

# リーマン多様体上の最適化理論とその周辺

佐藤 寛之

京都大学 大学院情報学研究科 数理工学専攻

連続最適化および関連分野に関する夏季学校 2021

2021 年 8 月 24 日

# 講演内容

- 1 はじめに
- 2 応用例
- 3 球面上の最急降下法の考え方
- 4 位相空間とリーマン多様体
- 5 リーマン多様体上の一般の最適化手法と最急降下法
- 6 リーマン多様体上の共役勾配法 (R-CG)
- 7 まとめ

# 講演内容

- 1 はじめに
- 2 応用例
- 3 球面上の最急降下法の考え方
- 4 位相空間とリーマン多様体
- 5 リーマン多様体上の一般の最適化手法と最急降下法
- 6 リーマン多様体上の共役勾配法 (R-CG)
- 7 まとめ

# 講義の構成

- 13:00–14:00
  - リーマン多様体上の最適化問題の例
  - 球面上の最急降下法の概観
  - 位相空間（・リーマン多様体）
- 14:10–15:10
  - （位相空間・）リーマン多様体
  - リーマン多様体上の最適化アルゴリズムとそこで用いられる写像
  - リーマン多様体上の最急降下法とその大域的収束性
- 15:20–16:20
  - リーマン多様体上の共役勾配法とその性質
  - 最新の研究動向

## 記法のまとめ

- $\mathbb{R}^{m \times n}$ :  $m \times n$  実行列全体;  $\mathbb{R}_r^{m \times n}$ : ランク  $r$  の  $m \times n$  実行列全体
  - 集合  $X, Y$  に対して,  $X - Y := \{x \in X \mid x \notin Y\}$ :  $X$  における  $Y$  の補集合
  - $n$  次単位行列を  $I_n$ , また, 集合  $X$  上の恒等写像を  $\text{id}_X$  または単に  $\text{id}$  と書く
- このページの以下では  $A \in \mathbb{R}^{n \times n}$  とする.
- $\text{tr}(A) := \sum_{i=1}^n a_{ii}$ :  $A$  のトレース
  - $\|A\|_F := \sqrt{\text{tr}(A^T A)}$ :  $A$  のフロベニウスノルム  
 $(\|a\|_2 := \sqrt{a^T a}: a \in \mathbb{R}^n \text{ の } 2\text{-ノルム (ユークリッドノルム)})$
  - $A$  が正則であるとき,  $A^{-T} := (A^{-1})^T = (A^T)^{-1}$ :  $A$  の転置の逆行列
  - $\text{sym}(A) := (A + A^T)/2$ :  $A$  の対称部分;  $\text{skew}(A) := (A - A^T)/2$ :  $A$  の反対称部分
  - $\text{Sym}(n) := \{A \in \mathbb{R}^{n \times n} \mid A = A^T\}$ : 実  $n$  次対称行列全体
  - $\text{Skew}(n) := \{A \in \mathbb{R}^{n \times n} \mid A = -A^T\}$ : 実  $n$  次反対称行列全体
  - $\mathcal{O}(n) := \{A \in \mathbb{R}^{n \times n} \mid A^T A = A A^T = I_n\}$ : 実  $n$  次直交行列全体 (直交群)
  - $\text{diag}(d_1, d_2, \dots, d_n)$ : 第  $i$  対角成分が  $d_i$  の  $n \times n$  対角行列

## 記法のまとめ（詳細はこの後のスライドで説明）

- 球面  $S^{n-1} = \{x \in \mathbb{R}^n \mid x^T x = 1\}$
- シュテューフェル多様体  $\text{St}(p, n) = \{X \in \mathbb{R}^{n \times p} \mid X^T X = I_p\}$  ( $p \leq n$ )
- グラスマン多様体  $\text{Grass}(p, n) = \{W \subset \mathbb{R}^n \mid W \text{ は } p \text{ 次元部分空間}\}$  ( $p \leq n$ )
- $\text{Sym}(n)^+ = \{A \in \text{Sym}(n) \mid A \text{ は正定値}\}$
- $\nabla f(x) = \left( \frac{\partial f}{\partial x_i}(x) \right)$ : 関数  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  の点  $x \in \mathbb{R}^n$  におけるユークリッド勾配
- $\mathfrak{F}_x(M)$ : 多様体  $M$  上の点  $x$  の近傍で定義される滑らかな実数値関数全体の集合
- $T_x M$ : 多様体  $M$  上の点  $x$  における接空間
- $DF(x)$ : 多様体  $M, N$  の間の写像  $F: M \rightarrow N$  の点  $x \in M$  における微分
- $\text{grad } f(x)$ : リーマン多様体  $(M, g)$  上の関数  $f: M \rightarrow \mathbb{R}$  の点  $x \in M$  における勾配

# よく使う関係式

- $A \in \mathbb{R}^{n \times n}$  に対して,  $\text{tr}(A) = \text{tr}(A^T)$ .
- $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times m}$  に対して,  $\text{tr}(AB) = \text{tr}(BA)$ .
- $A = (a_{ij}), B = (b_{ij}) \in \mathbb{R}^{m \times n}$  に対して,  $\text{tr}(A^T B) = \sum_{i,j} a_{ij} b_{ij}$ .  
 $\rightarrow a = (a_i), b = (b_i) \in \mathbb{R}^n$  に対する標準内積  $a^T b = \sum_{i=1}^n a_i b_i$  の行列空間への拡張

## 幾何学的な最適化

詳細な文献情報はスライド末尾の参考文献リストを参照.

1970 年代・1980 年代・1990 年代の研究：

- [Luenberger 1972] The gradient projection method along geodesics
- [Tanabe 1979] Differential geometric methods in nonlinear programming
- [Tanabe 1980] A geometric method in nonlinear programming
- [Gabay 1982] Minimizing a differentiable function over a differential manifold
- [Udriște 1994] Convex Functions and Optimization Methods on Riemannian Manifolds
- [Smith 1994] Optimization techniques on Riemannian manifolds
- [Edelman et al. 1998] The geometry of algorithms with orthogonality constraints

etc.

近年の研究で用いられる道具等が体系的にまとまった書籍

- [Absil et al. 2008] Optimization Algorithms on Matrix Manifolds



# 講演内容

- 1 はじめに
- 2 応用例
- 3 球面上の最急降下法の考え方
- 4 位相空間とリーマン多様体
- 5 リーマン多様体上の一般の最適化手法と最急降下法
- 6 リーマン多様体上の共役勾配法 (R-CG)
- 7 まとめ

- 今回の講義では、イメージを掴みやすくするために、一般論を、主に球面の例を交えながら説明する。
- ただし、球面はシンプルな多様体であるため、多様体上で最適化を行うことの意義を感じられにくいかもしれない。
- そこで、最初に球面より複雑な多様体上の最適化問題を、応用例を通していくつか紹介する。

## 例 1: 主成分分析—第一主成分を求める問題—

中心化されたデータ  $x_1, x_2, \dots, x_N \in \mathbb{R}^n$  に対する主成分分析:

射影した際にデータの分散が最大になる方向 (第一主成分)  $u \in \mathbb{R}^n$  を求めたい。  
方向を求めたいので  $\|u\|_2 = 1$  とし, また,  $X := (x_1, x_2, \dots, x_N) \in \mathbb{R}^{n \times N}$  とすると,

$$\sum_{i=1}^N \|(u^T x_i)u\|_2^2 = \sum_{i=1}^N \|uu^T x_i\|_2^2 = \|uu^T X\|_F^2 = \text{tr}(X^T uu^T X) = u^T X X^T u$$

を最大化したい。ノルム 1 の  $u \in \mathbb{R}^n$  全体は単位超球面 (以降, 単に球面という)  $S^{n-1} := \{u \in \mathbb{R}^n \mid \|u\|_2 = 1\}$  をなすから, 以下の最適化問題が得られる。

### 問題 2.1 (第一主成分を求めるための球面上の最適化問題)

$$\begin{aligned} & \text{maximize} && f(u) = u^T X X^T u, \\ & \text{subject to} && u \in S^{n-1}. \end{aligned}$$

## 例 1: 主成分分析—第二主成分まで求める問題—

中心化されたデータ  $X = (x_1, x_2, \dots, x_N) \in \mathbb{R}^{n \times N}$  に対する主成分分析 (続き):  
 第一主成分  $u_1 \in S^{n-1}$  に対して, 第二主成分  $u_2 \in S^{n-1}$  は  $u_1$  と直交するように,  
 かつ部分空間  $\text{span}\{u_1, u_2\}$  へのデータの射影の分散が最大になるように求めたい.

$$\sum_{i=1}^N \|(u_1 u_1^T + u_2 u_2^T) x_i\|_2^2 = \|(u_1 u_1^T + u_2 u_2^T) X\|_F^2 = u_1^T X X^T u_1 + u_2^T X X^T u_2$$

を最大化したい. さらに,  $u_1, u_2$  を同時に求めるには,  $u_1$  の項が重視されるよう  
 定数  $\mu_1 > \mu_2 > 0$  により重み付けをし,  $\mu_1 u_1^T X X^T u_1 + \mu_2 u_2^T X X^T u_2$  を最大化したい.

### 問題 2.2 (第二主成分までを求めるための2つの球面上の最適化問題)

$$\begin{aligned} & \text{maximize} && \mu_1 u_1^T X X^T u_1 + \mu_2 u_2^T X X^T u_2, \\ & \text{subject to} && u_1, u_2 \in S^{n-1}, u_1^T u_2 = 0. \end{aligned}$$

## 例 1: 主成分分析—第 $p$ 主成分まで求める問題—

中心化されたデータ  $X = (x_1, x_2, \dots, x_N) \in \mathbb{R}^{n \times N}$  に対する主成分分析 (続き):  
 同様に, 第  $p$  主成分 ( $p \leq n$ ) まで  $U = (u_1, u_2, \dots, u_p) \in \mathbb{R}^{n \times p}$  を同時に求めるには,  
 重み  $\mu_1 > \mu_2 > \dots > \mu_p > 0$  からなる対角行列  $D = \text{diag}(\mu_1, \mu_2, \dots, \mu_p)$  を用いて

$$\sum_{i=1}^p \mu_i u_i^T X X^T u_i = \sum_{i=1}^p (UD)_i^T (X X^T U)_i = \text{tr}((UD)^T (X X^T U)) = \text{tr}(U^T X X^T U D)$$

を条件  $u_i^T u_j = \delta_{ij} \iff U^T U = I_p$  の下で最大化すれば良い. これを満たす  $U$  全体  
 $\text{St}(p, n) := \{U \in \mathbb{R}^{n \times p} \mid U^T U = I_p\}$  をシュティーフェル多様体という.

### 問題 2.3 (第 $p$ 主成分までのためのシュティーフェル多様体上の問題)

$$\begin{aligned} & \text{maximize} \quad \text{tr}(U^T X X^T U D), \\ & \text{subject to} \quad U \in \text{St}(p, n). \end{aligned}$$

## 例 1: 主成分分析—第 $p$ 主成分までの張る部分空間を求める問題—

中心化されたデータ  $X = (x_1, x_2, \dots, x_N) \in \mathbb{R}^{n \times N}$  に対する主成分分析 (別の設定):  
 第  $p$  主成分までが張る部分空間  $W \subset \mathbb{R}^n$  を求めるには,  $W$  のある正規直交基底  $\{u_1, u_2, \dots, u_p\}$  について  $U := (u_1, u_2, \dots, u_p) \in \text{St}(p, n)$  を用いて, 重みなしの和  $\sum_{i=1}^p u_i^T X X^T u_i = \text{tr}(U^T X X^T U)$  を最大化すれば良い.  
 この値は  $W$  の正規直交基底の選び方によらない. 実際,  $Q \in \mathcal{O}(p)$  に対して,

$$\text{tr}((UQ)^T X X^T (UQ)) = \text{tr}(Q^T U^T X X^T U Q) = \text{tr}(U^T X X^T U).$$

つまり  $W$  の正規直交基底  $U$  を一つ選ぶと  $f(W) := \text{tr}(U^T X X^T U)$  は well-defined.  
 $\text{Grass}(p, n) := \{W \mid W \text{ は } \mathbb{R}^n \text{ の } p \text{ 次元部分空間}\}$  をグラスマン多様体という.

### 問題 2.4 (第 $p$ 主成分までのためのグラスマン多様体上の問題)

$$\begin{aligned} & \text{maximize} && f(W), \\ & \text{subject to} && W \in \text{Grass}(p, n). \end{aligned}$$

## グラスマン多様体 $\text{Grass}(p, n)$ についての補足

$U \in \text{St}(p, n)$  と, 任意の  $Q \in \mathcal{O}(p)$  についての  $UQ$  は, 同じ部分空間を張る.  
そこで,  $\text{St}(p, n)$  に同値関係  $\sim$  を,  $U_1, U_2 \in \text{St}(p, n)$  に対して

$$U_1 \sim U_2 : \Longleftrightarrow \exists Q \in \mathcal{O}(p), U_2 = U_1 Q$$

と定める. すると,  $U \in \text{St}(p, n)$  の同値類

$$[U] := \{UQ \mid Q \in \mathcal{O}(p)\}$$

は,  $U$  が張る部分空間と同一視できる. このような同値類の全体

$$\text{St}(p, n)/\mathcal{O}(p) := \text{St}(p, n)/\sim := \{[U] \mid U \in \text{St}(p, n)\}$$

は,  $\mathbb{R}^n$  の  $p$  次元部分空間全体, つまり  $\text{Grass}(p, n)$  と同一視できる.

$\text{Grass}(p, n) = \text{St}(p, n)/\mathcal{O}(p)$  には**商多様体**の構造が入る (が, 今回は深入りしない).

## 例 2: 低ランク行列補完

行列  $A = (a_{ij}) \in \mathbb{R}^{m \times n}$  の要素はいくつかのみが既知であるとし,  
 $\Omega := \{(i, j) \mid a_{ij} \text{ は既知} \}$  とする.  
 $A$  のランクを  $r$  として補完する最適化問題は以下のように書ける.

### 問題 2.5 (低ランク行列補完問題)

$$\begin{aligned} & \text{minimize} && \sum_{(i,j) \in \Omega} (x_{ij} - a_{ij})^2, \\ & \text{subject to} && X = (x_{ij}) \in \mathbb{R}_r^{m \times n}. \end{aligned}$$

ここで,  $\mathbb{R}_r^{m \times n} := \{X \in \mathbb{R}^{m \times n} \mid \text{rank}(X) = r\}$  は  $\mathbb{R}^{m \times n}$  の部分多様体である.



## 例 2: 低ランク行列補完

- 仮定  $\text{rank}(A) = r$  は, 分解  $A = LR^T$ ,  $L \in \mathbb{R}_r^{m \times r}$ ,  $R \in \mathbb{R}_r^{n \times r}$  の存在の仮定と等価.
- たとえば,  $A$  の要素がユーザ  $i$  のアイテム  $j$  に対する評価  $a_{ij}$  からなるならば, 潜在的な  $r$  個の特徴を仮定して, 各ユーザや各アイテムがそれぞれの特徴とどのような関係かを数値化したものが行列  $L, R$  であると解釈できる.
- このような  $L$  と  $R$  を決定変数にして最適化問題を考えることもできるが, 分解  $A = LR^T$  は一意ではない. ただし,  $A = L_1 R_1^T = L_2 R_2^T$  ならば, ある正則行列  $G \in \mathbb{R}^{r \times r}$  が存在して,  $L_2 = L_1 G$ ,  $R_2 = R_1 G^{-T}$  と書ける. そこで,  $\mathbb{R}_r^{m \times r} \times \mathbb{R}_r^{n \times r}$  に同値関係  $\sim$  を

$$(L_1, R_1) \sim (L_2, R_2) : \iff \exists G \in \mathbb{R}^{r \times r}, L_2 = L_1 G, R_2 = R_1 G^{-T}$$

と定めれば,  $\mathbb{R}_r^{m \times n}$  を商多様体  $(\mathbb{R}_r^{m \times r} \times \mathbb{R}_r^{n \times r}) / \mathbb{R}^{r \times r}$  と同一視することもできる.

