日本オペレーションズ・リサーチ学会
2022年 秋季研究発表会

# 一般化 Cayley 変換による動的パラメータ表現法に基づく直交制約付き最適化アルゴリズムのための統一的収束解析

05001596　東京工業大学工学院情報通信系　*久米啓太　KUME Keita
東京工業大学工学院情報通信系　山田功　YAMADA Isao

## 1. はじめに

Stiefel 多様体 $\mathrm{St}(p,N) := \{\boldsymbol{U} \in \mathbb{R}^{N\times p} \mid \boldsymbol{U}^{\mathrm{T}}\boldsymbol{U} = \boldsymbol{I}_p\}$ 上の点 $\boldsymbol{U}$ は低次元部分空間の正規直交基底を与える．

**問題 1.1** ($\mathrm{St}(p,N)$ 上最適化問題)**.** 連続関数 $f : \mathbb{R}^{N\times p} \to \mathbb{R}$ に対して[1]

$$\text{find } \boldsymbol{U}^\star \in \mathrm{St}(p,N) \text{ s.t. } f(\boldsymbol{U}^\star) \leqslant f(\boldsymbol{U})\ (\forall \boldsymbol{U} \in \mathrm{St}(p,N))$$

は特別な性質を持つ $p$ 次元部分空間とその正規直交基底の探索問題であり，行列の低ランク補間 [1]，同時対角化 [5] 等のモデルになっている他，深層学習の汎化性能向上にも応用されている [3]．

$\mathrm{St}(p,N)$ は強い非線形性[2]を持つため，問題 1.1 の解法には工夫が必要である．筆者らは $\mathrm{St}(p,N)$ 上の一般化 Cayley 変換を利用することで，問題 1.1 をベクトル空間上の最適化問題に緩和する手法を検討してきた [4,7]．一般化 Cayley 変換 $\Phi_{\boldsymbol{S}} : \mathrm{St}(p,N)\setminus E_{N,p}(\boldsymbol{S}) \to Q_{N,p}(\boldsymbol{S})$ は，中心点 $\boldsymbol{S} \in \mathrm{O}(N) := \mathrm{St}(N,N)$ 毎に定義され，$\mathrm{St}(p,N)$ の稠密な開部分集合 $\mathrm{St}(p,N)\setminus E_{N,p}(\boldsymbol{S})$ を歪対称行列全体の線形部分空間 $Q_{N,p}(\boldsymbol{S})$ に写す．$\Phi_{\boldsymbol{S}}$ の特異点集合 $E_{N,p}(\boldsymbol{S})$ ($\Phi_{\boldsymbol{S}}$ が定義されない $\mathrm{St}(p,N)$ の部分集合) は $\boldsymbol{S}$ 毎に異なる．また，$\Phi_{\boldsymbol{S}}$ の逆写像 $\Phi_{\boldsymbol{S}}^{-1} : Q_{N,p}(\boldsymbol{S}) \to \mathrm{St}(p,N)\setminus E_{N,p}(\boldsymbol{S})$ はパラメータ空間 $Q_{N,p}(\boldsymbol{S})$ を $\mathrm{St}(p,N)\setminus E_{N,p}(\boldsymbol{S})$ に対応づけるのに利用され (定義 2.1 と図 1 参照)，問題 1.1 は以下の問題に緩和される．

**問題 1.2.** 連続関数 $f : \mathbb{R}^{N\times p} \to \mathbb{R}$ と任意に与えられる $\epsilon > 0$ に対して[3]

$$\text{find } \boldsymbol{V}^\star \in Q_{N,p}(\boldsymbol{S}) \text{ s.t. } f\circ\Phi_{\boldsymbol{S}}^{-1}(\boldsymbol{V}^\star) < \min_{\boldsymbol{U}\in\mathrm{St}(p,N)} f(\boldsymbol{U}) + \epsilon$$

問題 1.2 は $Q_{N,p}(\boldsymbol{S})$ 上で定式化された最適化問題であるため，ベクトル空間上で設計された最適化アルゴリズムを利用することで $\boldsymbol{V}^\star$ を逐次推定できる．問題 1.2 による問題 1.1 を解く方法を Cayley Parametrization(CP) 法と呼ぶ [4,7]．

[1] $\mathrm{St}(p,N)$ は $\mathbb{R}^{N\times p}$ のコンパクトな部分集合であるから，問題 1.1 の解 $\boldsymbol{U}^\star$ の存在性は保証される．

