

シュティーフェル多様体上のニュートン法と その収束性解析

東京理科大学・工学部経営工学科 佐藤 寛之

東京理科大学・理学部数理情報科学科 相原 研輔

Hiroyuki Sato¹

Department of Management Science, Tokyo University of Science

Kensuke Aihara

Department of Mathematical Information Science, Tokyo University of Science

概要

ユークリッド空間における最適化アルゴリズムは多数あるが、その中でも制約条件なしの問題に対しては、最急降下法やニュートン法、あるいは共役勾配法など、性質の良いものが知られている。一方、制約条件付きの最適化問題に対しては、これらのアルゴリズムにより生成される点列は制約条件を満たさないため、制約条件付き問題に特化した他の最適化手法が用いられることが多い。しかし、制約条件付きの最適化問題であっても、その制約条件を満たす点全体がリーマン多様体をなす場合には、ニュートン法などをその多様体上に一般化したアルゴリズムを用いることができる。

本稿では、特異値分解と密接な関連を持つシュティーフェル多様体上の最適化問題を扱い、この問題に対するニュートン法について詳しく議論する。特に、ニュートン法の各反復において現れるニュートン方程式が線形方程式であることに着目し、この方程式に対する線形反復解法を提案する。また、ニュートン方程式を求解する際の精度とニュートン法自体の収束率の関係についても議論する。

1 序論

本稿では次の最適化問題を考える。

$$\begin{aligned} & \text{maximize} \quad \text{tr}(U^T A V N), \\ & \text{subject to} \quad U^T U = V^T V = I_p, \quad U \in \mathbb{R}^{m \times p}, \quad V \in \mathbb{R}^{n \times p}. \end{aligned} \quad (1.1)$$

ここで、正の整数 m, n, p は $m \geq n \geq p$ を満たし、 $A \in \mathbb{R}^{m \times n}$, $N = \text{diag}(\mu_1, \mu_2, \dots, \mu_p)$, $\mu_1 > \mu_2 > \dots > \mu_p > 0$ である。

数値線形代数の観点からは、問題 (1.1) を解くことは A の打ち切り特異値分解 [4] を求めることと等価である。すなわち、問題 (1.1) の最適解を (U_*, V_*) とすると、 U_*, V_* の列ベ

¹e-mail: hsato[at]ms.kagu.tus.ac.jp

クトルがそれぞれ A の大きい方から p 個の特異値に属する左特異ベクトル, 右特異ベクトルであることが示されている [7].

最適化の観点からは, この問題はユークリッド空間 $\mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p}$ における $U^T U = V^T V = I_p$ という制約条件付きの最適化問題であり, このままではニュートン法や共役勾配法などの無制約最適化手法 [5] を適用することができない. しかし, シュティーフェル多様体と呼ばれる多様体

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

を考えると, 問題 (1.1) の制約条件は $U \in \text{St}(p, m), V \in \text{St}(p, n)$ と書けるから, 問題 (1.1) は次のように積多様体 $\text{St}(p, m) \times \text{St}(p, n)$ 上の無制約最適化問題と見なすことができる.

$$\begin{aligned} & \text{minimize} && F(U, V) = -\text{tr}(U^T A V N), \\ & \text{subject to} && (U, V) \in \text{St}(p, m) \times \text{St}(p, n). \end{aligned} \quad (1.2)$$

本稿では, [7] で提案された, この問題に対するニュートン法を紹介するとともに, 一般の p に対するニュートン方程式の難しさについて述べ, それを解決する工夫について議論する. 具体的には, 複雑な行列方程式の形で表されるニュートン方程式を等価な連立一次方程式に変換し, クリロフ部分空間法的一种である共役残差法 [3] を適用する. さらに, ニュートン法の各反復におけるニュートン方程式の解の精度とニュートン法の収束率の関係についても議論する. 最後に, 提案手法の収束性を数値実験により検証する.

2 シュティーフェル多様体上のニュートン法

ユークリッド空間 $\mathbb{R}^N$ における制約条件なしの問題

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

に対するニュートン法では, 最適解の十分近くの初期点 x_0 から始め, k 回目の反復では $\eta \in \mathbb{R}^N$ に対するニュートン方程式

$$\nabla^2 f(x_k) \eta = -\nabla f(x_k) \quad (2.1)$$

を解き, その解を η_k として次の点 x_{k+1} を

$$x_{k+1} = x_k + \eta_k \quad (2.2)$$

によって計算する. ここで, ∇f および $\nabla^2 f$ はそれぞれ f の勾配, ヘッセ行列であり, η_k は探索方向と呼ばれる. この手法を $\text{St}(p, m) \times \text{St}(p, n)$ 上に拡張し, 問題 (1.2) に適用できるようにしたい. ここでは, [7] で提案されたアルゴリズムを紹介する.

まず, ニュートン方程式 (2.1) に現れる勾配やヘッセ行列は目的関数のみの情報しか有していないため, これらを多様体 $\text{St}(p, m) \times \text{St}(p, n)$ 上に拡張する必要がある. そのために, リーマン計量を多様体 $\text{St}(p, m) \times \text{St}(p, n)$ に導入する必要があるが, $\text{St}(p, m) \times \text{St}(p, n)$

は $\mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p}$ の部分多様体であるため, $\mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p}$ の標準内積から誘導される計量を考えることができる. すなわち, $\text{St}(p, m) \times \text{St}(p, n)$ 上の各点 (U, V) における接空間 $T_{(U, V)}(\text{St}(p, m) \times \text{St}(p, n))$ において,

$$\langle (\xi, \eta), (\xi', \eta') \rangle_{(U, V)} = \text{tr}(\xi^\top \xi') + \text{tr}(\eta^\top \eta'), \quad (\xi, \eta), (\xi', \eta') \in T_{(U, V)}(\text{St}(p, m) \times \text{St}(p, n)) \quad (2.3)$$

によって定まる内積 $\langle \cdot, \cdot \rangle_{(U, V)}$ を定義し, これをすべての点 (U, V) について集めた族を $\text{St}(p, m) \times \text{St}(p, n)$ のリーマン計量とすることができる. このリーマン計量を与えることで, $\text{St}(p, m) \times \text{St}(p, n)$ はリーマン多様体となり, その幾何学的構造がもたらす情報を含むものとして目的関数 F の勾配 $\text{grad } F$ やヘシアン $\text{Hess } F$ を定義することができる. 実際, リーマン計量により点 (U, V) における $\text{St}(p, m) \times \text{St}(p, n)$ の法空間を定義することができ, 接空間 $T_{(U, V)}(\text{St}(p, m) \times \text{St}(p, n))$ への直交射影 $P_{(U, V)} : \mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p} \rightarrow T_{(U, V)}(\text{St}(p, m) \times \text{St}(p, n))$ が

