数理解析研究所講究録
第 1829 巻 2013 年 39-53

リーマン多様体上の共役勾配法 およびその特異値分解問題への応用

京都大学大学院情報学研究科数理工学専攻
佐藤 寛之 (Hiroyuki Sato), 岩井 敏洋 (Toshihiro Iwai)
Department of Applied Mathematics and Physics,
Kyoto University

概 要

行列の特異値分解は，異なるサイズの 2 つのシュティーフェル多様体の積からなる多様体の上の最適化問題に帰着される．本稿では，その問題に対する最適化アルゴリズムとして最急降下法，共役勾配法，ニュートン法を導出し，それらの特徴について詳しく述べる．また，共役勾配法とニュートン法を組み合わせることによって，新しい特異値分解手法を提案する．

1 序論

m, n を整数とし，$m \geq n$ とする．$m \times n$ 行列 A は

$$A = U_0 \Sigma_0 V_0^T, \qquad U_0 \in O(m),\ V_0 \in O(n),\ \Sigma_0 = \begin{pmatrix} \Sigma_1 \\ 0 \end{pmatrix} \tag{1.1}$$

と分解される．ここで，$\Sigma_1 = \operatorname{diag}(\sigma_1, \ldots, \sigma_n)$，$\sigma_1 \geq \cdots \geq \sigma_n \geq 0$ である．この分解 (1.1) を A の特異値分解といい，σ_i, $i = 1, \ldots, n$ を A の特異値という [4, 7]．$u_1, \ldots, u_m$ と $v_1, \ldots, v_n$ をそれぞれ U_0 と V_0 の列ベクトルとする．つまり，$U_0 = (u_1, \ldots, u_m)$，$V_0 = (v_1, \ldots, v_n)$ である．u_i, v_i をそれぞれ A の左特異ベクトル，右特異ベクトルという．これらの列ベクトル u_i, v_i を使えば A の特異値分解は

$$A = \sum_{i=1}^{n} \sigma_i u_i v_i^T \tag{1.2}$$

と書くこともできる．

特異値分解は，次の最適化問題と密接な関係がある．

問題 1.1.

$$\begin{aligned} &\text{maximize} \quad \operatorname{tr}\left(U^T A V N\right), && (1.3) \\ &\text{subject to} \quad U \in \mathbb{R}^{m \times p},\ V \in \mathbb{R}^{n \times p},\ U^T U = V^T V = I_p. && (1.4) \end{aligned}$$

ここで，$1 \leq p \leq n$ であり，また $N = \mathrm{diag}(\mu_1, \ldots, \mu_p)$，$\mu_1 > \cdots > \mu_p > 0$ である．実際，次の命題が成り立つ．

命題 1.1. $m \geq n$ とする．$m \times n$ 行列 A の特異値分解 は (1.2) の形をしており，$\sigma_1 \geq \cdots \geq \sigma_n \geq 0$ であるとする．(U_*, V_*) を問題 *1.1* の最適解とする．このとき，U_* と V_* の列ベクトルたちは，それぞれ A の小さい方から p 個の特異値に属する左特異ベクトル，右特異ベクトルとなる．つまり，

$$U_* = (u_1, \ldots, u_p)\,,\ V_* = (v_1, \ldots, v_p) \tag{1.5}$$

が成り立つ．

この命題の証明は，ラグランジュの未定乗数法による．詳しくは [9] を参照されたい．

ここで，問題 1.1 の制約条件について注目する．一つの行列変数 U を見ると，これは集合 $\left\{U \in \mathbb{R}^{n \times p} \,|\, U^T U = I_p\right\}$ の上を動く．ところで，この集合には多様体の構造が入ることが知られており，シュティーフェル多様体と呼ばれている．つまり，シュティーフェル多様体 $\mathrm{St}(p, n)$ は

$$\mathrm{St}(p, n) = \left\{Y \in \mathbb{R}^{n \times p} \,|\, Y^T Y = I_p\right\} \tag{1.6}$$

と定義される．すると，問題 1.1 の制約条件 $U \in \mathbb{R}^{m \times p}$，$V \in \mathbb{R}^{n \times p}$，$U^T U = V^T V = I_p$ は，$(U, V) \in \mathrm{St}(p, m) \times \mathrm{St}(p, n)$ と書き換えることができる．さらに，最適化の分野では最小化問題を扱うのが普通であるから，目的関数に -1 を乗じて，問題 1.1 を次の最適化問題に変換する．

問題 1.2.

$$\begin{aligned} \text{minimize} \quad & F(U, V) = -\operatorname{tr}\left(U^T A V N\right), && (1.7) \\ \text{subject to} \quad & (U, V) \in \mathrm{St}(p, m) \times \mathrm{St}(p, n). && (1.8) \end{aligned}$$

以降の節では問題 1.1 の解法の導出およびその適用について議論する．その準備として，一般のリーマン多様体上の最適化手法についても簡単に紹介する．

2 リーマン多様体上の最適化手法

本節では，[1] に従ってリーマン多様体上の最適化手法の一般論を展開する．

2.1 ユークリッド空間からリーマン多様体へ

最初に，次のユークリッド空間 $\mathbb{R}^N$ における制約条件なしの最適化問題を考える．

問題 2.1.

$$\begin{aligned} &\text{minimize} && f(x), \\ &\text{subject to} && x \in \mathbb{R}^N. \end{aligned}$$

制約条件なしの問題に対しては，最急降下法や共役勾配法，ニュートン法などのアルゴリズムが知られているが，それらのアルゴリズムは次のように共通する枠組みを持つ．

Algorithm 2.1 $\mathbb{R}^N$ 上の最適化アルゴリズム

```
初期点 x_0 ∈ R^N を選ぶ.
for k = 0, 1, 2, ... do
    探索方向 η_k ∈ R^N とステップサイズ t_k > 0 を計算する.
    次の点 x_{k+1} を x_{k+1} := x_k + t_k η_k によって計算する.
end for
```

