修士論文 要旨

Sequential Quadratic Optimization for Nonlinear Optimization Problems on Riemannian Manifolds (リーマン多様体上の非線形最適化問題に対する逐次２次最適化法)

数理情報学専攻 48196210　小原 光暁

指導教員　武田 朗子 教授

1　はじめに

本研究では，リーマン多様体上の非線形最適化問題 (Riemannian nonlinear optimization problem; RNLO)

$$\begin{aligned}\underset{x\in\mathcal{M}}{\text{minimize}} \quad & f(x)\\ \text{subject to} \quad & g_i(x)\le 0, \quad (i\in\mathcal{I}),\\ & h_j(x)=0, \quad (j\in\mathcal{E})\end{aligned}$$

を考える．ただし $\mathcal{M}$ は d 次元リーマン多様体とする．リーマン多様体とはリーマン計量を備えた C^∞ 級可微分多様体である．点 x での接空間 $T_x\mathcal{M}$ 上の内積 $\langle\cdot,\cdot\rangle_x : T_x\mathcal{M}\times T_x\mathcal{M}\to\mathbb{R}$ を各点 x に与える，滑らかな写像 $\langle\cdot,\cdot\rangle : x\mapsto\langle\cdot,\cdot\rangle_x$ をリーマン計量という．$\mathcal{I}=\{1,\ldots,m\},\mathcal{E}=\{1,\ldots,n\}$ とする．実数値関数 $f,\{g_i\}_{i\in\mathcal{I}},\{h_j\}_{j\in\mathcal{E}}:\mathcal{M}\to\mathbb{R}$ を C^1 級とする．

RNLO の応用例として，非負値主成分分析や人型ロボットの姿勢計算などが挙げられる．RNLO の先行研究として，Yang ら [3] による Karush-Kuhn-Tucker (KKT) 条件 の導出や, Liu-Boumal [2] による拡張ラグランジュ法と正確なペナルティ法の提案が挙げられる．

本研究では，ユークリッド空間上の非線形最適化問題に対するアルゴリズムとしてよく知られる逐次 2 次最適化法を，RNLO に対して拡張した．さらに提案手法の大域的収束性と局所的収束性を証明し，数値実験を通じて提案手法による高精度な求解を確認した．

2　準備

$\{(\mathcal{U}_\alpha,\varphi_\alpha)\}_{\alpha\in\Lambda}$ を $\mathcal{M}$ の座標近傍系とする．各点 x に $R_x:\mathrm{T}_x\mathcal{M}\to\mathcal{M}$ を与える滑らかな写像 $R:x\mapsto R_x$ をレトラクションという．ただし R_x は，局所的に少なくとも 1 次近似の精度で，点 x での指数写像と一致する滑らかな写像である．

$$\langle \mathrm{grad} f(x),\xi\rangle_x=\mathrm{D}f(x)[\xi], \quad (\xi\in T_x\mathcal{M})$$

を満たす唯一の $\mathrm{grad} f(x)\in T_x\mathcal{M}$ を，f の点 x での勾配と定義する．ただし $\mathrm{D}f(x)[\xi]$ は f の点 x における ξ に沿った方向微分である．点 x でのヘシアン $\mathrm{Hess} f(x):T_x\mathcal{M}\to T_x\mathcal{M}$ を，

$$\mathrm{Hess} f(x)[\xi]\coloneqq\nabla_\xi \mathrm{grad} f(x)$$

とする．ただし ∇ は Levi-Civita 接続である．

ラグランジュ乗数 $\mu\in\mathbb{R}^m,\lambda\in\mathbb{R}^n$ を用いてラグランジュ関数 $\mathcal{L}_{\mu,\lambda}(x)$ を次式で定義する：

$$\mathcal{L}_{\mu,\lambda}(x)\coloneqq f(x)+\sum_{i\in\mathcal{I}}\mu_i g_i(x)+\sum_{j\in\mathcal{E}}\lambda_j h_j(x).$$

実行可能点 x で 1 次独立制約想定が成立するとは，

$$\{\mathrm{grad} g_i(x),\mathrm{grad} h_j(x), i\in\mathcal{I}_0(x), j\in\mathcal{E}\}$$

が 1 次独立であることと定義する．ただし $\mathcal{I}_0(x)\coloneqq\{i\mid i\in\mathcal{I}, g_i(x)=0\}$ は点 x における有効不等式制約の添字集合である．点 x^* において，

$$\begin{aligned}&\mathrm{grad}\mathcal{L}_{\mu^*,\lambda^*}(x^*)=0,\\ &g_i(x^*)\le 0,\ \mu_i^* g_i(x^*)=0, \quad (i\in\mathcal{I}),\\ &h_j(x^*)=0, \quad (j\in\mathcal{E})\end{aligned}$$

