

論 文

リーマン多様体上の最適化の理論と応用

佐藤 寛之

1 はじめに

近年、リーマン多様体上の最適化理論とその応用が盛んに研究されている。1994年にSmithはリーマン多様体上のニュートン法や共役勾配法を提案した[26]。リーマン多様体の中でもシュテューフェル多様体やグラスマン多様体は直交制約付きの行列空間における最適化問題を記述するのに適しており、1998年にはEdelmanらによってこうした多様体における最適化の多数の応用例がその理論とともに示されている[6]。こうした幾何学的な最適化手法が提案された当初は、リーマン多様体上の指数写像や平行移動など数学的に自然な概念を用いる手法が研究されていたが、数値計算の効率を上げるためにこれらを近似した概念であるレトラクションやvector transportなどが整備され、2008年のAbsilらの著書[1]にはこうした幾何学的な概念や最適化アルゴリズムが体系的にまとめられている。本稿では、こうしたリーマン多様体上の最適化の理論とその応用について、基本的な事柄から最近の研究結果までを紹介する。ユークリッド空間 \mathbb{R}^n における最適化アルゴリズムをリーマン多様体上に拡張することを念頭に置いて議論を進めるために、まず \mathbb{R}^n における十分滑らかな実数値関数 f に対する無制約最適化

問題

$$\text{minimize } f(x), \quad x \in \mathbb{R}^n$$

を考える[3][20]。最小化しようとする f を目的関数という。この問題に対する直線探索に基づく反復解法では初期点 $x_0 \in \mathbb{R}^n$ を選び、更新式

$$x_{k+1} = x_k + t_k \eta_k, \quad k=0, 1, 2, \dots \quad (1)$$

によって \mathbb{R}^n 上の点列 $\{x_k\}$ を生成する。 $\eta_k \in \mathbb{R}^n$ は探索方向と呼ばれ、目的関数が減少する方向として計算されることが多い。 $t_k > 0$ はステップ幅と呼ばれるスカラーで、探索方向 η_k が定まったときにその方向にどれくらい進むと目的関数値が十分減少するかを調べることで決定されることが多い。とくに、 f を x_k において η_k 方向に微分すると、

$$\left. \frac{d}{dt} f(x_k + t\eta_k) \right|_{t=0} = \nabla f(x_k)^T \eta_k \quad (2)$$

となる。ここで、 ∇f は f のユークリッド勾配、上付きの T は転置を表す。式(2)より、 η_k が

$$\nabla f(x_k)^T \eta_k < 0 \quad (3)$$

を満たすとき、十分小さい正の実数 t が存在して $f(x_k + t\eta_k) < f(x_k)$ が成り立つ。不等式(3)を満たす探索方向 η_k を f の x_k における降下方向という。探索方向とステップ幅をどう決めるかが最適化手法を特徴づける。たとえば $\eta_k = -\nabla f(x_k)$

[筆者紹介]



さとう ひろゆき。2009年3月京都大学工学部情報学科卒業。2011年3月同大学大学院情報学研究科数理工学専攻修士課程修了。2013年11月同博士後期課程修了。博士(情報学)。同年12月より日本学術振興会特別研究員(PD)。2014年4月より東京理科大学工学部第一部経営工学科助教。2016年4月より東京理科大学工学部情報工学科助教。多様体上の最適化手法の理論およびその応用の研究に従事。日本応用数理学会、日本オペレーションズ・リサーチ学会、IEEE、SIAM各会員。

としたアルゴリズムは最急降下法と呼ばれ、 $\nabla f(x_k) \neq 0$ ならば (3) が明らかに成り立つが、実用上、その収束は遅いことが知られている。同様に目的関数の 1 階微分の情報のみを用いるが、より速い収束性をもつ共役勾配法も提案されていて、これは本稿の主題であり 4 節で詳しく議論する。一方、ニュートン法では目的関数の 2 階微分の情報をもつヘッセ行列 $\nabla^2 f(x)$ を用いて探索方向を $\eta_k = -\nabla^2 f(x_k)^{-1} \nabla f(x_k)$ とする。

本稿ではこうした最適化アルゴリズムのリーマン多様体上への拡張を議論するが、多様体上の最適化問題の簡単な具体例を示すために、まずユークリッド空間における次の最適化問題を考える。

$$\text{minimize } f(x) = \frac{x^T A x}{x^T x}, \quad x \in \mathbb{R}^n - \{0\}. \quad (4)$$

ここで、 A は n 次実対称行列であり、 f はレイリー商と呼ばれる。Courant-Fischer の定理より、この問題の最適解 x_* は行列 A の最小固有値に属する固有ベクトルである [8]。目的関数 f の勾配 ∇f およびヘッセ行列 $\nabla^2 f$ は、それぞれ

$$\nabla f(x) = \frac{2(A - f(x)I_n)x}{x^T x}$$

および

$$\nabla^2 f(x) = \frac{2H_x(A - f(x)I_n)H_x}{x^T x}$$

と計算できる。ここで、 I_n は n 次単位行列であり、 $H_x := I_n - 2(xx^T)/(x^T x)$ は x に関するハウスホルダー行列である。 H_x は正則であるから、 $f(x)$ が A の固有値でない限り、すなわち行列 $A - f(x)I_n$ が正則である限り、 $\nabla^2 f(x)$ は正則で、 $d \in \mathbb{R}^n$ についての線形方程式 $\nabla^2 f(x)d = -\nabla f(x)$ は一意解をもつ。その解が $d = x$ であることは容易に確かめられる。簡単のためステップ幅を $t_k \equiv 1$ と固定すると、ニュートン法の各反復では $x_{k+1} = x_k + d_k = 2x_k$ という計算が行われ、任意の局所的最適解のいくら近くに初期点 x_0 をとっても収束しないことがある。