$$P_{(U, V)}(X, Y) = (X - U \text{sym}(U^\top X), Y - V \text{sym}(V^\top Y))$$

により定まる. ここで, $\text{sym}(\cdot)$ は行列の対称部分を返す写像, すなわち $\text{sym}(Z) = (Z + Z^\top)/2$ である. 目的関数 F の $(U, V) \in \text{St}(p, m) \times \text{St}(p, n)$ における勾配 $\text{grad } F$ は $\mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p}$ におけるユークリッド勾配を直交射影することで計算でき,

$$\text{grad } F(U, V) = P_{(U, V)}(\nabla \bar{F}(U, V)) \quad (2.4)$$

となる. ここで, $\bar{F}$ は $\mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p}$ 上で $\bar{F}(U, V) := -\text{tr}(U^\top A V N)$ と定義される関数である. すなわち, $\bar{F}|_{\text{St}(p, m) \times \text{St}(p, n)} = F$ である. さらに, 勾配 $\text{grad } F$ の微分をさらに射影することで, $(U, V) \in \text{St}(p, m) \times \text{St}(p, n)$ におけるヘシアン $\text{Hess } F(U, V)$ の $(\xi, \eta) \in T_{(U, V)}(\text{St}(p, m) \times \text{St}(p, n))$ の作用を

$$\text{Hess } F(U, V)[(\xi, \eta)] = P_{(U, V)}(D(\text{grad } F)(U, V)[(\xi, \eta)]) \quad (2.5)$$

と表すことができる.

(2.4) および (2.5) から勾配とヘシアンを具体的に計算すると, $(U, V) \in \text{St}(p, m) \times \text{St}(p, n)$, $(\xi, \eta) \in T_{(U, V)}(\text{St}(p, m) \times \text{St}(p, n))$ に対して,

$$\text{grad } F(U, V) = (U S_1 - A V N, V S_2 - A^\top U N),$$

$$\begin{aligned} \text{Hess } F(U, V)[(\xi, \eta)] = & \left(\xi S_1 - A \eta N - U \text{sym}(U^\top (\xi S_1 - A \eta N)), \right. \\ & \left. \eta S_2 - A^\top \xi N - V \text{sym}(V^\top (\eta S_2 - A^\top \xi N)) \right) \end{aligned} \quad (2.6)$$

となる. ここで,

$$S_1 = \text{sym}(U^\top A V N), \quad S_2 = \text{sym}(V^\top A^\top U N)$$

である. これらを用いると, ユークリッド空間上のニュートン法で現れるニュートン方程式 (2.1) を $\text{St}(p, m) \times \text{St}(p, n)$ 上に拡張した方程式

$$\text{Hess } F(U_k, V_k)[(\xi_k, \eta_k)] = -\text{grad } F(U_k, V_k)$$

を具体的に書き下すことができる。簡単のため反復回数 k を省略すると、問題 (1.2) に対するニュートン方程式は

$$\begin{cases} \xi S_1 - A\eta N - U \operatorname{sym}(U^\top(\xi S_1 - A\eta N)) = AVN - US_1, \\ \eta S_2 - A^\top \xi N - V \operatorname{sym}(V^\top(\eta S_2 - A^\top \xi N)) = A^\top UN - VS_2 \end{cases} \quad (2.7)$$

と書ける。

一般の p に対してニュートン方程式 (2.7) を解くのは一見すると困難で、それを解決するための工夫についての議論が本稿の主題の一つであり、次節以降で詳しく述べる。ここでは、第 k 反復で点 (U_k, V_k) におけるニュートン方程式が解けたとする。その解を $(\xi_k, \eta_k) \in T_{(U_k, V_k)}(\operatorname{St}(p, m) \times \operatorname{St}(p, n))$ とすると、ユークリッド空間における点列の更新式 (2.2) を $\operatorname{St}(p, m) \times \operatorname{St}(p, n)$ 上に拡張しておけば、ニュートン法の 1 反復が完成する。一般にユークリッド空間における最適化手法では、点 x_k から探索方向 η_k が定義する半直線上に次の点を定めることになるが、多様体 $\operatorname{St}(p, m) \times \operatorname{St}(p, n)$ 上では半直線の代わりに、現在の点 (U_k, V_k) において探索方向 (ξ_k, η_k) に接する曲線を考えれば良い。そこで、行列を引数としてその QR 分解の Q 成分を返す写像 $\operatorname{qf}(\cdot)$ を用いて

$$(U_{k+1}, V_{k+1}) = (\operatorname{qf}(U_k + \xi_k), \operatorname{qf}(V_k + \eta_k))$$

とすれば適切な更新式になることが示されている [1, 7]。これは、 $(U_k, V_k) + (\xi_k, \eta_k) = (U_k + \xi_k, V_k + \eta_k) \in \mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p}$ を $\operatorname{St}(p, m) \times \operatorname{St}(p, n)$ に引き戻すような写像、すなわちレトラクションと呼ばれる写像を考えていることになる。なお、一般のリーマン多様体上の最適化については [1] が詳しい。

ここまでの議論をアルゴリズムの形でまとめておく。

アルゴリズム 1 問題 (1.2) に対するニュートン法

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

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

を $(\xi_k, \eta_k) \in T_{(U_k, V_k)}(\operatorname{St}(p, m) \times \operatorname{St}(p, n))$ について解く。ここで、 $S_{1,k} = \operatorname{sym}(U_k^\top AV_k N)$, $S_{2,k} = \operatorname{sym}(V_k^\top A^\top U_k N)$ である。

- 4: 次の点を $(U_{k+1}, V_{k+1}) = (\operatorname{qf}(U_k + \xi_k), \operatorname{qf}(V_k + \eta_k))$ により計算する。
 - 5: **end for**
-

3 シュティーフェル多様体上のニュートン方程式の解法

本節では問題 (1.2) に対するニュートン方程式 (2.7) の解法について議論する。

3.1 ニュートン方程式の標準的な線形方程式への変換

方程式 (2.7) はそれ自体が複雑な形をしているが、さらに $(\xi, \eta) \in T_{(U,V)}(\text{St}(p, m) \times \text{St}(p, n))$ という条件があるため、これも考慮する必要がある。シュテューフェル多様体 $\text{St}(p, n)$ 上の点 Y における接空間 $T_Y \text{St}(p, n)$ は

$$T_Y \text{St}(p, n) = \{\xi \mid \xi^\top Y + Y^\top \xi = 0\}$$

と表せるため [2], $(\xi, \eta) \in T_{(U,V)}(\text{St}(p, m) \times \text{St}(p, n))$ は $\xi^\top U + U^\top \xi = \eta^\top V + V^\top \eta = 0$ を満たさなければならない。この条件を扱いやすくするため、本稿では [2] で示されている接空間の別の表現

$$T_Y \text{St}(p, n) = \{YB + Y_\perp C \mid B \in \text{Skew}(p), C \in \mathbb{R}^{(n-p) \times p}\} \quad (3.1)$$

