

Optimization algorithms on the Grassmann and the Stiefel manifolds

佐藤 寛之, 岩井 敏洋

京都大学大学院情報学研究科数理工学専攻

Hiroyuki Sato¹, Toshihiro Iwai

Department of Applied Mathematics and Physics, Kyoto University

概要

与えられた目的関数がある制約条件の下で最小化する問題は最適化問題と呼ばれる。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。制約条件なしの問題に対するアルゴリズムは、制約条件付きの問題に対するアルゴリズムより単純である。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である場合には、従来の最急降下法やニュートン法を多様体上に拡張したものをを用いることができる。

本講演では、こうした問題の例として、行列の固有値問題や特異値分解が、グラスマン多様体や、2つのシュティーフエル多様体の積多様体上での最適化問題に帰着されることを紹介し、一般のリーマン多様体上の最適化問題の解法について概説する。その後、具体的な問題に対する解法アルゴリズムについて論じる。

1 序論

与えられた関数がある領域において最小化する問題を最適化問題といい、最小化する関数を目的関数という。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、まずユークリッド空間における最適化アルゴリズムの一般論を概説し、次に多様体上の最適化問題の例を紹介する。その後、リーマン多様体上の最適化問題の解法アルゴリズムの一般論について述べる。最後に、それらのアルゴリズムの具体例への適用について紹介する。具体的には、固有値問題に関係のあるグラスマン多様体上の最適化問題と、特異値分解に関係のあるシュティーフエル多様体上の最適化問題について議論する。

¹e-mail: [hsato\[at\]amp.i.kyoto-u.ac.jp](mailto:hsato[at]amp.i.kyoto-u.ac.jp)

2 ユークリッド空間における制約条件なしの最適化アルゴリズム

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

問題 2.1.

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

制約条件なしの最適化問題に対する手法としては、最急降下法、ニュートン法などが知られている [6] が、これらの手法は次のアルゴリズムのように共通する構造を持つ.

Algorithm 1 $\mathbb{R}^N$ における制約条件なしの最適化手法

- 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 $\eta_k \in \mathbb{R}^N$ とステップサイズ $t_k > 0$ を求める.
- 4: 次の点 x_{k+1} を,

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

によって定める.

- 5: **end for**
-

2.1 探索方向の決め方

先のアルゴリズムの枠組みにおいて、探索方向の決め方とステップサイズの決め方が、個々のアルゴリズムを特徴づける. 探索方向 $\eta_k \in \mathbb{R}^N$ の決め方によって、アルゴリズムの名称は異なる. 特に基本的で有名なものは、最急降下法とニュートン法である. 最急降下法では、探索方向 η_k を $\eta_k := -\text{grad } f(x_k)$ として選ぶ. ここで、 grad はユークリッド勾配を表す. また、ニュートン法では、探索方向 η_k は未知ベクトル $\eta \in \mathbb{R}^N$ についてのニュートン方程式

$$\text{Hess } f(x_k) \eta = -\text{grad } f(x_k) \tag{2.2}$$

の解として計算する. ここで、 $\text{Hess } f(x_k)$ は、 f の点 x_k におけるヘッセ行列である.

2.2 ステップサイズの決め方

ステップサイズ $t_k \in \mathbb{R}$ の決め方の一つに、正確な直線探索によるものがある. これは、 $t \in \mathbb{R}$ の関数

$$f(x_k + t \eta_k) \tag{2.3}$$

を最小にする t を t_k とするものである。たとえば, A を N 次正値対称行列, b を N 次数ベクトルとすると, $\mathbb{R}^N$ 上の実数値関数 $f(x) = \frac{1}{2}x^T Ax - bx$ に最急降下法を適用し, かつステップサイズを正確な直線探索によって定めるというアルゴリズムでは, 更新式は次のようにして導出される。すなわち, まず点 x_k における探索方向 η_k はその点における f の逆勾配に等しく,

$$\eta_k = -\text{grad } f(x_k) = -(Ax_k - b) \quad (2.4)$$

となる。そして, 正確な直線探索によるステップサイズ t_k は, 点 x_k から探索方向 η_k の方向に伸びる半直線の上で f を最小にするように決めるのだから,

$$\begin{aligned} \frac{d}{dt} f(x_k + t\eta_k)|_{t=t_k} &= \text{grad } f(x_k + t_k\eta_k)^T \eta_k \\ &= (A(x_k + t_k\eta_k) - b)^T \eta_k \\ &= 0 \end{aligned} \quad (2.5)$$

により,

$$t_k = -\frac{(Ax_k - b)^T \eta_k}{\eta_k^T A \eta_k} \quad (2.6)$$

と具体的に定まる。こうして, 更新式

$$\begin{aligned} x_{k+1} &= x_k + t_k \eta_k \\ &= x_k + \frac{(Ax_k - b)^T \eta_k}{\eta_k^T A \eta_k} (Ax_k - b) \end{aligned} \quad (2.7)$$