この議論から、最適化問題 (4) に対してニュー

トン法は局所的 2 次収束性を必ずしももたないことがわかる。とくに、問題 (4) の大域的最小解を x_* として x_* の近傍において f を観察してみる。ニュートン法の局所的 2 次収束性の証明においては通常、目的関数のヘッセ行列が局所的最小解において正定値であることを仮定するが、問題 (4) においては、 $f(x_*)$ が A の最小固有値であることから上での議論より $\nabla^2 f(x_*)$ は特異行列であり、正定値性をもたない。したがって、ニュートン法の 2 次収束性を保証できない。

$\nabla^2 f(x_*)$ が正定値行列でないことは次のようにして確認することもできる。まず、

$$f(tx) = f(x), \quad t \in \mathbb{R} - \{0\} \quad (5)$$

に注意して、この式の両辺において $x = x_*$ と固定して t で 2 回微分した後 $t = 1$ とおくことで、 $x_*^T \nabla^2 f(x_*) x_* = 0$ を得る。 $x_* \neq 0$ であるから、これは $\nabla^2 f(x_*)$ が正定値でないことを意味している。この議論より、目的関数がレイリー商である場合に限らない一般の最適化問題に対しても、目的関数 f が 0 次同次性 (5) をもち、かつ $x_* \neq 0$ ならば、ニュートン法の局所的 2 次収束性を必ずしも保証できないことがわかる。

ところで、式 (5) が成り立つとき、 f の値は x の方向のみで決定され x のノルムの大きさには依存しないことがわかる。そこで、ユークリッドノルムを 1 に固定して、 $(n-1)$ 次元球面 $S^{n-1} := \{x \in \mathbb{R}^n | x^T x = 1\}$ 上で f を最小化しても同じことである。このように考えると、問題 (4) は次の問題と等価である。

$$\text{minimize } f(x) = x^T A x, \quad x \in S^{n-1}. \quad (6)$$

球面 S^{n-1} はリーマン多様体の最も簡単な例の一つであり、この問題はリーマン多様体上の最適化問題である。 $f(x) = x^T A x$ をユークリッド空間上で制約条件 $x^T x = 1$ の下で最小化すると考えるのではなく、リーマン多様体 S^{n-1} 上の無制約最小化問題に変換することで、以降の節で述べる、リーマン多様体上に拡張されたニュートン法や共役勾配法などの無制約最適化手法を適用することが

できる.

2 リーマン多様体について

本節では, 多様体の基本的な事柄をまとめ, 本稿で扱う最適化問題の舞台となるリーマン多様体を定義する. 詳細は[15][16]を参照されたい.

定義 2.1 ハウスドルフ空間 M の任意の点 p について, p を含む開集合 U および U から \mathbb{R}^n の開集合 U' への同相写像 $\varphi: U \rightarrow U'$ が存在するとき, M を n 次元位相多様体という.

すなわち, n 次元位相多様体 M 上の各点に対して, その近傍であって \mathbb{R}^n と同相なものが存在する. 定義 2.1 における (U, φ) を n 次元座標近傍という. $p \in M$ は一般には座標で表すことができないが, $p \in U$ のとき $\varphi(p) \in \mathbb{R}^n$ であるから, p は φ を通して n 個の実数の組として表現できる. そこで, φ を U 上の局所座標系という. また定義から, M が n 次元位相多様体であるとき, 添え字集合 Λ および M の n 次元座標近傍からなる族 $\{(U_\lambda, \varphi_\lambda)\}_{\lambda \in \Lambda}$ が存在して $M = \bigcup_{\lambda \in \Lambda} U_\lambda$ となる. このとき族 $\{(U_\lambda, \varphi_\lambda)\}_{\lambda \in \Lambda}$ を M の座標近傍系という.

最急降下法やニュートン法のように, 微分を用いた無制約最適化手法を多様体上で考えるためには, 多様体上の目的関数に対して微分操作ができなければならない. そのためには考えている多様体に微分構造を入れることができる必要がある.

定義 2.2 r を正の整数または ∞ とする. M を n 次元位相多様体とし, $\{(U_\lambda, \varphi_\lambda)\}_{\lambda \in \Lambda}$ を M の座標近傍系とする. このとき, $U_\alpha \cap U_\beta \neq \emptyset$ なる任意の $\alpha, \beta \in \Lambda$ について, $\varphi_\beta \circ \varphi_\alpha^{-1}: \varphi_\alpha(U_\alpha \cap U_\beta) \rightarrow \varphi_\beta(U_\alpha \cap U_\beta)$ が C^r 級写像であるとき M を n 次元 C^r 級可微分多様体という.

座標近傍 $(U_\alpha, \varphi_\alpha), (U_\beta, \varphi_\beta)$ に対して, $U_\alpha \cap U_\beta$ に属している点 p は局所座標系 φ_α による局所座

標 $\varphi_\alpha(p)$ および φ_β による局所座標 $\varphi_\beta(p)$ をもつが, その間の座標変換 $\varphi_\beta \circ \varphi_\alpha^{-1}$ が C^r 級であることを上の定義では要請している. また, $0 \leq s \leq r$ である整数 s に対して, n 次元 C^r 級可微分多様体 M 上の実数値関数 $f: M \rightarrow \mathbb{R}$ が C^s 級であるとは, M の座標近傍系に属する任意の座標近傍 (U, φ) に対して, $f \circ \varphi^{-1}$ が $\varphi(U)$ 上で C^s 級であることをいう. f は多様体 M 上の関数であるが, $f \circ \varphi^{-1}$ を考えることで φ を通して表現され, 通常の実数空間での微分と同様の操作を考えることができる. 以降では, $r = \infty$ とし C^∞ 級可微分多様体を扱う.