を満たす $\mu^*\in\mathbb{R}^m_{\ge 0},\lambda^*\in\mathbb{R}^n$ が存在するとき，(x^*,μ^*,λ^*) は KKT 条件を満たすという [3]．また KKT 条件の成立を前提として，任意の $i\in\mathcal{I}_0(x^*)$ について $\mu_i^*>0$ が成立するとき，点 x^* で狭義相補性が成立するという．さらに $\mathcal{F}(x^*)$ を

$$\xi\in\mathcal{F}(x^*) \Leftrightarrow \begin{cases}\xi\in T_{x^*}\mathcal{M},\\ \langle\xi,\mathrm{grad} h_j(x^*)\rangle_{x^*}=0,\ (j\in\mathcal{E}),\\ \langle\xi,\mathrm{grad} g_i(x^*)\rangle_{x^*}=0,\ \begin{pmatrix} i\in\mathcal{I}_0(x^*)\\ \text{with } \mu_i^*>0\end{pmatrix},\\ \langle\xi,\mathrm{grad} g_i(x^*)\rangle_{x^*}\le 0,\ \begin{pmatrix} i\in\mathcal{I}_0(x^*)\\ \text{with } \mu_i^*=0\end{pmatrix}\end{cases}$$

より定義する．KKT 条件の成立を前提として，

$$\langle \mathrm{Hess}\mathcal{L}_{\mu^*,\lambda^*}(x^*)[\xi],\xi\rangle_{x^*}>0, \quad (\xi\in\mathcal{F}(x^*)\setminus\{0\})$$

が成立するとき，点 x^* は 2 次の十分条件を満たすという．リーマン多様体上の最適化理論に関する詳細は [1, 3] を参照されたい．

Algorithm 1 リーマン多様体上の逐次 2 次最適化法

Require: リーマン多様体 $\mathcal{M}$; リーマン計量 $\langle\cdot,\cdot\rangle$; メリット関数 $P_\rho : \mathcal{M} \to \mathbb{R}$; レトラクション R; C^1 級関数 $f, \{g_i\}_{i\in\mathcal{I}}, \{h_j\}_{j\in\mathcal{E}} : \mathcal{M} \to \mathbb{R}$; $\tau > 0, \rho_{-1} > 0, \beta \in (0,1), \gamma \in (0,1)$.

Input: 初期点 $x_0 \in \mathcal{M}$; $B_0 : T_{x_0}\mathcal{M} \to T_{x_0}\mathcal{M}$.

for $k = 0, 1, \ldots$ **do**

部分問題 (QO) の最適解 Δx_k^* と対応するラグランジュ乗数 μ_k^*, λ_k^* を計算する.

$\nu = \max\left\{\max_{i\in\mathcal{I}} \mu_{ki}^*, \max_{j\in\mathcal{E}} \left|\lambda_{kj}^*\right|\right\}$.

$$\rho_k = \begin{cases} \rho_{k-1}, & (\text{if } \rho_{k-1} \geq \nu), \\ \nu + \tau, & (\text{otherwise}). \end{cases}$$

$\alpha_k = \beta^r$. ただし r は

$$P_{\rho_k}(x_k) - P_{\rho_k}\left(R_{x_k}\left(\beta^r \Delta x_k^*\right)\right) \geq \gamma\beta^r \langle B_k\left[\Delta x_k^*\right], \Delta x_k^*\rangle_{x_k}$$

を満たす最小非負整数.

$x_{k+1} = R_{x_k}(\alpha_k \Delta x_k^*)$, $\mu_{k+1} = \mu_k^*$, $\lambda_{k+1} = \lambda_k^*$.

$B_{k+1} : T_{x_{k+1}}\mathcal{M} \to T_{x_{k+1}}\mathcal{M}$ を計算する.

end for

3 リーマン多様体上の逐次 2 次最適化法

逐次 2 次最適化法は反復法の一種である．各反復で部分問題を解いて探索方向を求め，その方向に進むことを繰り返して RNLO の解を得る．部分問題は各反復の現在の点でのラグランジュ関数の 2 次のモデルと制約条件の 1 次近似により得られる，接空間上の 2 次最適化問題である．具体的には $B_k : T_{x_k}\mathcal{M} \to T_{x_k}\mathcal{M}$ を正定値対称な線形写像として，部分問題 (QO) を以下のように定義する：

$$\begin{aligned} \underset{\Delta x_k \in T_{x_k}\mathcal{M}}{\text{minimize}} \quad & \frac{1}{2}\langle B_k\left[\Delta x_k\right], \Delta x_k\rangle_{x_k} + \langle \operatorname{grad} f(x_k), \Delta x_k\rangle_{x_k} \\ \text{subject to} \quad & g_i(x_k) + \langle \operatorname{grad} g_i(x_k), \Delta x_k\rangle_{x_k} \leq 0, \quad (i \in \mathcal{I}), \\ & h_j(x_k) + \langle \operatorname{grad} h_j(x_k), \Delta x_k\rangle_{x_k} = 0, \quad (j \in \mathcal{E}). \end{aligned}$$

提案手法を Algorithm 1 に示す.