を得る。

しかし, 一般には, 上の例のように正確な直線探索によるステップサイズを陽に書くことができるとは限らない。そこで用いられるのが, アルミホのステップサイズやウルフのステップサイズである [6]。ここでは, アルミホのステップサイズのみを紹介することにする。点 x_k における探索方向 η_k が求まったとする。このとき, 次のアルミホのルールを満たす $t \in \mathbb{R}$ を 1 つ見つけて, それを t_k とするのである。アルミホのルールとは,

$$f(x_k + t\eta_k) \leq f(x_k) + ct \text{grad } f(x_k)^T \eta_k \quad (2.8)$$

である。ここで, $c \in (0, 1)$ は定数であり, 実装の際に任意に与える。

以上がユークリッド空間における制約条件なしの最適化アルゴリズムの概要である。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はそのままでは用いることができないが, これらのアルゴリズムを多様体上に拡張するというアイデアがある。これについては後で述べることにして, そもそもなぜ多様体上の最適化問題を考えるのか, その動機を示すために, 次の節ではいくつか問題の例を挙げることにする。

3 リーマン多様体上の最適化問題の例

この節では, リーマン多様体上の最適化問題の例をいくつか挙げる。特に, 問題 3.10 と 3.11 は, 本稿で主として扱うものである。

3.1 レイリー商最小化問題と球面上の制約条件なし最適化問題

A を n 次対称行列として、次の最適化問題を考える。

問題 3.1.

$$\begin{aligned} & \text{minimize} && \frac{x^T A x}{x^T x}, \\ & \text{subject to} && x \in \mathbb{R}_*^n. \end{aligned}$$

ここで、 $\mathbb{R}_*^n$ は $\mathbb{R}^n$ から原点を除いたものである。

この問題の目的関数はレイリー商と呼ばれ、レイリー商の最小点は、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。したがって、この問題は応用上も意味のあるものである。さて、この問題をユークリッド空間上の制約条件なしの最適化問題と見なしてニュートン法を適用してみる。ニュートン方程式

$$\text{Hess } f(x_k) \eta = -\text{grad } f(x_k) \quad (3.1)$$

を書き下すために、一般の点 x における f の勾配とヘッセ行列を計算すると、それぞれ

$$\text{grad } f(x) = \frac{2}{x^T x} (A - f(x) I_n) x, \quad (3.2)$$

$$\text{Hess } f(x) = \frac{2}{x^T x} \left(I_n - \frac{2xx^T}{x^T x} \right) (A - f(x) I_n) \left(I_n - \frac{2xx^T}{x^T x} \right) \quad (3.3)$$

となる。よって、ニュートン方程式は η についての線形方程式

$$\frac{2}{x_k^T x_k} \left(I_n - \frac{2x_k x_k^T}{x_k^T x_k} \right) (A - f(x_k) I_n) \left(I_n - \frac{2x_k x_k^T}{x_k^T x_k} \right) \eta = -\frac{2}{x_k^T x_k} (A - f(x_k) I_n) x_k \quad (3.4)$$

となる。 $\text{grad } f(x_k) = 0$ でなければ、この方程式の解は $\eta_k = x_k$ となる。したがって、この問題にニュートン法を適用しても、各反復で得られる点は常に原点と点 x_0 を結ぶ直線上にあり、一般の初期点に対してはいくら反復を行っても最適解は決して得られない。

そこで、少し見方を変えてみる。問題 3.1 の目的関数は、 x の二次形式 $x^T A x$ を x の標準ノルム $x^T x$ で割ったものであるから、あらかじめ x の標準ノルムが 1 であるという制約条件を付けた次の問題を解くことで、問題 3.1 の解が得られる。

問題 3.2.

$$\begin{aligned} & \text{minimize} && x^T A x, \\ & \text{subject to} && x \in \mathbb{R}_*^n, x^T x = 1. \end{aligned}$$

これは、制約条件付きの最適化問題であり、たとえば拡張ラグランジュ法と呼ばれる制約条件付き最適化アルゴリズムを適用することで解くことができる。しかし、実際にこの問題に拡張ラグランジュ法を適用すると、最適解に収束するまでに相当の反復回数を要することが分かり、実用的ではない。そこで、制約条件 $x^T x = 1$ の見方を変えて、この問題を次のように球面 $S^{n-1} = \{x \in \mathbb{R}^n | x^T x = 1\}$ 上の制約条件なしの最適化問題であると捉える。

問題 3.3.

$$\begin{aligned} & \text{minimize} && x^T A x, \\ & \text{subject to} && x \in S^{n-1}. \end{aligned}$$

こうして多様体上の最適化問題の例が一つ浮かび上がった。そこで、最急降下法やニュートン法を多様体上に拡張しておくことによって、それらのアルゴリズムを用いて問題 3.3 が解けることが期待される。

3.2 レイリー商の重み付き和の最小化問題とシュティーフェル多様体上の制約条件なしの最適化問題

次に、 $p \leq n$ とし、 $0 < \mu_1 < \dots < \mu_p$ を重みを表す定数として、レイリー商の p 個の重み付き和 $\sum_{i=1}^p \mu_i x_i^T A x_i$ を最小化する問題を考える。

問題 3.4.

