

# $n$ 次元 Fisher-Bingham 分布のホロノミック勾配降下法を用いた最尤推定とその実装について

中山 洋将

NAKAYAMA HIROMASA

神戸大学大学院理学研究科/ JST CREST

DEPARTMENT OF MATHEMATICS, GRADUATE SCHOOL OF SCIENCE, KOBE UNIVERSITY \* †

小山 民雄

KOYAMA TAMIO

神戸大学大学院理学研究科/ JST CREST

DEPARTMENT OF MATHEMATICS, GRADUATE SCHOOL OF SCIENCE, KOBE UNIVERSITY ‡

西山 絢太

NISHIYAMA KENTA

大阪大学大学院情報科学研究科/ JST CREST

GRADUATE SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY, OSAKA UNIVERSITY §

高山 信毅

TAKAYAMA NOBUKI

神戸大学大学院理学研究科/ JST CREST

DEPARTMENT OF MATHEMATICS, GRADUATE SCHOOL OF SCIENCE, KOBE UNIVERSITY ¶

## 1 概要

ホロノミック勾配降下法を  $n$  次元球面上 Fisher-Bingham 分布の最尤推定に適用した. 今までの方法では 2 次元までしか計算可能ではなかった ([5]). それは微分方程式系を Pfaff 系への書き換える部分 (具体的には微分作用素環のグレブナー基底計算) が計算困難なためであった. この書き換えをグレブナー基底を計算せずに手で行うことで, 高次元球面上の Fisher-Bingham 分布について, 最尤推定を行うことができるようになった ([2]).

---

\*nakayama@math.kobe-u.ac.jp

†本研究の一部は JSPS 科研費 24740064 の助成を受けたものです.

‡tkoyama@math.kobe-u.ac.jp

§k-nishiyama@cr.math.sci.osaka-u.ac.jp

¶takayama@math.kobe-u.ac.jp

## 2 Fisher-Bingham 分布とその最尤推定

この論文で対象とする Fisher-Bingham 分布とは次のような球面上の確率分布である.

### 定義 1 (Fisher-Bingham 分布)

Fisher-Bingham 分布とは, 半径  $r > 0$  の  $n$  次元球面  $S^n(r)$  上に

$$\mu(x, y, r; t) |dt| = \frac{1}{Z(x, y, r)} \exp \left( \sum_{1 \leq i \leq j \leq n+1} x_{ij} t_i t_j + \sum_{i=1}^{n+1} y_i t_i \right) |dt|$$

により定義される確率分布である. ここで,  $Z$  は正規化定数であり, パラメータは,  $x = (x_{ij})$  ( $1 \leq i \leq j \leq n+1$ ),  $y = (y_i)$  ( $1 \leq i \leq n+1$ ),  $r$  である.

### 定義 2 (Fisher-Bingham 積分)

Fisher-Bingham 積分を

$$Z(x, y, r) = \int_{S^n(r)} \exp \left( \sum_{1 \leq i \leq j \leq n+1} x_{ij} t_i t_j + \sum_{i=1}^{n+1} y_i t_i \right) |dt|.$$

で定義する. これは, Fisher-Bingham 分布の正規化定数である.

Fisher-Bingham 分布の最尤法を行う. そのためには尤度関数

$$L(x, y, r) = \frac{1}{Z(x, y, r)^N} \prod_{k=1}^N \exp \left( \sum_{1 \leq i \leq j \leq n+1} x_{ij} t_i^{(k)} t_j^{(k)} + \sum_{i=1}^{n+1} y_i t_i^{(k)} \right)$$

の最大点を求めることが必要になる. ただし,  $t^{(1)}, \dots, t^{(N)} \in S^n(r)$  は与えられた球面上の点である. 尤度関数は, パラメータ  $x, y, r$  を仮定した時に, 与えられたデータ  $t^{(1)}, \dots, t^{(N)}$  が出現する確率を表している. この最大値を求めるのに, ホロノミック勾配降下法 (Holonomic Gradient Descent method) を用いる.

### 例 1 (1 次元 Fisher-Bingham 分布とその最尤推定)

次は神戸市 2011 年 8 月 1 日の 10 分ごとの平均風向 (16 方位) のデータ ( $N = 144$ ) の表である.