この枠組みにおいて，探索方向の決め方が各最適化アルゴリズムを特徴付ける．探索方向 $\eta_k \in \mathbb{R}^N$ は，たとえば最急降下法では x_k における f の逆勾配として，ニュートン法ではニュートン方程式の解として定める．共役勾配法については後述する．

これらの概念をリーマン多様体上に拡張する．つまり，次のリーマン多様体 (M, g) 上の制約条件なしの最適化問題の解法を考えたい．

問題 2.2.

$$\begin{aligned} &\text{minimize} && f(x), \\ &\text{subject to} && x \in M. \end{aligned}$$

まず，k 反復目における探索方向 η_k は点 $x_k \in M$ における接ベクトルとして選ぶ．つまり，$\eta_k \in T_{x_k}M$ となるようにする．

また，探索方向 $\eta_k \in T_{x_k}M$ とステップサイズ $t_k > 0$ を何らかの方法で求めたとして，ユークリッド空間上のアルゴリズムにおける更新の式

$$x_{k+1} := x_k + t_k\eta_k \tag{2.1}$$

は多様体上では一般には意味をなさない．実際，多様体はユークリッド空間に埋め込まれているとは限らないので加法が一般には定義されないし，仮にユークリッド空間 $\mathbb{R}^N$ に埋め込まれていて $x_k + t_k\eta_k \in \mathbb{R}^N$ が定まるとしても，これは一般には多様体 M 上の点ではない．そこで，$\gamma(0) = x_k$，$\dot{\gamma}(0) = \eta_k$ なる M 上の曲線 γ に沿って次の点 x_{k+1} を探索す

る．そのために,「探索をする上で妥当な」曲線を定める写像 $R: TM \to M$ が見つかれば，$R_x := R|_{T_xM}$ として，

$$x_{k+1} := R_{x_k}(t_k\eta_k)$$

なる更新式を得られる．R に課すべき条件は，

$$R_{x_k}(0) = x_k \tag{2.2}$$

かつ

$$\frac{d}{dt}R_{x_k}(t\eta_k)|_{t=0} = \eta_k \tag{2.3}$$

であることが曲線の初期点と初速度についての考察から分かる．このことを踏まえて，レトラクション $R: TM \to M$ を次のように定義する．

定義 2.1. 写像 $R: TM \to M$ が次を満たすとき，R をレトラクションという．$R_x := R|_{T_xM}$ とする．

- $R_x(0_x) = x$, ここで，0_x は T_xM の零ベクトルである．
- $T_{0_x}T_xM \simeq T_xM$ なる同一視の下で，

$$\mathrm{D}R_x(0_x) = \mathrm{id}_{T_xM}.$$

2.2 リーマン多様体上の最適化アルゴリズム

さて，リーマン多様体 M 上のレトラクション R を用いて，M 上の最適化アルゴリズムは次のように記述される：

Algorithm 2.2 リーマン多様体 M 上の最適化アルゴリズム

初期点 $x_0 \in M$ を選ぶ．
for $k = 0, 1, 2, \ldots$ **do**
　探索方向 $\eta_k \in T_{x_k}M$ とステップサイズ $t_k > 0$ を求める．
　次の点 x_{k+1} を，$x_{k+1} := R_{x_k}(t_k\eta_k)$ によって計算する．
end for

ステップサイズの決め方としては，アルミホの方法が有名であるが，これもリーマン多様体上に拡張することができる．

定義 2.2. f をリーマン多様体 M 上の目的関数とし，R をレトラクションとする．スカラー $\bar{\alpha} > 0$, β, $\sigma \in (0, 1)$ は与えられた定数とする．
与えられた点 $x \in M$ と接ベクトル $\eta \in T_xM$ に対して，アルミホポイントを $\eta^A = \beta^m\bar{\alpha}\eta$ によって定義する．ここで，m は

$$f(x) - f(R_x(\beta^m\bar{\alpha}\eta)) \geq -\sigma\langle \mathrm{grad}\, f(x),\ \beta^m\bar{\alpha}\eta\rangle_x.$$

を満たす最小の非負の整数である．
また，$t^A = \beta^m\bar{\alpha}$ をアルミホステップサイズという．

これらを用いて，最急降下法やニュートン法，共役勾配法などをリーマン多様体上に拡張することができるが，それらについては実際に問題 1.2 に対するアルゴリズムを導出する際に述べることにする．

3 積多様体 $\mathrm{St}(p,m)\times\mathrm{St}(p,n)$ の幾何

$\mathrm{St}(p,m)\times\mathrm{St}(p,n)$ 上の問題 1.2 に対する解法を導出するために，$\mathrm{St}(p,m)\times\mathrm{St}(p,n)$ と目的関数 $F(U,V)=\mathrm{tr}(U^TAVN)$ の幾何について議論する．$\mathrm{St}(p,n)$ そのものの幾何については [1,3] を参照されたい．

3.1 接空間

$Y\in\mathrm{St}(p,n)$ における接空間 $T_Y\mathrm{St}(p,n)$ は

$$T_Y\mathrm{St}(p,n)=\left\{\xi\in\mathbb{R}^{n\times p}\,|\,\xi^TY+Y^T\xi=0\right\} \tag{3.1}$$

と書けるから，$\mathrm{St}(p,m)\times\mathrm{St}(p,n)$ 上の点 (U,V) における接空間は

$$\begin{aligned}T_{(U,V)}(\mathrm{St}(p,m)\times\mathrm{St}(p,n))&\simeq T_U\mathrm{St}(p,m)\times T_V\mathrm{St}(p,n)\\&=\left\{(\xi,\eta)\in\mathbb{R}^{m\times p}\times\mathbb{R}^{n\times p}\,|\,\xi^TU+U^T\xi=\eta^TV+V^T\eta=0\right\}\end{aligned} \tag{3.2}$$