## その他の応用例（ごく簡単に）

- 混合ガウスモデル [Hosseini & Sra 2015, Boumal 2020]:  
分散共分散行列の推定の際, 多様体  $\text{Sym}(d)^+ := \{X \in \text{Sym}(d) \mid X \text{ は正定値} \}$   
(と, 他の多様体の積多様体) 上の最適化問題が現れる.
- 辞書学習 [Sun et al. 2017, Boumal 2020]:  
決定変数 (辞書)  $b_1, b_2, \dots, b_n \in S^{d-1}$  についての最適化問題が現れる.  
 $b_1, b_2, \dots, b_n \in S^{d-1} \iff (b_1, b_2, \dots, b_n) \in S^{d-1} \times S^{d-1} \times \dots \times S^{d-1}$  より  
これは積多様体  $S^{d-1} \times S^{d-1} \times \dots \times S^{d-1}$  上の最適化問題.  
さらに,  $X = (b_1, b_2, \dots, b_n) \in \mathbb{R}^{d \times n}$  とおくと,  
 $b_1, b_2, \dots, b_n \in S^{d-1} \iff \text{diag}(X^T X) = I_n$  より,  $S^{d-1} \times S^{d-1} \times \dots \times S^{d-1}$  は  
**oblique manifold**  $\{X \in \mathbb{R}^{d \times n} \mid \text{diag}(X^T X) = I_n\}$  と同一視できる.
- 正準相関分析 [Yger et al. 2012, Sato & Aihara 2019]  
データ  $X \in \mathbb{R}^{N \times m}, Y \in \mathbb{R}^{N \times n}$  の分散共分散行列  $C_X, C_Y$  に対して  
 $\text{St}_{C_X}(p, m) \times \text{St}_{C_Y}(p, n)$  上の最適化問題が現れる. ここで, 正定値対称行列  $G$  に  
対して一般化シュティーフェル多様体  $\text{St}_G(p, n) := \{X \in \mathbb{R}^{n \times p} \mid X^T G X = I_p\}$ .

## 固有値問題に関連する $\mathbb{R}^n$ 上の最適化問題

$A \in \text{Sym}(n)$  とする.

### 問題 2.6 (レイリー商最小化問題)

$$\begin{aligned} &\text{minimize} && f(x) = \frac{x^T A x}{x^T x}, \\ &\text{subject to} && x \in \mathbb{R}^n - \{0\}. \end{aligned}$$

この問題の大域的最適解は  $A$  の最小固有値に属する固有ベクトルである.  
最適解を計算するためにニュートン法を適用してみると….

### 命題

点  $x$  が  $f$  の停留点でないならば,  $x$  におけるニュートン方向, すなわち,  
ニュートン方程式  $\nabla^2 f(x)[\eta] = -\nabla f(x)$  の解  $\eta$  は,  $\eta = x$  となる.

→  $x_k = x_{k-1} + \eta_{k-1} = 2x_{k-1} = 2^k x_0$  より, 初期点が停留点でなければ,  
いくら反復を繰り返しても停留点に到達できない.

## 固有値問題に関連する球面 $S^{n-1}$ 上の最適化問題

問題 2.6 は,  $\mathbb{R}^n$  上の制約条件付きの最適化問題と見なせる.

### 問題 2.7 ( $\mathbb{R}^n$ 上の制約付き最適化問題)

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

$n-1$  次元球面  $S^{n-1} := \{x \in \mathbb{R}^n \mid x^T x = 1\}$  (多様体) 上の  
制約条件なしの最適化問題と見なすこともできる.

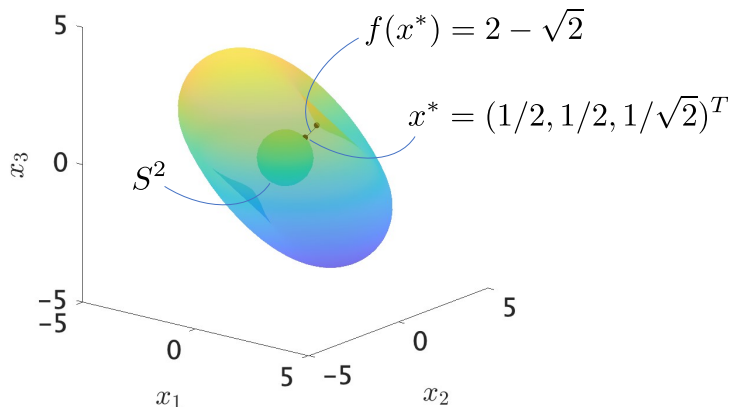
### 問題 2.8 ( $S^{n-1}$ 上の無制約最適化問題)

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

( $A = X X^T$  のとき, この問題は問題 2.1 と等価)

## 球面上の最適化問題の図示

例として,  $n = 3$ ,  $A = \begin{pmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 2 \end{pmatrix}$  の場合の図を示す.



## 参考：リーマン多様体 $S^{n-1}$ 上のニュートン法

詳細は割愛するが、この球面上の最適化問題 2.8 に対するニュートン法 ( $\mathbb{R}^n$  上のニュートン法をリーマン多様体  $S^{n-1}$  上に拡張したもの) では、レトラクション (後述) を  $R_x(\eta) := \frac{x + \eta}{\|x + \eta\|_2}$  とすると、

$$x_{k+1} = \frac{(A - f(x_k)I_n)^{-1}x_k}{\|(A - f(x_k)I_n)^{-1}x_k\|_2}$$

によって  $S^{n-1}$  上の点列  $\{x_k\}$  が生成される ( $f(x_k) = x_k^T A x_k$ ) .  
これは、数値線形代数でよく知られている**レイリー商反復法**と等価である.

## 補足

ユークリッド空間上の（等式・不等式）制約付き最適化問題と、  
リーマン多様体上の無制約最適化問題の関係について：

- ユークリッド空間上の制約付き最適化問題としても  
リーマン多様体上の無制約最適化問題としても定式化できる問題がある。  
例： $x \in \mathbb{R}^n, x^T x = 1 \iff x \in S^{n-1}$ （リーマン多様体）
- ユークリッド空間上の制約付き最適化問題だが、  
リーマン多様体上の無制約最適化問題として定式化できない問題もある。  
例： $x = (x_1, x_2)^T \in \mathbb{R}^2, x_1 x_2 = 0 \iff x \in \text{span}\{e_1, e_2\}$ （多様体でない）
- リーマン多様体上の無制約最適化問題だが、そのままでは  
ユークリッド空間上の制約付き最適化問題として定式化されない問題もある。  
例： $X \in \text{Sym}(n)^+ \iff X \in \mathbb{R}^{n \times n}, X^T = X, X \succ 0$ （（不）等式制約でない）  
例： $W \in \text{Grass}(p, n) \iff W$  は  $\mathbb{R}^n$  の  $p$  次元部分空間  
（ただし  $\text{Grass}(p, n) \simeq \{X \in \text{Sym}(n) \mid X^2 = X, \text{tr}(X) = p\}$  という表現は可能）

# 講演内容

- 1 はじめに
- 2 応用例
- 3 球面上の最急降下法の考え方**
- 4 位相空間とリーマン多様体
- 5 リーマン多様体上の一般の最適化手法と最急降下法
- 6 リーマン多様体上の共役勾配法 (R-CG)
- 7 まとめ



以下では、球面  $S^{n-1}$  上のレイリー商  $f(x) = x^T A x$  ( $A \in \text{Sym}(n)$ ) の最小化問題 (問題 2.8) を、一般論の説明のための最適化問題の例としてしばしば取り上げる.

## ユークリッド空間上の最急降下法の復習

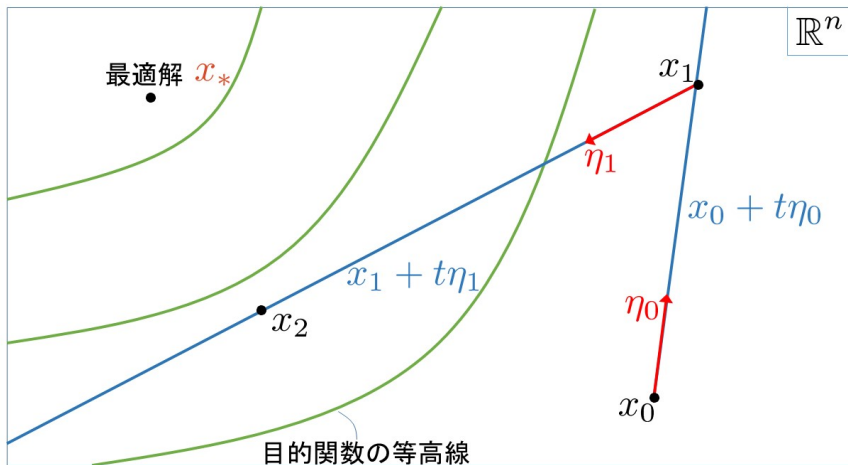
例として,  $A \in \text{Sym}(n)^+$  に対して以下の  $\mathbb{R}^n$  上の無制約最適化問題を考える.

### 問題 3.1 ( $\mathbb{R}^n$ 上の 2 次形式最小化問題)

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

この問題に対する最急降下法は以下の手順で  $\mathbb{R}^n$  の点列  $\{x_k\}$  を生成する:

- 初期点  $x_0 \in \mathbb{R}^n$  を与える.  $k = 0, 1, 2, \dots$  について以下を反復する.
- $x_k \in \mathbb{R}^n$  において, 探索方向を  $\eta_k := -\nabla f(x_k) = -2Ax_k \in \mathbb{R}^n$  とする.
- $\varphi_k(t) := f(x_k + t\eta_k)$  を  $t > 0$  の範囲で (近似的に) 最小にする  $t$  をステップ幅  $t_k$  とする (直線  $l_k(t) := x_k + t\eta_k$  上の探索).
- $x_{k+1} := x_k + t_k\eta_k$  とする.

$\mathbb{R}^n$  における最適化

# 球面上の最急降下法 (?)

例として,  $A \in \text{Sym}(n)$  に対して以下の球面  $S^{n-1}$  上の無制約最適化問題を考える.

## 問題 3.2 ( $S^{n-1}$ 上の 2 次形式 (レイリー商) 最小化問題)

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

この問題に対する最急降下法は以下の手順で  $S^{n-1}$  の点列  $\{x_k\}$  を生成する (?)

- 初期点  $x_0 \in S^{n-1}$  を与える.  $k = 0, 1, 2, \dots$  について以下を反復する.
- $x_k \in S^{n-1}$  において, 探索方向を  $\eta_k := -\nabla f(x_k) = -2Ax_k \in \mathbb{R}^n$  (?) とする.
- $\varphi_k(t) := f(x_k + t\eta_k)$  (?) を  $t > 0$  の範囲で (近似的に) 最小にする  $t$  をステップ幅  $t_k$  とする (直線  $l_k(t) := x_k + t\eta_k$  (?) 上の探索).
- $x_{k+1} := x_k + t_k\eta_k$  (?) とする.

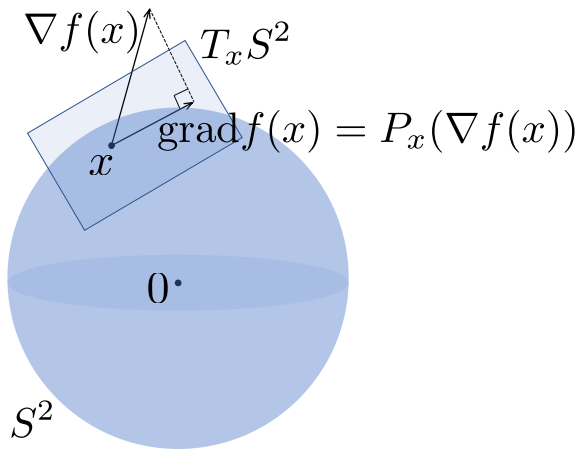
## 何が問題か？

- $x_k \in S^{n-1}$  において, 探索方向を  $\eta_k := -\nabla f(x_k) = -2Ax_k \in \mathbb{R}^n$  (?) とする.  
→  $-\nabla f(x_k)$  には球面  $S^{n-1}$  の情報が含まれていないので,  
一般には  $S^{n-1}$  と無関係な方向を向き得る.
- $\varphi_k(t) := f(x_k + t\eta_k)$  (?) を  $t > 0$  の範囲で (近似的に) 最小にする  $t$  を  
ステップ幅  $t_k$  とする (直線  $l_k(t) := x_k + t\eta_k$  (?) 上の探索).  
→ 球面  $S^{n-1}$  は曲がっているので直線  $l_k(t) = x_k + t\eta_k$  は  $S^{n-1}$  上に乗らない.
- $x_{k+1} := x_k + t_k\eta_k$  (?) とする.  
→ 特に,  $x_k + t_k\eta_k \notin S^{n-1}$  より  $S^{n-1}$  上の点列を生成できていない.

## どう解決するか？

- 球面  $S^{n-1}$  は曲がっているので直線  $l_k(t) = x_k + t\eta_k$  は  $S^{n-1}$  上に乗らない.  
 → 代わりに  $S^{n-1}$  上の曲線  $\gamma_k$  で  $\gamma_k(0) = x_k, \dot{\gamma}_k(0) = \eta_k$  となるものを考える.  
 そのためには,  $\eta_k$  は  $x_k$  における  $S^{n-1}$  の接ベクトルである必要がある.  
 $S^{n-1}$  上の点  $x$  における接空間を  $T_x S^{n-1}$  と書くと,  $\eta_k \in T_{x_k} S^{n-1}$  が必要.  
 このとき, たとえば  $\gamma_k(t) := \frac{x_k + t\eta_k}{\|x_k + t\eta_k\|_2}$  は  $\gamma_k(0) = x_k, \dot{\gamma}_k(0) = \eta_k$  を満たす.
- $-\nabla f(x_k)$  には球面  $S^{n-1}$  の情報が含まれていないので,  
 一般には  $S^{n-1}$  と無関係な方向を向き得る.  
 →  $-\nabla f(x_k)$  を  $T_{x_k} S^{n-1}$  に直交射影すれば良い. これを  $-\text{grad } f(x_k)$  と書く.  
 つまり,  $T_x S^{n-1}$  への直交射影を  $P_x: \mathbb{R}^n \rightarrow T_x S^{n-1}$  とすると,  
 $\text{grad } f(x) = P_x(\nabla f(x)).$   
 一般に,  $T_x S^{n-1} = \{\xi \in \mathbb{R}^n \mid x^T \xi = 0\}$  より,  $P_x(d) = (I_n - xx^T)d, d \in \mathbb{R}^n.$   
 → 探索方向は  $\eta_k = -\text{grad } f(x_k) = P_x(-\nabla f(x_k)) = -(I_n - xx^T)\nabla f(x).$

# 球面上の接空間での勾配 $\text{grad } f(x)$ のイメージ



# 球面上の最急降下法 (!)

## 問題 3.3 (再掲 ( $A \in \text{Sym}(n)$ ))

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

この問題に対する最急降下法は以下の手順で  $S^{n-1}$  の点列  $\{x_k\}$  を生成する：

- 初期点  $x_0 \in S^{n-1}$  を与える.  $k = 0, 1, 2, \dots$  について以下を反復する.
- $x_k \in S^{n-1}$  において, 探索方向を  
 $\eta_k := -\text{grad } f(x_k) = -2(I_n - x_k x_k^T) A x_k \in T_{x_k} S^{n-1}$  とする.
- $\varphi_k(t) := f(\gamma_k(t))$  を  $t > 0$  の範囲で (近似的に) 最小にする  $t$  を  
 ステップ幅  $t_k$  とする (曲線  $\gamma_k(t) := \frac{x_k + t\eta_k}{\|x_k + t\eta_k\|_2}$  上の探索).
- $x_{k+1} := \gamma_k(t_k) = \frac{x_k + t_k \eta_k}{\|x_k + t_k \eta_k\|_2}$  とする.



