

シュティーフェル多様体上の信頼領域法 の近似的同時特異値分解への応用*

佐藤 寛之

東京理科大学工学部経営工学科

Hiroyuki Sato[†]

Department of Management Science, Tokyo University of Science

概要

同じサイズの複数の長方形列の近似的同時特異値分解は、画像処理などの分野に応用を持ち、シュティーフェル多様体の2つの直積からなるリーマン多様体上の最適化問題として定式化できる。本稿では、この最適化問題に対する信頼領域法を導出することで、従来の手法より高速なアルゴリズムを提案する。提案手法は大域的収束性と超1次収束性を兼ね備え、部分問題を適当に解くと2次収束性が保証される。

1 はじめに

長方形列 $A \in \mathbb{R}^{m \times n}$ ($m \geq n$) に対して、

$$A = U \Sigma V^T, \quad U \in O(m), V \in O(n), \Sigma = \begin{pmatrix} \Sigma_0 \\ 0 \end{pmatrix}, \Sigma_0 = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n) \quad (1)$$

という分解を A の特異値分解という。ただし、 $O(n)$ は n 次直交群を表し、また、 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$ であり、 $\sigma_1, \sigma_2, \dots, \sigma_n$ を A の特異値という。また、 U, V の第 i 列 u_i, v_i を、それぞれ A の第 i 特異値 σ_i に対応する左特異ベクトル、右特異ベクトルという。行列の特異値分解は信号処理や制御理論など様々な分野への応用を持つ。

また、すべての特異値を求めるのではなく、大きい方から p ($\leq n$) 個の特異値だけに注目することも多い。そのような問題は次のような最適化問題として定式化できる：

問題 1.1.

$$\begin{aligned} & \text{minimize} && F(U, V) = -\text{tr}(U^T A V N), \\ & \text{subject to} && U^T U = V^T V = I_p, U \in \mathbb{R}^{m \times p}, V \in \mathbb{R}^{n \times p}. \end{aligned}$$

ここで、 $N = \text{diag}(\mu_1, \mu_2, \dots, \mu_p)$, $\mu_1 > \mu_2 > \dots > \mu_p > 0$ である。この問題の最適解を (U_*, V_*) とすると、 U_* と V_* の第 i 列はそれぞれ行列 A の第 i 特異値に属する左特異ベクトル、右特異ベクトルとなる [7].

一方、対称行列の固有値問題も重要な問題であるが、これを拡張したものに複数の対称行列の同時対角化問題がある。よく知られているように、 K 個の n 次対称行列 A_1, A_2, \dots, A_K

*本稿の結果は [8] によるものである。

[†]e-mail: hsato@ms.kagu.tus.ac.jp

が同時対角化できるための必要十分条件は、 A_1, A_2, \dots, A_K が互いに可換となることであるから、一般には複数の対称行列を厳密に同時対角化することはできない。したがって、近似的に同時対角化を実現するような直交行列を求めるアプローチが取られる。その定式化の一つに、

$$\sum_{l=1}^K \|\text{diag}(Y^T A_l Y)\|_F^2 \quad (2)$$

を最大にするような直交行列 $Y \in O(n)$ を見つけるというものがある [9].

この定式化のアイデアを複数の長方形行列 $A_1, A_2, \dots, A_K \in \mathbb{R}^{m \times n}$ の (近似的) 同時特異値分解に応用すると、次の最適化問題が得られる：

問題 1.2.

$$\begin{aligned} \text{minimize } f(U, V) &= - \sum_{l=1}^K \|\text{diag}(U^T A_l V)\|_F^2, \\ \text{subject to } U^T U &= V^T V = I_p, \quad U \in \mathbb{R}^{m \times p}, V \in \mathbb{R}^{n \times p}. \end{aligned}$$

これはユークリッド空間 $\mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p}$ 上の最適化問題であると思わせるが、制約条件を満たす点全体に注目すると、多様体上の最適化であるとも見なすことができる。すなわち、

$$\text{St}(p, n) := \{Y \in \mathbb{R}^{n \times p} \mid Y^T Y = I_p\} \quad (3)$$