と書かれる．$\mathrm{St}(p,n)$ はユークリッド空間 $\mathbb{R}^{n\times p}$ の部分多様体であるから，$\mathbb{R}^{n\times p}$ における通常の内積からの誘導計量

$$\langle\xi_1,\xi_2\rangle_Y:=\mathrm{tr}\left(\xi_1^T\xi_2\right),\qquad \xi_1,\xi_2\in T_Y\mathrm{St}(p,n) \tag{3.3}$$

を入れることができる．このリーマン計量を自然に拡張することで，$\mathrm{St}(p,m)\times\mathrm{St}(p,n)$ をリーマン多様体と見なす．つまり，

$$\begin{aligned}\langle(\xi_1,\eta_1),(\xi_2,\eta_2)\rangle_{(U,V)}:=\langle\xi_1,\xi_2\rangle_U+\langle\eta_1,\eta_2\rangle_V=\mathrm{tr}\left(\xi_1^T\xi_2\right)+\mathrm{tr}\left(\eta_1^T\eta_2\right),\\ (\xi_1,\eta_1),(\xi_2,\eta_2)\in T_{(U,V)}(\mathrm{St}(p,m)\times\mathrm{St}(p,n))\end{aligned} \tag{3.4}$$

というリーマン計量を入れる．

これを用いて接空間 $T_{(U,V)}(\mathrm{St}(p,m)\times\mathrm{St}(p,n))$ への射影を次のように記述できる．ここで，

$$T_{(U,V)}\left(\mathrm{St}(p,m)\times\mathrm{St}(p,n)\right)\simeq T_U\mathrm{St}(p,m)\times T_V\mathrm{St}(p,n) \tag{3.5}$$

に注意する．

命題 3.1. 任意の $(B,C)\in\mathbb{R}^{m\times p}\times\mathbb{R}^{n\times p}$ に対して，$(U,V)\in\mathrm{St}(p,m)\times\mathrm{St}(p,n)$ における接空間 $T_{(U,V)}(\mathrm{St}(p,m)\times\mathrm{St}(p,n))$ への直交射影 $P_{(U,V)}$ は，

$$P_{(U,V)}(B,C)=(P_U(B),P_V(C)) \tag{3.6}$$

で与えられる．ここで，

$$P_U(B)=B-U\,\mathrm{sym}\left(U^TB\right),\ P_V(C)=C-V\,\mathrm{sym}\left(V^TC\right) \tag{3.7}$$

はそれぞれ $T_U\mathrm{St}(p,m)$, $T_V\mathrm{St}(p,n)$ への直交射影である．なお，$\mathrm{sym}(B):=(B+B^T)/2$ は行列 B の対称部分を表す．

3.2 レトラクション

$\mathrm{St}(p,m) \times \mathrm{St}(p,n)$ 上のレトラクションの1つとして，QR分解に基づくものを紹介する．$n \times p$ のフルランク行列 B の QR 分解とは

$$B = QR, \qquad Q \in \mathrm{St}(p,n),\ R \in S_{\mathrm{upp}}^{+}(p) \tag{3.8}$$

なるものである．ここで，$S_{\mathrm{upp}}^{+}(p)$ は，対角成分がすべて正の $p \times p$ 上三角行列全体の集合である [4,7]．$\mathrm{qf}(\cdot)$ を QR 分解の Q 成分を返す写像とする．つまり，行列 B が (3.8) のように QR 分解されるなら，$\mathrm{qf}(B) = Q$ である．次のように，QR 分解を用いて $\mathrm{St}(p,m) \times \mathrm{St}(p,n)$ 上のレトラクションを定義することができる．

命題 3.2. 任意の点 $(U,V) \in \mathrm{St}(p,m) \times \mathrm{St}(p,n)$ において，$R_{(U,V)}$ を $T_{(U,V)}(\mathrm{St}(p,m) \times \mathrm{St}(p,n))$ から $\mathrm{St}(p,m) \times \mathrm{St}(p,n)$ への写像で

$$R_{(U,V)}(\xi,\eta) = (\mathrm{qf}(U+\xi), \mathrm{qf}(V+\eta)), \qquad (\xi,\eta) \in T_{(U,V)}(\mathrm{St}(p,m) \times \mathrm{St}(p,n)) \tag{3.9}$$

となるものと定義する．すると，$R_{(U,V)}$ をすべての $(U,V) \in \mathrm{St}(p,m) \times \mathrm{St}(p,n)$ にわたって集め合わせてできる $R : T(\mathrm{St}(p,m) \times \mathrm{St}(p,n)) \to \mathrm{St}(p,m) \times \mathrm{St}(p,n)$ は $\mathrm{St}(p,m) \times \mathrm{St}(p,n)$ 上のレトラクションとなる．

Proof. $(\mathrm{qf}(U+\xi), \mathrm{qf}(V+\eta)) \in \mathrm{St}(p,m) \times \mathrm{St}(p,n)$ は qf の定義から明らか．また，

$$R_{(U,V)}(0,0) = (\mathrm{qf}(U), \mathrm{qf}(V)) = (U,V). \tag{3.10}$$

さらに，$\mathrm{D}\,\mathrm{qf}(Y)[\Delta] = \Delta$ が任意の $\Delta \in T_Y \mathrm{St}(p,n)$ について成り立つ [1] から，

$$\mathrm{D}R_{(U,V)}(0,0)[(\xi,\eta)] = (\mathrm{D}\,\mathrm{qf}(U)[\xi], \mathrm{D}\,\mathrm{qf}(V)[\eta]) = (\xi,\eta) \tag{3.11}$$

