リーマン多様体上の最適化問題に対するブロイデン公式族に基づくメモリーレス準ニュートン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
（株）ラティス・テクノロジー 竹村 壮史 TAKEMURA Masashi
01702330 東京理科大学 矢部 博 YABE Hiroshi

1. はじめに

本発表では，リーマン多様体上の最適化問題:

$$\min_{x\in\mathcal{M}} \quad f(x), \quad f:\mathcal{M}\to\mathbb{R}$$

に対する数値解法を考える．ただし，$\mathcal{M}$ は n 次元のリーマン多様体とする．リーマン多様体上の最適化問題は多くの応用があり，特に機械学習の分野で生じる問題の多くがリーマン多様体上の最適化問題として記述できるため，近年，注目を集めている [4].

リーマン多様体上の最適化問題に対する数値解法の多くは通常のユークリッド空間上の無制約最適化問題に対する反復法を拡張したものであり，最急降下法，準ニュートン法や非線形共役勾配法など多くの方法がリーマン多様体上の最適化問題に拡張されている．一方，通常のユークリッド空間上の最適化問題に対して，近年，メモリーレス準ニュートン法が注目されている．メモリーレス準ニュートン法は非線形共役勾配法と同様に行列を陽に使用しない方法であり，大規模問題に有効な数値解法である．しかしながら，少なくとも著者らの知る限り，リーマン多様体上の最適化問題に対するメモリーレス準ニュートン法は提案されていない．したがって，本発表ではリーマン多様体上の最適化問題に対するメモリーレス準ニュートン法を考える．具体的には，Nakayama et al. [3] のメモリーレスブロイデン公式族にある種の正定値性を保証するような修正を加えた更新公式を考え，それに基づくアルゴリズムを提案し，さらに大域的収束性を議論する．

以降では，$T_x\mathcal{M}$ を点 $x\in\mathcal{M}$ における $\mathcal{M}$ の接空間とし，$T\mathcal{M}$ を $\mathcal{M}$ 上のすべての接ベクトル全体の集合であるとする．また，$\langle\cdot,\cdot\rangle_x : T_x\mathcal{M}\times T_x\mathcal{M}\to\mathbb{R}$ を $T_x\mathcal{M}$ の内積とし，$T_x\mathcal{M}$ の内積から誘導されるノルムを $\|\xi\|_x=\sqrt{\langle\xi,\xi\rangle_x}$ (for $\xi\in T_x\mathcal{M}$) とする．さらに，関数 $\theta:\mathcal{M}\to\mathbb{R}$ の点 $x\in\mathcal{M}$ における $\xi\in T_x\mathcal{M}$ 方向の方向微分を $D\theta(x)[\xi]$ で表す．同じく，関数 $\theta:\mathcal{M}\to\mathbb{R}$ の点 $x\in\mathcal{M}$ における勾配とヘッセ作用素をそれぞれ $\operatorname{grad}\theta(x)$ と $\operatorname{Hesse}\theta(x)$ で表すこととする．その他，以降で登場するレトラクションやベクトル輸送に関する詳細な定義は，例えば [4] などを参照されたい．

2. アルゴリズム

リーマン多様体上の反復法は，初期点 $x_0\in\mathcal{M}$ からスタートし，反復式：

$$x_{k+1}=R_{x_k}(\alpha_k\eta_k) \tag{1}$$

によって点列を更新する．ここで，$x_k\in\mathcal{M}$ は k 回目の近似解，$\alpha_k>0$ をステップ幅，$\eta_k\in T_{x_k}\mathcal{M}$ を探索方向であるとする．さらに，R_{x_k} はリーマン多様体 $\mathcal{M}$ 上のレトラクション $R:T\mathcal{M}\to\mathcal{M}$ の $T_{x_k}\mathcal{M}$ への制限である．なお，以降では $g_k=\operatorname{grad} f(x_k)\in T_{x_k}\mathcal{M}$ と表記することとする．このとき，リーマン多様体上の準ニュートン法の探索方向は

$$\eta_k=-\mathcal{H}_k g_k \tag{2}$$

によって与えられる．ただし，$\mathcal{H}_k : T_{x_k}\mathcal{M}\to T_{x_k}\mathcal{M}$ はヘッセ作用素の逆作用素 $\operatorname{Hesse} f(x_k)^{-1}$ の近似である．今回，リーマン多様体上のメモリーレスブロイデン公式族を提案するために，まずは，Nakayama et al. [3] と同様に，Chen and Cheng [1] のスペクトラルスケーリングセカント条件 (以下，SS セカント条件) に基づくブロイデン公式族を考え，それをリーマン多様体上に拡張すると，