# 講演内容

- 1 はじめに
- 2 応用例
- 3 球面上の最急降下法の考え方
- 4 位相空間とリーマン多様体**
- 5 リーマン多様体上の一般の最適化手法と最急降下法
- 6 リーマン多様体上の共役勾配法 (R-CG)
- 7 まとめ

## 最適化のためには何が必要か？

問題 4.1 (ある集合  $M$  上の最適化問題)

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

- 目標：局所的最適解  $x_* \in M$  を求めること。では  $x_*$  が局所的最適解であるとは？  
 $x_*$  のある近傍  $U \subset M$  が存在して、任意の  $x \in U$  について  $f(x_*) \leq f(x)$ .  
 ここで、「 $x_*$  のある近傍  $U \subset M$ 」は  $x_*$  を含む  $M$  の開集合  
 → 集合  $M$  における開集合の概念が必要  
 →  $M$  は位相空間であることが望ましい.
- 手法： $f$  が滑らかなら、 $f$  の勾配などの微分の情報を用いる。  
 →  $M$  上の関数  $f$  の「滑らかさ」や「微分」の定義が必要  
 →  $M$  は可微分多様体であることが望ましい。  
 さらに勾配などを用いるには  $M$  はリーマン多様体である必要がある.

# 位相空間

## 定義 4.2 (位相空間)

$X$  を空でない集合とする.  $X$  の部分集合族  $\mathcal{O}$  が  $X$  の位相であるとは,  $\mathcal{O}$  が以下の条件を満たすことをいう.

1.  $X \in \mathcal{O}$ ,  $\emptyset \in \mathcal{O}$ ,
2.  $O_1, O_2, \dots, O_k \in \mathcal{O}$  ならば,  $\bigcap_{i=1}^k O_i \in \mathcal{O}$ ,
3.  $O_\lambda \in \mathcal{O}$ ,  $\lambda \in \Lambda$  ならば,  $\bigcup_{\lambda \in \Lambda} O_\lambda \in \mathcal{O}$ .

このとき,  $(X, \mathcal{O})$  を位相空間,  $O \in \mathcal{O}$  を  $(X, \mathcal{O})$  の開集合という.

## 位相空間の例

### 例 4.3 (ユークリッド位相)

数ベクトル空間  $\mathbb{R}^n$  の 2-ノルム  $\|\cdot\|_2$  を用いて,

$$\mathcal{O} := \{U \subset \mathbb{R}^n \mid \forall p \in U, \exists \varepsilon > 0, \text{ s.t., } \{x \in \mathbb{R}^n \mid \|x - p\|_2 < \varepsilon\} \subset U\}$$

と定義すると,  $\mathcal{O}$  は  $\mathbb{R}^n$  の位相である.

### 例 4.4 (密着位相)

集合  $X \neq \emptyset$  に対して  $\mathcal{O}_i := \{X, \emptyset\}$  は  $X$  の位相である.

### 例 4.5 (離散位相)

集合  $X \neq \emptyset$  に対して  $\mathcal{O}_d := 2^X$  ( $X$  のべき集合) は  $X$  の位相である.

密着位相や離散位相は極端な位相であり議論の対象から外したい.

## 位相空間における点列の収束

### 定義 4.6 (点列の収束)

位相空間  $(X, \mathcal{O})$  における点列  $\{x_k\}$  が点  $x \in X$  に**収束**するとは、 $x$  の任意の開近傍  $O \in \mathcal{O}$  に対して、ある自然数  $N$  が存在して、 $k \geq N$  ならば  $x_k \in O$  となることをいう。

一般の位相空間における点列の収束先は必ずしも一意的でない。

### 例 4.7 (密着空間における点列の収束)

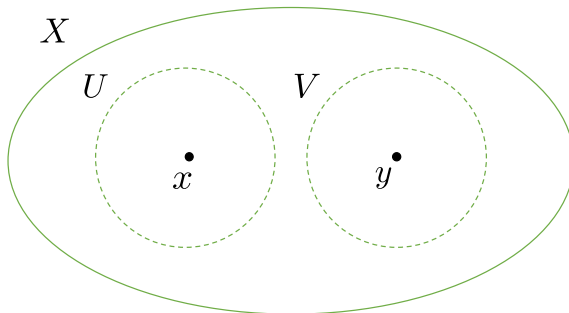
集合  $\mathbb{R}$  に密着位相  $\mathcal{O}_i = \{\mathbb{R}, \emptyset\}$  を与え、任意の点列  $\{x_k\}$  を考える。このとき、任意の実数  $x \in \mathbb{R}$  について、密着位相  $\mathcal{O}_i$  の下での  $x$  の開近傍は  $\mathbb{R}$  のみである。しかも、任意の自然数  $k$  について  $x_k \in \mathbb{R}$  であるから、 $\{x_k\}$  は  $x$  に収束する。すなわち、任意の実数列が任意の実数に収束する。

開集合が少なすぎる！

# ハウスドルフ空間

## 定義 4.8 (ハウスドルフ空間)

位相空間  $(X, \mathcal{O})$  において, 任意の相異なる 2 点  $x, y \in X$  に対し, 開集合  $U, V \in \mathcal{O}$  が存在して,  $x \in U, y \in V$ , および  $U \cap V = \emptyset$  が成り立つとき,  $\mathcal{O}$  はハウスドルフの分離公理を満たすといい,  $(X, \mathcal{O})$  を**ハウスドルフ空間**という.



## ハウスドルフ空間における点列の収束

### 例 4.9

集合  $X$  が 2 つ以上の要素をもつとき、密着空間  $(X, \mathcal{O}_i)$  はハウスドルフ空間でない。

### 例 4.10

通常の位相  $\mathcal{O}$  をもつユークリッド空間  $(\mathbb{R}^n, \mathcal{O})$  はハウスドルフ空間である。

### 命題 4.11

ハウスドルフ空間  $(X, \mathcal{O})$  における点列が収束するとき、その極限は一意である。

証明.  $(X, \mathcal{O})$  の点列  $\{x_k\}$  が相異なる  $x, y \in X$  に収束すると仮定すると、開集合  $U, V \in \mathcal{O}$  が存在して、 $x \in U, y \in V, U \cap V = \emptyset$  が成り立つ。

一方、 $\{x_k\}$  が  $x, y$  に収束することから、ある自然数  $N$  が存在して、 $k \geq N$  ならば  $x_k \in U, x_k \in V$  が成り立つが、これは  $U \cap V = \emptyset$  に矛盾する。 □

## 第2可算空間

離散空間のように開集合が多すぎるものも議論の対象外としたい。

### 定義 4.12 (第2可算空間)

位相空間  $(X, \mathcal{O})$  において,  $\mathcal{B} \subset \mathcal{O}$  が開基であるとは,  $X$  の任意の開集合  $O \in \mathcal{O}$  に対して  $\mathcal{B}_0 \subset \mathcal{B}$  が存在し,  $O = \bigcup \{W \mid W \in \mathcal{B}_0\}$  と書けることをいう.  
さらに, 位相  $\mathcal{O}$  に対して,  $\mathcal{O}$  の有限個または可算無限個の要素からなる開基  $\mathcal{B}$  が存在するとき,  $\mathcal{O}$  は第2可算公理を満たすといい,  $(X, \mathcal{O})$  を第2可算空間という.

### 例 4.13

集合  $X$  が非可算無限個の要素をもつとき, 離散空間  $(X, \mathcal{O}_d)$  は第2可算空間でない.

### 例 4.14

ユークリッド空間  $(\mathbb{R}^n, \mathcal{O})$  は第2可算空間である.



## 写像の連続性・同相写像

### 定義 4.15 (連続写像・同相写像)

位相空間  $(X_1, \mathcal{O}_1)$ ,  $(X_2, \mathcal{O}_2)$  に対して, 写像  $f: X_1 \rightarrow X_2$  が**連続**であるとは,  $X_2$  の任意の開集合  $U$  の  $f$  による逆像  $f^{-1}(U) := \{x \in X_1 \mid f(x) \in U\}$  が  $X_1$  の開集合であることをいう.

$f: X_1 \rightarrow X_2$  が全単射で,  $f$  も  $f^{-1}$  も連続であるとき,  $f$  を**同相写像**という. このとき,  $X_1$  と  $X_2$  は**同相**であるという.

この連続写像の定義はやや抽象的であるが, 通常之位相を備えたユークリッド空間の間の写像の場合には, 微分積分学においてよく知られる連続性の定義と一致する.

## 位相多様体

位相空間にさらなる構造が入る場合を考え、多様体の概念を導入する。

### 定義 4.16 (位相多様体)

次の条件を満たす位相空間  $M$  を  $n$  次元**位相多様体**という。

1.  $M$  はハウスドルフ空間であり、かつ、第 2 可算空間である。
2.  $M$  上の任意の点  $p$  に対して、 $p$  の開近傍  $U$  と  $\mathbb{R}^n$  の開集合  $V$  および同相写像  $\varphi: U \rightarrow V$  が存在する。 $(\varphi$  は全単射で、 $\varphi$  も  $\varphi^{-1}$  も連続)

定義 4.16 において、 $(U, \varphi)$  を**座標近傍**、 $\varphi$  を**局所座標系**という。

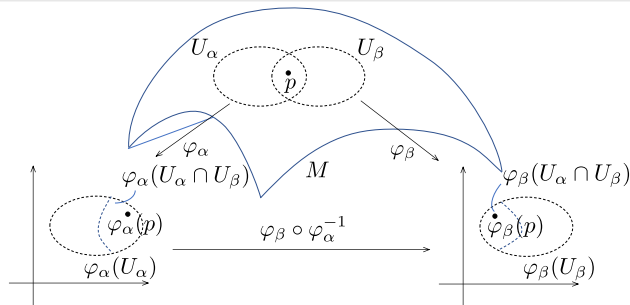
すなわち、 $p \in M$  は局所座標系  $\varphi$  により  $\mathbb{R}^n$  の点として  $\varphi(p) = (x_1, x_2, \dots, x_n)^T$  と表現でき、 $\varphi(p)$  を  $p$  の**局所座標**という。

また、 $M$  のすべての座標近傍からなる族を**座標近傍系**という。

# 可微分多様体

## 定義 4.17 (可微分多様体)

$r$  を正の整数または  $\infty$  とする. 位相多様体  $M$  の座標近傍系  $\{(U_\lambda, \varphi_\lambda)\}_{\lambda \in \Lambda}$  について,  $U_\alpha \cap U_\beta \neq \emptyset$  であるような任意の  $\alpha, \beta \in \Lambda$  に対して, 座標変換  $\varphi_\beta \circ \varphi_\alpha^{-1}: \varphi_\alpha(U_\alpha \cap U_\beta) \rightarrow \varphi_\beta(U_\alpha \cap U_\beta)$  が  $C^r$  級するとき,  $M$  を  $C^r$  級可微分多様体という.



## 可微分多様体 $M$ 上の関数の微分可能性

### 定義 4.18 (関数の微分可能性)

$M$  を  $C^r$  級可微分多様体 ( $r \geq 0$ ) とし,  $0 \leq s \leq r$  とする.

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, 任意の点  $p \in M$  に対し, ある座標近傍  $(U, \varphi)$  が存在して,  $f \circ \varphi^{-1}$  が  $\varphi(p)$  において  $C^s$  級であることをいう.

$f \circ \varphi^{-1}: \varphi(U) \rightarrow \mathbb{R}$  は  $\mathbb{R}^n$  ( $n = \dim M$ ) の開集合上の実数値関数であるから, その微分可能性については微分積分学で定義されている.

定義 4.18 が座標近傍  $(U, \varphi)$  の選び方によらないことは, 可微分多様体では座標変換が  $C^r$  級であることから保証される.

## 可微分多様体 $M$ から $N$ への写像の微分可能性

多様体  $M, N$  の間の一般の写像  $f: M \rightarrow N$  の微分可能性も同様に定義される。

### 定義 4.19 (写像の微分可能性)

$C^r$  級可微分多様体  $M, N$  はそれぞれ  $m$  次元および  $n$  次元であるとする。

写像  $f: M \rightarrow N$  が連続であるとき,  $f$  は  $C^0$  級であるという。

また,  $1 \leq s \leq r$  を満たす  $s$  に対して, 連続写像  $f: M \rightarrow N$  が  $p \in M$  において  $C^s$  級であるとは,  $p \in M, f(p) \in N$  それぞれの  $C^r$  級座標近傍  $(U, \varphi), (V, \psi)$  が存在して,  $f(U) \subset V$  かつ  $\psi \circ f \circ \varphi^{-1}: \varphi(U) \rightarrow \mathbb{R}^n$  が  $C^s$  級であることをいう。

写像  $f$  が  $M$  上のすべての点で  $C^s$  級であるとき  $f$  は  $C^s$  級であるという。

これ以降,  $C^\infty$  級可微分多様体を単に多様体と呼ぶことにする。

## 接ベクトル ( $\mathbb{R}^n$ の場合)

滑らかな曲線  $\gamma: I \rightarrow \mathbb{R}^n$  が点  $\gamma(0) =: p \in \mathbb{R}^n$  を通るとする ( $I$  は  $0$  を含む  $\mathbb{R}$  の区間). このとき, 基本ベクトル  $e_i \in \mathbb{R}^n$  (第  $i$  成分のみ  $1$ , 他の成分は  $0$ ) を用いて,  $\gamma$  は

$$\gamma(t) = (x_1(t), x_2(t), \dots, x_n(t))^T = \sum_{i=1}^n x_i(t) e_i$$

の形で表される. ここで,  $x_i(0) = p_i$ . この曲線に対して,

$$\gamma'(0) := \lim_{t \rightarrow 0} \frac{\gamma(t) - \gamma(0)}{t} = \left. \frac{d}{dt} \gamma(t) \right|_{t=0} = \sum_{i=1}^n \frac{dx_i}{dt}(0) e_i$$

は  $\gamma$  の  $t = 0$  における接ベクトルである.

この極限が定義されるのは  $\gamma(t) - \gamma(0) \in \mathbb{R}^n$  であるからということに注意する.

(多様体  $M$  上の曲線  $\gamma$  については一般には  $\gamma(t) - \gamma(0)$  が定義できない!)

## 接ベクトル (多様体 $M$ の場合)

- 滑らかな曲線  $\gamma: I \rightarrow M$  が点  $\gamma(0) =: p \in M$  を通るとする.  
点  $p$  の周りの座標近傍  $(U, \varphi)$  を用いれば,  $\varphi(U) \subset \mathbb{R}^n$  内で  $\gamma$  を

$$\varphi(\gamma(t)) = (x_1(t), x_2(t), \dots, x_n(t))^T$$

と表せるが,  $x(t) := \varphi(\gamma(t)) \in \mathbb{R}^n$  に対する前ページのような  $x'(0)$  (×) は局所座標系の選び方に依存してしまい ( $x$  自身が  $\varphi$  に依存するため),  $M$  上の曲線  $\gamma$  の  $t=0$  での接ベクトルとはみなせない.

- もちろん,  $\lim_{t \rightarrow 0} \frac{\gamma(t) - \gamma(0)}{t}$  (×) という量を考えることもできない.  
( $\gamma(t) - \gamma(0)$  が  $M$  上では定義できないため)

## 接ベクトル (多様体 $M$ の場合)

解決策:

$M$  上の任意の実数値関数  $f$  に対して  $f \circ \gamma: I \rightarrow \mathbb{R}$  の  $t = 0$  での微分を考える。  
これは微分積分学の範囲で微分が定義でき、局所座標系の選び方にも依存しない。  
一方、この微分は座標近傍  $(U, \varphi) = (U; x_1, x_2, \dots, x_n)$  を用いて書くこともできる:

$$\left. \frac{d}{dt}(f \circ \gamma)(t) \right|_{t=0} = \left. \frac{d}{dt}(f \circ \varphi^{-1})(\varphi(\gamma(t))) \right|_{t=0} = \sum_{i=1}^n \frac{dx_i}{dt}(0) \left. \frac{\partial}{\partial x_i} \right|_p f.$$

ここで、 $\left. \frac{\partial}{\partial x_i} \right|_p : \mathfrak{F}_p(M) \rightarrow \mathbb{R}$  を  $\left. \frac{\partial}{\partial x_i} \right|_p f := \frac{\partial(f \circ \varphi^{-1})}{\partial x_i}(\varphi(p))$  と定義した。

( $\mathfrak{F}_p(M)$  は  $M$  上の点  $p$  の近傍で定義される滑らかな実数値関数全体の集合)

$\sum_{i=1}^n \frac{dx_i}{dt}(0) \left. \frac{\partial}{\partial x_i} \right|_p$  は (局所座標系を用いて書けるが) 局所座標系の選び方によらない。

→ この微分作用素を  $\dot{\gamma}(0)$  と書き、曲線  $\gamma$  の  $t = 0$  での接ベクトルと見なそう!