が成り立つ．したがって R はレトラクションである． □

この R を QR レトラクションと呼ぶことにする．

3.3 目的関数の勾配とヘシアン

目的関数 F の $(U,V) \in \mathrm{St}(p,m) \times \mathrm{St}(p,n)$ における勾配 $\mathrm{grad}\, F(U,V)$ は

$$\langle \mathrm{grad}\, F(U,V), (\xi,\eta) \rangle_{(U,V)} = \mathrm{D}F(U,V)[(\xi,\eta)],\ (\xi,\eta) \in T_{(U,V)}(\mathrm{St}(p,m) \times \mathrm{St}(p,n)) \tag{3.12}$$

を満たす接ベクトルとして一意に定義される．また，ヘシアン $\mathrm{Hess}\, F(U,V)$ は，$T_{(U,V)}(\mathrm{St}(p,m) \times \mathrm{St}(p,n))$ 上の線形変換であり，共変微分 $\nabla_{(\xi,\eta)}\, \mathrm{grad}\, F$ の (U,V) における値として定義される．つまり，

$$\mathrm{Hess}\, F(U,V)[(\xi,\eta)] := \nabla_{(\xi,\eta)}\, \mathrm{grad}\, F, \qquad (\xi,\eta) \in T_{(U,V)}(\mathrm{St}(p,m) \times \mathrm{St}(p,n)). \tag{3.13}$$

ここで，共変微分は $\mathrm{St}(p,m) \times \mathrm{St}(p,n)$ 上のレヴィ・チビタ接続 ∇ によって定義される．

さて，ここでは F は問題 1.2 の目的関数

$$F(U,V) = -\operatorname{tr}(U^T AVN) \tag{3.14}$$

とする．

まずは勾配についてである．

命題 3.3. 目的関数 (3.14) の $(U,V) \in \mathrm{St}(p,m) \times \mathrm{St}(p,n)$ における勾配は

$$\operatorname{grad} F(U,V) = \left(U \operatorname{sym}\left(U^T AVN\right) - AVN, V \operatorname{sym}\left(V^T A^T UN\right) - A^T UN\right) \tag{3.15}$$

と書かれる．

Proof. $\operatorname{grad} F(U,V)$ は F の (U,V) におけるユークリッド勾配 $F_{(U,V)}$ を $T_{(U,V)}(\mathrm{St}(p,m) \times \mathrm{St}(p,n))$ に直交射影すれば得られるが，直交射影 $P_{(U,V)}$ は式 (3.6) と式 (3.7) で分かっているから，

$$\begin{aligned}
&\operatorname{grad} F(U,V) = P_{(U,V)}(F_{(U,V)}) = P_{(U,V)}\left(-AVN, -A^T UN\right) \\
&= \left(-P_U(AVN), -P_V(A^T UN)\right) \\
&= \left(U \operatorname{sym}\left(U^T AVN\right) - AVN, V \operatorname{sym}\left(V^T A^T UN\right) - A^T UN\right).
\end{aligned} \tag{3.16}$$

よって命題は示された． □

続いてヘシアンについてである．

命題 3.4. (ξ,η) を $(U,V) \in \mathrm{St}(p,m) \times \mathrm{St}(p,n)$ における接ベクトルとする．$S_1 = \operatorname{sym}\left(U^T AVN\right)$, $S_2 = \operatorname{sym}\left(V^T A^T UN\right)$ とする．目的関数 (3.14) の (U,V) におけるヘシアンは $T_{(U,V)}(\mathrm{St}(p,m) \times \mathrm{St}(p,n))$ 上の線形変換であり，

$$\begin{aligned}
\operatorname{Hess} F(U,V)[(\xi,\eta)] = \Big(&\xi S_1 - A\eta N - U \operatorname{sym}\left(U^T(\xi S_1 - A\eta N)\right), \\
&\eta S_2 - A^T \xi N - V \operatorname{sym}\left(V^T\left(\eta S_2 - A^T \xi N\right)\right)\Big)
\end{aligned} \tag{3.17}$$

のように作用する．

Proof. $(U(t), V(t))$ を $(U(0), V(0)) = (U,V)$, $(\dot U(0), \dot V(0)) = (\xi,\eta)$ を満たす $\mathrm{St}(p,m) \times \mathrm{St}(p,n)$ 上の測地線とする．$U(t)$ と $V(t)$ は

$$\ddot U(0) = -U\xi^T\xi, \qquad \ddot V(0) = -V\eta^T\eta \tag{3.18}$$

を満たす [1,9]．$\langle \operatorname{Hess} F(U,V)[(\xi,\eta)], (\xi,\eta)\rangle_{(U,V)}$ は $\dfrac{d^2}{dt^2}F(U,V)$ の $t=0$ における値であるから，

$$\begin{aligned}
&\langle \operatorname{Hess} F(U,V)[(\xi,\eta)], (\xi,\eta)\rangle_{(U,V)} = \left.\frac{d^2}{dt^2} F\left(U(t), V(t)\right)\right|_{t=0} \\
&= \operatorname{tr}\left(\xi^T\xi U^T AVN + U^T AV\eta^T\eta N - 2\xi^T A\eta N\right).
\end{aligned} \tag{3.19}$$