を用いる。ここで、 $Y_\perp \in \mathbb{R}^{n \times (n-p)}$ は $Y^\top Y_\perp = 0$ かつ $Y_\perp^\top Y_\perp = I_{n-p}$ を満たす任意の行列である。また、 $\text{Skew}(p)$ は p 次反対称行列全体の集合を表す。なお、 $Y_\perp$ を固定すると、 $\xi \in T_Y \text{St}(p, n)$ に対して $\xi = YB + Y_\perp C$ となる $B \in \text{Skew}(p), C \in \mathbb{R}^{(n-p) \times p}$ が一意に存在する。

以後、 $U^\top U_\perp = 0, U_\perp^\top U_\perp = I_{m-p}, V^\top V_\perp = 0, V_\perp^\top V_\perp = I_{n-p}$ を満たす $U_\perp \in \mathbb{R}^{m \times (m-p)}$ および $V_\perp \in \mathbb{R}^{n \times (n-p)}$ を固定しておく。式 (3.1) から、 $(\xi, \eta) \in T_{(U,V)}(\text{St}(p, m) \times \text{St}(p, n))$ は

$$(\xi, \eta) = (UB_x + U_\perp C_x, VD_x + V_\perp E_x) \quad (3.2)$$

と書ける。ここで、 $B_x, D_x \in \text{Skew}(p)$, $C_x \in \mathbb{R}^{(m-p) \times p}$, $E_x \in \mathbb{R}^{(n-p) \times p}$ である。さらに、勾配 $\text{grad } F(U, V)$ もまた (U, V) における接ベクトルであるので、 $B_g, D_g \in \text{Skew}(p)$, $C_g \in \mathbb{R}^{(m-p) \times p}$, $E_g \in \mathbb{R}^{(n-p) \times p}$ を用いて

$$\text{grad } F(U, V) = (UB_g + U_\perp C_g, VD_g + V_\perp E_g)$$

と書ける。したがって、ニュートン方程式 (2.7) を解くことは、ヘシアン の作用によって $(-B_g, -C_g, -D_g, -E_g)$ に移されるような (B_x, C_x, D_x, E_x) を求めることと等価である。そのためにはヘシアン の作用を詳しく解析する必要があるが、ヘシアン は線形性を持つため適当な基底の下での表現行列がわかれば良いので、 (B_x, C_x, D_x, E_x) の成分を 1 列に並べてベクトル表示することを考える。ただし、 B_x と D_x は反対称行列であるため、狭義下三角部分の成分のみが独立であると考えことにする。そこで、行列 $X = (x_{i,j}) \in \mathbb{R}^{m \times n}$ に対して作用素 $\text{vec}(\cdot)$ を

$$\text{vec}(X) = [x_{1,1} \ x_{2,1} \ \cdots \ x_{m,1} \mid x_{1,2} \ x_{2,2} \ \cdots \ x_{m,2} \mid \cdots \mid x_{1,n} \ x_{2,n} \ \cdots \ x_{m,n}]^\top \in \mathbb{R}^{mn}$$

により定義する。同様に、反対称行列 $Z = (z_{i,j}) \in \text{Skew}(p)$ に対して作用素 $\text{veck}(\cdot)$ を

$$\text{veck}(Z) = [z_{2,1} \ z_{3,1} \ \cdots \ z_{p,1} \mid z_{3,2} \ z_{4,2} \ \cdots \ z_{p,2} \mid \cdots \mid z_{p,(p-1)}]^\top \in \mathbb{R}^{p(p-1)/2}$$

と定める．すなわち， $\text{veck}(Z)$ は Z の狭義下三角部分の成分のみを左の列から順に縦に並べて得られる $p(p-1)/2$ 次元ベクトルである．これらの作用素を用いて

$$\mathbf{x} := \begin{bmatrix} \text{veck}(B_x) \\ \text{vec}(C_x) \\ \text{veck}(D_x) \\ \text{vec}(E_x) \end{bmatrix}, \quad \mathbf{g} := - \begin{bmatrix} \text{veck}(B_g) \\ \text{vec}(C_g) \\ \text{veck}(D_g) \\ \text{vec}(E_g) \end{bmatrix}$$

とすることで，ニュートン方程式 (2.7) を

$$H\mathbf{x} = \mathbf{g} \quad (3.3)$$

と書くことができる．ここで， H は $\text{Hess } F(U, V)$ の表現行列である．なお，線形方程式 (3.3) の次元 $K := p(m+n-p-1)$ は接空間 $T_{(U,V)}(\text{St}(p, m) \times \text{St}(p, n))$ の次元と一致していることに注意する．すなわち，(3.3) は $T_{(U,V)}(\text{St}(p, m) \times \text{St}(p, n))$ をベクトル空間 $\mathbb{R}^K$ と同一視したときの標準的な線形方程式である．

3.2 目的関数のヘシアン表現行列とその性質

方程式 (3.3) の解法を探るために，ヘシアンの表現行列 H を詳しく調べる必要がある．そのためには，まず $B_H, D_H \in \text{Skew}(p), C_H \in \mathbb{R}^{(m-p) \times p}, E_H \in \mathbb{R}^{(n-p) \times p}$ により

$$\text{Hess } F(U, V)[(\xi, \eta)] = (UB_H + U_\perp C_H, VD_H + V_\perp E_H) \quad (3.4)$$

とし， (B_x, C_x, D_x, E_x) が (B_H, C_H, D_H, E_H) にどのように変換されるかを調べれば良い．式 (2.6) から，

$$\begin{cases} UB_H + U_\perp C_H = \xi S_1 - A\eta N - U \text{sym}(U^\top (\xi S_1 - A\eta N)), \\ VD_H + V_\perp E_H = \eta S_2 - A^\top \xi N - V \text{sym}(V^\top (\eta S_2 - A^\top \xi N)) \end{cases} \quad (3.5)$$

を得る．さて，

$$W := U^\top AV, W_{\bullet\perp} := U^\top AV_\perp, W_{\perp\bullet} := U_\perp^\top AV, W_{\perp\perp} := U_\perp^\top AV_\perp$$

とおく．式 (3.2) より $U^\top \xi = B_x, U_\perp^\top \xi = C_x$ であることに注意して，(3.5) の第 1 式の両辺に左から $U^\top$ を乗じると，

$$B_H = \text{skew}(B_x S_1 - U^\top A(VD_x + V_\perp E_x)N) = \text{skew}(B_x S_1 - WD_x N - W_{\bullet\perp} E_x N) \quad (3.6)$$

となる．ここで， $\text{skew}(\cdot)$ は行列の反対称部分を返す写像，すなわち $\text{skew}(Z) = (Z - Z^\top)/2$ である．同様に (3.5) の第 1 式の両辺に左から $U_\perp^\top$ を乗じると，

$$C_H = C_x S_1 - U_\perp^\top A(VD_x + V_\perp E_x)N = C_x S_1 - W_{\perp\bullet} D_x N - W_{\perp\perp} E_x N$$

を得る. (3.5) の第2式についても同様に両辺に左から $V^\top$ および $V_\perp^\top$ を乗じることで,

$$D_H = \text{skew}(D_x S_2 - V^\top A^\top (U B_x + U_\perp C_x) N) = \text{skew}(D_x S_2 - W^\top B_x N - W_{\perp \bullet}^\top C_x N),$$

$$E_H = E_x S_2 - V_\perp^\top A^\top (U B_x + U_\perp C_x) N = E_x S_2 - W_{\bullet \perp}^\top B_x N - W_{\perp \perp}^\top C_x N \quad (3.7)$$

となる. 後は, これらの式の両辺に vec や veck を作用させれば表現行列 H の作用の様子がわかる.