## 接ベクトルと接空間

あらゆる曲線の接ベクトル（微分作用素）を考え、その全体を接空間と定義する。

### 定義 4.20 (接空間)

多様体  $M$  上の点  $p$  の近傍で定義された滑らかな実数値関数の全体を  $\mathfrak{F}_p(M)$  とする。写像  $\xi: \mathfrak{F}_p(M) \rightarrow \mathbb{R}$  に対して、 $M$  上の曲線  $\gamma$  が存在して、 $\gamma(0) = p$  かつ

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

となるとき、 $\xi$  を点  $p$  における  $M$  の接ベクトルといい  $\xi = \dot{\gamma}(0)$  と書く。点  $p$  における接ベクトル全体の集合を接空間といい  $T_p M$  と書く。

$M$  が  $n$  次元多様体のとき、接空間  $T_p M$  は  $n$  次元ベクトル空間であることが示せる。

## 多様体 $\mathbb{R}^n$ の接空間

$\mathbb{R}^n$  は、その座標系をただ一つの局所座標系として多様体と考えられる。このとき、  
 曲線  $\gamma(t) = (x_1(t), x_2(t), \dots, x_n(t))^T = \sum_{i=1}^n x_i(t)e_i$  の  $t = 0$  での接ベクトル  $\dot{\gamma}(0)$  は

$$\dot{\gamma}(0)f = \left. \frac{d}{dt}(f \circ \gamma)(t) \right|_{t=0} = Df(\gamma(0))[\gamma'(0)], \quad f \in \mathfrak{F}_{\gamma(0)}(\mathbb{R}^n) \quad (\text{最右辺は方向微分})$$

を満たす。ここで、ある曲線  $\gamma$  についての  $v = \gamma'(0) \in \mathbb{R}^n$  に対して  $\xi = \dot{\gamma}(0)$  を返す写像  $v \mapsto \xi$  は  $\gamma$  の選び方によらず定義され、 $\mathbb{R}^n$  と  $T_{\gamma(0)}\mathbb{R}^n$  の間の同型写像となる。そこで、 $\gamma'(0) \in \mathbb{R}^n$  と  $\dot{\gamma}(0) \in T_{\gamma(0)}\mathbb{R}^n$  を同一視する。

この同一視を任意の点  $p$  を通る任意の曲線について行い、 $T_p\mathbb{R}^n \simeq \mathbb{R}^n$  を得る。

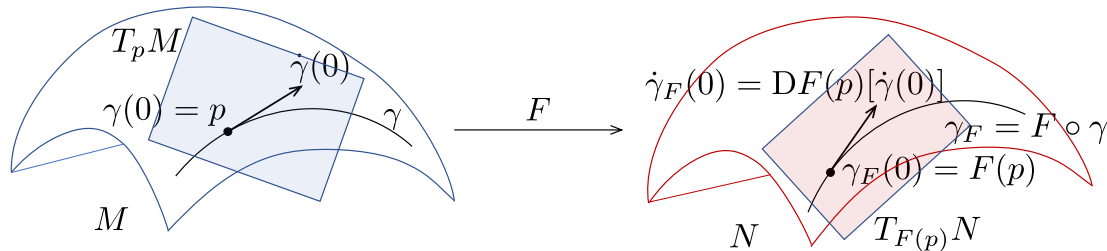
$$\gamma'(0) = \sum_{i=1}^n \frac{dx_i}{dt}(0)e_i, \quad \dot{\gamma}(0) = \sum_{i=1}^n \frac{dx_i}{dt}(0) \left. \frac{\partial}{\partial x_i} \right|_p \quad \text{より, } e_i \text{ と } \left. \frac{\partial}{\partial x_i} \right|_p \text{ を同一視している.}$$

## 多様体間の写像の微分

### 定義 4.21 (写像の微分)

多様体  $M$  から  $N$  への滑らかな写像  $F: M \rightarrow N$  の点  $p \in M$  における微分  $DF(p): T_p M \rightarrow T_{F(p)} N$  とは,  $DF(p)[\dot{\gamma}(0)] = \dot{\gamma}_F(0)$  を満たすものをいう. ここで,  $\gamma$  は  $\gamma(0) = p$  を満たす  $M$  上の曲線で,  $\gamma_F := F \circ \gamma$  は  $N$  上の曲線である.

上の定義で,  $DF(p)[\dot{\gamma}(0)]$  および  $\dot{\gamma}_F(0)$  は曲線  $\gamma$  の選び方によらない.



## 合成写像の微分

滑らかな写像  $F: M \rightarrow N$  の点  $p \in M$  における微分  $DF(p): T_pM \rightarrow T_{F(p)}N$  は、以下でも特徴付けられる（前ページの  $\dot{\gamma}_F(0)$  の  $g \in \mathfrak{F}_{F(p)}(N)$  への作用を調べよ）：

$$(DF(p)[\xi])(g) = \xi(g \circ F), \quad \xi \in T_pM, \quad g \in \mathfrak{F}_{F(p)}(N).$$

これを用いると、以下の重要な公式を容易に証明できる。

### 命題 4.22 (合成写像の微分)

多様体  $M, N, Q$  について、滑らかな写像  $F: N \rightarrow Q, G: M \rightarrow N$  の合成写像  $F \circ G$  の点  $p \in M$  における微分は、任意の  $\xi \in T_pM$  に対して次を満たす：

$$D(F \circ G)(p)[\xi] = DF(G(p))[DG(p)[\xi]].$$

**証明.** 任意の  $h \in \mathfrak{F}_{F(G(p))}(Q)$  に対して、

$$(D(F \circ G)(p)[\xi])(h) = \xi(h \circ F \circ G) = (DG(p)[\xi])(h \circ F) = (DF(G(p))[DG(p)[\xi]])(h). \quad \square$$

## 記号 $D$ について

$F: \mathbb{R}^m \rightarrow \mathbb{R}^n$  のとき,  $\gamma(0) = p$  なる  $\mathbb{R}^m$  上の曲線  $\gamma$  について  $\gamma_F := F \circ \gamma$  とする.

- $F$  を多様体間の写像と見なすと微分  $DF$  は

$$DF(p)[\dot{\gamma}(0)] = \dot{\gamma}_F(0).$$

- 一方,  $F$  の方向微分も同じ記号  $DF$  で書くと,

$$DF(p)[\gamma'(0)] := \left. \frac{d}{dt} F(p + t\gamma'(0)) \right|_{t=0} = \left. \frac{d}{dt} (F \circ \gamma)(t) \right|_{t=0} = \gamma'_F(0).$$

$\dot{\gamma}(0)$  と  $\gamma'(0)$ ,  $\dot{\gamma}_F(0)$  と  $\gamma'_F(0)$  を同一視するから, 2種類の微分  $DF$  も同一視できる.

## 部分多様体

- 球面  $S^{n-1}$  は  $\mathbb{R}^n$  の部分多様体
- シュテューフェル多様体  $\text{St}(p, n)$  は  $\mathbb{R}^{n \times p}$  の部分多様体

「単なる部分集合」ではない.

### 定義 4.23 (部分多様体)

$m$  次元多様体  $M$  の部分集合  $N$  が  $n$  次元**部分多様体**であるとは,  
任意の点  $p \in N$  について,  $p$  を含む  $M$  の局所座標系  $(U; x_1, x_2, \dots, x_m)$  が存在して,  
 $N \cap U = \{q \in U \mid x_{n+1}(q) = x_{n+2}(q) = \dots = x_m(q) = 0\}$  と書けるときをいう.

$N \subset M$  が多様体  $M$  の部分多様体であるための十分条件を与える定理があり,  
それを用いて,  $S^{n-1} \subset \mathbb{R}^n$  や  $\text{St}(p, n) \subset \mathbb{R}^{n \times p}$  が部分多様体であることを証明できる.  
(たとえば [Absil et al. 2008, Sato 2021, Tu 2011] を参照)

## 部分多様体の接空間

$N$  が多様体  $M$  の部分多様体ならば, 包含写像  $\iota: N \rightarrow M$  は任意の  $\xi \in T_x N$  について  $D\iota(x)[\xi] \in T_x M$  であり,  $\bar{f}: M \rightarrow \mathbb{R}$  とその  $N$  への制限  $f := \bar{f}|_N: N \rightarrow \mathbb{R}$  に対して

$$(D\iota(x)[\xi])\bar{f} = \xi(\bar{f} \circ \iota) = \xi f.$$

よって,  $\xi$  と  $D\iota(x)[\xi]$  を同一視でき,  $T_x N$  は  $T_x M$  の部分空間と見なせる.

### 例 4.24 (球面 $S^{n-1}$ の接空間)

$\mathbb{R}^n$  の部分多様体  $S^{n-1}$  の接空間  $T_x S^{n-1}$  は  $T_x \mathbb{R}^n \simeq \mathbb{R}^n$  の部分空間と同一視できる.  
具体的には,  $T_x S^{n-1} = \{\xi \in \mathbb{R}^n \mid x^T \xi = 0\}.$

### 例 4.25 (シュティーフェル多様体 $\text{St}(p, n)$ の接空間)

$\mathbb{R}^{n \times p}$  の部分多様体  $\text{St}(p, n)$  の接空間  $T_X \text{St}(p, n)$  は  $\mathbb{R}^{n \times p}$  の部分空間と同一視できる.  
具体的には,  $T_X \text{St}(p, n) = \{\xi \in \mathbb{R}^{n \times p} \mid X^T \xi + \xi^T X = 0\}.$

# リーマン多様体

## 定義 4.26 (リーマン多様体)

多様体  $M$  上の任意の点  $x \in M$  に対して,  $g_x$  を接空間  $T_x M$  における内積とする. さらに,  $g: x \mapsto g_x$  が滑らかであるとき,  $g$  を  $M$  上の**リーマン計量**といい,  $(M, g)$  を**リーマン多様体**という.

- 第2可算公理から, 任意の多様体  $M$  についてリーマン計量の存在が保証される.
- $\xi \in T_x M$  のノルムを  $\|\xi\| := \sqrt{g_x(\xi, \xi)}$  と定義すれば  $T_x M$  はノルム空間となる.
- $N$  がリーマン多様体  $(M, g^M)$  の部分多様体のとき,  $N$  のリーマン計量  $g^N$  を,

$$g_x^N(\xi, \eta) := g_x^M(\xi, \eta), \quad x \in N \subset M, \quad \xi, \eta \in T_x N \subset T_x M$$

と定義でき,  $g^N$  を**誘導計量**,  $(N, g^N)$  を  $(M, g^M)$  の**リーマン部分多様体**という.  
以下では  $T_x M$  の内積  $g_x(\cdot, \cdot)$  を  $\langle \cdot, \cdot \rangle_x$  と書くことにする.



# リーマン多様体上の関数の勾配

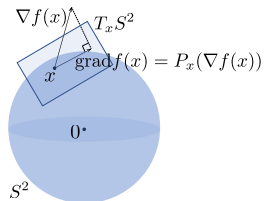
## 定義 4.27 (勾配)

リーマン多様体  $(M, g)$  上の滑らかな関数  $f$  の点  $x \in M$  における**勾配**  $\text{grad } f(x)$  とは、次を満たす、 $T_x M$  の一意的な元をいう：

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

cf.  $\mathbb{R}^n$  上の関数  $f$  の通常の勾配 (ユークリッド勾配)  $\nabla f$  は、

$Df(x)[\xi] = \langle \nabla f(x), \xi \rangle$ ,  $\xi \in \mathbb{R}^n$  を満たす. (左辺は方向微分,  $\langle \cdot, \cdot \rangle$  は標準内積)



$M = S^2 := \{x \in \mathbb{R}^3 \mid \|x\|_2 = 1\}$  の例.

$\mathbb{R}^3$  に埋め込まれた部分多様体  $S^2$  上の関数  $f$  の勾配は、 $f$  を外側の  $\mathbb{R}^3$  に拡張した関数  $\bar{f}$  のユークリッド勾配  $\nabla \bar{f}$  を、 $T_x S^2$  に直交射影することで得られる.

より一般の部分多様体についても同様 (次ページ).

# リーマン部分多様体上の関数の勾配

## 命題 4.28

リーマン多様体  $(\bar{M}, \bar{g})$  とそのリーマン部分多様体  $(M, g)$  に対し,  $P_x: T_x\bar{M} \rightarrow T_xM$  を  $T_xM$  への**直交射影**とする. すなわち, 任意の  $\chi \in T_x\bar{M}$  に対して,  $P_x(\chi) \in T_xM$  かつ,  $\chi - P_x(\chi) \in T_x\bar{M}$  は  $T_xM$  に直交する:  $\bar{g}_x(\chi - P_x(\chi), \xi) = 0, \xi \in T_xM$ . このとき, 滑らかな関数  $\bar{f}: \bar{M} \rightarrow \mathbb{R}$  の  $M$  への制限  $f: M \rightarrow \mathbb{R}$  の勾配は,

$$\text{grad } f(x) = P_x(\text{grad } \bar{f}(x)), \quad x \in M \subset \bar{M}$$

を満たす.

**証明.**  $\zeta_x := P_x(\text{grad } \bar{f}(x)) \in T_xM, \eta_x := \text{grad } \bar{f}(x) - \zeta_x \in N_xM$  とすると, 任意の  $\xi \in T_xM$  に対し  $g_x(\zeta_x, \xi) = \bar{g}_x(\zeta_x + \eta_x, \xi) = \bar{g}_x(\text{grad } \bar{f}(x), \xi) = D\bar{f}(x)[\xi] = Df(x)[\xi]$  より, 勾配の定義から  $\zeta_x = \text{grad } f(x)$  が成り立つ. □

球面  $S^{n-1}$  上の関数の勾配の例

## 例 4.29

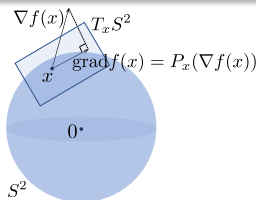
球面  $S^{n-1} = \{x \in \mathbb{R}^n \mid x^T x = 1\}$  に  $\mathbb{R}^n$  の標準計量からの誘導計量

$$g_x(\xi, \eta) := \xi^T \eta, \quad \xi, \eta \in T_x S^{n-1}$$

を入れる. このとき,  $\bar{f}: \mathbb{R}^n \rightarrow \mathbb{R}$  の (ユークリッド) 勾配を  $\nabla \bar{f}$  を用いて, 関数  $\bar{f}: \mathbb{R}^n \rightarrow \mathbb{R}$  の  $S^{n-1}$  への制限  $f: S^{n-1} \rightarrow \mathbb{R}$  の (リーマン) 勾配は

$$\text{grad } f(x) = P_x(\nabla \bar{f}(x)), \quad x \in S^{n-1}$$

となる. ここで,  $P_x: \mathbb{R}^n \rightarrow T_x S^{n-1}$  は接空間への直交射影で  $P_x = I_n - xx^T$ .