ヘシアンは対称かつ線形だから，接ベクトル (ξ,η) と (ζ,χ) に対しては，

$$
\begin{aligned}
&\langle \operatorname{Hess} F(U,V)[(\xi,\eta)],(\zeta,\chi)\rangle_{(U,V)} \\
=&\frac{1}{2}\Big(\langle \operatorname{Hess} F(U,V)[(\xi,\eta)+(\zeta,\chi)],(\xi,\eta)+(\zeta,\chi)\rangle_{(U,V)} \\
&\quad -\langle \operatorname{Hess} F(U,V)[(\xi,\eta)],(\xi,\eta)\rangle_{(U,V)}-\langle \operatorname{Hess} F(U,V)[(\zeta,\chi)],(\zeta,\chi)\rangle_{(U,V)}\Big) \\
=&\left\langle P_{(U,V)}\left((\xi S_1-A\eta N),(\eta S_2-A^T\xi N)\right),(\zeta,\chi)\right\rangle_{(U,V)} \qquad (3.20)
\end{aligned}
$$

となる．よって，

$$
\begin{aligned}
&\operatorname{Hess} F(U,V)[(\xi,\eta)]=P_{(U,V)}\left((\xi S_1-A\eta N),(\eta S_2-A^T\xi N)\right) \\
=&\Big(\xi S_1-A\eta N-U\operatorname{sym}\left(U^T(\xi S_1-A\eta N)\right),\eta S_2-A^T\xi N-V\operatorname{sym}\left(V^T\left(\eta S_2-A^T\xi N\right)\right)\Big). \qquad (3.21)
\end{aligned}
$$

これで示せた．□

4 $\mathrm{St}(p,m)\times\mathrm{St}(p,n)$ 上の最適化アルゴリズム

前の節で得られた結果を用いて，問題 1.2 の解法を考える．具体的には，最急降下法・共役勾配法・ニュートン法を導出する．

4.1 $\mathrm{St}(p,m)\times\mathrm{St}(p,n)$ 上の最急降下法

問題 2.2（リーマン多様体における制約なし問題）では，x_k における探索方向 $\Delta_k\in T_{x_k}M$ を f の $x_k\in M$ における逆勾配として選ぶ．つまり，$\Delta_k=-\operatorname{grad} f(x_k)$ とする．すると，レトラクション R を用いて更新の式は

$$
x_{k+1}=R_{x_k}(t_k\Delta_k) \qquad (4.1)
$$

と書ける．なお，t_k はアルミホのステップサイズである．

したがって最急降下法を問題 1.2 に対して導出すると，

$$
-\operatorname{grad} F(U_k,V_k)=\left(AV_kN-U_k\operatorname{sym}\left(U_k^TAV_kN\right),A^TU_kN-V_k\operatorname{sym}\left(V_k^TA^TU_kN\right)\right) \quad (4.2)
$$

に注意して，次のようなアルゴリズムが得られる．

Algorithm 4.1 問題 1.2 に対する最急降下法

1: 初期点 $(U_0, V_0) \in \mathrm{St}(p,m) \times \mathrm{St}(p,n)$ を選ぶ.
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: 探索方向 $(\xi_k, \eta_k) \in T_{(U,V)}(\mathrm{St}(p,m) \times \mathrm{St}(p,n))$ を

$$\xi_k = AV_kN - U_k \,\mathrm{sym}\left(U_k^T AV_kN\right), \ \eta_k = A^T U_kN - V_k \,\mathrm{sym}\left(V_k^T A^T U_k N\right) \tag{4.3}$$

によって計算する.
4: アルミホのステップサイズ $t_k > 0$ を計算する.
5: 次の点を $(U_{k+1}, V_{k+1}) = R_{(U_k,V_k)}(t_k(\xi_k, \eta_k))$ によって計算する. ここで, R は $\mathrm{St}(p,m) \times \mathrm{St}(p,n)$ 上のレトラクションである.
6: **end for**

リーマン多様体 M がコンパクトで目的関数 f が滑らかなら, 最急降下法で得られる点列 $\{x_k\}$ は

$$\lim_{k\to\infty} \|\,\mathrm{grad}\, f(x_k)\|_{x_k} = 0 \tag{4.4}$$

を満たすことが知られている [1]. $\mathrm{St}(p,m) \times \mathrm{St}(p,n)$ はコンパクトだから, アルゴリズム 4.1 によって得られる点列は F の臨界点に収束する. ところが, 実際に数値実験してみると [9], その収束は非常に遅く, 問題 1.2 に対しては実用的とはいえないことが分かる. そこで, 共役勾配法を考えてみる.

4.2 $\mathrm{St}(p,m) \times \mathrm{St}(p,n)$ 上の共役勾配法

まずユークリッド空間 $\mathbb{R}^N$ における問題 2.1 に対する (非線形) 共役勾配法を簡単に復習しておく. まず, $x_k \in \mathbb{R}^N$ における探索方向 $\Delta_k \in \mathbb{R}^N$ は,

$$\Delta_k = -\,\mathrm{grad}\, f(x_k) + \beta_k \Delta_{k-1}, \qquad k \geq 1 \tag{4.5}$$

として選ぶ. ただし, $\Delta_0 = -\,\mathrm{grad}\, f(x_0)$ とする. ここで, $\beta_k \in \mathbb{R}$ の選び方は多数提案されているが, 本稿では Polak-Ribiere による

$$\beta_k = \frac{\mathrm{grad}\, f(x_k)^T (\mathrm{grad}\, f(x_k) - \mathrm{grad}\, f(x_{k-1}))}{\mathrm{grad}\, f(x_{k-1})^T\, \mathrm{grad}\, f(x_{k-1})} \tag{4.6}$$

を採用する.

