

シュティーフェル多様体上の同時対角化問題 に対するニュートン法

佐藤 寛之

京都大学大学院情報学研究科数理工学専攻

Hiroyuki Sato¹

Department of Applied Mathematics and Physics, Kyoto University

概要

従来、ユークリッド空間上での最適化問題が盛んに研究され、様々な解法アルゴリズムが知られている。無制約最適化問題に対するアルゴリズムは、制約付き最適化問題に対するアルゴリズムより簡明であることが多い。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である場合には、従来の最急降下法やニュートン法、あるいは共役勾配法を多様体上に拡張したものを用いることができる。

本稿では、リーマン多様体上の最適化問題の様々な例とその解法について概説した後、実対称行列の同時対角化問題がシュティーフェル多様体上の最適化問題に帰着されることを紹介し、そのニュートン法について論じる。

1 序論

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、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。制約条件付きの問題に対して仮にこれらのアルゴリズムを用いると、制約条件が満たされなくなってしまうので、最急降下法やニュートン法を制約条件付きの問題にそのまま適用することはできない。そのため、制約条件付きアルゴリズムは一般に複雑である。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、最急降下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、リーマン多様体上の最適化問題の様々な例を紹介し、一般のリーマン多様体上の最適化問題に対するニュートン法を説明する。そして、実対称行列の近似的同時対角化問題をシュティーフェル多様体上の最適化問題として定式化した上で、特にシュティーフェル多様体が直交群に帰着される場合について、対象となる最適化問題に対するニュートン法を導出する。また、その中でニュートン方程式をどう解くべきかについて、詳しく論じる。

¹e-mail: hsato@amp.i.kyoto-u.ac.jp

2 リーマン多様体上の最適化問題とその解法アルゴリズム

2.1 リーマン多様体上の最適化問題の例

まず初めに、次のユークリッド空間上の制約条件付きの最適化問題を取り上げる。すなわち、 $n \times n$ 実対称行列 A に対して、次の問題を考える。

問題 2.1.

$$\text{minimize } \frac{x^T A x}{x^T x}, \quad (2.1)$$

$$\text{subject to } x \in \mathbb{R}^n, x \neq 0. \quad (2.2)$$

この問題における目的関数はレイリー商と呼ばれ、最適解は行列 A の最小固有値に属する固有ベクトルであることが知られている。問題 2.1 は \mathbb{R}_*^n 上の無制約最適化問題である。ここで、 \mathbb{R}_*^n は $\mathbb{R}^n - \{0\}$ を表す。ところが、無制約最適化問題 2.1 に、最適解の近傍内の点を初期点としてニュートン法を適用しても、一般には最適解に収束する点列は得られない。実際、 k 番目の反復での点 $x_k \in \mathbb{R}^n$ における探索方向 $\xi_k \in \mathbb{R}^n$ に対するニュートン方程式は、

$$\frac{2}{x_k^T x_k} \left(I_n - 2 \frac{x_k x_k^T}{x_k^T x_k} \right) \left(A - \frac{x_k^T A x_k}{x_k^T x_k} I_n \right) \left(I_n - 2 \frac{x_k x_k^T}{x_k^T x_k} \right) \xi_k = - \frac{2}{x_k^T x_k} \left(A - \frac{x_k^T A x_k}{x_k^T x_k} I_n \right) x_k \quad (2.3)$$

となるが、この方程式は、 $x_k^T A x_k / x_k^T x_k$ が A の固有ベクトルでない限り、ただ一つの解 $\xi_k = x_k$ を持つ [1]。したがって、そのような場合、ニュートン法によって得られる次の点 x_{k+1} は $x_{k+1} = x_k + \xi_k = 2x_k$ となる。つまり、この問題に対するニュートン法によって得られる点列は、一般にはいかなる停留点 x_* にも収束せず、最適解を得ることができない。

ところで、目的関数 (2.1) は x のスケール変換によって値を変えないから、 x のノルムを固定しても差し支えない。そこで、 $x^T x = 1$ という制約条件を付けて次の等価な問題に変換することができる。

問題 2.2.

$$\text{minimize } x^T A x, \quad (2.4)$$

$$\text{subject to } x^T x = 1, \quad (2.5)$$

$$x \in \mathbb{R}^n. \quad (2.6)$$

この問題は \mathbb{R}^n 上の制約付き最適化問題である。

ここで、制約条件 $x^T x = 1$ ，すなわち $\|x\| = 1$ に注目すると、探索領域はユークリッドノルムが1の点全体、すなわち $(n-1)$ -次元の単位超球面 $S^{n-1} := \{x \in \mathbb{R}^n \mid \|x\| = 1\}$ である。こうして、問題 2.2 は次の問題に書き換えられる。