## 講演内容

- 1 はじめに
- 2 応用例
- 3 球面上の最急降下法の考え方
- 4 位相空間とリーマン多様体
- 5 リーマン多様体上の一般の最適化手法と最急降下法**
- 6 リーマン多様体上の共役勾配法 (R-CG)
- 7 まとめ

## 多様体上の最適化問題

多様体  $M$  上の目的関数  $f: M \rightarrow \mathbb{R}$  の無制約最適化問題を考える：

### 問題 5.1 (多様体 $M$ 上の無制約最適化問題)

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

$M$  上の点  $x$  における接空間を  $T_x M$  と書き、以下を仮定する：

- $M$  はリーマン計量  $\langle \cdot, \cdot \rangle$  を備えたリーマン多様体；
- $f$  は滑らかで下に有界である。特に，  
 $M$  上のレトラクション  $R$  に対して，定数  $L > 0$  が存在して

$$|\mathrm{D}(f \circ R_x)(t\eta)[\eta] - \mathrm{D}(f \circ R_x)(0)[\eta]| \leq Lt, \quad t \geq 0$$

が任意の  $\eta \in T_x M$  ( $\|\eta\|_x = 1$ ) と  $x \in M$  について成り立つ。  
このためには  $f \circ R_x$  がリプシッツ連続的微分可能ならば十分。

## 最適性の 1 次の必要条件

### 定理 5.2 (最適性の 1 次の必要条件)

リーマン多様体  $M$  上の  $C^1$  級関数  $f$  の無制約最小化問題 5.1 の局所的最適解を  $x_* \in M$  とすると,  $\text{grad } f(x_*) = 0$  が成り立つ.

証明. 任意の  $\xi \in T_{x_*}M$  に対して,  $M$  上の曲線  $\gamma$  で  $\gamma(0) = x_*$ ,  $\dot{\gamma}(0) = \xi$  なるものを選ぶ.  $x_*$  は局所的最適解だから,  $t = 0$  は  $g(t) := f(\gamma(t))$  の極小点で,  $g'(0) = 0$ .  
 $g'(0) = Df(\gamma(0))[\dot{\gamma}(0)] = \langle \text{grad } f(x_*), \xi \rangle_{x_*}$  で,  $\xi$  は任意だから,  $\text{grad } f(x_*) = 0$ .  $\square$

そこで,  $\mathbb{R}^n$  の場合と同様に,  $\text{grad } f(x_*) = 0$  となる  $x_* \in M$  (**臨界点**) を求めたい.

## $\mathbb{R}^n$ における最適化

$M = \mathbb{R}^n$  の場合の無制約最適化問題に対する解法：  
最急降下法, ニュートン法, (非線形) 共役勾配法など.

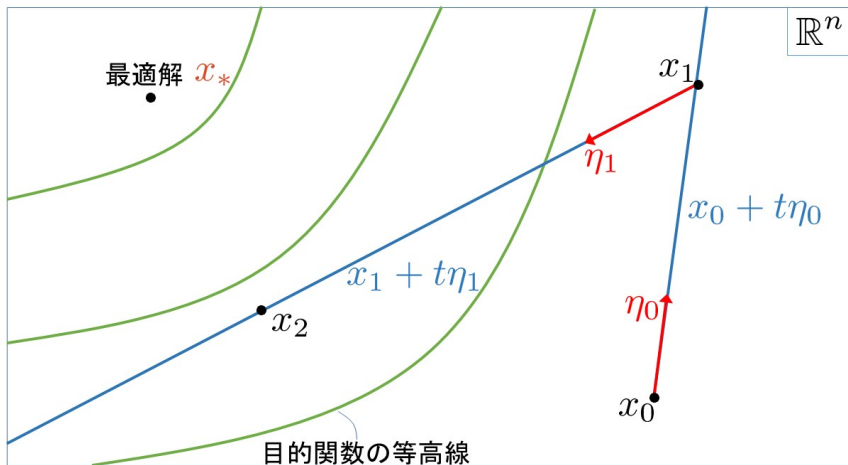
---

### アルゴリズム 5.1 $\mathbb{R}^n$ における最適化 (直線探索アルゴリズム)

---

- 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  $\eta_k \in \mathbb{R}^n$  とステップ幅  $t_k > 0$  を求める.
  - 4:   次の点  $x_{k+1}$  を  $x_{k+1} := x_k + t_k \eta_k$  によって計算する.
  - 5: **end for**
- 

探索方向  $\eta_k$  とステップ幅  $t_k$  の決め方がポイント.

$\mathbb{R}^n$  における最適化



## $\mathbb{R}^n$ における探索方向 $\eta_k$ の決め方

$\nabla f, \nabla^2 f$  をそれぞれ  $f$  の勾配, ヘッセ行列とする.

- **最急降下法** :  $\eta_k := -\nabla f(x_k)$ .
- **ニュートン法** :  $\eta_k$  を  $\eta \in \mathbb{R}^n$  についてのニュートン方程式

$$\nabla^2 f(x_k)[\eta] = -\nabla f(x_k)$$

の解として計算する.

- (非線形) **共役勾配法** : 
$$\begin{cases} \eta_0 := -\nabla f(x_0), \\ \eta_{k+1} := -\nabla f(x_{k+1}) + \beta_{k+1}\eta_k, \end{cases} \quad k \geq 0.$$
  
 $\beta_k \in \mathbb{R}$  の計算方法は多数提案されている.

## ステップ幅に関する条件

探索方向  $\eta_k$  は  $x_k$  における降下方向である ( $\Leftrightarrow \nabla f(x_k)^T \eta_k < 0$ ) と仮定する.

- アルミホ条件 :

$$f(x_k + t_k \eta_k) \leq f(x_k) + c_1 t_k \nabla f(x_k)^T \eta_k. \quad (1)$$

ここで,  $c_1$  は  $0 < c_1 < 1$  を満たす定数である.

- ウルフ条件 : (1) かつ

$$\nabla f(x_k + t_k \eta_k)^T \eta_k \geq c_2 \nabla f(x_k)^T \eta_k. \quad (2)$$

ここで,  $c_2$  は  $0 < c_1 < c_2 < 1$  を満たす定数である.

- 強ウルフ条件 : (1) かつ

$$|\nabla f(x_k + t_k \eta_k)^T \eta_k| \leq c_2 |\nabla f(x_k)^T \eta_k|. \quad (3)$$

## アルミホ条件やウルフ条件の解釈

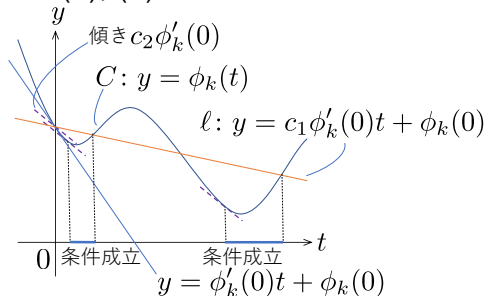
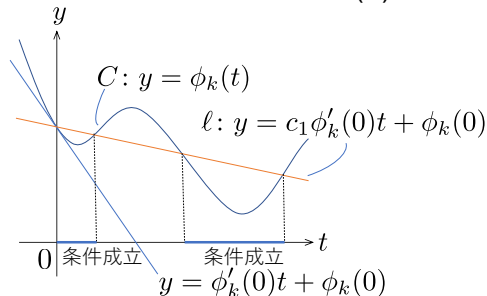
$\phi_k(t) := f(x_k + t\eta_k)$  とすると, (1), (2), (3) は, それぞれ

$$\phi_k(t_k) \leq c_1 \phi'_k(0)t_k + \phi_k(0), \quad (4)$$

$$\phi'_k(t_k) \geq c_2 \phi'_k(0), \quad (5)$$

$$|\phi'_k(t_k)| \leq c_2 |\phi'_k(0)| \quad (6)$$

と書ける. 【左：アルミホ条件 (4), 右：ウルフ条件 (4), (5)】



## リーマン多様体上の最適化への拡張

以下ではリーマン多様体  $M$  上の無制約最適化問題

### 問題 5.3 (リーマン多様体 $M$ 上の制約なし最適化問題)

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

点  $x_k \in M$  における探索方向を  $\eta_k \in T_{x_k}M$  とする.

- $M = \mathbb{R}^n$  の場合の直線探索:  $x_{k+1} = x_k + t_k \eta_k$ .
- 一般のリーマン多様体  $M$  の場合の探索:  $x_{k+1} = x_k + t_k \eta_k$ ?  
 $x_k \in M$  と  $t_k \eta_k \in T_{x_k}M$  は足し合わせることができない!

多様体  $M$  上では直線が定義できないので, その代わりに, 点  $x_k$  から  $\eta_k$  の方向に延びる曲線, すなわち  $\gamma_k(0) = x_k, \dot{\gamma}_k(0) = \eta_k$  なる曲線  $\gamma_k$  の上で探索を行う.

# レトラクション

多様体上の探索を行うための曲線を定義するための写像がレトラクションである.

## 定義 5.4 (レトラクション)

多様体  $M$  上のすべての接ベクトル全体の集合 (接束と呼ばれる) を  $TM$  と書く. 写像  $R: TM \rightarrow M$  が  $M$  上のレトラクションであるとは, 任意の  $x \in M$  について,  $R$  の  $T_x M$  への制限  $R_x: T_x M \rightarrow M$  が以下の条件を満たすときをいう.

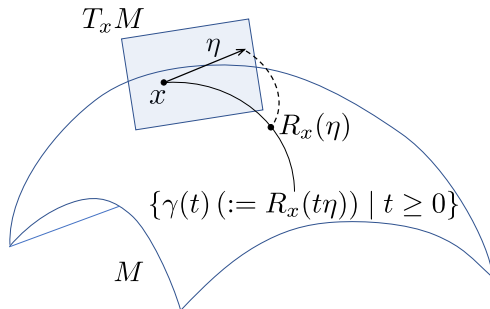
1.  $R_x(0_x) = x$ , ( $0_x$  は  $T_x M$  の零ベクトル)
2.  $DR_x(0_x) = \text{id}_{T_x M}$ .

## レトラクションと多様体上の曲線

多様体  $M$  上の点  $x \in M$  と接ベクトル  $\eta \in T_x M$  が与えられたとき、曲線  $\gamma$  をレトラクション  $R$  によって  $\gamma(t) = R_x(t\eta)$  と定義すると、

$$\gamma(0) = R_x(0) = x, \quad \dot{\gamma}(0) = DR_x(0)[\eta] = \eta$$

が成り立つ。この性質により、探索方向が与えられた際に、次の点の探索のために適切な曲線を得ることができる。



## リーマン多様体 $M$ 上の最適化

以降では、多様体  $M$  上のレトラクション  $R$  が与えられているとする。

---

### アルゴリズム 5.2 リーマン多様体 $M$ 上の曲線上の探索を用いる最適化アルゴリズム

---

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

探索方向  $\eta_k$  とステップ幅  $t_k$  の決め方がポイント.

## リーマン多様体上のアルミホ条件とウルフ条件

点  $x_k \in M$  における探索方向  $\eta_k \in T_{x_k} M$  に対して, レトラクション  $R$  によって定まる曲線  $\{R_{x_k}(t\eta_k) \in M \mid t \geq 0\}$  の上で  $x_{k+1}$  の探索を行う. そこで,

$$\phi_k(t) := f(R_{x_k}(t\eta_k))$$

と定義する. このとき, 式 (4)–(6) はそれぞれ

$$f(R_{x_k}(t_k\eta_k)) \leq f(x_k) + c_1 t_k \langle \text{grad } f(x_k), \eta_k \rangle_{x_k}, \quad (7)$$

$$\langle \text{grad } f(R_{x_k}(t_k\eta_k)), DR_{x_k}(t_k\eta_k)[\eta_k] \rangle_{x_{k+1}} \geq c_2 \langle \text{grad } f(x_k), \eta_k \rangle_{x_k}, \quad (8)$$

$$|\langle \text{grad } f(R_{x_k}(t_k\eta_k)), DR_{x_k}(t_k\eta_k)[\eta_k] \rangle_{x_{k+1}}| \leq c_2 |\langle \text{grad } f(x_k), \eta_k \rangle_{x_k}| \quad (9)$$

となる. リーマン多様体  $M$  上の**アルミホ条件**を (7),  
**ウルフ条件**を (7), (8), **強ウルフ条件**を (7), (9) と定義する.



## リーマン多様体上の最急降下法

リーマン多様体  $M$  上の最急降下法は、ユークリッド空間の場合と同様に、点  $x_k \in M$  における探索方向を  $\eta_k = -\text{grad } f(x_k) \in T_{x_k}M$  とする。

---

### アルゴリズム 5.3 リーマン多様体 $M$ 上の最急降下法

---

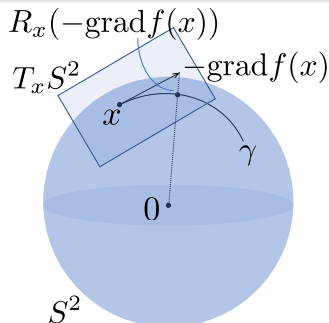
- 1: 初期点  $x_0 \in M$  を選ぶ.
  - 2: **for**  $k = 0, 1, 2, \dots$  **do**
  - 3:    $\eta_k = -\text{grad } f(x_k)$ .
  - 4:   ステップ幅  $t_k$  を計算する.
  - 5:    $x_{k+1} = R_{x_k}(t_k \eta_k)$ .
  - 6: **end for**
-

# 球面上のレトラクションの例と最急降下法のイメージ

## 例 5.5

$S^{n-1}$  において, 以下の  $R$  はレトラクションの一例となっている.

$$R_x(\xi) = \frac{x + \xi}{\|x + \xi\|_2}, \quad x \in S^{n-1}, \xi \in T_x S^{n-1}$$



## Zoutendijk の定理のリーマン多様体への拡張

$T_{x_k}M$  において, 最急降下方向  $-\text{grad } f(x_k)$  と探索方向  $\eta_k$  のなす角を  $\theta_k$  とする:

$$\cos \theta_k = \frac{\langle -\text{grad } f(x_k), \eta_k \rangle_{x_k}}{\|-\text{grad } f(x_k)\|_{x_k} \|\eta_k\|_{x_k}} = -\frac{\langle \text{grad } f(x_k), \eta_k \rangle_{x_k}}{\|\text{grad } f(x_k)\|_{x_k} \|\eta_k\|_{x_k}}.$$

### 定理 5.6

リーマン多様体  $M$  上で  $x_{k+1} = R_{x_k}(t_k \eta_k)$  によって  $\{x_k\}$  を生成し,  
探索方向  $\eta_k \in T_{x_k}M$  は  $\langle \text{grad } f(x_k), \eta_k \rangle_{x_k} < 0$  を満たし,  
ステップ幅  $t_k$  はウルフ条件を満たすとする. さらに, 目的関数  $f$  は下に有界かつ  
 $C^1$  級であり, 定数  $L > 0$  が存在して,  $\|\eta\|_x = 1$  を満たすすべての  $\eta \in T_x M$  について

$$|D(f \circ R_x)(t\eta)[\eta] - D(f \circ R_x)(0)[\eta]| \leq Lt, \quad x \in M, t \geq 0 \quad (10)$$

が成り立つとする. このとき, 次が成り立つ:

$$\sum_{k=0}^{\infty} \cos^2 \theta_k \|\text{grad } f(x_k)\|_{x_k}^2 < \infty.$$

## 証明 (1/2)

## ウルフ条件より

$\langle \text{grad } f(x_{k+1}), DR_{x_k}(t_k \eta_k)[\eta_k] \rangle_{x_{k+1}} - \langle \text{grad } f(x_k), \eta_k \rangle_{x_k} \geq (c_2 - 1) \langle \text{grad } f(x_k), \eta_k \rangle_{x_k}$   
 が成り立つ。一方、仮定 (10) を用いて左辺の量を評価すると、

$$\begin{aligned} & \langle \text{grad } f(x_{k+1}), DR_{x_k}(t_k \eta_k)[\eta_k] \rangle_{x_{k+1}} - \langle \text{grad } f(x_k), \eta_k \rangle_{x_k} \\ &= D(f \circ R_{x_k})(t_k \eta_k)[\eta_k] - D(f \circ R_{x_k})(0)[\eta_k] \\ &= \|\eta_k\|_{x_k} \left| D(f \circ R_{x_k}) \left( t_k \|\eta_k\|_{x_k} \frac{\eta_k}{\|\eta_k\|_{x_k}} \right) \left[ \frac{\eta_k}{\|\eta_k\|_{x_k}} \right] - D(f \circ R_{x_k})(0) \left[ \frac{\eta_k}{\|\eta_k\|_{x_k}} \right] \right| \\ &\leq L t_k \|\eta_k\|_{x_k}^2 \end{aligned}$$

が成り立つから、これらをあわせて次を得る：

$$t_k \geq - \frac{(1 - c_2) \langle \text{grad } f(x_k), \eta_k \rangle_{x_k}}{L \|\eta_k\|_{x_k}^2}.$$

## 証明 (2/2)

さらにアルミホ条件と  $\langle \text{grad } f(x_k), \eta_k \rangle_{x_k} < 0$  に注意すると,

$$\begin{aligned} f(x_{k+1}) &\leq f(x_k) + c_1 t_k \langle \text{grad } f(x_k), \eta_k \rangle_{x_k} \\ &\leq f(x_k) - \frac{c_1(1-c_2) \langle \text{grad } f(x_k), \eta_k \rangle_{x_k}^2}{L \|\eta_k\|_{x_k}^2} = f(x_k) - \frac{c_1(1-c_2)}{L} \cos^2 \theta_k \|\text{grad } f(x_k)\|_{x_k}^2 \end{aligned}$$

となる. この不等式を繰り返し用いることで,

$$\begin{aligned} f(x_{k+1}) &\leq f(x_0) - \frac{c_1(1-c_2)}{L} \sum_{j=0}^k \cos^2 \theta_j \|\text{grad } f(x_j)\|_{x_j}^2 \\ \iff \sum_{j=0}^k \cos^2 \theta_j \|\text{grad } f(x_j)\|_{x_j}^2 &\leq \frac{L}{c_1(1-c_2)} (f(x_0) - f(x_{k+1})). \end{aligned}$$

$f$  は下に有界なので,  $\sum_{k=0}^{\infty} \cos^2 \theta_k \|\text{grad } f(x_k)\|_{x_k}^2 < \infty$  が示せた.