共役勾配法をリーマン多様体 M 上に拡張する際に問題となる部分を見ていこう. まず, β_k についてであるが, (4.6) の右辺の分子において, $\mathrm{grad}\, f(x_k) - \mathrm{grad}\, f(x_{k-1})$ という部分がある. リーマン多様体 M 上では $\mathrm{grad}\, f(x_k) \in T_{x_k}M$, $\mathrm{grad}\, f(x_{k-1}) \in T_{x_{k-1}}M$ は異なる接空間の元であるから, この 2 つの差をとることはできない. 同様に, (4.5) の右辺においても, M 上では $\mathrm{grad}\, f(x_k) \in T_{x_k}M$, $\Delta_{k-1} \in T_{x_{k-1}}M$ となるから, やはり (4.5) の右辺は計算できない (定義されない). そこで, Smith [10] は平行移動を用いてこの問題点を解決することを提案した. ところが, 平行移動は必ずしも計算できるとは限らないし,

計算できたとしてもさほど効率的ではない．そこで，Absil [1] は vector transport という概念を導入した．

定義 4.1. 写像

$$\mathcal{T} : TM \oplus TM \to TM : (\eta_x, \xi_x) \mapsto \mathcal{T}_{\eta_x}(\xi_x) \in TM \tag{4.7}$$

が M 上の *vector transport* であるとは，任意の $x \in M$ に対して次の条件が満たされるときをいう．

1. レトラクション R が存在して，

$$\pi\left(\mathcal{T}_{\eta_x}(\xi_x)\right) = R\left(\eta_x\right) \tag{4.8}$$

   が成り立つ．ここで，$\pi\left(\mathcal{T}_{\eta_x}(\xi_x)\right)$ は接ベクトル $\mathcal{T}_{\eta_x}(\xi_x)$ の足を表す．

2. 任意の $\xi_x \in T_xM$ に対して $\mathcal{T}_{0_x}(\xi_x) = \xi_x$ が成り立つ．

3. $\mathcal{T}_{\eta_x}(a\xi_x + b\zeta_x) = a\mathcal{T}_{\eta_x}(\xi_x) + b\mathcal{T}_{\eta_x}(\zeta_x)$ が成り立つ．

この vector transport を用いて，リーマン多様体 M 上の共役勾配法の探索方向を

$$\eta_{k+1} = -\operatorname{grad} f(x_{k+1}) + \beta_{k+1}\mathcal{T}_{\alpha_k\eta_k}(\eta_k) \tag{4.9}$$

のように決める．また，β_k は

$$\beta_k = \frac{\langle \operatorname{grad} f(x_k), \operatorname{grad} f(x_k) - \mathcal{T}_{\alpha_{k-1}\eta_{k-1}}(\operatorname{grad} f(x_{k-1}))\rangle_{x_k}}{\langle \operatorname{grad} f(x_{k-1}), \operatorname{grad} f(x_{k-1})\rangle_{x_{k-1}}} \tag{4.10}$$

とする．

さて，$\mathrm{St}(p,m) \times \mathrm{St}(p,n)$ 上の vector transport $\mathcal{T}$ の一つは次のようにして計算できる．

命題 4.1. $M = \mathrm{St}(p,m) \times \mathrm{St}(p,n)$ に対して，$\mathcal{T} : TM \oplus TM \to TM$ を

$$\mathcal{T}_{(\xi,\eta)}(\zeta,\chi) := \left(\zeta - \mathrm{qf}(U+\xi)\,\mathrm{sym}\left(\mathrm{qf}(U+\xi)^T\zeta\right), \chi - \mathrm{qf}(V+\eta)\,\mathrm{sym}\left(\mathrm{qf}(V+\eta)^T\chi\right)\right) \tag{4.11}$$

で定義する．ここで，$(\xi,\eta), (\zeta,\chi) \in T_{(U,V)}\left(\mathrm{St}(p,m) \times \mathrm{St}(p,n)\right)$ である．このとき，$\mathcal{T}$ は $\mathrm{St}(p,m) \times \mathrm{St}(p,n)$ 上の *vector transport* である．

Proof. 一般に，M がユークリッド空間の部分多様体であり，レトラクション R を持つならば，

$$\mathcal{T}_{\eta_x}(\xi_x) = P_{R_x(\eta_x)}(\xi_x), \qquad \eta_x, \xi_x \in T_xM, \tag{4.12}$$

によって定義される $\mathcal{T}$ は M 上の vector transport となる [1]． ここで，$P_{R_x(\eta_x)}$ は M 上の $R_x(\eta_x)$ における接空間への射影である．よって $M = \mathrm{St}(p,m) \times \mathrm{St}(p,n)$ に QR レトラクション R ((3.9)) を付随させて考えれば，既知である直交射影の式から (4.11) は従う．したがって，(4.11) によって定義される $\mathcal{T}$ は vector transport である． □

また，ユークリッド空間における通常の共役勾配法ではウルフのステップサイズが用いられることが多い．そこで，リーマン多様体上にもウルフのステップサイズを拡張し，それを用いて共役勾配法を行うことが望ましいと考えられる．このことについても，詳細は [9] を参照されたい．

これで問題 1.2 に対しても共役勾配法を適用することができる．ただし紙面の都合上，詳細なアルゴリズムは割愛する．数値実験によれば，最急降下法よりかなり速く臨界点に収束することが分かる [9].