多様体は曲面の概念を抽象化したものである. 滑らかな曲面上の各点で接平面を考えることができるように, C^∞ 級可微分多様体においても接空間と呼ばれるベクトル空間を定義することができる. たとえば \mathbb{R}^3 内の 2 次元曲面上の点 p を通る曲線 c で $c(0) = p$ となるものを考えると, 点 p において

$$c'(0) = \lim_{t \rightarrow 0} \frac{c(t) - c(0)}{t} \quad (7)$$

は曲面の接ベクトルである. $c(t) - c(0) \in \mathbb{R}^3$ であるから (7) の計算は可能であるが, 多様体 M 上の曲線 c を考えたときに $c(0), c(t) \in M$ の差は一般には定義されない. そこで, 多様体上の接ベクトルを次のように実数値関数に対する微分作用素として定義する. 以降では, $p \in M$ の近傍で定義される滑らかな実数値関数全体の集合を $\mathfrak{F}_p(M)$ と書く.

定義 2.3 ξ を $\mathfrak{F}_p(M)$ から \mathbb{R} への写像とする. M 上の曲線 c が存在して, $c(0) = p$ かつ

$$\xi f = \left. \frac{d(f(c(t)))}{dt} \right|_{t=0}, \quad f \in \mathfrak{F}_p(M)$$

が成り立つとき, ξ を M の点 p における接ベクトルといい $\dot{c}(0)$ と表す. また, このような ξ 全体の集合を M の点 p における接空間といい $T_p M$ で表す.

次に、最適化において重要となる接ベクトルの内積やノルムを考えるために、リーマン計量およびリーマン多様体を導入する。

定義 2.4 C^∞ 級可微分多様体 M 上の 2 次の対称テンソル場 $\langle \cdot, \cdot \rangle$ が任意の点 $p \in M$ において正定値であるとき、 $\langle \cdot, \cdot \rangle$ を M 上のリーマン計量といい、リーマン計量を備えた多様体をリーマン多様体という。

テンソル場の定義は割愛するが、 M 上の各点 p における接空間 $T_p M$ において内積 $\langle \cdot, \cdot \rangle_p$ が与えられていて、これが点 p について滑らかであるとき、各点 p に $\langle \cdot, \cdot \rangle_p$ を対応させる写像がリーマン計量である。リーマン計量 $\langle \cdot, \cdot \rangle$ が与えられると、 M の任意の点 p における接ベクトル ξ のノルムを

$$\|\xi\|_p := \sqrt{\langle \xi, \xi \rangle_p}$$

によって定義でき、さらに M 上の曲線 $c : [a, b] \rightarrow M$ の長さ $L(c)$ を

$$L(c) := \int_a^b \|\dot{c}(t)\|_{c(t)} dt$$

によって定義できる。リーマン多様体 M が連結であるとき、任意の $p, q \in M$ に対して p と q を結ぶあらゆる曲線 c の長さ $L(c)$ の下限を $d(p, q)$ とすると、 $d : M \times M \rightarrow \mathbb{R}$ は M 上の距離関数となり、したがって M は距離空間となる。

この節の最後に、いくつかのリーマン多様体の例を紹介する。以下では p, n を $p \leq n$ を満たす正の整数とする。シュティーフエル多様体 $\text{St}(p, n) := \{Y \in \mathbb{R}^{n \times p} | Y^T Y = I_p\}$ は、たとえば以下のようにリーマン計量を入れることでリーマン多様体となる。 $\text{St}(p, n) \subset \mathbb{R}^{n \times p}$ であり、任意の $Y \in \text{St}(p, n)$ における接空間 $T_Y \text{St}(p, n)$ は $\mathbb{R}^{n \times p}$ の部分空間と見なすことができる。そこで、 $\mathbb{R}^{n \times p}$ の標準内積

$$\langle Z, W \rangle = \text{tr}(Z^T W), \quad Z, W \in \mathbb{R}^{n \times p}$$

に基づいて、 $T_Y \text{St}(p, n)$ の内積 $\langle \cdot, \cdot \rangle_Y$ を

$$\langle \xi, \eta \rangle_Y = \text{tr}(\xi^T \eta), \quad \xi, \eta \in T_Y \text{St}(p, n)$$

と定義すると、これにより $\text{St}(p, n)$ のリーマン計量が定まる。1 節で紹介した球面 S^{n-1} や n 次直交群 $O(n)$ については $\text{St}(1, n) = S^{n-1}$, $\text{St}(n, n) = O(n)$ であるから、これらはシュティーフエル多様体の特別な場合としてやはりリーマン多様体となる。たとえば、球面には \mathbb{R}^n の標準内積

$$\langle z, w \rangle = z^T w, \quad z, w \in \mathbb{R}^n$$

からリーマン計量

$$\langle \xi, \eta \rangle_x = \xi^T \eta, \quad \xi, \eta \in T_x S^{n-1}, \quad x \in S^{n-1} \quad (8)$$

が誘導される。リーマン多様体の他の例としては、 \mathbb{R}^n の p 次元部分空間全体からなるグラスマン多様体 $\text{Grass}(p, n)$ などがある。

シュティーフエル多様体 $\text{St}(p, n)$ は $\mathbb{R}^{n \times p}$ の部分多様体であるが、各点が \mathbb{R}^n の部分空間であるグラスマン多様体にはこのように自明な外側のユークリッド空間があるわけではない。しかし、こうしたユークリッド空間の部分多様体であるとは限らない一般のリーマン多様体上の最適化問題も扱えるよう、多くの最適化アルゴリズムがある程度一般的な多様体上の枠組みで議論されている。

3 リーマン多様体上の最適化手法の一般論

本節では、最適化で必要となる関数の勾配などの概念を一般のリーマン多様体 M 上で定義するとともに、それらの概念を具体的に球面の例を通して説明する。最後に M 上の最適化アルゴリズムについて議論する。以降では、 C^∞ 級のリーマン多様体 M 上の滑らかな目的関数 f の無制約最適化問題

$$\text{minimize } f(x), \quad x \in M \quad (9)$$

を考える。また、球面 S^{n-1} にはリーマン計量 (8) が導入されているとする。

多様体 M_1, M_2 と写像 $F : M_1 \rightarrow M_2$ に対して、 F の点 $x \in M_1$ における微分 $DF(x)$ は $T_x M_1$ から $T_{F(x)} M_2$ への写像であって、 $\xi \in T_x M_1$ に対して $DF(x)[\xi] \in T_{F(x)} M_2$ が

$$(DF(x)[\xi])h = \xi(h \circ F), \quad \forall h \in \mathfrak{F}_{F(x)}(M_2)$$

を満たすものとして定義される. 点 $x \in M$ において目的関数は $f \in \mathfrak{F}_x(M)$ と見なせるから, $T_{f(x)}\mathbb{R}$ と \mathbb{R} を同一視すると, とくに $\xi \in T_x M$ に対して $Df(x)[\xi] = \xi f$ である.