$$\begin{aligned} & \text{minimize} && \sum_{i=1}^p \mu_i x_i^T A x_i, \\ & \text{subject to} && x_1, \dots, x_p \in \mathbb{R}_*^n, \quad x_i^T x_j = \delta_{ij}. \end{aligned}$$

この問題の最適解 $x_1^*, \dots, x_p^*$ は、それぞれ、行列 A の小さい方から p 個の固有値に属する p 個の固有ベクトル $v_1, \dots, v_p$ である：

$$x_i^* = v_i, \quad i = 1, \dots, p.$$

さて、行列 $N = \text{diag}(\mu_1, \dots, \mu_p)$ と、 $x_1, \dots, x_p$ を並べてできる行列 $Y = (x_1, \dots, x_p)$ を導入すると、この問題の目的関数と制約条件はそれぞれ $F(Y) = \text{tr}(Y^T A Y N)$, $Y^T Y = I_p$

と書けるから、次のように $n \times p$ 行列を変数とする最適化問題であると見なすことができる。

問題 3.5.

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

この問題も、一見するとユークリッド空間における制約条件付きの最適化問題であるが、見方を変えるとシュティーフエル多様体 $\text{St}(p, n) = \{Y \in \mathbb{R}^{n \times p} | Y^T Y = I_p\}$ 上の制約条件なしの最適化問題と捉えることができる。つまり、次の問題と等価である。

問題 3.6.

$$\begin{aligned} & \text{minimize } F(Y) = \text{tr}(Y^T A Y N), \\ & \text{subject to } Y \in \text{St}(p, n). \end{aligned}$$

3.3 レイリー商の和の最小化問題とグラスマン多様体上の制約条件なしの最適化問題

今度は重みを付けずにレイリー商の和を考え、その最小化問題を考える。

問題 3.7.

$$\begin{aligned} & \text{minimize } \sum_{i=1}^p x_i^T A x_i, \\ & \text{subject to } x_1, \dots, x_p \in \mathbb{R}_*^n, x_i^T x_j = \delta_{ij}. \end{aligned}$$

この問題の最小解 $(x_1^*, \dots, x_p^*)$ は、行列 A の小さい方から p 個の固有値に属する p 個の固有ベクトル $v_1, \dots, v_p$ と次の関係にある：

$$\text{span}\{x_1^*, \dots, x_p^*\} = \text{span}\{v_1, \dots, v_p\}.$$

この問題も、シュティーフエル多様体 $\text{St}(p, n)$ 上の問題として次のように書くことができる。

問題 3.8.

$$\begin{aligned} & \text{minimize } F(Y) = \text{tr}(Y^T AY), \\ & \text{subject to } Y \in \text{St}(p, n). \end{aligned}$$

しかし、目的関数 F は $O(p)$ 不変であるから、この問題は実は次のように $\text{St}(p, n)/O(p) \simeq \text{Grass}(p, n)$ 上の最適化問題と見なすことができる。

問題 3.9.

$$\begin{aligned} & \text{minimize } F(Y) = \text{tr}(Y^T AY), \\ & \text{subject to } [Y] \in \text{Grass}(p, n). \end{aligned}$$

グラスマン多様体上の最適化問題 3.9 の解法アルゴリズムは、[1, 2] において導出されている。しかし、この問題に対するニュートン法は、ニュートン方程式が解きにくいという難点がある。そこで、我々は

$$\text{Grass}(p, n) \simeq \{X \in \mathbb{R}^{n \times n} | X^2 = X, \text{rank}(X) = p\} \quad (3.5)$$

として最適化アルゴリズムを導出する。これが本稿の一つ目の主題である。我々の観点からは、問題 3.9 を次のように書き直すことができる。

問題 3.10.

$$\begin{aligned} & \text{minimize } F(X) = \text{tr}(AX), \\ & \text{subject to } X \in \text{Grass}(p, n). \end{aligned}$$

3.4 特異値分解とシュティーフェル多様体上の最適化問題

特異値分解

$m \geq n$ とする。行列 $A \in \mathbb{R}^{m \times n}$ に対して、

$$A = U\Sigma V^T,$$

$$U \in O(m), V \in O(n), \Sigma = \begin{pmatrix} \text{diag}(\sigma_1, \dots, \sigma_n) \\ 0 \end{pmatrix}, \sigma_1 \geq \dots \geq \sigma_n \geq 0$$

なる分解を, A の特異値分解という. $\sigma_1, \dots, \sigma_n$ を特異値といい, U や V の列ベクトルをそれぞれ左特異ベクトル, 右特異ベクトルという.

シュティーフェル多様体上の最適化問題

m, n を $m \geq n$ なる整数として, $A \in \mathbb{R}^{m \times n}$ とする. $N = \text{diag}(\mu_1, \dots, \mu_n)$, $\mu_1 > \dots > \mu_n > 0$ として, 次の問題を考える.

問題 3.11.

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

この問題の最適解を (U_*, V_*) とすると, U_*, V_* の列ベクトルはそれぞれ, 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.

この問題の解法アルゴリズムに関する議論が, 本稿の二つ目の主題である.