□

# リーマン多様体上の最急降下法の大域的収束性

定理 5.6 を用いて最急降下法の大域的収束性を示すことができる.

## 定理 5.7

式 (10) を仮定し,  $\{x_k\} \subset M$  をリーマン多様体上の最急降下法によって生成される点列とする. ただし, ステップ幅はウルフ条件を満たすように選ぶとする. このとき,

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

が成り立つ.

## 証明

最急降下法での探索方向は  $\eta_k = -\operatorname{grad} f(x_k)$  であるから、  
これと  $-\operatorname{grad} f(x_k)$  のなす角は  $\theta_k \equiv 0$  であり、 $\cos \theta_k \equiv 1$  が成り立つ。  
よって、定理 5.6 より

$$\sum_{k=0}^{\infty} \|\operatorname{grad} f(x_k)\|_{x_k}^2 < \infty$$

を得る。したがって

$$\lim_{k \rightarrow \infty} \|\operatorname{grad} f(x_k)\|_{x_k} = 0$$

が成り立つ。



# 講演内容

- 1 はじめに
- 2 応用例
- 3 球面上の最急降下法の考え方
- 4 位相空間とリーマン多様体
- 5 リーマン多様体上の一般の最適化手法と最急降下法
- 6 リーマン多様体上の共役勾配法 (R-CG)**
- 7 まとめ



## リーマン多様体上の共役勾配法

- 最急降下法は基本的な手法で最適化における重要な概念を含む一方で、その収束は非常に遅い。
- 共役勾配法は最急降下法に似ているが、収束はずっと速い。
- 共役勾配法は大規模な問題に対して有効であり、現在でも、ユークリッド空間においても研究が続けられている。
- ユークリッド空間での共役勾配法は盛んに研究されているが、リーマン多様体上の共役勾配法についての研究は少ない。
- 以降ではユークリッド空間とリーマン多様体での共役勾配法の基礎事項を述べ、歴史的背景や最新の研究動向を紹介する。

## 線形共役勾配法

- **線形共役勾配法** (Hestenes & Stiefel 1952) は以下の最適化問題に対する解法:

**問題 6.1 (狭義凸 2 次関数の最小化問題 ( $A \in \text{Sym}(n)^+, b \in \mathbb{R}^n$ ))**

$$\begin{aligned} &\text{minimize} \quad f(x) = \frac{1}{2}x^T Ax - b^T x, \\ &\text{subject to} \quad x \in \mathbb{R}^n. \end{aligned}$$

- この問題を解くことは線形方程式  $Ax = b$  を解くことと等価.
- 線形共役勾配法は  $x_{k+1} = x_k + t_k \eta_k$  により点列を生成する.

ここで,  $\eta_k = -\nabla f(x_k) + \beta_k \eta_{k-1}$  (ただし,  $\eta_0 = -\nabla f(x_0)$ ).

また,  $t_k = \arg \min_{t \geq 0} f(x_k + t \eta_k) = -\frac{\nabla f(x_k)^T \eta_k}{\eta_k^T A \eta_k}$  (正確な直線探索)

かつ  $\beta_k = \frac{\|\nabla f(x_k)\|_2^2}{\|\nabla f(x_{k-1})\|_2^2}$  (これにより  $\eta_k^T A \eta_l = \delta_{kl}$  が保証される).

## 線形共役勾配法における $\beta_k$ の性質

$f(x) = \frac{1}{2}x^T Ax - b^T x$  に対して  $\beta_k$  は複数の表現ができる：

$$\begin{aligned}\beta_k &= \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}} = \frac{g_k^T g_k}{\eta_{k-1}^T y_k} = \frac{g_k^T g_k}{-g_{k-1}^T \eta_{k-1}} \\ &= \frac{g_k^T y_k}{g_{k-1}^T g_{k-1}} = \frac{g_k^T y_k}{\eta_{k-1}^T y_k} = \frac{g_k^T y_k}{-g_{k-1}^T \eta_{k-1}},\end{aligned}$$

ここで,  $g_k := \nabla f(x_k)$ ,  $y_k := g_k - g_{k-1}$ .

## 非線形共役勾配法

- $x_{k+1} = x_k + t_k \eta_k$  (ただし  $\eta_k = -\nabla f(x_k) + \beta_k \eta_{k-1}$ ) を反復するというアイデアは、より一般の最適化問題を解くためにも用いることができる：

### 問題 6.2

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

- ステップ幅  $t_k$  は近似的に（ウルフ条件等を満たすように）求めることができる。
- $\beta_k$  はどう計算すれば良いか？

## 非線形共役勾配法における $\beta_k$

- $g_k := \nabla f(x_k)$ ,  $y_k := g_k - g_{k-1}$  とする.
- 線形共役勾配法では ( $f(x) = \frac{1}{2}x^T Ax - b^T x$ ),

$$\begin{aligned}\beta_k &= \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}} = \frac{g_k^T g_k}{\eta_{k-1}^T y_k} = \frac{g_k^T g_k}{-g_{k-1}^T \eta_{k-1}} \\ &= \frac{g_k^T y_k}{g_{k-1}^T g_{k-1}} = \frac{g_k^T y_k}{\eta_{k-1}^T y_k} = \frac{g_k^T y_k}{-g_{k-1}^T \eta_{k-1}}.\end{aligned}$$

- **非線形共役勾配法** (単に**共役勾配法**という) では ( $f$  は一般の滑らかな関数), 上の 6 種類の公式は一般には一致せず, 個別に研究されてきた.

$$\begin{aligned}\beta_k^{\text{FR}} &= \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}, & \beta_k^{\text{DY}} &= \frac{g_k^T g_k}{\eta_{k-1}^T y_k}, & \beta_k^{\text{CD}} &= \frac{g_k^T g_k}{-g_{k-1}^T \eta_{k-1}}, \\ \beta_k^{\text{PRP}} &= \frac{g_k^T y_k}{g_{k-1}^T g_{k-1}}, & \beta_k^{\text{HS}} &= \frac{g_k^T y_k}{\eta_{k-1}^T y_k}, & \beta_k^{\text{LS}} &= \frac{g_k^T y_k}{-g_{k-1}^T \eta_{k-1}}.\end{aligned}$$

## $\beta_k$ についての他の公式

- Hager and Zhang (2005) による  $\beta_k$ :

$$\beta_k^{\text{HZ}} = \frac{1}{\eta_{k-1}^T y_k} \left( y_k - 2\eta_{k-1} \frac{\|y_k\|_2^2}{\eta_{k-1}^T y_k} \right)^T g_k.$$

- その他, 詳細は [Narushima & Yabe, 2014] などを参照.

$\mathbb{R}^n$  上の共役勾配法

---

アルゴリズム 6.1  $\mathbb{R}^n$  上の共役勾配法

---

- 1: 初期点  $x_0 \in \mathbb{R}^n$  を選ぶ.
- 2:  $\eta_0 := -\nabla f(x_0)$ .
- 3: **while**  $\nabla f(x_k) \neq 0$  **do**
- 4:    $t_k$  を計算し,  $x_{k+1} := x_k + t_k \eta_k$  とする.
- 5:    $\beta_{k+1}$  を計算し,

$$\eta_{k+1} := -\nabla f(x_{k+1}) + \beta_{k+1} \eta_k \quad (11)$$

とする.

- 6:    $k := k + 1$ .
  - 7: **end while**
-

# リーマン多様体上の共役勾配法 (R-CG)

リーマン多様体  $M$  上の最適化問題に戻る：

## 問題 6.3

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

$\mathbb{R}^n$  上の共役勾配法における

$$\eta_{k+1} := -\nabla f(x_{k+1}) + \beta_{k+1} \eta_k \tag{11}$$

をリーマン多様体  $M$  上に拡張する際, (11) の「+」が実行できない！

$\text{grad } f(x_{k+1}) \in T_{x_{k+1}}M, \eta_k \in T_{x_k}M \rightarrow$  **和を計算できない.**



## Smith (1994) による R-CG

- Smith はリーマン多様体上で探索方向を

$$\eta_{k+1} = -\text{grad } f(x_{k+1}) + \beta_{k+1} P(\eta_k)$$

と計算することを提案した.

ここで,  $P$  は,  $x_k$  から  $x_{k+1}$  への測地線に沿っての平行移動.

- ここで,

$$\beta_{k+1} = \frac{\langle g_{k+1} - P(g_k), g_{k+1} \rangle_{x_{k+1}}}{-\langle g_k, \eta_k \rangle_{x_k}}.$$

これはユークリッド空間の場合の Liu and Storey による  $\beta_{k+1}^{\text{LS}} = \frac{y_{k+1}^T g_{k+1}}{-g_k^T \eta_k}$  を  
拡張したものに相当する.

# Edelman, Arias, and Smith (1998) による R-CG

- Edelman らも同様に

$$\eta_{k+1} = -\text{grad } f(x_{k+1}) + \beta_{k+1}P(\eta_k)$$

について議論した.

- ここで,

$$\beta_{k+1} = \frac{\langle g_{k+1} - P(g_k), g_{k+1} \rangle_{x_{k+1}}}{\langle g_k, g_k \rangle_{x_k}}.$$

これはユークリッド空間における  $\beta_{k+1}^{\text{PRP}} = \frac{y_{k+1}^T g_{k+1}}{g_k^T g_k}$  を拡張したものに相当する.

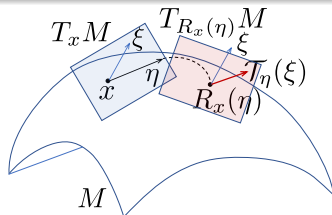
# Absil, Mahony, and Sepulchre (2008) による vector transport

$TM \oplus TM := \{(\eta, \xi) \mid \eta, \xi \in T_x M, x \in M\}$  と定義する.

## 定義 6.4

写像  $\mathcal{T}: TM \oplus TM \rightarrow TM: (\eta, \xi) \mapsto \mathcal{T}_\eta(\xi)$  が任意の  $x \in M$  に対し次の条件を満たすとき,  $\mathcal{T}$  を  $M$  上の **vector transport** という.

1.  $\mathcal{T}_\eta(\xi) \in T_{R_x(\eta)}M$ ,  $\eta, \xi \in T_xM$ ,
2.  $\mathcal{T}_{0_x}(\xi) = \xi$ ,  $\xi \in T_xM$ , ( $0_x$  は  $T_xM$  の零ベクトル)
3.  $\mathcal{T}_\eta(a\xi + b\zeta) = a\mathcal{T}_\eta(\xi) + b\mathcal{T}_\eta(\zeta)$ ,  $a, b \in \mathbb{R}$ ,  $\eta, \xi, \zeta \in T_xM$ .



# Absil et al. による vector transport を用いた R-CG

Vector transport  $\mathcal{T}$  を用いると, R-CG は次のように書ける:

---

## アルゴリズム 6.2 リーマン多様体 $M$ 上の共役勾配法 (R-CG)

---

- 1: 初期点  $x_0 \in M$  を選び,  $\eta_0 := -\text{grad } f(x_0)$  を計算する.
  - 2: **while**  $\text{grad } f(x_k) \neq 0$  **do**
  - 3:     $t_k > 0$  を計算し  $x_{k+1} := R_{x_k}(t_k \eta_k)$  とする.
  - 4:     $\beta_{k+1}$  を計算し  $\eta_{k+1} := -\text{grad } f(x_{k+1}) + \beta_{k+1} \mathcal{T}_{t_k \eta_k}(\eta_k)$  とする.
  - 5:     $k := k + 1$ .
  - 6: **end while**
- 

$\beta_k \in \mathbb{R}$  をどう計算するか?

# ユークリッド空間での共役勾配法における $\beta_k$ (再掲)

$g_k := \nabla f(x_k)$ ,  $y_{k+1} := g_{k+1} - g_k$  とする.

- $\beta_{k+1}^{\text{HS}} = \frac{g_{k+1}^T y_{k+1}}{\eta_k^T y_{k+1}}$ . [Hestenes & Stiefel, 1952]
- $\beta_{k+1}^{\text{FR}} = \frac{\|g_{k+1}\|_2^2}{\|g_k\|_2^2}$ . [Fletcher & Reeves, 1964]
- $\beta_{k+1}^{\text{PRP}} = \frac{g_{k+1}^T y_{k+1}}{\|g_k\|_2^2}$ . [Polak & Ribière 1969], [Polyak 1969]
- $\beta_{k+1}^{\text{CD}} = \frac{\|g_{k+1}\|_2^2}{-\eta_k^T g_k}$ . [Fletcher, 1987]
- $\beta_{k+1}^{\text{LS}} = \frac{g_{k+1}^T y_{k+1}}{-\eta_k^T g_k}$ . [Liu & Storey, 1991]
- $\beta_{k+1}^{\text{DY}} = \frac{\|g_{k+1}\|_2^2}{\eta_k^T y_{k+1}}$ . [Dai & Yuan, 1999]

$\beta_k$  のリーマン多様体への拡張

$g_k := \nabla f(x_k)$ ,  $y_{k+1} := g_{k+1} - g_k$  とする.

- Fletcher–Reeves (FR):  $\mathbb{R}^n$  では,  $\beta_{k+1}^{\text{FR}} = \frac{\|g_{k+1}\|_2^2}{\|g_k\|_2^2}$ .

→  $M$  上では,

$$\beta_{k+1} = \frac{\|\text{grad } f(x_{k+1})\|_{x_{k+1}}^2}{\|\text{grad } f(x_k)\|_{x_k}^2}.$$

- Dai–Yuan (DY):  $\mathbb{R}^n$  では,  $\beta_{k+1}^{\text{DY}} = \frac{\|g_{k+1}\|_2^2}{\eta_k^T y_{k+1}}$ .