定義 3.1 リーマン多様体 M 上の関数 f の点 $x \in M$ における勾配 $\text{grad } f(x)$ は

$$\langle \text{grad } f(x), \xi \rangle_x = Df(x)[\xi], \quad \forall \xi \in T_x M \quad (10)$$

を満たすものとして一意に定まる $T_x M$ の元である.

式(10)は \mathbb{R}^n における事実(2)に対応している. 本稿ではユークリッド空間における通常の勾配, 定義 3.1 による勾配をそれぞれ ∇ および grad を用いて表す. たとえば, 問題(6)の目的関数をユークリッド空間全体で定義されていると考えて勾配を計算すると $\nabla f(x) = 2Ax$ となるが, S^{n-1} 上では $\text{grad } f(x) = 2(I_n - xx^T)Ax$ となる. これが点 $x \in S^{n-1}$ における接ベクトルであることおよび式(10)を満たすことは, $T_x S^{n-1} = \{\xi \in \mathbb{R}^n \mid \xi^T x = 0\}$ に注意すると容易に確かめられる.

以降では, リーマン多様体 M 上の点 x_k における f の勾配 $\text{grad } f(x_k)$ を g_k と書くことにする. 勾配を定義したので, たとえば M 上の最急降下法では点 x_k における探索方向を $\eta_k = -g_k$ とすれば良い. しかし, ステップ幅 t_k を適当に選んだとしても, 式(1)を用いて点列を生成することはできない. なぜなら右辺の和が M 上では一般には定義されないからである. ここで, 更新式(1)の右辺は点 x_k から η_k の方向に伸びる半直線上で次の点を探索した結果を計算する式と見なせる. t_k を適切に決定するところが探索に相当する. これに倣って, 多様体 M 上で点 x_k と探索方向 η_k が得られているとき, x_k から η_k の方向に伸びる曲線, すなわち

$$c_k(0) = x_k, \quad \dot{c}_k(0) = \eta_k \quad (11)$$

なる曲線 c_k 上で次の点を探索すれば良い. その

ための道具としてレトラクション^{†1}と呼ばれる写像を次のように定義する[1]. ここで, $TM := \bigcup_{x \in M} \{x\} \times T_x M$ は M の接バンドルと呼ばれる.

定義 3.2 滑らかな写像 $R: TM \rightarrow M$ が

$$R_x(0) = x, \quad DR_x(0) = \text{id}_{T_x M}$$

を満たすとき, R をレトラクションという. ただし, R_x は R の $T_x M$ への制限を表し, $\text{id}_{T_x M}$ は $T_x M$ における恒等写像を表す.

M 上のレトラクション R が与えられたとき, 点 x_k と接ベクトル $\eta_k \in T_{x_k} M$ に対して曲線 c_k を $c_k(t) := R_{x_k}(t\eta_k)$ によって定めれば, レトラクションの定義より c_k は(11)を満たす望ましい曲線となる.

測地線を定義する指数写像はレトラクションの特別な例であるが[‡], 数値的には必ずしも簡単に計算できるとは限らない. レトラクションは一意に定まるものではなく, たとえば球面 S^{n-1} 上で

$$R_x(\eta) = \frac{x + \eta}{\|x + \eta\|}, \quad \eta \in T_x S^{n-1}, \quad x \in S^{n-1} \quad (12)$$

によって定まる R はレトラクションの一例である. ここで, $\|\cdot\|$ は \mathbb{R}^n のユークリッドノルムである.

また, 4 節で述べる共役勾配法では異なる接空間に属する 2 つのベクトルに対して, 片方を他方の接空間に移す必要が生じる. 測地線に沿った平行移動を用いればそれは可能であるが, 計算上の観点から平行移動を一般化した次の vector transport という概念が提案された[1]. ここで, $TM \oplus TM := \{(\eta, \xi) \mid \eta, \xi \in T_x M, x \in M\}$ である.

定義 3.3 M 上の写像 $\mathcal{T}: TM \oplus TM \rightarrow TM$ について $\mathcal{T}(\eta, \xi)$ を $\mathcal{T}_\eta(\xi)$ と書くことにする. 任意の点 $x \in M$ についてレトラクション R が存在して, $\eta, \xi, \zeta \in T_x M$ に対して次の 3 つの条件が成り立つとき \mathcal{T} を M 上の vector transport という.

^{†1} 位相幾何学分野においてもレトラクションと呼ばれる概念があるが, ここでのものとは異なる[16].

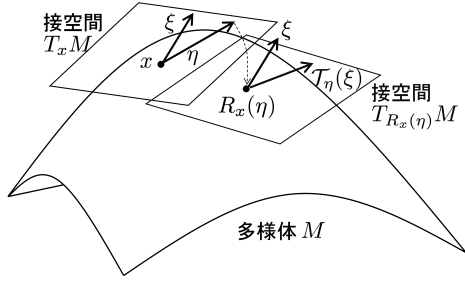


図1 レトラクションと vector transport の概念図

- 1) $\mathcal{T}_\eta(\xi) \in T_{R_x(\eta)}M$,
- 2) $\mathcal{T}_0(\xi) = \xi$,
- 3) $\mathcal{T}_\eta(a\xi + b\zeta) = a\mathcal{T}_\eta(\xi) + b\mathcal{T}_\eta(\zeta)$, $a, b \in \mathbb{R}$.