問題 2.3.

$$\text{minimize } x^T A x, \quad (2.7)$$

$$\text{subject to } x \in S^{n-1}. \quad (2.8)$$

問題 2.3 は、 S^{n-1} 上の無制約最適化問題である。こうして、球面という多様体上の最適化問題が得られた。したがって、ニュートン法をあらかじめ一般論として多様体上に拡張しておけば、収束性の良いアルゴリズムが得られることが期待される。実際、この最適化問題に対する（多様体上の）ニュートン法は、実はレイリー商反復法と全く等価であることが確認できる [1]。

数値線形代数における問題がリーマン多様体上の最適化問題に書き直される例は他にもある。問題 2.3 をシュティーフエル多様体 $\text{St}(p, n) := \{Y \in \mathbb{R}^{n \times p} \mid Y^T Y = I_p\}$ ($p \leq n$) に拡張すると、次の問題を得る：

問題 2.4.

$$\text{minimize } \text{tr}(Y^T A Y N), \quad (2.9)$$

$$\text{subject to } Y \in \text{St}(p, n). \quad (2.10)$$

ここで、 N は $N = \text{diag}(\mu_1, \dots, \mu_p)$ 、 $0 < \mu_1 < \dots < \mu_p$ である。この問題の最適解 Y_* の第 j 列は行列 A の小さい方から j 番目の固有値に属する単位固有ベクトルであることが示される。

さらに、行列の固有ベクトルそのものを知る必要はなく、小さい方から p 個の固有値に属する固有空間だけが必要なときは、問題 2.4 から定数の対角行列 N を除くことができ、目的関数は $\text{tr}(Y^T A Y)$ となる。ところで、この目的関数は $O(p)$ 不変である。つまり、 p 次直交行列 $Q \in O(p)$ に対して、

$$\text{tr}((YQ)^T A (YQ)) = \text{tr}(Q^T Y^T A Y Q) = \text{tr}(Y^T A Y). \quad (2.11)$$

ゆえに、探索領域としては商多様体 $\text{St}(p, n)/O(p)$ を考えれば十分で、この $\text{St}(p, n)/O(p)$ はグラスマン多様体 $\text{Grass}(p, n)$ として知られている。つまり、次のグラスマン多様体上の最適化問題を得る：

問題 2.5.

$$\text{minimize } \text{tr}(Y^T A Y), \quad (2.12)$$

$$\text{subject to } [Y] \in \text{Grass}(p, n) := O(n)/\text{St}(p, n). \quad (2.13)$$

ここで, $[Y] \in \text{Grass}(p, n)$ は $Y \in \text{St}(p, n)$ を代表元とする同値類である. 最適解 $[Y_*]$ は, 行列 A の小さい方から p 個の固有値に属する固有ベクトルが張る空間に対応する.

グラスマン多様体を $\text{Grass}(p, n) = O(n)/\text{St}(p, n)$ と見ると, ユークリッド空間に埋め込まれていない. 一方, 先の2つの例では球面 S^{n-1} やシュティーフエル多様体 $\text{St}(p, n)$ はそれぞれユークリッド空間 \mathbb{R}^n や $\mathbb{R}^{n \times p}$ に自然に埋め込まれていた. 実は, グラスマン多様体 $\text{Grass}(p, n)$ も, 階数 p の直交射影行列全体と見なすことで, ユークリッド空間 $\mathbb{R}^{n \times n}$ の部分多様体となり得る [5, 6]. すなわち,

$$\text{Grass}(p, n) \simeq \{X \in \mathbb{R}^{n \times n} | X^T = X, X^2 = X, \text{rank}(X) = p\} \quad (2.14)$$

$$= \{X = Y Y^T | Y \in \text{St}(p, n)\}. \quad (2.15)$$

ここで, $X = Y Y^T$ のとき,

$$\text{tr}(Y^T A Y) = \text{tr}(A Y Y^T) = \text{tr}(A X) \quad (2.16)$$

が成り立つことに注意すると, 問題 2.5 を次のように変換することもできる:

問題 2.6.

$$\text{minimize } \text{tr}(A X), \quad (2.17)$$

$$\text{subject to } X \in \text{Grass}(p, n) := \{X \in \mathbb{R}^{n \times n} | X^T = X, X^2 = X, \text{rank}(X) = p\}. \quad (2.18)$$

行列の特異値分解についても同様の考察が可能で, 結果として2つのシュティーフエル多様体の積 $\text{St}(p, m) \times \text{St}(p, n)$ 上の次の最適化問題に帰着される.