4 リーマン多様体上の最適化アルゴリズム

4.1 ユークリッド空間における最適化アルゴリズムの多様体上への拡張

ユークリッド空間における最適化アルゴリズムでは, 現在の点 x_k における探索方向 η_k とステップサイズ t_k を求めた後,

$$x_{k+1} = x_k + t_k \eta_k \tag{4.1}$$

によって次の点を定めるのであった. しかし, 更新式 (4.1) は一般の多様体 M 上では意味をなさない. そこで, まず探索方向 η_k は現在の点 $x_k \in M$ における接ベクトルとして選ぶことにする. すなわち, $\eta_k \in T_{x_k} M$ とする. 次に, $\gamma(0) = x_k, \dot{\gamma}(0) = \eta_k$ 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 $R: TM \rightarrow M$ が見つければ, $R_x := R|_{T_x M}$ として,

$$x_{k+1} := R_{x_k}(t_k \eta_k) \tag{4.2}$$

なる更新式を用いることができる. R に課すべき条件は,

$$R_{x_k}(0) = x_k \tag{4.3}$$

と

$$\frac{d}{dt}R_{x_k}(t\eta_k)|_{t=0} = \eta_k \quad (4.4)$$

から自然に導かれる。そして、その条件を満たす R をレトラクションと呼ぶことにする。
以上を踏まえて、多様体 M 上のレトラクション R は次のように定義される。

定義 4.1. 写像 $R : TM \rightarrow M$ が次の条件を満たすとき、 R をレトラクションという。
ただし、 $R_x := R|_{T_x M}$ とする。

- $R_x(0_x) = x$. ここで、 0_x は $T_x M$ 上の零ベクトルである。
- $T_{0_x}T_x M \simeq T_x M$ なる同一視の下で、 $DR_x(0_x) = \text{id}_{T_x M}$.

4.2 アルミホのルール

ステップサイズの決め方として、2.2 節で紹介したアルミホのルールをリーマン多様体上に拡張する。

定義 4.2. 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 $\bar{\alpha} > 0, \beta, \sigma \in (0, 1)$ は与えられた定数とする。現在の点 $x_k \in M$ と探索方向 $\eta_k \in T_{x_k} M$ 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

$$f(x_k) - f(R_{x_k}(\beta^m \bar{\alpha} \eta_k)) \geq -\sigma \langle \text{grad } f(x_k), \beta^m \bar{\alpha} \eta_k \rangle_{x_k} \quad (4.5)$$

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

4.3 リーマン多様体上の最適化アルゴリズム

以上から、リーマン多様体 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 リーマン多様体 M における制約なし最適化手法

- 1: 初期点 $x_0 \in M$ を選ぶ.
- 2: **for** $k = 0, 1, 2, \dots$ **do**
- 3: 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.
- 4: 次の点 x_{k+1} を,

$$x_{k+1} := R_{x_k}(t_k \eta_k) \quad (4.6)$$

によって定める.

- 5: **end for**
-

このアルゴリズムの様子は図 4.1 のようである。

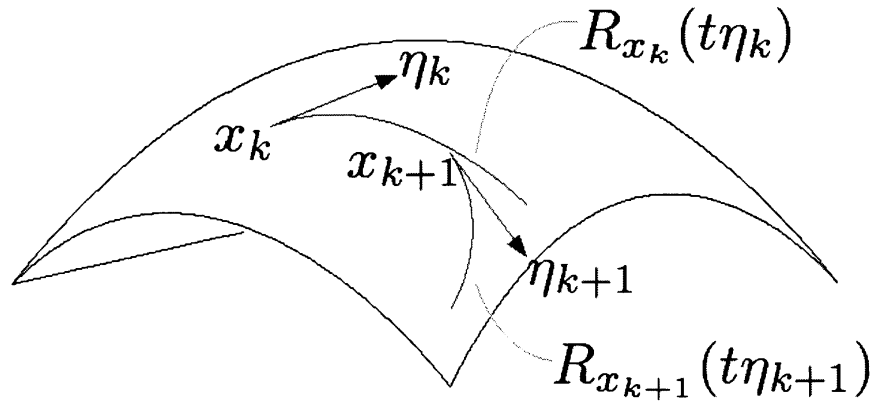


図 4.1: リーマン多様体上の最適化アルゴリズム

ステップサイズはたとえばアルミホのステップサイズを採用すれば良い。また、ニュートン法においてはステップサイズを常に 1 とする方法も用いられる。後は、探索方向の決め方さえ与えてやれば良い。

4.4 最急降下法とニュートン法

探索方向 $\eta_k \in T_{x_k}M$ の決め方として代表的なものは、最急降下法とニュートン法である。最急降下法においては

$$\eta_k = -\text{grad } f(x_k) \quad (4.7)$$

によって定める。

一方、ニュートン法においては、 $\eta \in T_{x_k}M$ を未知のベクトルとするニュートン方程式

$$\text{Hess } f(x_k)[\eta] = -\text{grad } f(x_k) \quad (4.8)$$

の解として定める。本稿ではヘシアン $\text{Hess } f$ は M 上のレヴィ・チビタ接続により定まるものを用いる。

5 グラスマン多様体とその上のレイリー商の幾何

5.1 レトラクション

グラスマン多様体上のレトラクションを 2 通り述べる。

指数レトラクション

一般に、リーマン多様体 M 上の指数写像はレトラクションである。そこで、グラスマン多様体 M 上の指数写像を求めることを考える。そのためには、測地線方程式を導出して解けば良い。