M 上のレトラクション R に対して, 任意の点 $x \in M$ について

$$\mathcal{T}_\eta^R(\xi) = DR_x(\eta)[\xi], \quad \eta, \xi \in T_x M \quad (13)$$

によって定義される \mathcal{T}^R は vector transport の一例であることが確かめられる[1].

図1に, レトラクションと vector transport の概念を図示する. 図において, 多様体 M 上の点 x から $\eta \in T_x M$ の方向に進んだ M 上の点がレトラクションによって $R_x(\eta)$ と得られている. また, ξ は $T_x M$ の元であるが, 一般には $T_{R_x(\eta)} M$ の元ではない. そこで, vector transport \mathcal{T} によって ξ を η の方向に移動させた, 点 $R_x(\eta)$ における接ベクトルが $\mathcal{T}_\eta(\xi)$ と得られている.

レトラクションを用いることで, \mathbb{R}^n での直線探索に対応して M 上の曲線上の探索に基づく問題(9)に対する反復アルゴリズムを記述することができる. すなわち, 適当な初期点 $x_0 \in M$ から出発し,

$$x_{k+1} = R_{x_k}(t_k \eta_k), \quad k = 0, 1, 2, \dots \quad (14)$$

によって M 上の点列 $\{x_k\}$ を生成する. t_k はステップ幅であり, $\eta_k \in T_{x_k} M$ は探索方向である. とくに共役勾配法については次の節で詳しく述べる.

詳細は割愛するが, ヘッセ行列に対応する概念としてリーマン多様体上の f のヘシアン $\text{Hess } f$ も定義される. たとえば問題(6)における球面上で

の点 x における $\text{Hess } f(x)$ の接ベクトル η への作用は, $T_x S^{n-1}$ への射影 $P_x = I_n - xx^T$ を用いて, $\text{Hess } f(x)[\eta] = 2(P_x A P_x - f(x)I_n)\eta$ となる. これよりニュートン法の探索方向を定義する線形方程式 $\text{Hess } f(x)[\eta] = -\text{grad } f(x)$ は, $\eta \in T_x S^{n-1}$, すなわち $x^T \eta = 0$ から,

$$\eta = -x + \frac{(A - f(x)I_n)^{-1}x}{x^T(A - f(x)I_n)^{-1}x}$$

と解ける. したがって, レトラクション(12)を用いた場合のニュートン法の更新式は

$$x_{k+1} = R_{x_k}(\eta_k) = \frac{(A - f(x_k)I_n)^{-1}x_k}{\|(A - f(x_k)I_n)^{-1}x_k\|}$$

となり, これは数値線形代数の分野でよく知られているレイリー商反復法と等価である[8]. この事実から, レイリー商反復法の局所的2次収束性などはリーマン多様体上のニュートン法の一般論からただちに導かれる.

4 リーマン多様体上の共役勾配法

共役勾配法は, 数値線形代数の分野では正定値対称な係数行列 A をもつ線形方程式 $Ax = b$ の解法として知られている[10]. 最適化の観点からは, これは狭義凸な目的関数 $x^T A x / 2 - b^T x$ の \mathbb{R}^n での無制約最小化問題に対する反復解法と見ることができる. この目的関数を一般の滑らかな関数とした場合に拡張されたアルゴリズムを非線形共役勾配法といい, 線形方程式の解法としての共役勾配法をとくに線形共役勾配法という. 本稿では, 非線形共役勾配法のことを単に共役勾配法と呼ぶことにする. \mathbb{R}^n での共役勾配法における探索方向は, 初期点においては最急降下法と同じく $\eta_0 = -\nabla f(x_0)$ とし, $k \geq 1$ のときは $\beta_k \in \mathbb{R}$ に対して

$$\eta_k = -\nabla f(x_k) + \beta_k \eta_{k-1}, \quad (15)$$

すなわち最急降下方向に前の探索方向をスカラー倍して加えたものとする. β_k の計算方法はたとえば

$$\beta_k^{\text{FR}} = \frac{\|\nabla f(x_k)\|^2}{\|\nabla f(x_{k-1})\|^2}, \quad \beta_k^{\text{DY}} = \frac{\|\nabla f(x_k)\|^2}{\eta_{k-1}^T y_k}$$

(16)

がそれぞれ Fletcher と Reeves[7], Dai と Yuan[5] によって提案されている. ただし, $y_k := \nabla f(x_k) - \nabla f(x_{k-1})$ とおいた. 他にも多数の β_k の計算方法が提案されており, それぞれの方法を用いた共役勾配法の性質が研究されている. また, 式(15)に修正項を付け加えた3項共役勾配法も提案されている[2][19]. ユークリッド空間における共役勾配法の詳細については[9][18]を参照されたい.

さて, 共役勾配法のリーマン多様体 M 上への拡張について議論する. 式(15)の右辺の第1項および第2項は, M 上ではそれぞれ x_k および x_{k-1} における接ベクトルであり, 異なる接空間に属するので足し合わせることができない. [26]では Liu と Storey[17]による β_k を M 上に拡張したものに対し, 指数写像や平行移動を用いたアルゴリズムが提案され, 局所的収束性の解析が行われている. その後, [1]ではレトラクションや vector transport を用いた共役勾配法が提案された. 探索方向 $\eta_k \in T_{x_k}M$ の更新式としては, 式(15)に vector transport \mathcal{T} による修正を施した

$$\eta_k = -g_k + \beta_k \mathcal{T}_{t_{k-1}\eta_{k-1}}(\eta_{k-1}) \quad (17)$$

を用いる. [21]では(17)において, \mathcal{T} を(13)で定義される \mathcal{T}^R とし, Fletcher-Reeves の β_k を M 上に拡張した

$$\beta_k^{\text{FR}} = \frac{\|g_k\|_{x_k}^2}{\|g_{k-1}\|_{x_{k-1}}^2} \quad (18)$$