→  $M$  上では, FR の場合より一般化について注意が必要.

分母の  $\eta_k^T y_{k+1} = \eta_k^T (g_{k+1} - g_k)$  をどのように一般化するか?

- $\langle \mathcal{T}_{t_k \eta_k}(\eta_k), \text{grad } f(x_{k+1}) - \mathcal{T}_{t_k \eta_k}(\text{grad } f(x_k)) \rangle_{x_{k+1}}$  ?
- $\langle \eta_k, (\mathcal{T}_{t_k \eta_k})^{-1}(\text{grad } f(x_{k+1})) - \text{grad } f(x_k) \rangle_{x_k}$  ?
- $\langle \mathcal{T}_{t_k \eta_k}(\eta_k), \text{grad } f(x_{k+1}) \rangle_{x_{k+1}} - \langle \eta_k, \text{grad } f(x_k) \rangle_{x_k}$  ?

# Ring and Wirth (2012) による FR-type R-CG

- Ring and Wirth は FR-type の R-CG について議論した：

$$\beta_{k+1} = \frac{\|\text{grad } f(x_{k+1})\|_{x_{k+1}}^2}{\|\text{grad } f(x_k)\|_{x_k}^2}.$$

- 大域的収束性の解析においては、以下の仮定が必要：

$$\|\mathcal{T}_{t_k \eta_k}(\eta_k)\|_{x_{k+1}} \leq \|\eta_k\|_{x_k}. \quad (12)$$

しかしながら、この条件は必ずしも成り立たない。

たとえば、標準内積を備えた  $\mathbb{R}^n$  のリーマン部分多様体としての球面  $S^{n-1}$  上のレトラクション  $R_x(\eta) := \sqrt{1 - \eta^T \eta} x + \eta$  ( $\eta^T \eta < 1$ ) の微分により定義される vector transport  $\mathcal{T}_\eta^R(\xi) := DR_x(\eta)[\xi]$  は条件 (12) を満たさない。実際、

$$\mathcal{T}_\eta^R(\xi) = -\frac{\eta^T \xi}{\sqrt{1 - \eta^T \eta}} x + \xi$$

より、 $x^T x = 1$  と  $x^T \xi = 0$  から、 $\eta^T \xi \neq 0$  ならば  $\|\mathcal{T}_\eta^R(\xi)\|_{R_x(\eta)} > \|\xi\|_x$ .

# Sato and Iwai (2015) による scaled vector transport

- 条件 (12) を取り除くための一つの方策：

Vector transport によって移されたベクトルのノルムが大きくならないようにスケーリングする.

Sato and Iwai は,  $\mathcal{T}_\eta^R(\xi) := DR_x(\eta)[\xi]$  について scaled vector transport  $\mathcal{T}^0$  を次のように定義した：

$$\mathcal{T}_\eta^0(\xi) = \frac{\|\xi\|_x}{\|\mathcal{T}_\eta^R(\xi)\|_{R_x(\eta)}} \mathcal{T}_\eta^R(\xi), \quad \xi, \eta \in T_x M.$$

- 実践的には,

$$\mathcal{T}_{t_k \eta_k}^{(k)}(\eta_k) := \begin{cases} \mathcal{T}_{t_k \eta_k}^R(\eta_k), & \text{if } \|\mathcal{T}_{t_k \eta_k}^R(\eta_k)\|_{x_{k+1}} \leq \|\eta_k\|_{x_k}, \\ \mathcal{T}_{t_k \eta_k}^0(\eta_k), & \text{otherwise} \end{cases}$$

とし, 探索方向を  $\eta_{k+1} := -\text{grad } f(x_{k+1}) + \beta_{k+1} \mathcal{T}_{t_k \eta_k}^{(k)}(\eta_k)$  と計算する.



# FR-type R-CG の大域的収束性

## 定理 6.1 (Sato & Iwai, 2015)

$f$  の有界性と滑らかさの仮定の下で, ステップ幅が常に**強ウルフ条件**を満たす  
(ただし  $0 < c_1 < c_2 < 1/2$ ) とする.

このとき, R-CG (ただし  $\beta_{k+1} = \beta_{k+1}^{\text{R-FR}} := \|\text{grad } f(x_{k+1})\|_{x_{k+1}}^2 / \|\text{grad } f(x_k)\|_{x_k}^2$ ) に  
より生成される点列  $\{x_k\}$  は

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

を満たす.

## Sato (2016) による DY-type R-CG

- Sato は Dai and Yuan (1999) による  $\beta_k^{\text{DY}}$  を一般化した.
- $\mathbb{R}^n$  では,  $g_k = \nabla f(x_k)$ ,  $y_{k+1} = g_{k+1} - g_k$  として,

$$\beta_{k+1}^{\text{DY}} = \frac{\|g_{k+1}\|_2^2}{\eta_k^T y_{k+1}} = \frac{g_{k+1}^T \eta_{k+1}}{g_k^T \eta_k}.$$

- リーマン多様体  $M$  上では,  $g_k = \text{grad } f(x_k)$  として,

$$\beta_{k+1} = \frac{\langle g_{k+1}, \eta_{k+1} \rangle_{x_{k+1}}}{\langle g_k, \eta_k \rangle_{x_k}}$$

から始める. ここで,  $\eta_{k+1}$  は  $\beta_{k+1}$  を用いて計算される.  
 $\rightarrow \beta_{k+1}$  に対するこの式はアルゴリズム中で利用できない.

# DY 法のリーマン多様体への拡張

$$\begin{aligned}
 \beta_{k+1} &= \frac{\langle g_{k+1}, \eta_{k+1} \rangle_{x_{k+1}}}{\langle g_k, \eta_k \rangle_{x_k}} \\
 &= \frac{\langle g_{k+1}, -g_{k+1} + \beta_{k+1} \mathcal{T}_{t_k \eta_k}^{(k)}(\eta_k) \rangle_{x_{k+1}}}{\langle g_k, \eta_k \rangle_{x_k}} \\
 &= \frac{-\|g_{k+1}\|^2 + \beta_{k+1} \langle g_{k+1}, \mathcal{T}_{t_k \eta_k}^{(k)}(\eta_k) \rangle_{x_{k+1}}}{\langle g_k, \eta_k \rangle_{x_k}}.
 \end{aligned}$$

よって,  $\beta_{k+1}$  について解くことで,

$$\beta_{k+1} = \frac{\|g_{k+1}\|_{x_{k+1}}^2}{\langle g_{k+1}, \mathcal{T}_{t_k \eta_k}^{(k)}(\eta_k) \rangle_{x_{k+1}} - \langle g_k, \eta_k \rangle_{x_k}}.$$

## DY 法のリーマン多様体への拡張

- $\mathbb{R}^n$  の場合 :

$$\beta_{k+1}^{\text{DY}} = \frac{g_{k+1}^T \eta_{k+1}}{g_k^T \eta_k} = \frac{\|g_{k+1}\|^2}{\eta_k^T y_{k+1}}, \quad y_k = g_{k+1} - g_k.$$

- リーマン多様体  $M$  の場合 :

$$\beta_{k+1}^{\text{R-DY}} = \frac{\langle g_{k+1}, \eta_{k+1} \rangle_{x_{k+1}}}{\langle g_k, \eta_k \rangle_{x_k}} = \frac{\|g_{k+1}\|_{x_{k+1}}^2}{\langle \mathcal{T}_{t_k \eta_k}^{(k)}(\eta_k), y_{k+1} \rangle_{x_{k+1}}}.$$

ここで,

$$y_{k+1} = g_{k+1} - \frac{\langle g_k, \eta_k \rangle_{x_k}}{\langle \mathcal{T}_{t_k \eta_k}^{(k)}(g_k), \mathcal{T}_{t_k \eta_k}^{(k)}(\eta_k) \rangle_{x_{k+1}}} \mathcal{T}_{t_k \eta_k}^{(k)}(g_k).$$

## DY-type R-CG の大域的収束性

### 定理 6.2 (Sato, 2016)

$f$  の有界性と滑らかさの仮定の下で, ステップ幅が常に**ウルフ条件**を満たすとき,  
R-CG (ただし

$$\beta_{k+1} = \beta_{k+1}^{\text{R-DY}} := \frac{\|\text{grad } f(x_{k+1})\|_{x_{k+1}}^2}{\langle \text{grad } f(x_{k+1}), \mathcal{T}_{t_k \eta_k}^{(k)}(\eta_k) \rangle_{x_{k+1}} - \langle \text{grad } f(x_k), \eta_k \rangle_{x_k}}$$

である) により生成される点列  $\{x_k\}$  は次を満たす:

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

## Sakai and Iiduka (2020) による R-CG のクラス

- Dai and Yuan (2001) の結果を一般化し, Sakai and Iiduka は  $-\sigma \leq r_k \leq 1$  を満たす  $\beta_k$  を用いる手法を提案し, その大域的収束性を証明した.  
ここで,  $r_k := \beta_k / \beta_k^{\text{R-DY}}$  である. ただし,  $\sigma := (1 - c_2) / (1 + c_2)$  であり,  $c_2$  は強ウルフ条件  $|\phi'_k(t_k)| \leq c_2 |\phi'_k(0)|$  ( $\phi_k(t) := f(R_{x_k}(t\eta_k))$ ) での定数である.
- この手法は例として  $\beta_k = \max\{0, \min\{\beta_k^{\text{R-DY}}, \beta_k^{\text{R-HS}}\}\}$  や  $\beta_k = \max\{-\sigma \beta_k^{\text{R-DY}}, \min\{\beta_k^{\text{R-DY}}, \beta_k^{\text{R-HS}}\}\}$  を含む. ここで,

$$\beta_{k+1}^{\text{R-HS}} = \frac{\langle g_{k+1}, g_{k+1} - \mathcal{T}_{t_k \eta_k}^{(k)}(g_k) \rangle_{x_{k+1}}}{\langle g_{k+1}, \mathcal{T}_{t_k \eta_k}^{(k)}(\eta_k) \rangle_{x_{k+1}} - \langle g_k, \eta_k \rangle_{x_k}}$$

は Hestenes and Stiefel による  $\beta_{k+1}^{\text{HS}} = \frac{g_{k+1}^T y_{k+1}}{y_{k+1}^T \eta_k}$  の一般化である.

# Sakai and Iiduka (2021) による HZ-type R-CG

- Hager and Zhang (2005) の結果を一般化し, Sakai and Iiduka は

$$\beta_{k+1}^{\text{R-HZ}} = \beta_k^{\text{R-HS}} - \mu \frac{\|y_{k+1}\|_{x_{k+1}}^2 \langle g_{k+1}, \mathcal{T}_{t_k \eta_k}^{(k)}(\eta_k) \rangle_{x_{k+1}}}{\left( \langle g_{k+1}, \mathcal{T}_{t_k \eta_k}^{(k)}(\eta_k) \rangle_{x_{k+1}} - \langle g_k, \eta_k \rangle_{x_k} \right)^2}$$

を提案した. ここで,  $\mu > 1/4$  である.

これは Hager and Zhang による  $\beta_k^{\text{HZ}}$  の拡張に相当する.

# 数値実験 1

## 問題 6.5

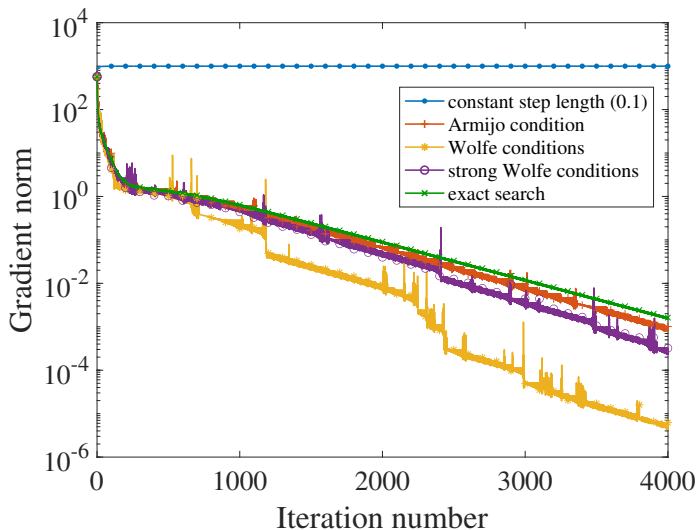
$$\begin{array}{ll} \text{minimize} & f(x) := x^T A x, \\ \text{subject to} & x \in S^{n-1}. \end{array}$$

ここで,  $n = 100$ ,  $A := \text{diag}(1, 2, \dots, n)$ .

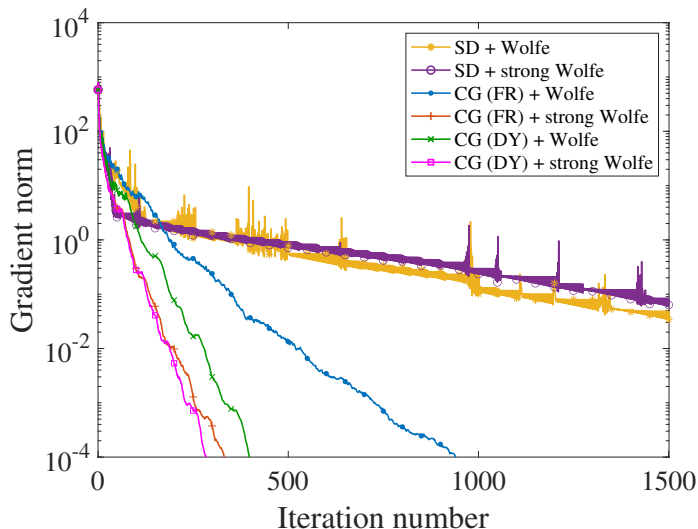
初期点  $x_0 \in S^{n-1}$  はランダムに生成.



# 数値計算結果 (最急降下法 (SD))



# 数値計算結果 (最急降下法 (SD) と共役勾配法 (CG))



## 数値実験 2

### 問題 6.6

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

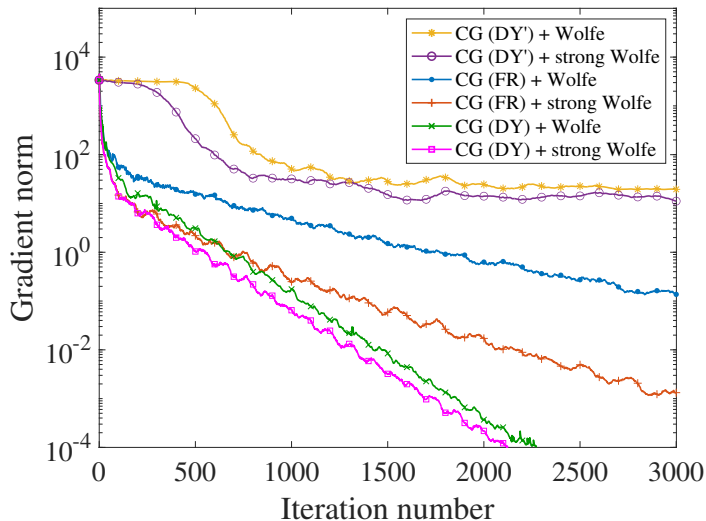
ここで,  $n = 300$ ,  $p = 10$ ,  $A := \text{diag}(1, 2, \dots, n)$ ,  $N := \text{diag}(p, p-1, \dots, 1)$ .

初期点  $X_0 \in \text{St}(p, n)$  はランダムに生成.

なお,  $\beta_{k+1}^{\text{R-DY}} := \frac{\|\text{grad } f(x_{k+1})\|_{x_{k+1}}^2}{\langle \text{grad } f(x_{k+1}), \mathcal{T}_{t_k \eta_k}^{(k)}(\eta_k) \rangle_{x_{k+1}} - \langle \text{grad } f(x_k), \eta_k \rangle_{x_k}}$  を導出する際に不採用とした,

$\beta_{k+1}^{\text{R-DY}'} := \frac{\|\text{grad } f(x_{k+1})\|_{x_{k+1}}^2}{\langle \text{grad } f(x_{k+1}) - \mathcal{T}_{t_k \eta_k}^{(k)}(\text{grad } f(x_k)), \mathcal{T}_{t_k \eta_k}^{(k)}(\eta_k) \rangle_{x_{k+1}}}$  についても実験を行った.

