．このとき，

$$\mathrm{Prox}_h^{B_k}(\bar{x}) = \mathrm{Prox}_h(\xi(\alpha^*)) \quad (8)$$

が成立する． □

上述したように，Prox_h は容易に計算できることが多い．したがって，(8) より，2 次元の非線形方程式系 $\mathcal{L}(\alpha) = 0$ を解くことで $\mathrm{Prox}_h^{B_k}$ が計算できるため，(2) を直接反復法で解くよりも時間効率性の改善が期待できる．$\mathcal{L}$ には Prox_h が含まれているため，微分可能ではないが，リプシッツ連続であり，Prox_h が半平滑の場合には $\mathcal{L}$ も半平滑となる[4]．したがって，$\mathcal{L}(\alpha) = 0$ を解くために，以下のような一般化ニュートン法を考えることができる．

アルゴリズム 2.

Step 3.0: 初期点 $\alpha_0 \in \mathbb{R}^2$ を与え，$j = 0$ とする．

Step 3.1: 終了条件を満たすならばアルゴリズムは停止して α_j を最適解とする．

Step 3.2: $V_j \in \partial \mathcal{L}(\alpha_j)$ を選び，ニュートン方程式 $\mathcal{L}(\alpha_j) + V_j p_j = 0$ を解いて探索方向 p_j を求める[5]．

Step 3.3: $\alpha_{j+1} = \alpha_j + p_j$ によって，点を更新し，$j = j+1$ として，Step 3.1 に戻る． □

Step 3.1 の終了判定では，外部反復の Step 3. の条件 (6), (7) を満たすような条件が必要となる．そのために以下の命題を与える．

命題 2. $U = [-u_1, u_2] \in \mathbb{R}^{n \times 2}$ とする．このとき，任意の α に対して，$U\mathcal{L}(\alpha) \in \nabla g(x_k) + B_k(\mathrm{Prox}_h(\xi(\alpha)) - x_k) + \partial h(\mathrm{Prox}_h(\xi(\alpha)))$ が成り立つ． □

命題 2 より，j 回目の内部反復において

$$x_k^{+(j)} = \mathrm{Prox}_h(\xi(\alpha_j)), \quad r_k^{(j)} = U\mathcal{L}(\alpha_j)$$

と置けば，Step 3.1 の終了判定条件として (6) を用いることができる．

最後にアルゴリズム 2 の局所的な収束率に関して以下の定理を与える．

定理 3. Prox_h が半平滑であると仮定する．解 α^* のある近傍が存在して，初期点 α_0 がその近傍内から選択されるとき，アルゴリズム 2 によって生成される点列 $\{\alpha_j\}$ は α^* に超一次収束する．さらに，Prox_h が強半平滑ならば，$\{\alpha_j\}$ は α^* に二次収束する． □

参考文献

[1] S. Becker, J. Fadili and P. Ochs, On quasi-Newton forward-backward splitting: proximal calculus and convergence, *SIAM Journal on Optimization*, **29** (2019), 2445–2481.

[2] S. Nakayama, Y. Narushima and H. Yabe Inexact proximal memoryless quasi-Newton methods based on the Broyden family for minimizing composite functions, submitted.

[2] (4) のパラメータ $\nu_k \geq 0$ をうまく選ぶことによって $s_{k-1}^T z_{k-1} > 0$ が保証できることを注意しておく．

[3] $\mathcal{L} : \mathbb{R}^2 \to \mathbb{R}^2$ は以下で定義される:

$$\mathcal{L}(\alpha) = \begin{pmatrix} u_1^T(\bar{x} - \alpha_2(I + u_1 u_1^T)^{-1} u_2 - \mathrm{Prox}_h(\xi(\alpha))) - \alpha_1 \\ u_2^T(\bar{x} - \mathrm{Prox}_h(\xi(\alpha))) - \alpha_2 \end{pmatrix}.$$

[4] 一般的に Prox_h は必ずしも半平滑ではないことが知られている．しかしながら，実用上の問題における h は Prox_h が半平滑（または強半平滑）となることが多い．

[5] $\partial \mathcal{L}(\alpha)$ は α における $\mathcal{L}$ の Clarke 劣微分とする．