$$\begin{aligned}\mathcal{H}_{k+1}=\tilde{\mathcal{H}}_k-\frac{\tilde{\mathcal{H}}_k y_k(\tilde{\mathcal{H}}_k y_k)^\flat}{(\tilde{\mathcal{H}}_k y_k)^\flat y_k}+\frac{1}{\tau_k}\frac{s_k s_k^\flat}{s_k^\flat y_k}\\ +\phi_k(\tilde{\mathcal{H}}_{k-1}y_k)^\flat y_k w_k w_k^\flat,\end{aligned} \tag{3}$$

$$w_k=\frac{s_k}{s_k^\flat y_k}-\frac{\tilde{\mathcal{H}}_k y_k}{(\tilde{\mathcal{H}}_k y_k)^\flat y_k}$$

となる．ただし，$a^\flat : T_{x_{k+1}}\mathcal{M}\to\mathbb{R}$ は任意の $v\in T_{x_{k+1}}\mathcal{M}$ に対して $a^\flat v\equiv\langle a,v\rangle_{x_{k+1}}$ によって定義されるものとする．さらに，$\tilde{\mathcal{H}}_k=\mathcal{T}_{\alpha_k\eta_k}\circ\mathcal{H}_k\circ(\mathcal{T}_{\alpha_k\eta_k})^{-1}$,

$$s_k=\mathcal{T}_{\alpha_k\eta_k}(\alpha_k\eta_k), \quad y_k=g_{k+1}-\mathcal{T}_{\alpha_k\eta_k}(g_k)$$

とし，写像 $\mathcal{T}:T\mathcal{M}\oplus T\mathcal{M}\to T\mathcal{M} : (\eta,\xi)\mapsto\mathcal{T}_\eta(\xi)$ をレトラクション R に対応するベクトル輸送であるとする．また，$\tau_k\in[\tau_{\min},\tau_{\max}]$ $(0<\tau_{\min}\le\tau_{\max})$ は SS セカント条件に関するスケーリングパラメータであり，$\phi_k\in[0,\bar{\phi}]$ $(0\le\bar{\phi})$ はブロイデン公式族のパラメータである．なお，(3) は $\phi_k=0$ の時には DFP 公式に，$\phi_k=1$ の時には BFGS 公式に対応する．今回，近似作

用素 $\mathcal{H}_k$ のリーマン多様体における正定値性に相当する性質を担保するとともに，ベクトル輸送に関する条件を緩和するために，y_k に正則化項を加えた

$$z_k = y_k + \nu_k s_k$$

を y_k の代わりに用いる．ただし，パラメータ $\nu_k \in [0, \bar{\nu}]$ $(0 < \bar{\nu})$ は，ある正の定数 $\underline{\nu} > 0$ が存在して，$s_k^\flat z_k \geq \underline{\nu} \|s_k\|^2_{x_{k+1}}$ を満たすように選ばれれるものとする．このとき，(3) において，y_k を z_k で，$\mathcal{H}_k$ を $\gamma_k \mathrm{id}_{T_{x_{k+1}}\mathcal{M}}$ で置き換えれば，修正SSセカント条件に基づくメモリーレスブロイデン公式族:

$$\begin{aligned}\mathcal{H}_{k+1} = \gamma_k \mathrm{id}_{T_{x_{k+1}}\mathcal{M}} - \gamma_k \frac{z_k z_k^\flat}{z_k^\flat z_k} + \frac{1}{\tau_k} \frac{s_k s_k^\flat}{s_k^\flat z_k} \\ + \phi_k \gamma_k z_k^\flat z_k w_k w_k^\flat,\end{aligned} \tag{4}$$

$$w_k = \frac{s_k}{s_k^\flat z_k} - \frac{z_k}{z_k^\flat z_k}$$

が得られる．なお，$\gamma_k \in [\gamma_{\min}, \gamma_{\max}]$ $(0 < \gamma_{\min} \leq \gamma_{\max})$ はサイジングパラメータであり，$\mathrm{id}_{T_{x_{k+1}}\mathcal{M}}$ を $T_{x_{k+1}}\mathcal{M}$ 上の単位元とする．なお，実際に (4) を用いる場合，内積の計算のみで探索方向 (2) の計算が可能であることを注意しておく．