問題 2.7.

$$\text{maximize } \text{tr}(U^T A V N), \quad (2.19)$$

$$\text{subject to } (U, V) \in \text{St}(p, m) \times \text{St}(p, n). \quad (2.20)$$

その詳細については [7] を参照されたい。

さて、実対称行列の同時対角化問題もシュティーフエル多様体 $\text{St}(p, n)$ 上の最適化問題に帰着される。互いに可換な実対称行列が同時対角化可能であることは、線形代数のよく知られた基本的な結果であるが、ここでは互いに可換とは限らない K 個の $n \times n$ 実対称行列 A_1, A_2, \dots, A_K を考える。これら K 個の行列に対して、近似的同時対角化問題は次のように定式化される [8]:

問題 2.8.

$$\text{maximize } \sum_{l=1}^K \|\text{diag}(Y^T A_l Y)\|_F^2, \quad (2.21)$$

$$\text{subject to } Y \in \text{St}(p, n). \quad (2.22)$$

ここで、 $\|\cdot\|_F$ はフロベニウスノルムを表し、 $\text{diag}(\cdot)$ は引数の行列と同じ対角成分を持つ対角行列を表す。同時対角化問題は、独立成分分析と密接な関係がある [2-4]。この問題については後ほど詳しく論じる。

2.2 リーマン多様体上のニュートン法

まず、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 2.9.

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

本稿では、ニュートン法に限定して話を進める。

Algorithm 1 \mathbb{R}^N 上のニュートン法

- 1: 初期点 $x_0 \in \mathbb{R}^N$ を選ぶ。
- 2: **for** $k = 0, 1, 2, \dots$ **do**
- 3: 探索方向 $\eta_k \in \mathbb{R}^N$ をニュートン方程式

$$f_{xx}(x_k)[\eta_k] = -f_x(x_k) \quad (2.23)$$

の解として計算する。

- 4: 次の点 x_{k+1} を $x_{k+1} := x_k + \eta_k$ によって計算する。
 - 5: **end for**
-

ここからニュートン法をリーマン多様体 M 上の最適化問題 2.10 に拡張する.

問題 2.10.

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

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. このことから, ニュートン方程式自体を多様体 M の接空間における線形方程式として扱う必要が生じ, M 上のニュートン方程式は

$$\text{Hess } f(x_k)[\eta_k] = -\text{grad } f(x_k) \quad (2.24)$$

となる. ここで, ヘシアン Hess と勾配 grad は, いずれもリーマン多様体 M に備わるリーマン計量に基いて定義されるもので, ヘッセ行列やユークリッド勾配とは一致しないことに注意されたい. また, 探索方向 $\eta_k \in T_{x_k}M$ が求まったとしても, ユークリッド空間上のアルゴリズムにおける更新の式

$$x_{k+1} := x_k + \eta_k \quad (2.25)$$

はリーマン多様体上では一般には意味をなさない. 実際, リーマン多様体上では加法が一般には定義されないし, もしユークリッド空間の部分多様体の場合で加法が定義されていても, $x_k + \eta_k$ は一般には多様体 M 上の点ではない. そこで, レトラクションと呼ばれる写像 R によって, $x_k + \eta_k$ に相当する多様体 M からはみ出た点を, M 上に引き戻す操作が不可欠である. こうして,

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

なる更新式を得る. なお, レトラクション R の詳細については [1] を参照されたい.

リーマン多様体 M 上のレトラクション R を用いると, M 上のニュートン法は次のように記述される:

Algorithm 2 リーマン多様体 M 上のニュートン法

初期点 $x_0 \in M$ を選ぶ.

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

探索方向 $\eta_k \in T_{x_k}M$ を

$$\text{Hess } f(x_k)[\eta_k] = -\text{grad } f(x_k) \quad (2.26)$$

として計算する.

次の点 x_{k+1} を, $x_{k+1} := R_{x_k}(\eta_k)$ によって計算する.

end for

3 シュティーフェル多様体上の同時対角化問題

この節ではシュティーフェル多様体上の同時対角化問題について考察するが、簡単のため、記号の煩雑さを避けるために $p = n$ の場合に限定して議論する。 $\text{St}(n, n) = O(n)$ に注意すると、対象となる問題は次のように書ける。

問題 3.1.

$$\text{maximize } f(Y) = \sum_{l=1}^K \|\text{diag}(Y^T A_l Y)\|_F^2, \quad (3.1)$$

$$\text{subject to } Y \in O(n). \quad (3.2)$$

この問題に対するニュートン法を導出する。シュティーフェル多様体上のレトラクションとしては、QR分解に基づく次のレトラクション

$$R_Y(\xi) = \text{qf}(Y + \xi), \quad Y \in \text{St}(p, n), \quad \xi \in T_Y \text{St}(p, n) \quad (3.3)$$