L をグラスマン多様体 $\text{Grass}(p, n)$ 上の自由粒子のラグランジアンとする。つまり、

$$L = \frac{1}{2} \langle \dot{X}, \dot{X} \rangle_X + \text{tr}(\Omega(X^2 - X)) + \lambda(\text{tr}(X) - p) \quad (5.1)$$

とする。ここで、 Ω と λ はラグランジュの未定乗数であり、 Ω は n 次実対称行列、 λ はスカラーである。変分を計算して、

$$\begin{aligned} \delta L &= \text{tr}(\delta \dot{X} \dot{X}) + \text{tr}(\delta \Omega(X^2 - X)) + \text{tr}(\Omega(2X\delta X - \delta X)) + \delta \lambda(\text{tr}(X) - p) + \lambda \text{tr}(\delta X) \\ &= \frac{d}{dt}(\text{tr}(\delta X \dot{X})) + \text{tr}(\delta X(-\ddot{X} + \Omega(2X - I_n) + \lambda I_n)) + \text{tr}(\delta \Omega(X^2 - X)) \\ &\quad + \delta \lambda(\text{tr}(X) - p). \end{aligned} \quad (5.2)$$

変分原理から、 $\delta X(t_1) = \delta X(t_2) = 0$ を満たす任意の δX に対して

$$\int_{t_1}^{t_2} \delta L dt = 0 \quad (5.3)$$

が成り立つから、

$$-\ddot{X} + \Omega(2X - I_n) + \lambda I_n = 0, \quad (5.4)$$

$$X^2 - X = 0, \quad (5.5)$$

$$\text{tr}(X) - p = 0. \quad (5.6)$$

これらの式から Ω と λ を消去して、測地線方程式

$$\ddot{X} + 2\dot{X}^2 - 4\dot{X}X\dot{X} = 0. \quad (5.7)$$

を得る。これを初期値 $X(0) = X$, $\dot{X}(0) = \xi$ の下で解いて、

$$X(t) = \frac{1}{2} \begin{pmatrix} YV & U \end{pmatrix} \begin{pmatrix} I_p + \cos 2\Sigma t & \sin 2\Sigma t \\ \sin 2\Sigma t & I_p - \cos 2\Sigma t \end{pmatrix} \begin{pmatrix} V^T Y^T \\ U^T \end{pmatrix} \quad (5.8)$$

を得る。ここで、 $Y \in \text{St}(p, n)$ は $X = YY^T$ を満たし（このような Y は一意には定まらない）、 U, Σ, V は ξY の特異値分解

$$\xi Y = U\Sigma V^T \quad (5.9)$$

の成分である。

よって、これら Y, U, Σ, V を用いて指数写像は

$$\text{Exp}_X(\xi) = \frac{1}{2} \begin{pmatrix} YV & U \end{pmatrix} \begin{pmatrix} I_p + \cos 2\Sigma & \sin 2\Sigma \\ \sin 2\Sigma & I_p - \cos 2\Sigma \end{pmatrix} \begin{pmatrix} V^T Y^T \\ U^T \end{pmatrix}, \quad (5.10)$$

と書くことができる。そこで、 $R := \text{Exp}$ とすれば R はレトラクションであり、これを指数レトラクションと呼ぶことにする。しかし、これを実際に計算するには各反復において U, Σ, V を求めるために特異値分解を計算する必要がある。したがって、その分の計算量が大きく、実用的とは言い難い。そこで、別のレトラクションを考える。

QR レトラクション

$n \times p$ 行列 B は

$$B = QR, \quad Q \in \text{St}(p, n), \quad R \in S_{\text{upp}}^+(p), \quad (5.11)$$

の形に分解することができる [4, 5]. ここで, $S_{\text{upp}}^+(p)$ は $p \times p$ の上三角かつ対角成分が全て正の行列全体の集合である. この分解 (5.11) を B の QR 分解という. さて, $\text{qf}(B)$ を QR 分解 $B = QR$ の Q 成分とする. つまり, $\text{qf}(B) = Q$.

このとき, $X \in \text{Grass}(p, n)$ に対して $Y \in \text{St}(p, n)$ を $X = YY^T$ を満たすものとして選んでおき, R を次のように定義すると, これもレトラクションの性質を満たすことが示される. すなわち,

$$R_X(\xi) = \text{qf}((I + \xi)Y) \text{qf}((I + \xi)Y)^T. \quad (5.12)$$

5.2 グラスマン多様体上の関数の勾配とヘシアン

グラスマン多様体 $\text{Grass}(p, n)$ 上の一般の滑らかな実数値関数 F を考える.

F の $X \in \text{Grass}(p, n)$ における勾配 $\text{grad } F(X)$ は,

$$\text{grad } F(X) = F_X X + X F_X - 2X F_X X, \quad (5.13)$$

となる. ここで, F_X は (i, j) 成分が $\partial F(X)/\partial X_{ij}$ であるような $n \times n$ 行列である.