ここで, クロネッカー積 $\otimes$ を導入する. すなわち, 行列 $X = (x_{ij}) \in \mathbb{R}^{m \times n}$, $Y \in \mathbb{R}^{p \times q}$ に対して

$$X \otimes Y = \begin{bmatrix} x_{11}Y & \cdots & x_{1n}Y \\ \vdots & & \vdots \\ x_{m1}Y & \cdots & x_{mn}Y \end{bmatrix}$$

である. このとき, $X \in \mathbb{R}^{m \times p}$, $Y \in \mathbb{R}^{p \times q}$, $Z \in \mathbb{R}^{q \times n}$ に対して,

$$\text{vec}(XYZ) = (Z^\top \otimes X) \text{vec}(Y)$$

であることが知られている [8]. さらに, (i, j) 成分のみ 1 で, その他の成分はすべて 0 であるような $m \times n$ 行列を $E_{i,j}^{(m \times n)}$ と表し,

$$D_p = \sum_{p \geq i \geq j \geq 1} \left(E_{p(j-1)+i, j(p-(j+1)/2)-p+i}^{(p^2 \times p(p-1)/2)} - E_{p(i-1)+j, j(p-(j+1)/2)-p+i}^{(p^2 \times p(p-1)/2)} \right)$$

とおくと, 反対称行列 $S \in \text{Skew}(p)$ に対して

$$\text{vec}(S) = D_p \text{veck}(S), \quad \text{veck}(S) = \frac{1}{2} D_p^\top \text{vec}(S)$$

が成り立つことが容易に証明される.

これらを用いると, 式 (3.6)–(3.7) により

$$\text{veck}(B_H) = \frac{1}{2} D_p^\top ((S_1 \otimes I_p) D_p \text{veck}(B_x) - (N \otimes W) D_p \text{veck}(D_x) - (N \otimes W_{\bullet \perp}) \text{vec}(E_x)),$$

$$\text{vec}(C_H) = (S_1 \otimes I_{m-p}) \text{vec}(C_x) - (N \otimes W_{\perp \bullet}) D_p \text{veck}(D_x) - (N \otimes W_{\perp \perp}) \text{vec}(E_x),$$

$$\begin{aligned} \text{veck}(D_H) = & \frac{1}{2} D_p^\top (-(N \otimes W^\top) D_p \text{veck}(B_x) - (N \otimes W_{\perp \bullet}^\top) \text{vec}(C_x) \\ & - (S_2 \otimes I_p) D_p \text{veck}(D_x)), \end{aligned}$$

$$\text{vec}(E_H) = -(N \otimes W_{\bullet \perp}^\top) D_p \text{veck}(B_x) - (N \otimes W_{\perp \perp}^\top) \text{vec}(C_x) + (S_2 \otimes I_{n-p}) \text{vec}(E_x)$$

と計算できる. したがって

$$\mathbb{D} := \begin{bmatrix} D_p & & & \\ & I_{p(m-p)} & & \\ & & D_p & \\ & & & I_{p(n-p)} \end{bmatrix}$$

とすると、表現行列 H は

$$H = \mathbb{D}^\top \begin{bmatrix} \frac{1}{2}S_1 \otimes I_p & 0_{p^2 \times p(m-p)} & -\frac{1}{2}N \otimes W & -\frac{1}{2}N \otimes W_{\bullet\perp} \\ 0_{p(m-p) \times p^2} & S_1 \otimes I_{m-p} & -N \otimes W_{\perp\bullet} & -N \otimes W_{\perp\perp} \\ -\frac{1}{2}N \otimes W^\top & -\frac{1}{2}N \otimes W_{\perp\bullet}^\top & \frac{1}{2}S_2 \otimes I_p & 0_{p^2 \times p(n-p)} \\ -N \otimes W_{\bullet\perp}^\top & -N \otimes W_{\perp\perp}^\top & 0_{p(n-p) \times p^2} & S_2 \otimes I_{n-p} \end{bmatrix} \mathbb{D} \quad (3.8)$$

と書ける.

一方, 方程式 (3.3) の右辺も簡単に書き下せて,

$$\mathbf{g} := \begin{bmatrix} \text{veck}(B_g) \\ \text{vec}(C_g) \\ \text{veck}(D_g) \\ \text{vec}(E_g) \end{bmatrix} = \begin{bmatrix} \text{veck}(\text{skew}(WN)) \\ \text{vec}(W_{\perp\bullet}N) \\ \text{veck}(\text{skew}(W^\top N)) \\ \text{vec}(W_{\bullet\perp}^\top N) \end{bmatrix}$$

となる. これで, (3.3) の係数行列 (ヘシアン表現行列) も右辺ベクトル (勾配の列ベクトル表示) も明らかになったので, H が正則であれば少なくとも理論的には

$$\mathbf{x} = H^{-1}\mathbf{g}$$

と解くことができる. これで $\text{veck}(B_x), \text{vec}(C_x), \text{veck}(D_x), \text{vec}(E_x)$ が求まるので, そこから B_x, C_x, D_x, E_x を構成でき, 元のニュートン方程式 (2.7) の解を

$$(\xi, \eta) = (UB_x + U_\perp C_x, VD_x + V_\perp E_x)$$

と計算できる. ただし, 実際に H を陽に構成せずとも, 任意のベクトル $\mathbf{x}$ に対して $H\mathbf{x}$ さえ計算できれば線形方程式 (3.3) を解くことができ, 次節ではその考えに基づき共役残差法による求解について述べる.

ところで, 共役残差法を線形方程式に適用する際の前提として, その方程式の係数行列が対称行列であるという条件が必要である. 実は, (3.3) の係数行列 H , すなわち (3.8) は対称ではない. しかし, 容易にわかるように, H の左から

$$D' := \begin{bmatrix} 2I_{p(p-1)/2} & & & \\ & I_{p(m-p)} & & \\ & & 2I_{p(p-1)/2} & \\ & & & I_{p(n-p)} \end{bmatrix}$$

を乗じると, $D'H$ は対称行列である. このことは, 接空間 $T_{(U,V)}(\text{St}(p, m) \times \text{St}(p, n))$ におけるヘシアン $\text{Hess } F(U, V)$ の線形作用素としての対称性からも次のように導かれる. まず, $(\xi_1, \eta_1), (\xi_2, \eta_2)$ を点 (U, V) における接ベクトルとし,

$$(\xi_1, \eta_1) = (UB_1 + U_\perp C_1, VD_1 + V_\perp E_1), \quad (\xi_2, \eta_2) = (UB_2 + U_\perp C_2, VD_2 + V_\perp E_2)$$