を用いる方法が提案され, $k=0, 1, 2, \dots$ に対して

$$\|\mathcal{T}_{t_k\eta_k}^R(\eta_k)\|_{x_{k+1}} \leq \|\eta_k\|_{x_k} \quad (19)$$

が成り立つという仮定の下でアルゴリズムの大域的収束性が証明されている. しかし, この仮定は \mathcal{T} が M 上の測地線に沿った平行移動であるときは成り立つが, そうでないときは一般には成り立たない. [24]では(17)の第2項を修正することで(19)を仮定せずとも大域的収束性が保証される Fletcher-Reeves 型のアルゴリズムが提案され, [22]ではさらに弱い仮定でも大域的収束性が保証される Dai-Yuan 型のアルゴリズムが提案された.

以降では[22][24]の結果を詳しく紹介する.

共役勾配法ではステップ幅の選び方が収束性に大きく影響する. \mathbb{R}^n においては, 点 x_k と f の降下方向であるような探索方向 η_k が既知のとき, $\phi(t) := f(x_k + t\eta_k)$ とおくことで, $t=t_k$ において ϕ の値が十分小さくなっている条件や ϕ のグラフの接線の傾きが十分緩やかになっている条件として, 定数 $0 < c_1 < c_2 < 1$ に対して

$$\phi(t_k) \leq \phi(0) + c_1 t \phi'(0), \quad (20)$$

$$\phi'(t_k) \geq c_2 \phi'(0), \quad (21)$$

$$|\phi'(t_k)| \leq c_2 |\phi'(0)| \quad (22)$$

といったものが考えられており, (20)をアルミホ条件, (20)と(21)をあわせてウルフ条件, (20)と(22)をあわせて強ウルフ条件という. M 上では点 $x_k \in M$ において降下方向であるような, すなわち $\langle g_k, \eta_k \rangle_{x_k} < 0$ であるような探索方向 $\eta_k \in T_{x_k}M$ が与えられたとしても, 直線探索は実行できない. そこで, [22][24]では代わりにレトラクション R が定める曲線上の探索を考え $\phi(t) := f(R_{x_k}(t\eta_k))$ とし, この ϕ について条件(20)-(22)を f について書き下すことで, \mathbb{R}^n におけるステップ幅に対する条件を

$$f(R_{x_k}(t_k\eta_k)) \leq f(x_k) + c_1 t \langle g_k, \eta_k \rangle_{x_k}, \quad (23)$$

$$\begin{aligned} & \langle \text{grad } f(R_{x_k}(t_k\eta_k)), \text{DR}_{x_k}(t_k\eta_k)[\eta_k] \rangle_{R_{x_k}(t_k\eta_k)} \\ & \geq c_2 \langle g_k, \eta_k \rangle_{x_k}, \end{aligned} \quad (24)$$

$$\begin{aligned} & |\langle \text{grad } f(R_{x_k}(t_k\eta_k)), \text{DR}_{x_k}(t_k\eta_k)[\eta_k] \rangle_{R_{x_k}(t_k\eta_k)}| \\ & \leq c_2 |\langle g_k, \eta_k \rangle_{x_k}| \end{aligned} \quad (25)$$

と M 上に拡張している. \mathbb{R}^n の場合と同様に, (23)をアルミホ条件, (23)と(24)をあわせてウルフ条件, (23)と(25)をあわせて強ウルフ条件と呼ぶ.

[24]では不等式(19)を仮定せず, 探索方向の更新式(17)を

$$\eta_k = -g_k + \beta_k \mathcal{T}^{(k-1)}(\eta_{k-1}) \quad (26)$$

に変更し, β_k として(18)の β_k^{FR} を用いるアルゴリズムが提案されている. (26)において, $\mathcal{T}^{(k)}$ は

$$\mathcal{T}^{(k)}(\eta_k) := \begin{cases} \mathcal{T}_{t_k\eta_k}^R(\eta_k) & ((19) \text{ が成り立つとき}), \\ \mathcal{T}_{t_k\eta_k}^0(\eta_k) & (\text{そうでないとき}) \end{cases}$$

と定義される. ここで,

$$\mathcal{J}_{t_k \eta_k}^0(\eta_k) := \frac{\|\eta_k\|_{x_k}}{\|\mathcal{J}_{t_k \eta_k}^R(\eta_k)\|_{x_{k+1}}} \mathcal{J}_{t_k \eta_k}^R(\eta_k)$$

である. すなわち, (19)が成り立たないときは \mathcal{J}^R によって移されたベクトルのノルムを調整している.

[22]でもやはり(19)を仮定せず, (26)を用いて(16)の Dai-Yuan の β_k を拡張した

$$\beta_k^{\text{DY}} = \frac{\|g_k\|_{x_k}^2}{\langle g_k, \mathcal{J}^{(k-1)}(\eta_{k-1}) \rangle_{x_k} - \langle g_{k-1}, \eta_{k-1} \rangle_{x_{k-1}}} \quad (27)$$

が提案されている.

いずれのアルゴリズムも以下のように大域的収束性が証明されている.

定理 4.1 リーマン多様体 M 上の最適化問題 (9)において目的関数 f は下に有界で C^1 級とし, 定数 $L > 0$ が存在して, 任意の $t \geq 0, x \in M$, および $\|\eta\|_x = 1$ なる $\eta \in T_x M$ に対して

$$|D(f \circ R_x)(t\eta)[\eta] - D(f \circ R_x)(0)[\eta]| \leq Lt$$

が成り立つとする. ステップ幅が強ウルフ条件 (23), (25)を満たすとき, Fletcher-Reeves 型の共役勾配法 (14), (18), (26)により生成される点列 $\{x_k\}$ は

$$\liminf_{k \rightarrow \infty} \|\text{grad } f(x_k)\|_{x_k} = 0 \quad (28)$$

を満たす. また, ステップ幅がウルフ条件 (23), (24)を満たすとき, Dai-Yuan 型の共役勾配法 (14), (26), (27)により生成される点列 $\{x_k\}$ も (28)を満たす.