次に, ξ を点 $X \in \text{Grass}(p, n)$ における接ベクトルとすると, X における F のヘシアン $\text{Hess } F(X)$ は, ξ に次のように作用する.

$$\text{Hess } F(X)[\xi] = 2 \text{sym} \left(X \text{sym}(F_{XX}[\xi] + 4F_X \xi X - 2F_X \xi) (I_n - X) \right). \quad (5.14)$$

ここで, $F_{XX}[\xi]$ は (i, j) 成分が $\sum_{k,l=1}^n (\partial^2 F(X)/\partial X_{ij} \partial X_{kl}) \xi_{kl}$ であるような $n \times n$ 行列である.

このことに注意して, 先に挙げた問題 3.10 に対する解法アルゴリズムを具体的に導出することができる.

5.3 グラスマン多様体上のレイリー商の勾配とヘシアン

問題 3.10 を再掲する.

問題 5.1.

$$\begin{aligned} & \text{minimize} && F(X) = \text{tr}(AX), \\ & \text{subject to} && X \in \text{Grass}(p, n). \end{aligned}$$

$F(X) = \text{tr}(AX)$ について、勾配とヘシアンはそれぞれ

$$\text{grad } F(X) = \text{sym}(AX) - XAX, \quad (5.15)$$

$$\text{Hess } F(X)[\xi] = 2 \text{sym}((X\xi A - XA\xi)(I_n - X)), \quad \xi \in T_X \text{Grass}(p, n) \quad (5.16)$$

と計算される。

5.4 最急降下法とニュートン法・ハイブリッド法

こうして問題 5.1 に対する解法アルゴリズムを書き下すことができる。ここでは個別のアルゴリズムをあらためて書かないが、いくつかコメントをする。まずリーマン多様体 M 上の最急降下法については、一般に M がコンパクトであれば、生成される点列は目的関数 f の臨界点に線形収束することが知られている。グラスマン多様体 $\text{Grass}(p, n)$ はコンパクトであり、特に、レイリー商 F の臨界点は最小点と最大点を除いて鞍点であることが示されるので、問題 5.1 に対する最急降下法によって生成される点列は F の最小点に線形収束する。

次に、ニュートン法についてであるが、ニュートン方程式は

$$\xi_k B_k + B_k \xi_k = C_k \quad (5.17)$$

となる。ここで、 $B_k := A - AX_k - X_k A$ 、 $C_k := 2(-\text{sym}(AX_k) + X_k AX_k)$ である。方程式 (5.17) はリヤプノフ方程式と呼ばれるものであり、その数値解法が盛んに研究されている [3]。一般に、目的関数 f の臨界点 $x_* \in M$ において $\text{Hess } f(x_*)^{-1}$ が存在するとき、 x_* の十分近くを初期点にとると、ニュートン法によって生成される点列 $\{x_k\}$ は x_* に 2 次収束する。

以上を踏まえて、最急降下法とニュートン法を組み合わせたハイブリッド法を適用すると、これによって生成される点列は大域的最適解に大域的に 2 次収束することが分かる。実際に最急降下法のための反復を用いた場合と、ハイブリッド法を用いた場合を比較したものが次の図 5.1 である。

ここで、 A の固有値が $\lambda_1 \leq \dots \leq \lambda_{p-1} < \lambda_p = \dots = \lambda_{p+r-1} < \lambda_{p+r} \leq \dots \leq \lambda_n$ である場合、つまり小さい方から p 番目の固有値が $r (\leq n - p)$ 重に縮退している場合を考える。このとき、レイリー商の大域的最小点 X_* の集合は実射影空間 $\mathbb{R}P^{r-1}$ に同相な部分多様体をなす。さらに、各 X_* においてヘシアン $\text{Hess } F(X_*)$ が退化しており、 $\xi \in T_{X_*} \text{Grass}(p, n)$ について、 $\langle \text{Hess } F(X_*)[\xi], \xi \rangle = 0$ であるための必要十分条件は、 ξ が上記の部分多様体に接することである。ところで、レイリー商最小化問題において、 $X \in \text{Grass}(p, n)$ に対するニュートン方程式はリヤプノフ方程式

$$\eta B + B\eta = C, \quad B = A - AX - XA, C = 2(-\text{sym}(AX) + XAX)$$

であった。この方程式が一意に解けるための必要十分条件は B と $-B$ が共通の固有値を持たないことである。 A が縮退する固有値 $\lambda_p = \lambda_{p+1}$ を持つとき、行列 $B = A - AX_* - X_* A$