で定義されるシュティーフエル多様体 $\text{St}(p, n)$ を用いると、問題 1.2 は次のように書き換えることができる。

問題 1.3.

$$\begin{aligned} \text{minimize } f(U, V) &= - \sum_{l=1}^K \|\text{diag}(U^T A_l V)\|_F^2, \\ \text{subject to } (U, V) &\in \text{St}(p, m) \times \text{St}(p, n). \end{aligned}$$

以下では、問題 1.3 に現れる目的関数 f と多様体 $\text{St}(p, m) \times \text{St}(p, n)$ に関する幾何学について述べた上で、この問題に対する信頼領域法を導出することで、同時特異値分解問題に対する解法アルゴリズムを与える。

なお、行列の同時特異値分解問題は画像処理などの分野において現れ、重要な応用を持つ [2, 4, 5, 6].

2 問題 1.3 の幾何学的な種々の量について

$\text{St}(p, m) \times \text{St}(p, n)$ の幾何学に関しては、[7] に詳しい。ここでは主要な結果をまとめることとする。いずれの結果も、単独のシュティーフエル多様体 $\text{St}(p, n)$ についての議論からの帰結である。

2.1 接空間とリーマン計量および射影

$\text{St}(p, m) \times \text{St}(p, n)$ の点 (U, V) における接空間は,

$$T_{(U,V)}(\text{St}(p, m) \times \text{St}(p, n)) \simeq T_U \text{St}(p, m) \times T_V \text{St}(p, n) \\ = \{(\xi, \eta) \in \mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p} \mid \xi^T U + U^T \xi = \eta^T V + V^T \eta = 0\} \quad (4)$$

である.

次に, $\text{St}(p, m) \times \text{St}(p, n) \subset \mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p}$ より

$$T_{(U,V)}(\text{St}(p, m) \times \text{St}(p, n)) \subset T_{(U,V)}(\mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p}) \simeq \mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p} \quad (5)$$

であり, $\mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p}$ には標準内積が

$$\langle (X_1, Y_1), (X_2, Y_2) \rangle = \text{tr}(X_1^T X_2) + \text{tr}(Y_1^T Y_2) \quad (6)$$

のように入るから, 接空間 $T_{(U,V)}(\text{St}(p, m) \times \text{St}(p, n))$ には (6) から誘導される内積

$$\langle (\xi_1, \eta_1), (\xi_2, \eta_2) \rangle_{(U,V)} = \text{tr}(\xi_1^T \xi_2) + \text{tr}(\eta_1^T \eta_2), \quad (\xi_1, \eta_1), (\xi_2, \eta_2) \in T_{(U,V)}(\text{St}(p, m) \times \text{St}(p, n)). \quad (7)$$

を入れることができ, この内積を任意の (U, V) における接空間に与える写像は $\text{St}(p, m) \times \text{St}(p, n)$ のリーマン計量となる. この誘導計量により $\text{St}(p, m) \times \text{St}(p, n)$ は $\mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p}$ のリーマン部分多様体となる.

このリーマン計量の下で, 点 $(U, V) \in \text{St}(p, m) \times \text{St}(p, n)$ における接空間 $T_{(U,V)}(\text{St}(p, m) \times \text{St}(p, n))$ への射影 $P : \mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p} \rightarrow T_{(U,V)}(\text{St}(p, m) \times \text{St}(p, n))$ は, 任意の $(B, C) \in \mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p}$ に対して

$$P_{(U,V)}(B, C) = (B - U \text{sym}(U^T B), C - V \text{sym}(V^T C)) \quad (8)$$

のように作用する. ここで, 行列 A に対して $\text{sym}(A) = (A + A^T)/2$ である.