ウルフ条件は強ウルフ条件より弱い条件であるから, Dai-Yuan 型のアルゴリズムは Fletcher-Reeves 型の改良版と考えられる. [22]では数値的にも Dai-Yuan 型の方が Fletcher-Reeves 型より収束が速いことが実験により確かめられている.

先述の通り共役勾配法はユークリッド空間でも盛んに研究されており, 最新の結果も引き続きリーマン多様体上に拡張されていくと期待される.

5 リーマン多様体上の最適化の応用

本節では, リーマン多様体上の最適化の応用として, 数値線形代数における話題から最近の研究結果までを紹介する.

1 節で述べたように, 最小固有値問題は球面上のレイリー商最小化問題と等価である. これを拡張し, p ($\leq n$) 個の正規直交ベクトルそれぞれに対する, n 次対称行列 A のレイリー商の重み付き和を最小化するシュティーフェル多様体上の最適化問題

$$\text{minimize } \text{tr}(Y^T A Y N), \quad Y \in \text{St}(p, n)$$

を考えよう. ただし, $N = \text{diag}(\mu_1, \mu_2, \dots, \mu_p)$, $0 < \mu_1 < \mu_2 < \dots < \mu_p$ である. この問題の最適解の第 i 列ベクトルは, 行列 A の i 番目に小さい固有値に属する固有ベクトルである [1]. これに類似した例で特異値分解への応用もある. $m \geq n$ とし, 任意の長方形行列 $A \in \mathbb{R}^{m \times n}$ と $N = \text{diag}(\mu_1, \mu_2, \dots, \mu_p)$, $\mu_1 > \mu_2 > \dots > \mu_p > 0$ に対して, 最適化問題

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

$$(U, V) \in \text{St}(p, m) \times \text{St}(p, n) \quad (29)$$

の最適解を (U_*, V_*) とすると, U_*, V_* の第 i 列ベクトルはそれぞれ A の i 番目に大きい特異値に属する左特異ベクトル, 右特異ベクトルである [23].

数値線形代数の分野以外への応用の研究も盛んに行われている. n 次正定値対称行列 G に対して $\text{St}_G(p, n) := \{Y \in \mathbb{R}^{n \times p} | Y^T G Y = I_p\}$ を一般化シュティーフェル多様体という. (29)を一般化した

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

$$(U, V) \in \text{St}_B(p, m) \times \text{St}_C(p, n) \quad (30)$$

を考えよう. 詳細は割愛するが, 2 組のデータ行列が与えられ, さらに時間とともに新しいデータが次々に追加されていくような状況において, A を相互共分散行列, B, C をそれぞれのデータの自己共分散行列としたとき, 問題 (30) は正準相関分析と呼ばれる統計的手法において現れる最適化問題となる [29]. 他の統計的手法として, トポグラフィック独立成分分析への応用も研究されている [14]. トポグラフィック独立成分分析では古典

的な独立成分分析において残留してしまう成分間の従属性を、トポグラフィックな順序を決めるのに用いる[11]。そこで現れる尤度の最大化問題はシュティーフェル多様体上の最適化問題となる。

また、[12]では機械学習にも用いられるテンソル補完問題への応用が提案されている。そこで扱われているのは、特定の成分だけが既知であるようなテンソル $\mathcal{X}^* \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ を良く近似するテンソル \mathcal{X} を、固定されたランク (r_1, r_2, r_3) のテンソル全体の中から見つける問題である。 \mathcal{X} のタッカー分解を用いることで、この問題は $\text{St}(r_1, n_1) \times \text{St}(r_2, n_2) \times \text{St}(r_3, n_3) \times \mathbb{R}^{r_1 \times r_2 \times r_3}$ を $\mathcal{O}(r_1) \times \mathcal{O}(r_2) \times \mathcal{O}(r_3)$ で割った商多様体上の最適化問題に書き直される。

さらに、制御理論における応用として、与えられた制御モデルを良く近似する低次元のモデルを求める問題が、シュティーフェル多様体やグラスマン多様体上で、元のモデルと近似モデルの伝達関数の誤差の H^2 ノルムを最小化する問題として定式化できる[25][27]。

他にも、二重確率行列の逆固有値問題は積多様体 $\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$ 上の最適化問題として定式化できる[28]。ここで、 $\mathcal{OB} := \{Z \in \mathbb{R}^{n \times n} | \text{diag}(ZZ^T) = I_n\}$ であり、 \mathcal{U} は n 次狭義上三角行列全体の部分集合である。

本節では、いくつかの応用例とその多様体上の最適化問題としての定式化について簡単に紹介したが、それぞれの問題に対してその特徴を考慮した最適化手法が用いられ、新たな解法アルゴリズムが提案されている。さらに、機械学習の分野で大規模な問題に対する有効な手法として近年注目を浴びている確率的勾配降下法も、リーマン多様体上への拡張の研究が進んでおり、今後の発展が期待される[4][13]。また、たとえば2016年に発表された[28]においては、上で述べた二重確率行列の逆固有値問題に対して、2015年の[24]で提案された、4節で紹介したリーマン多様体上のFletcher-Reeves型の共役勾配法に基づく解法が提案されており、このように比較的新しい理論研