と書かれているとする. ここで, $B_1, B_2, D_1, D_2 \in \text{Skew}(p)$, $C_1, C_2 \in \mathbb{R}^{(m-p) \times p}$, $E_1, E_2 \in \mathbb{R}^{(n-p) \times p}$ である. リーマン計量 (2.3) に基づいてこの2つの接ベクトルの内積を計算すると,

$$\begin{aligned}
 \langle (\xi_1, \eta_1), (\xi_2, \eta_2) \rangle_{(U,V)} &= \text{tr}(\xi_1^\top \xi_2) + \text{tr}(\eta_1^\top \eta_2) \\
 &= \text{tr}(B_1^\top B_2 + C_1^\top C_2) + \text{tr}(D_1^\top D_2 + E_1^\top E_2) \\
 &= 2 \cdot \text{veck}(B_1)^\top \text{veck}(B_2) + \text{vec}(C_1)^\top \text{vec}(C_2) \\
 &\quad + 2 \cdot \text{veck}(D_1)^\top \text{veck}(D_2) + \text{vec}(E_1)^\top \text{vec}(E_2) \\
 &= \begin{bmatrix} 2 \cdot \text{veck}(B_1) \\ \text{vec}(C_1) \\ 2 \cdot \text{veck}(D_1) \\ \text{vec}(E_1) \end{bmatrix}^\top \begin{bmatrix} \text{veck}(B_2) \\ \text{vec}(C_2) \\ \text{veck}(D_2) \\ \text{vec}(E_2) \end{bmatrix} \\
 &= \begin{bmatrix} \text{veck}(B_1) \\ \text{vec}(C_1) \\ \text{veck}(D_1) \\ \text{vec}(E_1) \end{bmatrix}^\top D' \begin{bmatrix} \text{veck}(B_2) \\ \text{vec}(C_2) \\ \text{veck}(D_2) \\ \text{vec}(E_2) \end{bmatrix}
 \end{aligned}$$

となる. 一方, ヘシアン $\text{Hess } F(U, V)$ はリーマン計量から定まる内積について対称な線形作用素として定義されるので,

$$\langle (\xi_1, \eta_1), \text{Hess } F(U, V)[(\xi_2, \eta_2)] \rangle_{(U,V)} = \langle \text{Hess } F(U, V)[(\xi_1, \eta_1)], (\xi_2, \eta_2) \rangle_{(U,V)}$$

が成り立つ. したがって,

$$\begin{bmatrix} \text{veck}(B_1) \\ \text{vec}(C_1) \\ \text{veck}(D_1) \\ \text{vec}(E_1) \end{bmatrix}^\top D' H \begin{bmatrix} \text{veck}(B_2) \\ \text{vec}(C_2) \\ \text{veck}(D_2) \\ \text{vec}(E_2) \end{bmatrix} = \begin{bmatrix} \text{veck}(B_1) \\ \text{vec}(C_1) \\ \text{veck}(D_1) \\ \text{vec}(E_1) \end{bmatrix}^\top H \begin{bmatrix} \text{veck}(B_2) \\ \text{vec}(C_2) \\ \text{veck}(D_2) \\ \text{vec}(E_2) \end{bmatrix} D'$$

となる. これは任意の $(\xi_1, \eta_1), (\xi_2, \eta_2) \in T_{(U,V)}(\text{St}(p, m) \times \text{St}(p, n))$ について成り立つから,

$$D' H = H^\top D' \quad (3.9)$$

を得る. 式 (3.9) は行列 $D' H$ が対称行列であることを示している. したがって, 線形方程式 (3.3) の両辺に左から D' を乗じることで, 係数行列が対称行列であるような方程式

$$D' H x = D' g \quad (3.10)$$

に書き直すことができる.

3.3 共役残差法による線形方程式の求解

$H' := D' H, g' := D' g$ において, 線形方程式 (3.3) および (3.10) と等価かつ, 係数行列 H' が対称行列である

$$H' x = g' \quad (3.11)$$

が得られた。この線形方程式に対して共役残差法を適用したい。まず、共役残差法のアルゴリズムを記述しておく。

アルゴリズム 2 線形方程式 (3.11) に対する共役残差法

- 1: 初期値 $\mathbf{x}_0$ を選び, $\mathbf{r}_0 = \mathbf{g}' - H'\mathbf{x}_0$, $\mathbf{p}_0 = \mathbf{r}_0$, $\mathbf{q}_0 = H'\mathbf{p}_0$ とする.
 - 2: **for** $l = 0, 1, 2, \dots$ **do**
 - 3: $\alpha_l = (\mathbf{r}_l, H'\mathbf{r}_l) / (\mathbf{q}_l, \mathbf{q}_l)$.
 - 4: $\mathbf{x}_{l+1} = \mathbf{x}_l + \alpha_l \mathbf{p}_l$.
 - 5: $\mathbf{r}_{l+1} = \mathbf{r}_l - \alpha_l \mathbf{q}_l$.
 - 6: $\beta_l = (\mathbf{r}_{l+1}, H'\mathbf{r}_{l+1}) / (\mathbf{r}_l, H'\mathbf{r}_l)$.
 - 7: $\mathbf{p}_{l+1} = \mathbf{r}_{l+1} + \beta_l \mathbf{p}_l$.
 - 8: $\mathbf{q}_{l+1} = H'\mathbf{r}_{l+1} + \beta_l \mathbf{q}_l$.
 - 9: **end for**
-

共役残差法においては H' は行列-ベクトル積の形でしか現れないことに注意する。したがって、行列 H' そのものを (3.8) に基づいて計算する必要はなく、任意のベクトル $\mathbf{x}$ に対して $H'\mathbf{x}$ が計算できさえすれば良い。 $(\xi, \eta), \text{Hess } F(U, V)[(\xi, \eta)] \in T_{(U, V)}(\text{St}(p, m) \times \text{St}(p, n))$ をそれぞれ式 (3.2), (3.4) のように表したとすると、式 (3.6)–(3.7) により B_H, C_H, D_H, E_H は既に B_x, C_x, D_x, E_x で表せているため、 $\mathbf{x}$ から $H\mathbf{x}$ への変換、すなわち

$$\begin{bmatrix} \text{veck}(B_x) \\ \text{vec}(C_x) \\ \text{veck}(D_x) \\ \text{vec}(E_x) \end{bmatrix} \mapsto (B_x, C_x, D_x, E_x) \mapsto (B_H, C_H, D_H, E_H) \mapsto \begin{bmatrix} \text{veck}(B_H) \\ \text{vec}(C_H) \\ \text{veck}(D_H) \\ \text{vec}(E_H) \end{bmatrix}$$

は計算できる。ところで、変換 $\mathbf{y} \mapsto D'\mathbf{y}$ は、 $\mathbf{y}$ の一部の成分を 2 倍するだけであるから容易に計算できるため、結局 $H' := D'H$ による行列-ベクトル積の計算は次のアルゴリズムのように書ける。