## 数値計算結果 (共役勾配法)



# Zhu and Sato (2020) による レトラクションの逆写像を用いた R-CG

- 動機：R-CG にはレトラクションと vector transport という 2 つの写像が必ず必要？
- $\eta_{k+1} = -\text{grad } f(x_{k+1}) + \beta_{k+1} \mathcal{T}_{t_k \eta_k}(\eta_k)$  に注目すると、この部分には必ず vector transport が必要？
- 目的は、 $\eta_k \in T_{x_k} M$  を  $T_{x_{k+1}} M$  に移すこと。
- ユークリッド空間  $\mathbb{R}^n$  の場合：  $R_x(\eta) = x + \eta$ ,  $x \in \mathbb{R}^n$ .  
 $x_{k+1} = x_k + t_k \eta_k$  より,  $x_k = x_{k+1} + (-t_k \eta_k) = R_{x_{k+1}}(-t_k \eta_k)$ .  
 したがって,  $T_{x_k} \mathbb{R}^n \ni \eta_k = -t_k^{-1} R_{x_{k+1}}^{-1}(x_k) \in T_{x_{k+1}} \mathbb{R}^n$  ( $T_{x_k} \mathbb{R}^n \simeq \mathbb{R}^n \simeq T_{x_{k+1}} \mathbb{R}^n$ ).
- リーマン多様体  $M$  の場合：  
 $\mathcal{T}_{t_k \eta_k}(\eta_k)$  の代わりに  $-t_k^{-1} R_{x_{k+1}}^{-1}(x_k) \in T_{x_{k+1}} M$  を利用できるはず！

# レトラクションは異なるものを用いても良い

- レトラクションは指数写像の 1 次近似なので、  
任意のレトラクション  $R^{\text{bw}}$  は  $R$  の 1 次近似である.
- 特に  $x_{k+1} \approx x_k$  のとき、**任意のレトラクション  $R^{\text{bw}}$  に対する**  
 $-t_k^{-1}(R_{x_{k+1}}^{\text{bw}})^{-1}(x_k) \in T_{x_{k+1}}M$  が vector transport の代わりに役割を果たす.
- $R = R^{\text{bw}}$  であっても、そうでなくても良い.

## レトラクションの逆写像を用いた R-CG

- $M$  上のレトラクション  $R^{\text{fw}}$  と  $R^{\text{bw}}$  ( $R^{\text{fw}} = R^{\text{bw}}$  でも良い) を用いて,  
 $x_{k+1} = R_{x_k}^{\text{fw}}(t_k \eta_k)$  かつ

$$\eta_{k+1} := -\text{grad } f(x_{k+1}) - \beta_{k+1} s_k t_k^{-1} (R_{x_{k+1}}^{\text{bw}})^{-1}(x_k)$$

とする. ここで,

$$s_k := \min \left\{ 1, \|\eta_k\|_{x_k} / \|t_k^{-1} (R_{x_{k+1}}^{\text{bw}})^{-1}(x_k)\|_{x_{k+1}} \right\} > 0.$$

- このアルゴリズムは vector transport を必要としない.
- $s_k > 0$  は, 得られた接ベクトルのノルムをスケーリングすることによって大域的収束性を保証するためのパラメータ.
- $\beta_k^{\text{R-FR}}$  と  $\beta_k^{\text{R-DY}}$  に対して大域的収束性を証明した.

# 講演内容

- 1 はじめに
- 2 応用例
- 3 球面上の最急降下法の考え方
- 4 位相空間とリーマン多様体
- 5 リーマン多様体上の一般の最適化手法と最急降下法
- 6 リーマン多様体上の共役勾配法 (R-CG)
- 7 **まとめ**



## まとめ

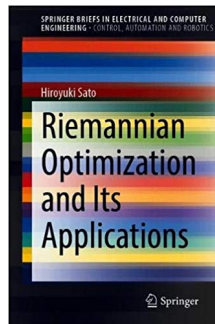
- 機械学習や信号処理など、様々な分野に現れる最適化問題の中には、リーマン多様体上の問題として定式化できるものが多い。
- リーマン多様体上の最適化アルゴリズムについて、球面上の最急降下法の例を交えながら、一般論を説明した。
- リーマン多様体上の共役勾配法の枠組みと最新の研究成果を紹介した。
- 様々な言語で、多様体上の最適化のためのライブラリが利用可能である：
  - Manopt: MATLAB [Boumal et al. 2014]
  - Pymanopt: Python [Townsend et al. 2016]
  - ManifoldOptim: R [Martin et al. 2016]
  - ROPTLIB: C++ [Huang et al. 2018]
  - Manopt.jl: Julia [Bergman 2019]

## 時間の都合上、本講義で扱えなかったこと

- 目的関数の **2 次** の情報（ヘシアン）を用いたアルゴリズム（ニュートン法など）
- **商多様体**（グラスマン多様体など）の理論
- リーマン多様体上の **制約付き** 最適化問題 [Liu & Boumal 2020]
- リーマン多様体上の信頼領域法 [Absil et al. 2007]
- リーマン多様体上の確率的最適化 [Sato et al. 2019]
- リーマン多様体上の近接勾配法 [Chen et al. 2020]

etc.

この分野の書籍として、  
[Absil et al. 2008, Boumal 2020, Sato 2021] がある。



## その他の応用例

### ● 数値線形代数への応用

- 固有値問題への応用 [Sato & Iwai 2014]
- 特異値分解への応用 [Sato & Iwai 2013]・その方法論の確立 [Aihara & Sato 2017]
- 近似的同時対角化（→ 独立成分分析）への応用 [Absil & Gallivan 2006, Sato 2017]
- 近似的同時特異値分解 [Sato 2015]
- 逆固有値問題への応用 [Yao et al. 2020]
- 低ランク行列補完 [Mishra et al. 2014]・テンソル補完 [Kasai & Mishra 2016]

### ● 制御分野への応用

- システム同定への応用 [Sato & Sato 2017, Sato et al. 2020]
- 制御システムの低次元化への応用 [Sato & Sato 2016, Yan et al. 1999]

### ● 統計手法への応用

- 正準相関分析への応用 [Yger et al. 2012]・その計算法の確立 [Sato & Aihara 2019]
- トポグラフィック独立成分分析への応用 [Kitao et al. 2015]

## 参考文献 I

- [1] Absil, P.A., Baker, C.G., Gallivan, K.A.: Trust-region methods on Riemannian manifolds. *Foundations of Computational Mathematics* **7**(3), 303–330 (2007)
- [2] Absil, P.A., Gallivan, K.: Joint diagonalization on the oblique manifold for independent component analysis. In: *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, vol. 5, pp. V–V. IEEE (2006)
- [3] Absil, P.A., Mahony, R., Sepulchre, R.: *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, NJ (2008)
- [4] Aihara, K., Sato, H.: A matrix-free implementation of Riemannian Newton's method on the Stiefel manifold. *Optimization Letters* **11**(8), 1729–1741 (2017)

## 参考文献 II

- [5] Bergmann, R.: manopt.jl — Optimization on Manifolds in Julia.  
<http://www.manoptjl.org> (2019)
- [6] Boumal, N.: An Introduction to Optimization on Smooth Manifolds (2020).  
URL <http://www.nicolasboumal.net/book>
- [7] Boumal, N., Mishra, B., Absil, P.A., Sepulchre, R.: Manopt, a matlab toolbox for optimization on manifolds. *Journal of Machine Learning Research* **15**(1), 1455–1459 (2014)
- [8] Chen, S., Ma, S., Man-Cho So, A., Zhang, T.: Proximal gradient method for nonsmooth optimization over the Stiefel manifold. *SIAM Journal on Optimization* **30**(1), 210–239 (2020)

## 参考文献 III

- [9] Dai, Y.H., Yuan, Y.: A nonlinear conjugate gradient method with a strong global convergence property. *SIAM Journal Optimization* **10**(1), 177–182 (1999)
- [10] Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications* **20**(2), 303–353 (1998)
- [11] Fletcher, R.: *Practical Methods of Optimization*. John Wiley & Sons (2013)
- [12] Fletcher, R., Reeves, C.M.: Function minimization by conjugate gradients. *The Computer Journal* **7**(2), 149–154 (1964)
- [13] Gabay, D.: Minimizing a differentiable function over a differential manifold. *J. Optim. Theory Appl.* **37**(2), 177–219 (1982)

## 参考文献 IV

- [14] Hager, W.W., Zhang, H.: A new conjugate gradient method with guaranteed descent and an efficient line search. *SIAM Journal on Optimization* **16**(1), 170–192 (2005)
- [15] Hestenes, M.R., Stiefel, E.: Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards* **49**(6), 409–436 (1952)
- [16] Hosseini, R., Sra, S.: Matrix manifold optimization for Gaussian mixtures. In: *Advances in Neural Information Processing Systems*, pp. 910–918 (2015)
- [17] Huang, W., Absil, P.A., Gallivan, K.A., Hand, P.: ROPTLIB: an object-oriented C++ library for optimization on Riemannian manifolds. *ACM Transactions on Mathematical Software* **44**(4), 43 (2018)

## 参考文献 V

- [18] Kasai, H., Mishra, B.: Low-rank tensor completion: a Riemannian manifold preconditioning approach. In: Proceedings of the 33rd International Conference on Machine Learning, pp. 1012–1021 (2016)
- [19] Kitao, A., Shiohama, T., Sato, H.: Financial news classification based on topographic independent component analysis: Optimization on the Stiefel manifold. In: Proceedings of the 16th Applied Stochastic Models and Data Analysis International Conference, pp. 403–415 (2015)
- [20] Liu, C., Boumal, N.: Simple algorithms for optimization on Riemannian manifolds with constraints. Applied Mathematics and Optimization (2019). DOI 10.1007/s00245-019-09564-3



## 参考文献 VI

- [21] Liu, Y., Storey, C.: Efficient generalized conjugate gradient algorithms, part 1: theory. *Journal of Optimization Theory and Applications* **69**(1), 129–137 (1991)
- [22] Luenberger, D.G.: The gradient projection method along geodesics. *Management Science* **18**(11), 620–631 (1972)
- [23] Martin, S., Raim, A.M., Huang, W., Adraghi, K.P.: ManifoldOptim: An R interface to the ROPTLIB library for Riemannian manifold optimization. arXiv preprint arXiv:1612.03930 (2016)
- [24] Mishra, B., Meyer, G., Bonnabel, S., Sepulchre, R.: Fixed-rank matrix factorizations and Riemannian low-rank optimization. *Computational Statistics* **29**(3-4), 591–621 (2014)

## 参考文献 VII

- [25] Narushima, Y., Yabe, H., et al.: A survey of sufficient descent conjugate gradient methods for unconstrained optimization. *SUT Journal of Mathematics* **50**(2), 167–203 (2014)
- [26] Polak, E., Ribiere, G.: Note sur la convergence de méthodes de directions conjuguées. *ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique* **3**(R1), 35–43 (1969)
- [27] Polyak, B.T.: The conjugate gradient method in extremal problems. *USSR Computational Mathematics and Mathematical Physics* **9**(4), 94–112 (1969)
- [28] Ring, W., Wirth, B.: Optimization methods on Riemannian manifolds and their application to shape space. *SIAM Journal on Optimization* **22**(2), 596–627 (2012)

## 参考文献 VIII

- [29] Sakai, H., Iiduka, H.: Hybrid Riemannian conjugate gradient methods with global convergence properties. *Computational Optimization and Applications* **77**, 811–830 (2020)
- [30] Sakai, H., Iiduka, H.: Sufficient descent Riemannian conjugate gradient methods. *Journal of Optimization Theory and Applications* **190**, 130–150 (2021)
- [31] Sato, H.: Joint singular value decomposition algorithm based on the Riemannian trust-region method. *JSIAM Letters* **7**, 13–16 (2015)
- [32] Sato, H.: A Dai–Yuan-type Riemannian conjugate gradient method with the weak Wolfe conditions. *Computational Optimization and Applications* **64**(1), 101–118 (2016)

## 参考文献 IX

- [33] Sato, H.: Riemannian Newton-type methods for joint diagonalization on the Stiefel manifold with application to independent component analysis. *Optimization* **66**(12), 2211–2231 (2017)
- [34] Sato, H.: Riemannian Optimization and Its Applications. Springer Nature (2021)
- [35] Sato, H., Aihara, K.: Cholesky QR-based retraction on the generalized Stiefel manifold. *Computational Optimization and Applications* **72**(2), 293–308 (2019)
- [36] Sato, H., Iwai, T.: A Riemannian optimization approach to the matrix singular value decomposition. *SIAM Journal Optimization* **23**(1), 188–212 (2013)

## 参考文献 X

- [37] Sato, H., Iwai, T.: Optimization algorithms on the Grassmann manifold with application to matrix eigenvalue problems. Japan Journal of Industrial and Applied Mathematics **31**(2), 355–400 (2014)
- [38] Sato, H., Iwai, T.: A new, globally convergent Riemannian conjugate gradient method. Optimization **64**(4), 1011–1031 (2015)
- [39] Sato, H., Kasai, H., Bamdev, M.: Riemannian stochastic variance reduced gradient algorithm with retraction and vector transport. SIAM Journal on Optimization **29**(2), 1444–1472 (2019)
- [40] Sato, H., Sato, K.: A new  $H^2$  optimal model reduction method based on Riemannian conjugate gradient method. In: 2016 IEEE 55th Conference on Decision and Control (CDC), pp. 5762–5768. IEEE (2016)

## 参考文献 XI

- [41] Sato, K., Sato, H., Damm, T.: Riemannian optimal identification method for linear systems with symmetric positive-definite matrix. *IEEE Transactions on Automatic Control* (2019)
- [42] Smith, S.T.: Optimization techniques on Riemannian manifolds. In: *Hamiltonian and Gradient Flows, Algorithms and Control*, pp. 113–135. American Mathematical Society (1994)
- [43] Sun, J., Qu, Q., Wright, J.: Complete dictionary recovery over the sphere ii: Recovery by riemannian trust-region method. *IEEE Transactions on Information Theory* **63**(2), 885–914 (2016)
- [44] Tanabe, K.: Differential geometric methods in nonlinear programming. In: *Applied Nonlinear Analysis*, pp. 707–719. Elsevier (1979)

## 参考文献 XII

- [45] Tanabe, K.: A geometric method in nonlinear programming. Journal of Optimization Theory and Applications **30**(2), 181–210 (1980)
- [46] Townsend, J., Koep, N., Weichwald, S.: Pymanopt: A python toolbox for optimization on manifolds using automatic differentiation. Journal of Machine Learning Research **17**(1), 4755–4759 (2016)
- [47] Tu, L.W.: An Introduction to Manifolds. Springer New York (2011)
- [48] Udriște, C.: Convex Functions and Optimization Methods on Riemannian Manifolds, vol. 297. Springer Science & Business Media (2013)
- [49] Yan, W.Y., Lam, J.: An approximate approach to  $H^2$  optimal model reduction. IEEE Transactions on Automatic Control **44**(7), 1341–1358 (1999)

## 参考文献 XIII

- [50] Yao, T.T., Bai, Z.J., Jin, X.Q., Zhao, Z.: A geometric Gauss–Newton method for least squares inverse eigenvalue problems. BIT Numerical Mathematics pp. 1–28 (2020)
- [51] Yger, F., Berar, M., Gasso, G., Rakotomamonjy, A.: Adaptive canonical correlation analysis based on matrix manifolds. In: Proceedings of the 29th International Conference on Machine Learning, pp. 1071–1078 (2012)
- [52] Zhu, X., Sato, H.: Riemannian conjugate gradient methods with inverse retraction. Computational Optimization and Applications **77**, 779–810 (2020)



**ご清聴ありがとうございました**