究の結果も応用研究において用いられている。このことから、リーマン多様体上の最適化は理論と応用の両方において研究が盛んに行われていることがわかる。

謝 辞

本研究は日本学術振興会科学研究費補助金(課題番号：16K17647)の助成を受けている。

参考文献

- [1] Absil, P.-A., Mahony, R. and Sepulchre, R., Optimization Algorithms on Matrix Manifolds, Princeton University Press, Princeton, 2008.
- [2] Al-Baali, M., Narushima, Y. and Yabe, H., A family of three-term conjugate gradient methods with sufficient descent property for unconstrained optimization, Computational Optimization and Applications, 60(2015), 89–110.
- [3] Bertsekas, D. P., Nonlinear Programming, 2nd Edition, Athena Scientific, Belmont, 1999.
- [4] Bonnabel, S., Stochastic gradient descent on Riemannian manifolds, IEEE Transactions on Automatic Control, 58(2013), 2217–2229.
- [5] Dai, Y. H. and Yuan, Y., A nonlinear conjugate gradient method with a strong global convergence property, SIAM Journal on Optimization, 10(1999), 177–182.
- [6] Edelman, A., Arias, T. A. and Smith, S. T., The geometry of algorithms with orthogonality constraints, SIAM Journal on Matrix Analysis and Applications, 20(1998), 303–353.
- [7] Fletcher, R. and Reeves, C. M., Function minimization by conjugate gradients, The Computer Journal, 7(1964), 149–154.
- [8] Golub, G. H. and Van Loan, C. F., Matrix Computations, 4th Edition, Johns Hopkins University Press, Baltimore, 2012.
- [9] Hager, W. W. and Zhang, H., A survey of nonlinear conjugate gradient methods, Pacific Journal of Optimization, 2(2006), 35–58.
- [10] Hestenes, M. R. and Stiefel, E., Methods of conjugate gradients for solving linear systems, Journal of Research of the National Bureau of Standards, 49(1952), 409–436.
- [11] Hyvärinen, A., Karhunen, J. and Oja, E., Independent component analysis, John Wiley & Sons, New York, 2001.
- [12] Kasai, H. and Mishra, B., Low-rank tensor completion: a Riemannian manifold preconditioning approach, Proceedings of the 33rd International Conference on Machine Learning, 2016.
- [13] Kasai, H., Sato, H. and Mishra, B., Riemannian stochastic variance reduced gradient on Grassmann manifold, The 9th NIPS Workshop on Optimization for Machine Learning, 2016.
- [14] Kitao, A., Shiohama, T. and Sato, H., Financial news

- classification based on topographic independent component analysis: Optimization on the Stiefel manifold, Proceedings of the 16th Applied Stochastic Models and Data Analysis International Conference, 2015, 403–415.
- [15] Lee, J. M., Riemannian Manifolds: An Introduction to Curvature, Springer-Verlag, New York, 1997.
- [16] Lee, J. M., Introduction to Smooth Manifolds, 2nd Edition, Springer-Verlag, New York, 2012.
- [17] Liu, Y. and Storey, C., Efficient generalized conjugate gradient algorithms, Part 1: Theory, Journal of Optimization Theory and Applications, 69 (1991), 129–137.
- [18] 成島康史, 大規模無制約最適化問題に対する非線形共役勾配法の最近の研究動向, 応用数理, 22 (2012), 27–39.
- [19] Narushima, Y., Yabe, H. and Ford, J. A., A three-term conjugate gradient method with sufficient descent property for unconstrained optimization, SIAM Journal on Optimization, 21 (2011), 212–230.
- [20] Nocedal, J. and Wright, S., Numerical Optimization, 2nd Edition, Springer-Verlag, New York, 2006.
- [21] Ring, W. and Wirth, B., Optimization methods on Riemannian manifolds and their applications to shape space, SIAM Journal on Optimization, 22 (2012), 596–627.
- [22] Sato, H., A Dai-Yuan-type Riemannian conjugate gradient method with the weak Wolfe conditions, Computational Optimization and Applications, 64 (2016), 101–118.
- [23] Sato, H. and Iwai, T., A Riemannian optimization approach to the matrix singular value decomposition, SIAM Journal on Optimization, 23 (2013), 188–212.
- [24] Sato, H. and Iwai, T., A new, globally convergent Riemannian conjugate gradient method, Optimization, 64 (2015), 1011–1031.
- [25] Sato, H. and Sato, K., Riemannian trust-region methods for H^2 optimal model reduction, Proceedings of the 54th IEEE Conference on Decision and Control, 2015, 4648–4655.
- [26] Smith, S. T., Optimization techniques on Riemannian manifolds, Fields Institute Communications, 3, Bloch, A. (ed.), American Mathematical Society, Providence, 1994, 113–136.
- [27] Xu, Y. and Zeng, T., Fast optimal \mathcal{H}_2 model reduction algorithms based on Grassmann manifold optimization, International Journal of Numerical Analysis and Modeling, 10 (2013), 972–991.
- [28] Yao, T. T., Bai, Z. J., Zhao, Z. and Ching, W. K., A Riemannian Fletcher-Reeves conjugate gradient method for doubly stochastic inverse eigenvalue problems, SIAM Journal on Matrix Analysis and Applications, 37 (2016), 215–234.
- [29] Yger, F., Berar, M., Gasso, G. and Rakotomamonjy, A., Adaptive canonical correlation analysis based on matrix manifolds, Proceedings of the 29th International Conference on Machine Learning, 2012, 1071–1078.

[Abstract]

This paper deals with Riemannian optimization, that is, optimization on Riemannian manifolds. Theories of Euclidean optimization and Riemannian manifolds are first briefly reviewed together with some simple and motivating examples, followed by the Riemannian optimization theory. Retractions and vector transports on Riemannian manifolds are introduced according to the literature to describe a general Riemannian optimization algorithm. Recent convergence analysis results of several types of Riemannian conjugate gradient methods, such as Fletcher-Reeves and Dai-Yuan-types, are then given and discussed in detail. Some applications of Riemannian optimization to problems of current interest, such as 1) singular value decomposition in numerical linear algebra; 2) canonical correlation analysis and topographic independent component analysis as statistical methods; 3) low-rank tensor completion for machine learning; 4) optimal model reduction in control theory; and 5) doubly stochastic inverse eigenvalue problem, are also introduced.