アルゴリズム 3 H' による行列-ベクトル積の計算

- 1: Given: $\mathbf{x} \in \mathbb{R}^K$.
- 2: $\mathbf{x} = \left[\text{veck}(B)^\top \quad \text{vec}(C)^\top \quad \text{veck}(D)^\top \quad \text{vec}(E)^\top \right]^\top$ となるように $B, D \in \text{Skew}(p)$, $C \in \mathbb{R}^{(m-p) \times p}$, $E \in \mathbb{R}^{(n-p) \times p}$ を構成する.
- 3: B_H, C_H, D_H, E_H を

$$\begin{cases} B_H = \text{skew}(B_x S_1 - U^\top A(V D_x + V_\perp E_x)N), \\ C_H = C_x S_1 - U_\perp^\top A(V D_x + V_\perp E_x)N, \\ D_H = \text{skew}(D_x S_2 - V^\top A^\top(U B_x + U_\perp C_x)N), \\ E_H = E_x S_2 - V_\perp^\top A^\top(U B_x + U_\perp C_x)N \end{cases}$$

により計算する。

- 4: $H'\mathbf{x} = \left[2 \cdot \text{veck}(B_H)^\top \quad \text{vec}(C_H)^\top \quad 2 \cdot \text{veck}(D_H)^\top \quad \text{vec}(E_H)^\top \right]^\top$.
-

これで、共役残差法を適用する準備が整った。なお、共役残差法では通常、相対残差ノルムがあらかじめ定めた定数 $\varepsilon (> 0)$ 未満になったところで線形方程式の十分な近似解が得られたと判断し、反復を終了する。すなわち、今回の場合であれば、

$$\frac{\|g' - H'x_l\|}{\|g'\|} < \varepsilon \quad (3.12)$$

を $H'x = g'$ に対する共役残差法の停止条件とする。ただし、初期値は $x_0 = 0$ とする。以上の事柄の数値実験による検証は5節で行う。また、共役残差法の停止条件として、(3.12) より良い条件の提案へと繋がる議論を次節で行う。

4 ニュートン方程式の解の精度とニュートン法の収束率

一般のリーマン多様体 M 上のニュートン法の局所的収束性について、次の定理が成り立つ [1, 7].

定理 4.1. リーマン多様体 M 上の滑らかな目的関数 f を最小化する最適化問題において、 $\text{grad } f(x_*) = 0$ かつ $\text{Hess } f(x_*)^{-1}$ が存在するような $x_* \in M$ があると仮定する。このとき、 x_* の M における近傍 $\mathcal{U}$ が存在して、任意の $x_0 \in \mathcal{U}$ を初期点としてこの問題に対するニュートン法により生成される点列 $\{x_k\}$ は x_* に2次収束する。

前節で述べたニュートン法の解法は、ニュートン方程式 (2.7) の代わりに標準形の線形方程式 (3.11) を解くというものであるが、数値誤差がなく厳密に解けるとすれば、理論的には (2.7) を解くことと等価である。したがって、定理 4.1 より、提案手法は局所的に2次収束する点列を生成する。

一方、ニュートン法の各反復におけるニュートン方程式を非厳密に解いた場合の収束性も議論されている。特に、リーマン多様体 M 上では次の定理が成り立つ [1].

定理 4.2. 定理 4.1 で述べたリーマン多様体 M 上のニュートン法の各反復において、ニュートン方程式

$$\text{Hess } f(x_k)[\eta_k] = -\text{grad } f(x_k) \quad (4.1)$$

の代わりに

$$\text{Hess } f(x_k)[\eta_k] = -\text{grad } f(x_k) + \varepsilon_k \quad (4.2)$$

を満たすように探索方向 η_k を計算したとする。ここで、 $x_k \in M$, $\eta_k, \varepsilon_k \in T_{x_k}M$ である。また、 $\text{grad } f(x_*) = 0$ かつ $\text{Hess } f(x_*)^{-1}$ が存在するような $x_* \in M$ があると仮定する。 $(\mathcal{U}', \phi)$ を $x_* \in \mathcal{U}'$ なる M の局所座標系とし、この局所座標系を用いた表現を $\hat{\cdot}$ により表すとする。さらに、定数 $\kappa > 0$ が存在して

$$\|\hat{\varepsilon}_k\| \leq \min\{\|\widehat{\text{grad } f(\hat{x}_k)}\|^\theta, \kappa\} \|\widehat{\text{grad } f(\hat{x}_k)}\|$$

が任意の k で成り立つとする。ここで、 $\theta > 0$ である。このとき、 x_* の M における近傍 $\mathcal{U}$ が存在して、任意の $x_0 \in \mathcal{U}$ を初期点としてこの非厳密なニュートン法により生成される点列 $\{x_k\}$ は x_* に超1次収束 ($\min\{1 + \theta, 2\}$ 次収束) する。特に、 $\theta = 1$ のときは2次収束する。

ε_k を残差と見なすと、方程式 (4.2) を満たす η_k はニュートン方程式 (4.1) の近似解であると考えられる。

定理 4.2 は本稿で考えている問題 (1.2) に適用できる。しかし、本稿では $\text{St}(p, m) \times \text{St}(p, n)$ の局所座標系を考えることなく、式 (3.1) に基づいて接空間における基底を適当にとること、その基底についての接ベクトルの成分表示やヘシアンの表現行列を用いてニュートン方程式を (3.11) に変換し、これを共役残差法で近似的に解くという方法を提案した。したがって、その差異についての議論は必要ではあるが、基本的には定理 4.2 の証明と同様にして、(3.11) を近似的に解いた場合のニュートン法の収束率を議論できる。すなわち、各反復 k における $H'_k \mathbf{x} = \mathbf{g}'_k$ の近似解 $\mathbf{x}_k$ が

$$H'_k \mathbf{x}_k = \mathbf{g}'_k + \varepsilon_k \quad (4.3)$$

を満たすとし、定数 $\kappa, \theta > 0$ に対して

$$\|\varepsilon_k\| \leq \min\{\|\mathbf{g}'_k\|^\theta, \kappa\} \|\mathbf{g}'_k\| \quad (4.4)$$

が成り立つとき、アルゴリズム 1 においてニュートン方程式の解 (ξ_k, η_k) の代わりに、(4.3) を満たす $\mathbf{x}_k$ から構成した $(\xi_k, \eta_k) \in T_{(U_k, V_k)}(\text{St}(p, m) \times \text{St}(p, n))$ を用いると、局所的に $\min\{1 + \theta, 2\}$ 次収束する点列が生成される。

特に、 $\theta = 1$ のとき、式 (4.4) は

$$\frac{\|\varepsilon_k\|}{\|\mathbf{g}'_k\|} \leq \min\{\|\mathbf{g}'_k\|, \kappa\}$$