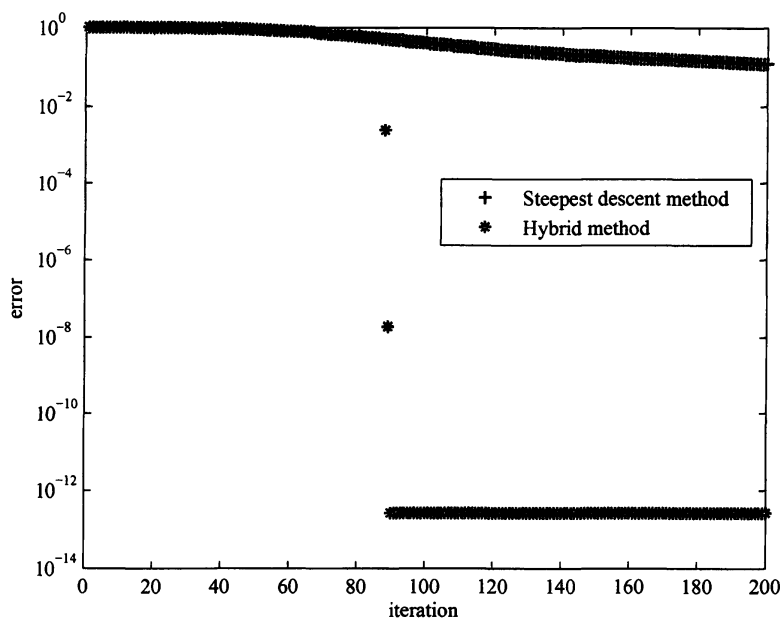


図 5.1: $n = 300, p = 150$.

は固有値 $-\lambda_p$ と λ_{p+1} を持つから, B と $-B$ は共通の固有値 $-\lambda_p = -\lambda_{p+1}$ を持つ. よって, X_* においてはニュートン方程式

$$\eta B + B\eta = 0, \quad B = A - AX_* - X_*A$$

は $\mathbb{R}^{n \times n}$ の行列方程式と見て一意に解けない. 一方, 大域的最適解 X_* について, 行列 $B = A - AX_* - X_*A$ の固有値は $-\lambda_1, \dots, -\lambda_p, \lambda_{p+1}, \dots, \lambda_n$ である. すると, $\lambda_p = \lambda_{p+1}$ なら $\text{Hess } F(X_*)$ は退化しており, ニュートン方程式 $\text{Hess } F(X_*)[\eta] = -\text{grad } F(X_*)$, つまり

$$\eta B + B\eta = 0, \quad \eta \in T_{X_*} \text{Grass}(p, n) \quad (5.18)$$

が非自明な解 $\eta \neq 0$ を持つから, 先ほどのリヤプノフ方程式が一意に解けないという事実と整合性が取れている.

6 シュティーフェル多様体上の最適化問題について

問題 3.11 を再掲する. ただし, 最適化の分野では普通, 目的関数は最小化するように選ぶので, 最大化すべき関数の -1 倍を最小化することにする.

問題 6.1.

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

6.1 レトラクション

シュティーフエル多様体の積 $\text{St}(p, m) \times \text{St}(p, n)$ 上のレトラクションも、グラスマン多様体のときと同様に指数レトラクションと QR レトラクションの2つを紹介する。まず指数写像によるレトラクションは次の通りである。

$$\begin{aligned} R_{(U,V)}(\xi, \eta) &= \text{Exp}_{(U,V)}(\xi, \eta) \\ &= \begin{pmatrix} U & \xi \end{pmatrix} \exp \left(\begin{pmatrix} U^T \xi & -\xi^T \xi \\ I_p & U^T \xi \end{pmatrix} \right) \begin{pmatrix} I_p \\ 0 \end{pmatrix} \exp(-U^T \xi), \\ &\quad \begin{pmatrix} V & \eta \end{pmatrix} \exp \left(\begin{pmatrix} V^T \eta & -\eta^T \eta \\ I_p & V^T \eta \end{pmatrix} \right) \begin{pmatrix} I_p \\ 0 \end{pmatrix} \exp(-V^T \eta). \end{aligned} \quad (6.1)$$

ここで、 $(\xi, \eta) \in T_{(U,V)}(\text{St}(p, m) \times \text{St}(p, n))$ である。

また、QR 分解に基づくレトラクションは次の通りである。 $(U, V) \in \text{St}(p, m) \times \text{St}(p, n)$ に対して、

$$R_{(U,V)}(\xi, \eta) = (\text{qf}(U + \xi), \text{qf}(V + \eta)), \quad (\xi, \eta) \in T_{(U,V)}(\text{St}(p, m) \times \text{St}(p, n)). \quad (6.2)$$

とすれば、 $R_{(U,V)}$ によって $\text{St}(p, m) \times \text{St}(p, n)$ 上のレトラクションが定義される。

6.2 問題 6.1 の目的関数の勾配とヘシアン

$\text{St}(p, m) \times \text{St}(p, n)$ 上の目的関数 $F(U, V) = -\text{tr}(U^T A V N)$ の勾配とヘシアンについて述べる。 F の $(U, V) \in \text{St}(p, m) \times \text{St}(p, n)$ における勾配は、

$$\text{grad } F(U, V) = (U \text{sym}(U^T A V N) - A V N, V \text{sym}(V^T A^T U N) - A^T U N), \quad (6.3)$$