2.2 目的関数の勾配とヘシアン

信頼領域法においては目的関数の勾配とヘシアンが必要であるが, 一般に, ユークリッド空間 E のリーマン部分多様体 M について, M 上の関数 $g : M \rightarrow \mathbb{R}$ の勾配とヘシアンは g の定義域を E 全体に拡張した関数の勾配やヘシアンおよび接空間への射影を用いて表すことができる. 実際, 目的関数 f に対して \bar{f} を, $\mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p}$ 全体で定義され, M への制限 $\bar{f}|_M$ が f と一致するような関数, 特に,

$$\bar{f}(U, V) = - \sum_{l=1}^K \|\text{diag}(U^T A_l V)\|_F^2, \quad (U, V) \in \mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p} \quad (9)$$

としよう. 一般に, 行列 $A = (a_{ij}), B = (b_{ij}) \in \mathbb{R}^{p \times p}$ に対して

$$\begin{aligned} \operatorname{tr}(\operatorname{diag}(A) \operatorname{diag}(B)) &= \sum_{i=1}^p a_{ii} b_{ii} = \sum_{i=1}^p a_{ii} b_{ii} + \sum_{i \neq j} a_{ij} \cdot 0 = \operatorname{tr}(A^T \operatorname{diag}(B)) \\ &= \operatorname{tr}(\operatorname{diag}(B)^T A) = \operatorname{tr}(\operatorname{diag}(B) A) = \operatorname{tr}(A \operatorname{diag}(B)) \end{aligned} \quad (10)$$

より

$$\|\operatorname{diag}(A)\|_F^2 = \operatorname{tr}(\operatorname{diag}(A)^2) = \operatorname{tr}(A \operatorname{diag}(A)) \quad (11)$$

が成り立つから,

$$\bar{f}(U, V) = - \sum_{l=1}^K \operatorname{tr}(U^T A_l V \operatorname{diag}(U^T A_l V)) \quad (12)$$

と書ける. したがって, \bar{f} を (U, V) において任意の $(\xi, \eta) \in \mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p}$ 方向に微分すると, (10) より

$$\begin{aligned} D\bar{f}(U, V)[(\xi, \eta)] &= - \sum_{l=1}^K \operatorname{tr}(\xi^T A_l V \operatorname{diag}(U^T A_l V) + U^T A_l V \operatorname{diag}(\xi^T A_l V) \\ &\quad + U^T A_l \eta \operatorname{diag}(U^T A_l V) + U^T A_l V \operatorname{diag}(U^T A_l \eta)) \\ &= - 2 \sum_{l=1}^K \operatorname{tr}(\xi^T A_l V \operatorname{diag}(U^T A_l V) + \eta^T A_l^T U \operatorname{diag}(U^T A_l V)) \end{aligned} \quad (13)$$

であるから, $D_l := \operatorname{diag}(U^T A_l V)$ とおくと, \bar{f} の勾配は

$$\nabla \bar{f}(U, V) = -2 \sum_{l=1}^K (A_l V D_l, A_l^T U D_l) \quad (14)$$

となる. また, $D'_l := \operatorname{diag}(\xi^T A_l V + U^T A_l \eta)$ とおくと, $(\xi, \eta) \in \mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p}$ に対して

$$\nabla^2 \bar{f}(U, V)[(\xi, \eta)] = -2 \sum_{l=1}^K (A_l (\eta D_l + V D'_l), A_l^T (\xi D_l + U D'_l)) \quad (15)$$

となる.

これら \bar{f} の勾配とヘシアンおよび射影作用素 (8) を用いて, 多様体 $\operatorname{St}(p, m) \times \operatorname{St}(p, n)$ 上の目的関数 f の勾配とヘシアンが次のように計算できる. まず, $S_l := U^T A_l V$ とおいて, (U, V) における勾配 $\operatorname{grad} f(U, V)$ は,

$$\begin{aligned} \operatorname{grad} f(U, V) &= P_{(U, V)}(\nabla \bar{f}(U, V)) \\ &= -2 \sum_{l=1}^K (A_l V D_l - U \operatorname{sym}(S_l D_l), A_l^T U D_l - V \operatorname{sym}(D_l S_l)) \end{aligned} \quad (16)$$