4.3 $\mathrm{St}(p,m) \times \mathrm{St}(p,n)$ 上のニュートン法

次にニュートン法について考える．一般のリーマン多様体 M 上の最適化問題 2.2 に対するニュートン法では，$x_k \in M$ における探索方向 $\Delta_k \in T_{x_k}M$ はニュートン方程式

$$\operatorname{Hess} f(x_k)[\Delta_k] = -\operatorname{grad} f(x_k) \tag{4.13}$$

の解として決める．なお，ステップサイズは $t_k = 1$ とすることにする．問題 1.2 の目的関数の勾配やヘシアンは既知であるから，それらを用いてニュートン法のアルゴリズムは次のようになる．

Algorithm 4.2 問題 1.2 に対するニュートン法

1: 初期点 $(U_0, V_0) \in \mathrm{St}(p,m) \times \mathrm{St}(p,n)$ を選ぶ.
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: ニュートン方程式

$$\begin{cases} \xi_k S_{1,k} - A\eta_k N - U_k \operatorname{sym}\left(U_k^T(\xi_k S_{1,k} - A\eta_k N)\right) = AV_k N - U_k S_{1,k}, \\ \eta_k S_{2,k} - A^T\xi_k N - V_k \operatorname{sym}\left(V_k^T\left(\eta_k S_{2,k} - A^T\xi_k N\right)\right) = A^T U_k N - V_k S_{2,k} \end{cases} \tag{4.14}$$

を $(\xi_k, \eta_k) \in T_{(U_k,V_k)}(\mathrm{St}(p,m) \times \mathrm{St}(p,n))$ について解く．ここで，$S_{1,k} = \operatorname{sym}\left(U_k^T A V_k N\right)$，$S_{2,k} = \operatorname{sym}\left(V_k^T A^T U_k N\right)$ である．
4: 次の点を $(U_{k+1}, V_{k+1}) = R_{(U_k,V_k)}((\xi_k, \eta_k))$ によって計算する．ここで，R は $\mathrm{St}(p,m) \times \mathrm{St}(p,n)$ 上のレトラクションである．
5: **end for**

ただし，このアルゴリズムにおけるニュートン方程式 (4.14) を解くのは困難である．そこで，$p = 1$ の場合を考える．この場合，ニュートン方程式 (4.14) は

$$S_k\xi_k - \left(I_m - U_k U_k^T\right) A\eta_k = \left(I_m - U_k U_k^T\right) AV_k, \tag{4.15}$$

$$S_k\eta_k - \left(I_n - V_k V_k^T\right) A^T\xi_k = \left(I_n - V_k V_k^T\right) A^T U_k \tag{4.16}$$

という簡単な形になる．ここで，$S_{1,k} = S_{2,k} =: S_k$ としており，さらに S_k はスカラーである．

もし $S_k \neq 0$ かつ $S_k^2 I_n - \left(I_n - V_k V_k^T\right) A^T \left(I_m - U_k U_k^T\right) A$ が正則であれば，方程式 (4.15), newtoneqp12 は

$$\eta_k = \left(S_k^2 I_n - \left(I_n - V_k V_k^T\right) A^T \left(I_m - U_k U_k^T\right) A\right)^{-1} \left(I_n - V_k V_k^T\right) A^T A V_k, \tag{4.17}$$

$$\xi_k = S_k^{-1} \left(I_m - U_k U_k^T\right) A \left(\eta_k + V_k\right) \tag{4.18}$$

と解くことができる．

こうして $p = 1$ のときのニュートン法は次のように書くことができる．

Algorithm 4.3 $p = 1$ のときの問題 1.2 に対するニュートン法

1: 初期点 $(u_0, v_0) \in \mathrm{St}(1, m) \times \mathrm{St}(1, n) = S^{m-1} \times S^{n-1}$ を選ぶ．
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: η_k と ξ_k を

$$\eta_k = \left(s_k^2 I_n - \left(I_n - v_k v_k^T\right) A^T \left(I_m - u_k u_k^T\right) A\right)^{-1} \left(I_n - v_k v_k^T\right) A^T A v_k, \tag{4.19}$$

$$\xi_k = s_k^{-1} \left(I_m - u_k u_k^T\right) A \left(\eta_k + v_k\right) \tag{4.20}$$

によって計算する．ここで，$s_k = u_k^T A v_k$ である．
4: 次の点を $(u_{k+1}, v_{k+1}) = R_{(u_k, v_k)}\left((\xi_k, \eta_k)\right)$ によって計算する．
5: **end for**

ニュートン法は，ヘシアンが退化していない臨界点の近傍に初期点を取れば，その臨界点に二次収束する点列を生成することが知られている．しかし，必ずしも最適解に収束するという保証はない．そこで，共役勾配法とニュートン法を組み合わせた方法を次節で提案する．

5 $\mathrm{St}(p, m) \times \mathrm{St}(p, n)$ 上の最適化手法に基づく新しい特異値分解手法

まず，$p = 1$ のときの問題 1.2 の解法から，行列 A の最大特異値とそれに属する特異ベクトルを求めるアルゴリズムを提案する．それは，前節で議論したように共役勾配法とニュートン法を組み合わせたものである．

Algorithm 5.1 $p = 1$ のときの問題1.2に対するハイブリッド法

初期点 $(U_0, V_0) \in S^{m-1} \times S^{n-1}$ を選び，パラメータ $\varepsilon > 0$ を決める．$k := 0$ とおく．

$$\begin{aligned}(\xi_0, \eta_0) &= -\operatorname{grad} F(U_0, V_0) \\ &= \left(AV_0N - U_0 \operatorname{sym}\left(U_0^T AV_0 N\right), A^T U_0 N - V_0 \operatorname{sym}\left(V_0^T A^T U_0 N\right)\right) \\ &= \left(AV_0 - U_0 U_0^T AV_0, A^T U_0 - V_0 V_0^T A^T U_0\right) \end{aligned} \tag{5.1}$$

とし，$\left(\bar{\xi}_0, \bar{\eta}_0\right) = (\xi_0, \eta_0)$ とする．

while $\| \operatorname{grad} F(U_k, V_k) \| > \varepsilon$ **do**

　共役勾配法の反復部分を実行する．

　$k := k + 1$.