[2] 任意の $\boldsymbol{U} \in \mathrm{St}(p,N)$ と $\boldsymbol{D} \in \mathbb{R}^{N\times p}\setminus\{\boldsymbol{0}\}$ に対し，$\boldsymbol{U} + t\boldsymbol{D} \in \mathrm{St}(p,N)$ を満たす $t \in \mathbb{R}\setminus\{0\}$ は高々1 つであることが簡単に示せる．

[3] $\mathrm{St}(p,N)\setminus E_{N,p}(\boldsymbol{S})$ は $\mathrm{St}(p,N)$ の稠密な部分集合であるから，問題 1.2 の解 $\boldsymbol{V}^\star$ の存在性は保証される．

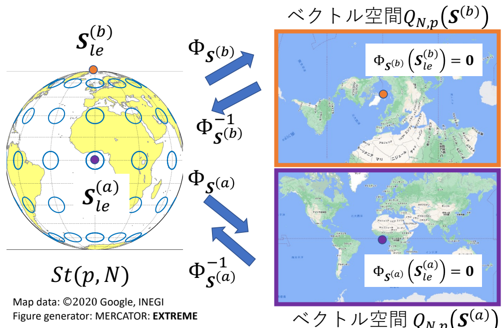


図 1: 異なる 2 つの中心点 $\boldsymbol{S}^{(a)}, \boldsymbol{S}^{(b)}$ を使って定義された一般化 Cayley 変換による $\mathrm{St}(p,N)$ の稠密部分集合のパラメータ表現

一方，問題 1.1 の解 $\boldsymbol{U}^\star$ が $\Phi_{\boldsymbol{S}}$ の特異点集合 $E_{N,p}(\boldsymbol{S})$ の近くに位置する場合に，「CP 法による逐次推定法」の $\boldsymbol{V}^\star$ への接近速度が遅くなる現象 (特異点問題と呼ぶ) が生じることが知られている [4,7]．特異点問題の発生を避けるために，筆者らは $\Phi_{\boldsymbol{S}}^{-1}$ の可動性解析を行い，中心点 $\boldsymbol{S}$(と $E_{N,p}(\boldsymbol{S})$) が適宜更新された緩和問題 1.2 を解く Adaptive Localized Cayley Parametrization(ALCP: 動的パラメータ表現) 法を提案した [4]．

CP 法と異なり，ALCP 法では中心点 $\boldsymbol{S}$ の更新に伴い，推定解の表現空間 $Q_{N,p}(\boldsymbol{S})$ も変化してしまうため，見通しのよい収束解析は与えられていなかった．

小文では，あるクラスの「ベクトル空間上最適化アルゴリズム」を組み込んだ ALCP 法の統一的収束解析を提案する．具体的には，ベクトル空間上最適化アルゴリズムの収束解析 (例えば [2]) でよく利用されてきた「生成点列に関する特別な性質」(仮定 3.1) を基に，ALCP 法の統一的収束解析 (定理 3.2) を紹介する．

以下では，$\boldsymbol{S} \in \mathrm{O}(N)$ の左ブロック行列と右ブロック行列をそれぞれ $\boldsymbol{S}_{\mathrm{le}} \in \mathbb{R}^{N\times p}, \boldsymbol{S}_{\mathrm{ri}} \in \mathbb{R}^{N\times(N-p)}$ と表す．また，$\mathrm{S_{kew}}(\boldsymbol{X}) = (\boldsymbol{X} - \boldsymbol{X}^{\mathrm{T}})/2 \in \mathbb{R}^{N\times N}$ は正方行列 $\boldsymbol{X} \in \mathbb{R}^{N\times N}$ の歪対称成分を表す．

## 2. 一般化 Cayley 変換と Stiefel 多様体の動的パラメータ表現法 (ALCP 法)

**定義 2.1** (一般化 Cayley 変換 [4])**.** $\boldsymbol{S} \in \mathrm{O}(N)$ 毎に決まる $E_{N,p}(\boldsymbol{S}) := \{\boldsymbol{U} \in \mathrm{St}(p,N) \mid \det(\boldsymbol{I}_p + \boldsymbol{S}_{\mathrm{le}}^{\mathrm{T}}\boldsymbol{U}) = 0\}$,

$$Q_{N,p}(\boldsymbol{S}) := Q_{N,p} := \left\{ \begin{bmatrix} \boldsymbol{A} & -\boldsymbol{B}^{\mathrm{T}} \\ \boldsymbol{B} & \boldsymbol{0} \end{bmatrix} \middle| \begin{array}{l} -\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A} \in \mathbb{R}^{p\times p}, \\ \boldsymbol{B} \in \mathbb{R}^{(N-p)\times p} \end{array} \right\}$$