4 収束解析

4.1 大域的収束性

本研究では ℓ_1 ペナルティ関数 $P_\rho : \mathcal{M} \to \mathbb{R}$,

$$P_\rho(x) \coloneqq f(x) + \rho\left(\sum_{i\in\mathcal{I}} \max\{0, g_i(x)\} + \sum_{j\in\mathcal{E}} |h_j(x)|\right)$$

をメリット関数として用いる.

定理 1. Algorithm 1 において，任意の k で (QO) は実行可能かつ生成された点列 $\{(x_k, \mu_k, \lambda_k)\}_{k=0}^\infty$ は有界であると仮定する．さらに任意の $k, \xi \in T_{x_k}\mathcal{M}$ に対して，

$$m\langle\xi,\xi\rangle_{x_k} < \langle B_k[\xi], \xi\rangle_{x_k} < M\langle\xi,\xi\rangle_{x_k}$$

を満たす $m > 0, M > 0$ が存在すると仮定する．このとき，生成された点列の任意の集積点 (x^*, μ^*, λ^*) は RNLO の KKT 条件を満たす.

4.2 局所的収束性

RNLO の解に十分近い初期点から出発した際の局所的収束性を示す．$\eta_k \coloneqq (\mu_k, \lambda_k) \in \mathbb{R}^{m+n}$ と表記する．点列 $\{(x_k, \eta_k)\}_{k=0}^\infty$ が点 (x^*, η^*) に 2 次収束するとは，ある $c \geq 0, K \geq 0$ が存在して，任意の $k \geq K$ に対して

$$\left\|\begin{pmatrix}\varphi(x_{k+1}) - \varphi(x^*) \\ \eta_{k+1} - \eta^*\end{pmatrix}\right\| \leq c\left\|\begin{pmatrix}\varphi(x_k) - \varphi(x^*) \\ \eta_k - \eta^*\end{pmatrix}\right\|^2$$

が成立することと定義する．このとき，2 次収束の成立は φ の選択に依存しない．つまり，c の値は φ に依存するが，不等式を満たす c, K が存在するか否かは φ によらず決定する.

定理 2. 定理 1 における仮定が成立し，Algotirhm 1 の生成する点列 $\{(x_k, \eta_k)\}_{k=0}^\infty$ が点 (x^*, η^*) に集積するとする．さらに $f, \{g_i\}_{i\in\mathcal{I}(x^*)}, \{h_j\}_{j\in\mathcal{E}}$ は C^3 級であり，点 (x^*, η^*) において 1 次独立制約想定・狭義相補性・2 次の十分条件が成立し，ある K_0 が存在して任意の $k \geq K_0$ について $B_k = \operatorname{Hess}\mathcal{L}_{\eta_k}(x_k)$ かつ $\alpha_k = 1$ が成立すると仮定する．このとき，生成された点列 $\{(x_k, \eta_k)\}_{k=0}^\infty$ は点 (x^*, η^*) に 2 次収束する.

5 数値実験

提案手法および既存手法 [2] を線形等式制約付き oblique 多様体上の最適化問題に適用した．その結果，既存手法は KKT 条件の残差が 10^{-6} 程度の解を求めた一方，提案手法は残差が 10^{-12} を下回る求解が出来た．これより提案手法は高精度な求解が可能であると分かる．同様に非負制約付き fixed-rank 多様体上の最適化問題に対して実験を行った．その結果，次元が大きくなるにつれて提案手法の計算時間が急増したものの，提案手法は既存手法に比べて高精度な求解を行った.

参考文献

[1] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds.* Princeton University Press, New York, 2008.

[2] C. Liu and N. Boumal. Simple algorithms for optimization on Riemannian manifolds with constraints. *Appl. Math. Optim.*, pages 1–33, 2019.

[3] W. H. Yang, L.-H. Zhang, and R. Song. Optimality conditions for the nonlinear programming problems on Riemannian manifolds. *Pac. J. Optim.*, 10(2):415–434, 2014.