となる. ヘシアン $\operatorname{Hess} f(U, V)$ の $(\xi, \eta) \in T_{(U, V)}(\operatorname{St}(p, m) \times \operatorname{St}(p, n))$ に対する作用は,

$$\begin{aligned} \operatorname{Hess} f(U, V)[(\xi, \eta)] &= P_{(U, V)}(D(\operatorname{grad} f)(U, V)[(\xi, \eta)]) \\ &= P_{(U, V)}(D(P_{(U, V)}(\nabla \bar{f}))(U, V)[(\xi, \eta)]) \\ &= P_{(U, V)}(D P_{(U, V)}[(\xi, \eta)](\nabla \bar{f}(U, V)) + P_{(U, V)}(D(\nabla \bar{f}))(U, V)[(\xi, \eta)]) \\ &= P_{(U, V)}(D P_{(U, V)}[(\xi, \eta)](\nabla \bar{f}(U, V)) + \nabla^2 \bar{f}(U, V)[(\xi, \eta)]) \end{aligned} \quad (17)$$

と書ける. ここで, 射影作用素の性質 $P_{(U,V)}^2 = P_{(U,V)}$ を用いた. また, 一般に $(B, C) \in \mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p}$ に対して

$$DP_{(U,V)}[(\xi, \eta)](B, C) = (-\xi \operatorname{sym}(U^T B) - U \operatorname{sym}(\xi^T B), -\eta \operatorname{sym}(V^T C) - V \operatorname{sym}(\eta^T C)) \quad (18)$$

である. さらに,

$$\begin{aligned} & P_{(U,V)}(DP_{(U,V)}[(\xi, \eta)](B, C)) \\ &= P_{(U,V)}(-\xi \operatorname{sym}(U^T B), -\eta \operatorname{sym}(V^T C)) - P_{(U,V)}(U \operatorname{sym}(\xi^T B), V \operatorname{sym}(\eta^T C)) \\ &= P_{(U,V)}(-\xi \operatorname{sym}(U^T B), -\eta \operatorname{sym}(V^T C)) \\ &\quad - (U \operatorname{sym}(\xi^T B) - U \operatorname{sym}(U^T U \operatorname{sym}(\xi^T B)), V \operatorname{sym}(\eta^T C) - V \operatorname{sym}(V^T V \operatorname{sym}(\eta^T C))) \\ &= P_{(U,V)}(-\xi \operatorname{sym}(U^T B), -\eta \operatorname{sym}(V^T C)). \end{aligned} \quad (19)$$

したがって, $(G, H) := \nabla \bar{f}(U, V)$ とおくと,

$$\operatorname{Hess} f(U, V)[(\xi, \eta)] = P_{(U,V)}((-\xi \operatorname{sym}(U^T G), -\eta \operatorname{sym}(V^T H)) + \nabla^2 \bar{f}(U, V)[(\xi, \eta)]) \quad (20)$$

を得る.

3 新しい同時特異値分解アルゴリズム

3.1 ユークリッド空間における最適化とリーマン多様体における最適化

リーマン多様体上の最適化についての詳細は [1] を参照されたい. ユークリッド空間 E における無制約最適化においては, 現在の点 $x_k \in E$ から $\xi_k \in E$ 方向に進んだ点を次の点 x_{k+1} とする際,

$$x_{k+1} = x_k + \xi_k \quad (21)$$

という更新式を用いるが, ユークリッド空間 E の部分多様体 M 上の無制約最適化においては, x_k に対して $\xi_k \in T_{x_k} M$ とすると, $x_k + \xi_k$ は一般に M 上の点でない. したがって, 式 (21) をそのまま用いることはできず, $x_k + \xi_k$ を多様体 M に移す, レトラクションと呼ばれる写像が必要となる [1]. すなわち, レトラクション $R: TM \rightarrow M$ が与えられたとき,

$$x_{k+1} = R_{x_k}(\xi_k) \quad (22)$$

によって次の点 x_{k+1} を定める.