に対して，一般化 Cayley 変換
$\Phi_{\boldsymbol{S}}: \mathrm{St}(p,N)\backslash E_{N,p}(\boldsymbol{S}) \to Q_{N,p}(\boldsymbol{S})$ を

$$\Phi_{\boldsymbol{S}}: \boldsymbol{U} \mapsto \begin{bmatrix} \boldsymbol{A}_{\boldsymbol{S}}(\boldsymbol{U}) & -\boldsymbol{B}_{\boldsymbol{S}}(\boldsymbol{U})^{\mathrm{T}} \\ \boldsymbol{B}_{\boldsymbol{S}}(\boldsymbol{U}) & \boldsymbol{0} \end{bmatrix}$$

で定義する．ただし，各ブロック行列は

$$\boldsymbol{A}_{\boldsymbol{S}}(\boldsymbol{U}) := 2(\boldsymbol{I}_p + \boldsymbol{S}_{\mathrm{le}}^{\mathrm{T}}\boldsymbol{U})^{-\mathrm{T}} \mathrm{S}_{\mathrm{kew}}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{S}_{\mathrm{le}})(\boldsymbol{I}_p + \boldsymbol{S}_{\mathrm{le}}^{\mathrm{T}}\boldsymbol{U})^{-1} \in Q_{p,p}$$

$\boldsymbol{B}_{\boldsymbol{S}}(\boldsymbol{U}) := -\boldsymbol{S}_{\mathrm{ri}}^{\mathrm{T}}\boldsymbol{U}(\boldsymbol{I}_p + \boldsymbol{S}_{\mathrm{le}}^{\mathrm{T}}\boldsymbol{U})^{-1} \in \mathbb{R}^{(N-p)\times p}$ である．

逆変換 $\Phi_{\boldsymbol{S}}^{-1}: Q_{N,p}(\boldsymbol{S}) \to \mathrm{St}(p,N)\backslash E_{N,p}(\boldsymbol{S})$ は

$$\Phi_{\boldsymbol{S}}^{-1}: \boldsymbol{V} \mapsto \boldsymbol{S}(\boldsymbol{I}-\boldsymbol{V})(\boldsymbol{I}+\boldsymbol{V})^{-1}\boldsymbol{I}_{le}$$

で与えられる．

**注意:** $p = N$ かつ $\boldsymbol{S} = \boldsymbol{I}$ のとき，$\Phi_{\boldsymbol{S}}$ は古典的な Cayley 変換 [6] に一致する．$Q_{N,p}(\boldsymbol{S}) = Q_{N,p}$ $(\boldsymbol{S} \in \mathrm{O}(N))$ であるが，$\boldsymbol{S} \in \mathrm{O}(N)$ 毎に集合 $\mathrm{St}(p,N)\backslash E_{N,p}(\boldsymbol{S})$ が変化するため，そのパラメータ表現空間にも区別した表記 $Q_{N,p}(\boldsymbol{S})$ を用いている．

ALCP 法の全体構成の概要を Algorithm 1 に示す．$\mathcal{A}_n: Q_{N,p}(\boldsymbol{S}_n) \to Q_{N,p}(\boldsymbol{S}_n)$ は ALCP 法に組み込まれる「ベクトル空間上最適化アルゴリズム」の $n$ 回目の推定値 $\boldsymbol{V}_n$ の更新則を表している．特異点問題の影響が検出されない間は共通の中心点 $\boldsymbol{S}_n$ が維持されるが，影響検出され次第，$\Phi_{\boldsymbol{S}_n}^{-1}(\boldsymbol{V}_{n+1})$ から新しい特異点集合 $E_{N,p}(\boldsymbol{S}_{n+1})$ が十分に離れるような中心点 $\boldsymbol{S}_{n+1}$ に更新される (影響検出法と中心点の更新法は [4] を参照)．