となり、 $\|\mathbf{g}'_k\| > \kappa$ のときは (3.12) と同様に相対残差ノルムが定数 κ 未満になったら共役残差法を終了するが、ニュートン法の反復が進んで $\|\mathbf{g}'_k\|$ が小さくなり $\|\mathbf{g}'_k\| < \kappa$ が成り立つようになると、相対残差ノルムが $\|\mathbf{g}'_k\|$ 未満になるまで共役残差法の反復をすることになる。すなわち、 $\|\mathbf{g}'_k\|$ が大きいうちは線形方程式 (3.11) を粗く解いても良いことになる。このように (3.11) を厳密に解かなくてもニュートン法の局所的な 2 次収束性は保たれるというわけである。

5 数値実験

本節では、3 節で提案した手法を用いた数値実験例を紹介する。実験は、Intel Core i7-4790 CPU 3.60 GHz, 16.0 GB RAM を搭載した PC により、MATLAB R2014b を用いて行った。

まず、列直交行列 $U_* \in \mathbb{R}^{m \times n}$, $V_* \in \mathbb{R}^{n \times n}$ を乱数を用いて生成し、 $\Sigma_* = \text{diag}(n, n-1, \dots, 1)$ として $A = U_* \Sigma_* V_*^\top$ とした。また、目的関数 F を定義する際に必要な定数の対角行列 N は、 $N = \text{diag}(p, p-1, \dots, 1)$ とした。以上から、 U_* , V_* のそれぞれ左から p 列目までを取り出してできる $m \times p$ 行列 U_{opt} および $n \times p$ 行列 V_{opt} に対して、 $(U_{\text{opt}}, V_{\text{opt}}) \in \text{St}(p, m) \times \text{St}(p, n)$ は問題 (1.2) の最適解である。さらに、標準正規分布に従う乱数を各成分に持つ $m \times p$ 行列 $R_{m \times p}$ および $n \times p$ 行列 $R_{n \times p}$ を用いて、ニュートン法の初期点を

$$(U_0, V_0) = (\text{qf}(U_{\text{opt}} + 0.001 R_{m \times p}), \text{qf}(V_{\text{opt}} + 0.001 R_{n \times p}))$$

により構成した．すなわち，最適解に摂動を加えた点を初期点として実験を行った．これは，ニュートン法が大域的収束性を持たず，任意の初期点に対しては収束性が保証できないためである．なお，最適解の情報が利用できない場合に最適解の近くの点を得るためには，たとえば共役勾配法などの大域的収束性を持つ最適化手法を適用すれば良く，実用的には最適化の途中で共役勾配法からニュートン法に切り替えるというハイブリッドな方法が考えられる [7]．ニュートン法の反復回数は3回とし，各反復におけるニュートン方程式を解くための共役残差法の停止条件は $\varepsilon = 10^{-12}$ とした (3.12) を用い，最大反復回数は500回とした．問題のサイズは $m = 3000, p = 5$ と固定し， n の値は 100, 500, 1000, 1500, 2000 としてそれぞれの場合で実験を行った．各問題に対するニュートン法についての勾配の相対誤差ノルム $\|\text{grad } F(U_k, V_k)\|_{(U_k, V_k)} / \|\text{grad } F(U_0, V_0)\|_{(U_0, V_0)}$ の収束履歴を表1に示す．表1より， n の値によらず勾配が滑らかに0に収束している様子がわかる．

表 1: $n = 100, 500, 1000, 1500, 2000$ での勾配の相対誤差ノルムの収束履歴

$n \backslash k$	0	1	2	3
100	1	1.729×10^{-3}	6.428×10^{-6}	3.321×10^{-11}
500	1	1.672×10^{-3}	3.710×10^{-6}	3.500×10^{-11}
1000	1	2.074×10^{-3}	1.795×10^{-6}	1.394×10^{-11}
1500	1	2.298×10^{-3}	1.813×10^{-6}	3.960×10^{-11}
2000	1	2.407×10^{-3}	1.166×10^{-6}	4.651×10^{-11}

また， $n = 500$ としたときのニュートン法の各反復におけるニュートン方程式に対する共役残差法の収束履歴を図 5.1 に， U_k, V_k の相対誤差ノルム $\|U_* - U_k\|_F / \|U_*\|_F, \|V_* - V_k\|_F / \|V_*\|_F$ ，目的関数値の相対誤差 $|F(U_*, V_*) - F(U_k, V_k)| / |F(U_*, V_*)|$ ，勾配の相対誤差ノルム $\|\text{grad } F(U_k, V_k)\|_{(U_k, V_k)} / \|\text{grad } F(U_0, V_0)\|_{(U_0, V_0)}$ の収束履歴を図 5.2–5.4 に示す．図 5.1 において， $r_{k,l}$ ， $k = 1, 2, 3$ は，ニュートン法の第 k 反復で現れるニュートン方程式に対する共役残差法の，第 l 反復での残差ベクトルを表す．

これらの結果から，提案手法は最適解に2次収束する点列を生成できていることがわかる．特に，ニュートン方程式を解く際の共役残差法による相対残差ノルムも滑らかに収束している様子が確認できる．

6 結論と今後の課題

本稿では問題 (1.2) に対するニュートン法の各反復で現れるニュートン方程式を連立一次方程式に変換し，その共役残差法による効率的な解法を提案した．また，数値実験により提案手法の有効性を検証した．さらに，共役残差法による線形方程式の求解の精度とニュートン方程式の収束率の関係についても議論した．この議論を活かし，より効率的なアルゴリズムを提案することが今後の課題である．また，問題 (1.2) は特異値分解と関係があるが，[6] で示されているように，目的関数を変えると近似的同時特異値分解に対応する $\text{St}(p, m) \times \text{St}(p, n)$ 上の最適化問題が得られ，この問題に対しても本稿の議論は同様に成り立つと期待される．他にもシュティーフェル多様体上の最適化問題として定式化される応用問題は多くあり，本稿の結果を利用して，それらに対する解法アルゴリズムも提案したい．

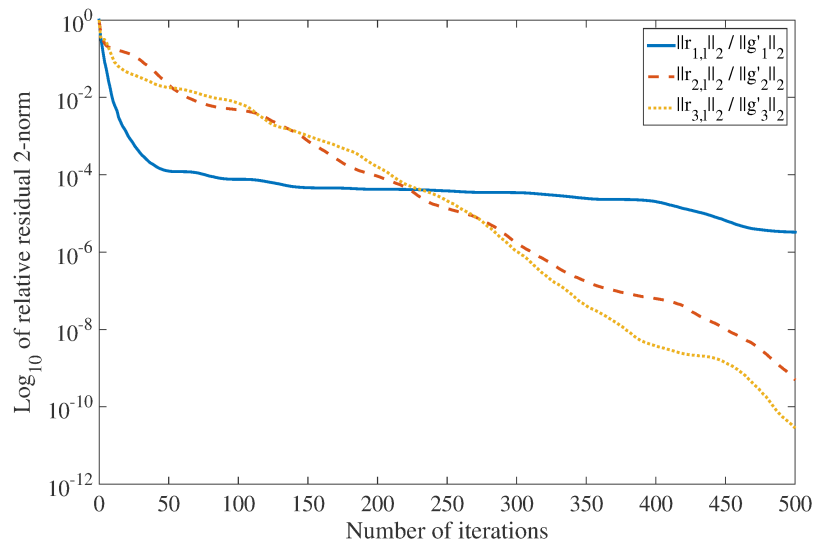


図 5.1: 共役残差法による $\|\mathbf{r}_{k,l}\|_2 / \|\mathbf{g}'_k\|_2$ の収束履歴 ($k = 1, 2, 3$)