がよく用いられる。ここで、 $\text{qf}(\cdot)$ は、引数のフルランク行列をQR分解したときのQ成分を表す。

したがって、後はニュートン方程式についてのみ議論すれば良い。目的関数 f の勾配とヘシアンは、それぞれ次のように計算される。

$$\text{grad } f(Y) = -4 \sum_{l=1}^N Y \text{skew}(Y^T A_l Y \text{diag}(Y^T A_l Y)), \quad (3.4)$$

$$\begin{aligned} & \text{Hess } f(Y)[\xi] \\ &= P_Y(D(\text{grad } \bar{f})(Y)[\xi] - \xi \text{sym}(Y^T \text{grad } \bar{f}(Y))) \\ &= -4 \sum_{l=1}^N P_Y(A_l \xi \text{diag}(Y^T A_l Y) + 2A_l Y \text{diag}(Y^T A_l \xi) - \xi \text{sym}(Y^T A_l Y \text{diag}(Y^T A_l Y))). \end{aligned} \quad (3.5)$$

ここで、 P は接空間への直交射影を表し、

$$P_Y(W) = Y \text{skew}(Y^T W), \quad Y \in O(n), \quad W \in \mathbb{R}^{n \times n} \quad (3.6)$$

である。こうしてニュートン方程式

$$\text{Hess } f(Y)[\xi] = -\text{grad } f(Y) \quad (3.7)$$

を次のように具体的に書き下すことができる.

$$\begin{aligned}
 & -4 \sum_{l=1}^N P_Y (A_l \xi \operatorname{diag}(Y^T A_l Y) + 2A_l Y \operatorname{diag}(Y^T A_l \xi) - \xi \operatorname{sym}(Y^T A_l Y \operatorname{diag}(Y^T A_l Y))) \\
 & = 4 \sum_{l=1}^N (A_l Y \operatorname{diag}(Y^T A_l Y) - Y \operatorname{sym}(Y^T A_l Y \operatorname{diag}(Y^T A_l Y))).
 \end{aligned} \tag{3.8}$$

しかし, この方程式を真正面から解くのは容易ではない.

ここで, $\xi \in T_Y O(n)$ は $\xi = YB$, $B \in \operatorname{Skew}(n)$ と書けることに注目する. $\operatorname{Hess} f(Y)[\xi]$ もまた接ベクトルだから, $Y^T \operatorname{Hess} f(Y)[\xi] \in \operatorname{Skew}(n)$ である. そこで, $T_Y O(n) \simeq \operatorname{Skew}(n)$ と見なして $\operatorname{Hess} f(Y)[\xi]$ を $\operatorname{Skew}(n)$ 上の線形写像と考えたものを H とする:

$$H(B) := Y^T \operatorname{Hess} f(Y)[YB] \tag{3.9}$$

$$= -4 \sum_{k=1}^K \operatorname{skew}(Z_k B \operatorname{diag}(Z_k) - B Z_k \operatorname{diag}(Z_k) + 2Z_k \operatorname{diag}(Z_k B)). \tag{3.10}$$

$\dim(\operatorname{Skew}(n)) = n(n-1)/2$ であるから, B の独立変数 (たとえば下三角部分) のみを並べて列ベクトルを作り, 表現行列を考えたい.

そこで, よく知られているように, 行列 $A = (\mathbf{a}_1, \dots, \mathbf{a}_n)$ に対して $\operatorname{vec}(A) = \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_n \end{pmatrix}$

と定義する. また, n 次正方行列 $A = (a_{ij})$ に対して $\operatorname{veck}(A) = \begin{pmatrix} a_{21} \\ \vdots \\ a_{n1} \\ a_{32} \\ \vdots \\ a_{n2} \\ \vdots \\ a_{n,n-1} \end{pmatrix}$ と定義する.