シュティーフエル多様体 $\operatorname{St}(p, n)$ 上の点 Y と接ベクトル ξ に対して, R を

$$R_Y(\xi) = \operatorname{qf}(Y + \xi) \quad (23)$$

と定義すると, R は $\operatorname{St}(p, n)$ 上のレトラクションとなる. ここで, R_Y は R の $T_Y \operatorname{St}(p, n)$ への制限 $R|_{T_Y \operatorname{St}(p, n)}$ であり, $\operatorname{qf}(\cdot)$ は行列の QR 分解の Q 成分を表す. したがって, 積多様体 $\operatorname{St}(p, m) \times \operatorname{St}(p, n)$ 上で R を

$$\begin{aligned} R_{(U,V)}(\xi, \eta) &= (\operatorname{qf}(U + \xi), \operatorname{qf}(V + \eta)), \\ (U, V) &\in \operatorname{St}(p, m) \times \operatorname{St}(p, n), (\xi, \eta) \in T_{(U,V)}(\operatorname{St}(p, m) \times \operatorname{St}(p, n)) \end{aligned} \quad (24)$$

によって定義すると、 R はレトラクションとなる。

3.2 信頼領域法に基づく同時特異値分解アルゴリズム

さて、リーマン多様体 M 上の信頼領域法はユークリッド空間の場合と同様、目的関数の2次近似によるモデルの最小化に基づく最適化手法である。問題 1.3 の目的関数 f の点 $(U, V) \in \text{St}(p, m) \times \text{St}(p, n)$ における2次モデル $\hat{m}_{(U, V)}$ は

$$\hat{m}_{(U, V)}(\xi, \eta) = f(U, V) + \langle \text{grad } f(U, V), (\xi, \eta) \rangle_{(U, V)} + \frac{1}{2} \langle \text{Hess } f(U, V)[(\xi, \eta)], (\xi, \eta) \rangle_{(U, V)} \quad (25)$$

によって定義される。点 (U_k, V_k) において探索方向の候補 $(\xi_k, \eta_k) \in T_{(U_k, V_k)}$ が求めたら、この接ベクトルが目的関数を十分小さくするかを評価するために、

$$\rho_k := \frac{f(R_{(U_k, V_k)}(0)) - f(R_{(U_k, V_k)}(\xi_k, \eta_k))}{\hat{m}_{(U_k, V_k)}(0) - \hat{m}_{(U_k, V_k)}(\xi_k, \eta_k)} \quad (26)$$

を計算し、この比の値によって、探索方向を採用するか否か、また現在の信頼領域半径 Δ_k が適当な値であるか否かを判定する。

結局、問題 1.3 に対する信頼領域法はアルゴリズム 1 のように書くことができる。

問題 1.3 に対する先行研究には、逆勾配 $-\text{grad } f$ が定める微分方程式

$$(\dot{U}, \dot{V}) = -\text{grad } f(U, V) \quad (28)$$

に沿って f の最小点を求めるものがある [4, 5]。これは、最適化の観点からは最急降下法を、指数写像が定めるレトラクションによって行うことに相当する。ただし、シュティーフエル多様体上の指数写像を直接計算しようとする行列の指数関数の計算が必要となり [3]、QR 分解に基づくレトラクション (24) に比べて計算量が大きく、また測地線方程式と呼ばれる測地線を定める微分方程式を数値的に解くことで指数写像による像を計算するのもやはり QR 分解に比べると大変である。

さらに、最急降下法は大域的収束性こそ持つものの、その収束は信頼領域法に比べると一般にずっと遅い。より正確には、最急降下法は1次収束性しか持たないが、信頼領域法は部分問題を適切に解けば2次収束することが示される [1]。たとえば [1] では truncated CG と呼ばれる手法を用いた一般的な収束解析が議論されている。

その収束速度の大きな違いは、実際に数値実験の結果によっても示される [8]。

4 終わりに

本稿では同時特異値分解問題をリーマン多様体 $\text{St}(p, m) \times \text{St}(p, n)$ 上の最適化問題として定式化し、信頼領域法によるアルゴリズムについて議論した。提案手法は、大域的収束性と超1次収束性を併せ持ち、既存の手法より速い収束速度を持つ。

しかしながら、同時特異値分解の研究は、多様体上の最適化によるアプローチか否かを問わずまだ十分になされていないのが現状である。本稿で提案した手法は、たとえばヤコ

Algorithm 1 問題 1.3 に対する信頼領域法

1: パラメータ $\bar{\Delta} > 0$, $\Delta_0 \in (0, \bar{\Delta})$, $\rho' \in [0, \frac{1}{4}]$ および, 初期点 $(U_0, V_0) \in \text{St}(p, m) \times \text{St}(p, n)$ を選ぶ.