で表される。また、 (U, V) におけるヘシアン $\text{Hess } F(U, V)$ については、接ベクトル $(\xi, \eta) \in T_{(U,V)}(\text{St}(p, m) \times \text{St}(p, n))$ に対して

$$\begin{aligned} \text{Hess } F(U, V)[(\xi, \eta)] &= \left(U \text{skew}(U^T(\xi S_1 - A \eta N)) + (I_m - U U^T)(\xi S_1 - A \eta N), \right. \\ &\quad \left. V \text{skew}(V^T(\eta S_2 - A^T \xi N)) + (I_n - V V^T)(\eta S_2 - A^T \xi N) \right). \end{aligned} \quad (6.4)$$

のように作用する。ここで、 S_1 と S_2 は、それぞれ行列 $\text{sym}(U^T A V N)$, $\text{sym}(V^T A^T U N)$ である。

6.3 最急降下法とニュートン法

以上で、問題 6.1 の最急降下法やニュートン法を導出するために必要なものは揃った。

最急降下法

最急降下法では点 $(U_k, V_k) \in \text{St}(p, m) \times \text{St}(p, n)$ における探索方向を

$$(\xi_k, \eta_k) = -\text{grad } F(U_k, V_k) = (AV_k N - U_k \text{sym}(U_k^T AV_k N), A^T U_k N - V_k \text{sym}(V_k^T A^T U_k N)) \quad (6.5)$$

として計算する。しかし、実際に具体例に対して数値計算を行うと、図 6.3 のように、最急降下法による収束はかなり遅いことが分かる。

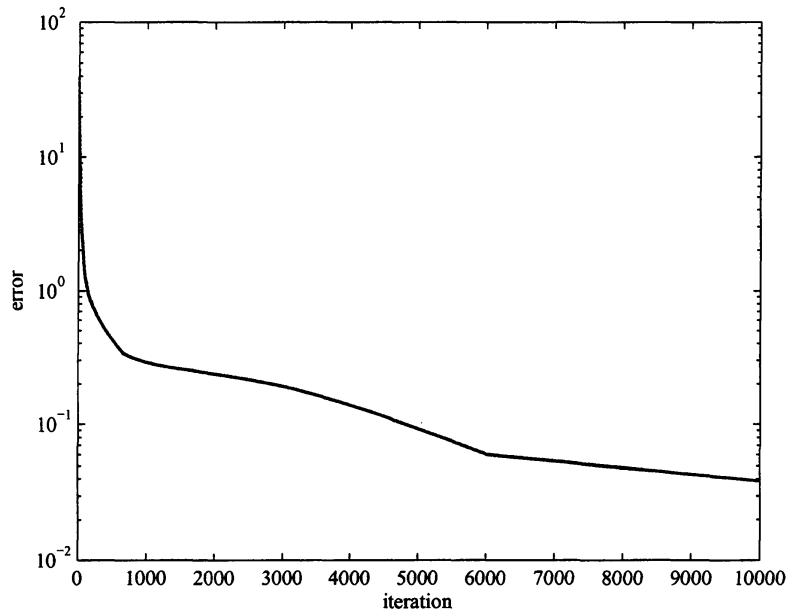


図 6.1: 最急降下法の数値計算結果. $m = 1000, n = 500, p = 10$.

ニュートン法

ニュートン法においては、ニュートン方程式は未知の $(\xi, \eta) \in T_{(U_k, V_k)}(\text{St}(p, m) \times \text{St}(p, n))$ に対して

$$\begin{cases} U_k \text{skew}(U_k^T(\xi S_{1,k} - A\eta N)) + (I_m - U_k U_k^T)(\xi S_{1,k} - A\eta N) \\ = AV_k N - U_k \text{sym}(U_k^T AV_k N), \\ V \text{skew}(V_k^T(\eta S_{2,k} - A^T \xi N)) + (I_n - V_k V_k^T)(\eta S_{2,k} - A^T \xi N) \\ = A^T U_k N - V_k \text{sym}(V_k^T A^T U_k N) \end{cases} \quad (6.6)$$

となる. ここで, $S_{1,k} = \text{sym}(U_k^T A V_k N)$, $S_{2,k} = \text{sym}(V_k^T A^T U_k N)$ である.

グラスマン多様体上のレイリー商に対するニュートン方程式がリヤプノフ方程式の形で書かれたのとは対照的に, この線形方程式を解くのは困難である. しかし, 最急降下法の収束の速さでは, 実用的とはとても言えないので, ニュートン方程式 (6.6) を真正面から解くかどうかはともかく, 何らかの形でニュートン法の二次収束の性質を取り入れたアルゴリズムを導出することが必要である. これは, 今後の課題である.

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] A. Edelman, T. A. Arias, and S. T. Smith, The geometry of algorithms with orthogonality constraints, *SIAM J. Matrix Anal. Appl.*, **20**(2):303-353, 1998.
- [3] Z. Gajic and M. T. J. Qureshi, *Lyapunov Matrix Equation in System Stability and Control*. Academic Press, Inc, 1995.
- [4] G. H. Golub and C. F. V. Loan, *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, third edition, 1996.
- [5] L. N. Trefethen and D. Bau, *Numerical Linear Algebra*. SIAM, 1997.
- [6] J. Nocedal and S. J. Wright, *Numerical Optimization*. Springer Series in Operations Research. Springer-Verlag, New York, 1999.