## 3. ALCP 法に基づく Stiefel 多様体上最適化アルゴリズムの統一的収束解析

ALCP 法の統一的収束解析を与えるために，$f$ に微分可能性と Algorithm 1 に下記条件の成立を仮定する．

**仮定 3.1.** ある定数 $c > 0$ が存在し，Algorithm 1 で生成される点列 $(\boldsymbol{V}_n)_{n=0}^{\infty}$ が $\boldsymbol{S}_k = \boldsymbol{S}_{k+1} = \cdots = \boldsymbol{S}_K$ となる任意の $k, K \in \mathbb{N}$ に対して以下の不等式を満たす．

$$(l \in \{k, k+1, \ldots, K\}) \quad \sum_{n=k}^{l} \|\nabla(f \circ \Phi_{\boldsymbol{S}_n}^{-1})(\boldsymbol{V}_n)\|_F^2 \leqslant c$$

実際に，Algorithm 1 内の $\mathcal{A}_n$ として，あるクラスの加速勾配法 [2](通常の最急降下法も含む) の更新則を採用すれば $\nabla f$ の Lipschitz 連続性の下で仮定 3.1 の成立が確認できる ( [4] の議論参照)．

**定理 3.2** (ALCP 法の統一的収束解析)**.** *Algorithm 1* が仮定 *3.1* とともに，同一中心点の継続利用回数に関するある条件[4]を満たすとする．このとき，任意の $\epsilon > 0$ に対し，$\|\nabla(f \circ \Phi_{\boldsymbol{S}_n}^{-1})(\boldsymbol{V}_n)\|_F^2 < \epsilon$ を達成するまでに必要な反復回数は高々$o(\epsilon^{-3/2})$ である．

[4]正確な条件は発表の中で説明する予定である．

**Algorithm 1** ALCP 法の全体構成の概要

**Input:** $\boldsymbol{S}_0 \in \mathrm{O}(N)$, $\boldsymbol{U}_0 \in \mathrm{St}(p,N)\backslash E_{N,p}(\boldsymbol{S}_0)$
  $\boldsymbol{V}_0 \leftarrow \Phi_{\boldsymbol{S}_0}(\boldsymbol{U}_0)$
  **for** $n = 0, 1, 2, \ldots, m-1$ **do**
    $\boldsymbol{V}_{n+1} \leftarrow \mathcal{A}_n(\boldsymbol{V}_n)$, $\boldsymbol{U}_{n+1} \leftarrow \Phi_{\boldsymbol{S}_n}^{-1}(\boldsymbol{V}_{n+1})$
    **if** 特異点の影響が検出されない **then**
      $\boldsymbol{S}_{n+1} \leftarrow \boldsymbol{S}_n$ (同一中心点を利用する)
    **else**
      $\boldsymbol{U}_{n+1} \leftarrow \boldsymbol{U}_n$
      $\boldsymbol{U}_{n+1}$ を用いて新たな $\boldsymbol{S}_{n+1}$ を計算
      $\boldsymbol{V}_{n+1} \leftarrow \Phi_{\boldsymbol{S}_{n+1}}(\boldsymbol{U}_{n+1})$
    **end if**
  **end for**
**Output:** $\boldsymbol{U}_m$

**謝辞** 本研究は，特別研究員奨励費 (21J21353) と JST SICORP(JPMJSC20C6) の助成を受けた．

## 参考文献

[1] N. Boumal and P.-A. Absil, “Low-rank matrix completion via preconditioned optimization on the Grassmann manifold,” *Linear Algebra Its Appl*, 2015.

[2] S. Ghadimi and G. Lan, “Accelerated gradient methods for nonconvex nonlinear and stochastic programming,” *Math Program*, 2016.

[3] K. Helfrich, D. Willmott, and Q. Ye, “Orthogonal recurrent neural networks with scaled Cayley transform,” in *ICML*, 2018.

[4] K. Kume and I. Yamada, “A Nesterov-type acceleration with adaptive localized Cayley parametrization for optimization over the Stiefel manifold,” in *EUSIPCO*, 2020.

[5] H. Sato, “Riemannian Newton-type methods for joint diagonalization on the Stiefel manifold with application to independent component analysis,” *Optimization*, 2017.

[6] H. Weyl, *The classical groups: their invariants and representations.* Princeton university press, 1946.

[7] I. Yamada and T. Ezaki, “An orthogonal matrix optimization by dual Cayley parametrization technique,” in *ICA*, 2003.