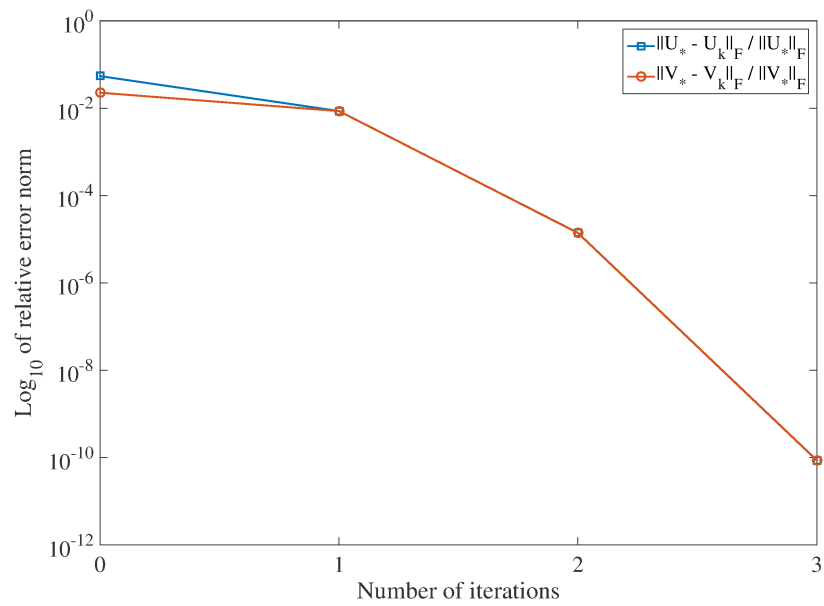


図 5.2: $\|U_* - U_k\|_F / \|U_*\|_F$ および $\|V_* - V_k\|_F / \|V_*\|_F$ の収束履歴

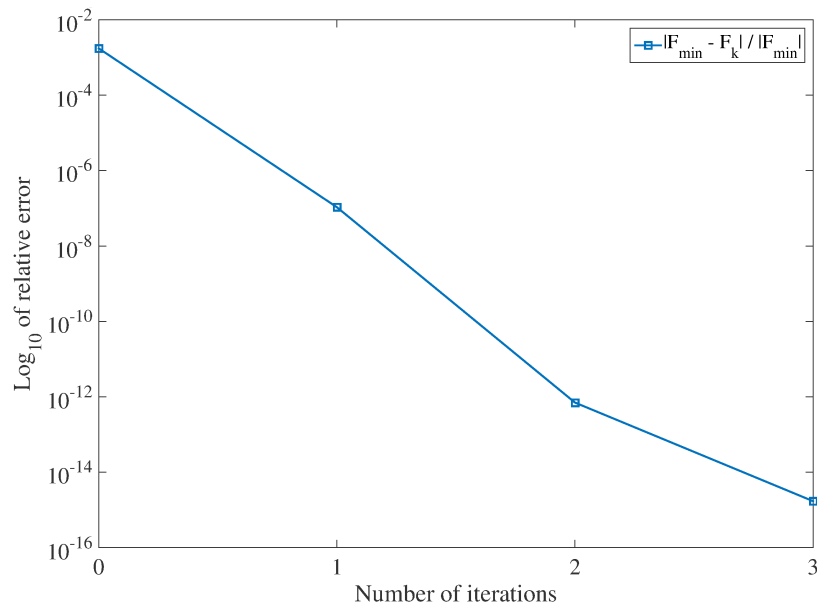


図 5.3: $|F(U_*, V_*) - F(U_k, V_k)| / |F(U_*, V_*)|$ の収束履歴

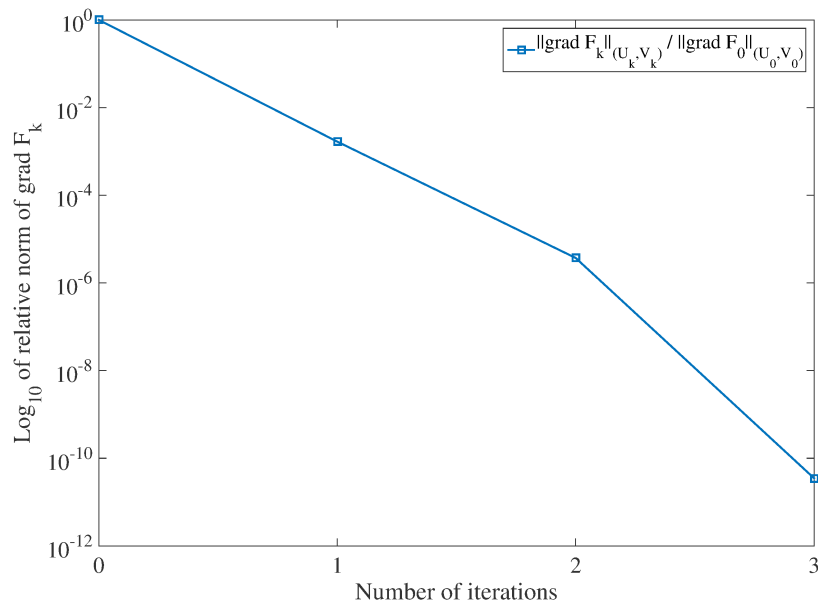


図 5.4: $\|\text{grad } F(U_k, V_k)\|_{(U_k, V_k)} / \|\text{grad } F(U_0, V_0)\|_{(U_0, V_0)}$ の収束履歴

参考文献

- [1] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, 2008.
- [2] A. Edelman, T.A. Arias, and S.T. Smith, The geometry of algorithms with orthogonality constraints, *SIAM J. Matrix Anal. Appl.*, **20**, 303–353, 1998.
- [3] S.C. Eisenstat, H.C. Elman, and M.H. Schultz, Variational iterative methods for non-symmetric systems of linear equations, *SIAM J. Numer. Anal.*, **20**, 345–357, 1983.
- [4] G.H. Golub and C.F. Van Loan, *Matrix Computations*, 4th ed., The Johns Hopkins University Press, Baltimore, 2013.
- [5] J. Nocedal and S.J. Wright, *Numerical Optimization*, 2nd ed., Springer, New York, 2006.
- [6] H. Sato, Joint singular value decomposition algorithm based on the Riemannian trust-region method, *JSIAM Letters*, **7**, 13–16, 2015.
- [7] H. Sato and T. Iwai, A Riemannian optimization approach to the matrix singular value decomposition, *SIAM J. Optim.*, **23**, 188–212, 2013.
- [8] J.R. Schott, *Matrix Analysis for Statistics*, 2nd ed., Wiley, New York, 2005.