$A \in \operatorname{Skew}(n)$ ならば, n にのみ依存する duplication matrix D が存在して,

$$D \operatorname{veck}(A) = \operatorname{vec}(A) \tag{3.11}$$

となる. $D^T D = 2I$ であることが容易に分かるから,

$$\operatorname{veck}(A) = \frac{1}{2} D^T \operatorname{vec}(A) \tag{3.12}$$

であることに注意する. $H : \operatorname{Skew}(n)$ を $\mathbb{R}^{\frac{n(n-1)}{2}}$ 上の線形変換と見なすために, 次の計算

をする.

$$\text{veck}(H(B)) \quad (3.13)$$

$$= -2D^T \sum_{k=1}^K \frac{1}{2} (I - U) \text{vec} (Z_k B \text{diag}(Z_k) - B Z_k \text{diag}(Z_k) + 2Z_k \text{diag}(Z_k B)) \quad (3.14)$$

$$= -D^T (I - U) \sum_{k=1}^K (\text{diag}(Z_k) \otimes Z_k - Z_k \text{diag}(Z_k) \otimes I + 2(I \otimes Z_k) E(I \otimes Z_k)) D \text{veck}(B). \quad (3.15)$$

結局, $H(B)$ は

$$\text{veck}(H(B)) = F_H \text{veck}(B), \quad (3.16)$$

$$F_H = -D^T (I - U) \sum_{k=1}^K (\text{diag}(Z_k) \otimes Z_k - Z_k \text{diag}(Z_k) \otimes I + 2(I \otimes Z_k) E(I \otimes Z_k)) D \quad (3.17)$$

を満たす. この F_H が表現行列である.

さて, ニュートン方程式

$$\text{Hess } f(Y)[\xi] = -\text{grad } f(Y) \quad (3.18)$$

の解 $\xi \in T_Y O(n)$ を求めたい. $\xi = YB$, $B \in \text{Skew}(n)$ と書けるから, これは B の方程式と考えられる. ところが, B は $n \times n$ 行列であるが, 独立変数は $n(n-1)/2 (= \dim(O(n)) = \dim(T_Y O(n)))$ 個である. そこで, $n(n-1)/2$ 次元のベクトル空間上の方程式を導出する. そのために, 先ほどの H を使って, 次の同値関係に注目する.

$$\text{Hess } f(Y)[YB] = -\text{grad } f(Y) \quad (3.19)$$

$$\Leftrightarrow Y^T \text{Hess } f(Y)[YB] = -Y^T \text{grad } f(Y) \quad (3.20)$$

$$\Leftrightarrow H(B) = -Y^T \text{grad } f(Y) \quad (3.21)$$

$$\Leftrightarrow \text{veck}(H(B)) = -\text{veck}(Y^T \text{grad } f(Y)) \quad (3.22)$$

$$\Leftrightarrow F_H \text{veck}(B) = -\text{veck}(Y^T \text{grad } f(Y)). \quad (3.23)$$

すると, ニュートン方程式

$$F_H \text{veck}(B) = -\text{veck}(Y^T \text{grad } f(Y)) \quad (3.24)$$

は容易に解けて,

$$\text{veck}(B) = -F_H^{-1} \text{veck}(Y^T \text{grad } f(Y)) \quad (3.25)$$

となるから, $\text{veck}(B)$ から $\text{vec}(B) = D^T \text{veck}(B)/2$ を得ることができ, その成分を n 個ずつ横に並べれば B が復元できる. 最後に $\xi = YB$ でニュートン方向が求まる.

4 終わりに

実対称行列の同時対角化問題をリーマン多様体上の最適化問題と見なして、そのニュートン法について論じた。JADE 法と呼ばれる独立成分分析の手法においては、信号のキュムラント行列の同時対角化がその本質的な部分を占める。したがって、提案手法は独立成分分析にも有用であると考えられ、今後、数値計算実験によってその有効性を実証したい。

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] Jean-François Cardoso. High-order contrasts for independent component analysis. *Neural computation*, 11(1):157–192, 1999.
- [3] Jean-François Cardoso and Antoine Souloumiac. Blind beamforming for non-gaussian signals. In *Radar and Signal Processing, IEE Proceedings F*, volume 140, pages 362–370. IET, 1993.
- [4] Andrzej Cichocki and Shunichi Amari. *Adaptive Blind Signal and Image Processing*. John Wiley Chichester, 2002.
- [5] Uwe Helmke, Knut Hüper, and Jochen Trumpf. Newton’s method on graßmann manifolds. *arXiv preprint arXiv:0709.2205*, 2007.
- [6] Uwe Helmke and John B Moore. *Optimization and Dynamical Systems*. Communications and Control Engineering Series, Springer-Verlag (London and New York), 1994.
- [7] Hiroyuki Sato and Toshihiro Iwai. A Riemannian optimization approach to the matrix singular value decomposition. *SIAM Journal on Optimization*, 23(1):188–212, 2013.
- [8] F.J. Theis, Thomas P. Cason, and P.-A. Absil. Soft dimension reduction for ica by joint diagonalization on the Stiefel manifold. In *Proceedings of the 8th International Conference on Independent Component Analysis and Signal Separation, ICA ’09*, pages 354–361, Berlin, Heidelberg, 2009. Springer-Verlag.