2: **for** $k = 0, 1, 2, \dots$ **do**

3: $(\xi_k, \eta_k) \in T_{(U_k, V_k)}(\text{St}(p, m) \times \text{St}(p, n))$ を信頼領域部分問題の解とする. すなわち,

$$(\xi_k, \eta_k) = \arg \min_{(\xi, \eta)} \{ \hat{m}_{(U_k, V_k)}(\xi, \eta) \mid (\xi, \eta) \in T_{(U_k, V_k)}(\text{St}(p, m) \times \text{St}(p, n)), \langle (\xi, \eta), (\xi, \eta) \rangle_{(U_k, V_k)} \leq \Delta_k^2 \}$$

(27)

とする.

4: $\rho_k := \frac{f(R_{(U_k, V_k)}(0)) - f(R_{(U_k, V_k)}(\xi_k, \eta_k))}{\hat{m}_{(U_k, V_k)}(0) - \hat{m}_{(U_k, V_k)}(\xi_k, \eta_k)}$ を計算する.

5: **if** $\rho_k < \frac{1}{4}$ **then**

6: $\Delta_{k+1} = \frac{1}{4} \Delta_k$.

7: **else if** $\rho_k > \frac{3}{4}$ かつ $\|(\xi_k, \eta_k)\|_{(U_k, V_k)} = \Delta_k$ **then**

8: $\Delta_{k+1} = \min(2\Delta_k, \bar{\Delta})$.

9: **else**

10: $\Delta_{k+1} = \Delta_k$.

11: **end if**

12: **if** $\rho_k > \rho'$ **then**

13: $(U_{k+1}, V_{k+1}) = R_{(U_k, V_k)}(\xi_k, \eta_k)$.

14: **else**

15: $(U_{k+1}, V_{k+1}) = (U_k, V_k)$.

16: **end if**

17: **end for**

び法に基づく数値線形代数からのアプローチと比較する必要や、実際的な問題データに対する数値実験を行うことで実用的にも有効な手段となることを検証する必要があると考えられる。

参考文献

- [1] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, Princeton, 2008.
- [2] M. Congedo, R. Phlypo, Approximate joint singular value decomposition of an asymmetric rectangular matrix set, *IEEE Trans. Signal Process.*, **59** (2011), 415–424.
- [3] A. Edelman, T. A. Arias, and S. T. Smith, The geometry of algorithms with orthogonality constraints, *SIAM J. Matrix Anal. Appl.*, **20**(2) (1998), pp. 303–353.
- [4] G. Hori, Comparison of two main approaches to joint SVD, in: *Proc. 8th Int. Conf. Independent Component Analysis and Signal Separation*, pp. 42-49, 2009.
- [5] G. Hori, Joint SVD and its application to factorization method, in: *Proc. 9th Int. Conf. Latent Variable Analysis and Signal Separation*, pp. 563-570, 2010.
- [6] B. Pesquet-Popescu, J.-C. Pesquet and A. P. Petropulu, Joint singular value decomposition—a new tool for separable representation of images, in: *Proc. 2001 IEEE Int. Conf. Image Processing*, pp. 569-572, 2001.
- [7] H. Sato, T. Iwai, A Riemannian optimization approach to the matrix singular value decomposition, *SIAM J. Optim.*, **23** (2013), 188–212.
- [8] H. Sato, Joint singular value decomposition algorithm based on the Riemannian trust-region method, *JSIAM Lett.*, to appear.
- [9] F.J. Theis, Thomas P. Cason, and P.-A. Absil, Soft dimension reduction for ica by joint diagonalization on the Stiefel manifold, in: *Proc. 8th Int. Conf. Independent Component Analysis and Signal Separation*, pp. 354–361, 2009.