end while

$(u_0, v_0) := (U_k, V_k)$ とおき，$k := 0$ とおく．

アルゴリズム 4.3 を実行する．

しかし，実際にはもちろん一般の p に対して問題の解法アルゴリズムを得たい．ただし，先に述べたように，ニュートン法をそのまま実行するのは $p = 1$ の場合を除いて困難だから，問題を分割してアルゴリズム 4.3 ($p = 1$ のニュートン法) を p 回反復するという方法を提案する．具体的には，まず最初に共役勾配法で最適解に十分近い点 $\left(\tilde{U}, \tilde{V}\right) \in \mathrm{St}(p, m) \times \mathrm{St}(p, n)$ を得ておく．$u_1, \ldots, u_p$ と $v_1, \ldots, v_p$ をそれぞれ $\tilde{U}$ と $\tilde{V}$ の列ベクトルとする．つまり，$\tilde{U} = (u_1, \ldots, u_p), \tilde{V} = (v_1, \ldots, v_p)$ である．すると，u_i や v_i は，大きい方から i 番目の特異値に属する特異ベクトル u_i^* や v_i^* に十分近い．すると，関数 $\operatorname{tr}\left(u_i^T A v_i\right)$ の値は特異値 σ_i と十分近くなる．そこで，i を一つ固定するたびに先ほどの $p = 1$ の場合のニュートン法を点 (u_i, v_i) に適用すると，得られる点列は (u_i^*, v_i^*) に二次収束する．こうしてニュートン法で u_i^* と v_i^* を $i = 1, \ldots, p$ について得ることができれば，それらを並べて U_* と V_* を作る．つまり，$U_* = \left(u_1^*, \ldots, u_p^*\right)$，$V_* = \left(v_1^*, \ldots, v_p^*\right)$ とする．特異値が互いに異なるとき，特異ベクトルは互いに直交するから，$U_* \in \mathrm{St}(p, m)$ と $V_* \in \mathrm{St}(p, n)$ が言える．アルゴリズムは次の通り．

Algorithm 5.2 問題 1.2 に対するハイブリッド法

初期点 $(U_0, V_0) \in \mathrm{St}(p, m) \times \mathrm{St}(p, n)$ を選び，パラメータ $\varepsilon > 0$ を決める． $k := 0$ とおく．

$$
\begin{aligned}
(\xi_0, \eta_0) &= -\operatorname{grad} F(U_0, V_0) \\
&= \left(AV_0N - U_0 \operatorname{sym}\left(U_0^T A V_0 N\right), A^T U_0 N - V_0 \operatorname{sym}\left(V_0^T A^T U_0 N\right)\right)
\end{aligned} \tag{5.2}
$$

とし，$\left(\bar{\xi}_0, \bar{\eta}_0\right) = (\xi_0, \eta_0)$ とする．

while $\|\operatorname{grad} F(U_k, V_k)\| > \varepsilon$ **do**

　共役勾配法の反復部分を実行する．

　$k := k + 1$.

end while

$\left(\tilde{U}, \tilde{V}\right) := (U_k, V_k)$ とおく．

for $i = 1, 2, \ldots, p$ **do**

　$(u_0, v_0) := \left(\tilde{U}_i, \tilde{V}_i\right)$ とおき，$k := 0$ とおく．ここで，$\tilde{U}_i$ と $\tilde{V}_i$ は，それぞれ $\tilde{U}$ と $\tilde{V}$ の i 番目の列ベクトルである．

　アルゴリズム 4.3 を実行する．

end for

6　終わりに

行列の特異値分解を，積多様体 $\mathrm{St}(p, m) \times \mathrm{St}(p, n)$ 上の最適化問題に帰着させ，その解法アルゴリズムを導出することによって，新たな特異値分解手法を得た．本稿では数値実験については扱わなかったが，詳細は [9] にある．その実験結果も踏まえて提案手法は次のような利点を持つ．まず一つには，通常の特異値分解手法に必要な前処理を必要としない．あえて言えば，アルゴリズム 5.2 において，共役勾配法で近似解を得るところが前処理と見なせるかもしれない．見方を変えれば，収束性の要はあくまでニュートン法であり，近似解を求めるところは共役勾配法でやらずとも良い．たとえば MATLAB の svd 関数で求めた特異値分解を初期点としてニュートン法を適用することで，より良い解を得るということもできる．これが第二の利点であり，精度の高い特異値分解を実現することができる．

また，行列 A の特異値が縮退している場合についてのアルゴリズムの振る舞いについても詳しく調べる必要があるが，そちらについては [9] を参照されたい．そのような場合でも，提案手法で特異値分解は達成されることが分かっている．

なお，最後に，本稿で扱ったリーマン多様体上の共役勾配法の収束性についてはさほど一般的な研究がなされておらず，詳しく解析されるべき事柄である．

参考文献

[1] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, Princeton, NJ, 2008.

[2] R. L. Adler, J. P. Dedieu, J. Y. Margulies, M. Martens, and M. Shub, Newton's method on Riemannian manifolds and a geometric model for the human spine, *IMA J. Numer. Anal.,* **22**(3) (2002), pp. 359–390.

[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. H. Golub and C. F. Van Loan, *Matrix Computations*, Third ed., The Johns Hopkins University Press, Baltimore, MD, 1996.

[5] U. Helmke, J. B. Moore, *Optimization and Dynamical Systems*, in: Communications and Control Engineering, Springer, London, 1994.

[6] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bur. Stand., **49** (1952), pp. 409–436.

[7] L. N. Trefethen and D. Bau, *Numerical Linear Algebra*, SIAM, 1997.

[8] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer Series in Operations Research. Springer-Verlag, New York, 1999.

[9] H. Sato and T. Iwai, A Riemannian optimization approach to the matrix singular value decomposition, *SIAM J. Optim.*, to appear.

[10] S. T. Smith, *Optimization techniques on Riemannian manifolds*, in Hamiltonian and gradient flows, algorithms and control, Fields Inst. Commun., **3**. Amerl Math. Soc., Providence, RI, 1994, pp. 113–136.