提案アルゴリズムでは，ステップ幅 α_k にウルフ条件：

$$f(R_{x_k}(\alpha_k \eta_k)) \leq f(x_k) + \delta_1 \alpha_k \langle g_k, \eta_k \rangle_{x_k}, \tag{5}$$

$$\begin{aligned}\langle \mathrm{grad} f(R_{x_k}(\alpha_k \eta_k)), \mathrm{D}R_{x_k}(\alpha_k \eta_k)[\eta_k] \rangle_{R_{x_k}(\alpha_k \eta_k)} \\ \geq \delta_2 \langle g_k, \eta_k \rangle_{x_k}\end{aligned} \tag{6}$$

を課すこととする．ただし，δ_1, δ_2 は $0 < \delta_1 < \delta_2 < 1$ を満たす定数である．

以上より，メモリーレスブロイデン公式族に基づく以下のアルゴリズムを提案する．

Algorithm 1 Memoryless RBroyden family

初期設定：初期点 $x_0 \in \mathcal{M}$，初期近似作用素 $\mathcal{H}_0 : T_{x_0}\mathcal{M} \to T_{x_0}\mathcal{M}$，$0 < \delta_1 < \delta_2 < 1$ を与える．

```
k ← 0.
while ‖g_k‖_{x_k} > 0 do
    (2) により探索方向 η_k を計算する．
    ウルフ条件 (5)–(6) を満たすステップ幅 α_k を計算する．
    (1) により x_{k+1} を計算する．
    (4) により H_{k+1} を計算する．
    k ← k + 1 とする．
end while
```

3. 大域的収束性

以降では，初期点における準位集合を

$$\mathcal{L} = \{x \in \mathcal{M} \mid f(x) \leq f(x_0)\}.$$

で表す．Algorithm 1 の大域的収束性を保証するために以下の仮定を設ける．

Assumption 1. 目的関数 f は連続微分可能かつ下に有界であるとする．さらに，以下が成り立つ．

1. 正の定数 L_1 が存在して，すべての $x \in \mathcal{L}$, $t \geq 0$ と $\|\eta\|_x = 1$ であるような $\eta \in T_x\mathcal{M}$ に対して以下が成り立つ：
$$|\mathrm{D}(f \circ R_x)(t\eta)[\eta] - \mathrm{D}(f \circ R_x)(0)[\eta]| \leq tL_1.$$
2. 正の定数 L_2 が存在して，すべての $x \in \mathcal{L}$ と $R_x(\eta) \in \mathcal{L}$ であるような η に対して以下が成り立つ：
$$\begin{aligned}\|\mathrm{grad} f(R_x(\eta)) - \mathcal{T}_\eta(\mathrm{grad} f(x))\|_{R_x(\eta)} \\ \leq L_2 \|\mathcal{T}_\eta(\eta)\|_{R_x(\eta)}.\end{aligned} \tag{7}$$

Assumption 1-1 はウルフ条件の下でズーテンダイク条件が成り立つための基本的な仮定である．一方，Assumption 1-2 は [2] において，レトラクションリプシッツ連続と呼ばれる性質に対応した条件である．レトラクションリプシッツ連続は (7) の右辺が $\|\eta\|_x$ で定義されているが，[2] ではさらに $\|\mathcal{T}_\eta(\eta)\|_{R_x(\eta)} = \|\eta\|_x$ となるような仮定も課されており，その仮定の下では (7) はレトラクションリプシッツ連続の定義に一致する．

Assumption 1 の下で，以下の大域的収束性の定理が得られる．

Theorem 3.1. Assumption 1 が成り立つとし，$\{x_k\}$ を Algorithm 1 によって生成される点列であるとする．このとき，正の定数 m, M が存在し，すべての $k \geq 0$ に対して

$$m\|\xi\|^2_{x_k} \leq \langle \mathcal{H}_k \xi, \xi \rangle_{x_k} \leq M\|\xi\|^2_{x_k}, \quad \text{for } {}^\forall \xi \in T_{x_k}\mathcal{M}$$

が成り立つ．さらに，Algorithm 1 はある k 回目の反復において $g_k = 0$ となり終了するか，$\lim_{k\to\infty} \|g_k\|_{x_k} = 0$ の意味で大域的に収束する．

4. 数値実験

数値実験の結果は当日発表する．

参考文献

[1] Z. Chen and W. Cheng, Spectral-scaling quasi-Newton methods with updates from the one parameter of the Broyden family, *Journal of Computational and Applied Mathematics*, **248** (2013), 88–98.

[2] W. Huang, P.-A. Absil, and K. A. Gallivan, A Riemannian BFGS method without differentiated retraction for nonconvex optimization problems, *SIAM Journal on Optimization*, **28** (2018), 470–495.

[3] 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** (2019), 1773–1793.

[4] H. Sato, Riemannian Optimization and Its Applications, Springer, (2021).