方位	ラジアン	頻度
北	1.570760	0
北北東	1.178070	0
北東	0.785380	34
東北東	0.392690	35
東	0.000000	5
東南東	5.890349	17
南東	5.497659	5
南南東	5.104969	0
南	4.712279	3
南南西	4.319589	4
南西	3.926899	28
西南西	3.534209	10
西	3.141519	4
西北西	2.748829	0
北西	2.356139	0
北北西	1.963450	0

(気象庁の過去の気象データ <http://www.data.jma.go.jp/obd/stats/etrn/> より引用)

このデータについて、1 次元 Fisher-Bingham 分布の尤度関数  $L$  の  $-\frac{1}{N}$  乗は、

$$L(x, y, 1)^{-\frac{1}{N}} = \exp(-(s_{11}x_{11} + s_{12}x_{12} + s_{22}x_{22} + s_1y_1 + s_2y_2))Z(x, y, 1)$$

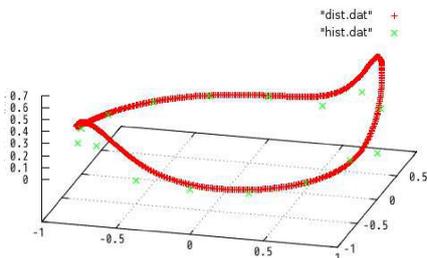
となる。ここで  $s_{ij}$  は、

$$s_{11} = 0.662125, s_{12} = 0.274563, s_{22} = 0.337875, s_1 = 0.317564, s_2 = -0.020188$$

という値で、これは上のデータより計算される値である。ホロノミック勾配降下法により  $L^{-\frac{1}{N}}$  の極小値 (近似値) を求めてやれば、

$$(x, y, 1) = (-0.616669, 3.46646, -2.23333, 0.918053, -0.834519, 1)$$

において、極小値 3.43216 をとることが分かる。すなわち、尤度関数  $L$  は上の点において極大値をとる。次の図は、上のデータと最尤推定で得られた Fisher-Bingham 分布を図示したものである。



このようなパラメータつき積分についての最適化問題を解く方法の一つとして、ホロノミック勾配降下法がある ([5])。次のような手順で実行される。

1. 目的関数の満たすホロノミック微分方程式系を計算する
2. ホロノミック微分方程式系を Pfaff 系に変換する
3. Pfaff 系を用いて、勾配降下法を実行し最適値を求める

特に Fisher-Bingham 分布の最尤推定の場合には、目的関数は Fisher-Bingham 分布の尤度関数であり、その尤度関数の満たすホロノミック微分方程式系は理論的に分かっている。あと必要になるのは、ホロノミック微分方程式系を Pfaff 系に変換すること、勾配降下法を実行する際の開始点における尤度関数の値の計算である。以下では説明を簡単にするために、目的関数を尤度関数の代わりに Fisher-Bingham 積分として、ホロノミック勾配降下法を説明する。

### 3 Fisher-Bingham 積分についての Pfaff 系の計算

Pfaff 系の計算は次のような微分作用素環において行われる。

#### 定義 3 (有理関数係数微分作用素環)

有理関数係数微分作用素環を  $R = \mathbb{C}(x_1, \dots, x_n)(\partial_1, \dots, \partial_n)$  とおく。ここで  $\partial_i$  は  $\frac{\partial}{\partial x_i}$  で  $x_i$  についての微分作用素を表す。左イデアル  $I \subset R$  が 0 次元イデアルとは、 $R/I$  が  $\mathbb{C}(x)$  ベクトル空間として有限次元の時をいう。0 次元イデアル  $I \subset R$  のホロノミックランクを、 $R/I$  の  $\mathbb{C}(x)$  ベクトル空間の次元 ( $\dim_{\mathbb{C}(x)} R/I$ ) として定義する。

**定義 4 (Pfaff 系)**

左イデアル  $I \subset R$  の Pfaff 系を

$$\partial_i F \equiv A_i F \pmod{I} \quad (i = 1, \dots, n)$$

で定義する. ここで,  $F = (P_1, \dots, P_s)^T$ ,  $P_i$  は  $\mathbb{C}(x)$  ベクトル空間  $R/I$  の基底とする.  $A_i$  は有理関数  $\mathbb{C}(x)$  を成分とする  $s \times s$  行列で積分可能条件

$$\frac{\partial A_i}{\partial x_j} + A_i A_j = \frac{\partial A_j}{\partial x_i} + A_j A_i \quad (1 \leq i, j \leq n)$$

を満たすものとする.

Pfaff 系について,

$$\partial_i F = A_i F \quad (i = 1, \dots, n), \quad F(a) = C \quad (a \text{ は点}, C \text{ は定数行列})$$

を満たす行列値べき級数  $F$  が一意に存在することを示すことができる.

**例 2 (Pfaff 系)**

1 変数の有理関数係数微分作用素環  $R = \mathbb{C}(x)\langle \partial \rangle$  とする.  $I = \langle P = \partial^3 + 2x^2\partial^2 + 3x\partial + 4 \rangle$  とすると,  $R/I \cong \mathbb{C}(x) + \mathbb{C}(x)\partial + \mathbb{C}(x)\partial^2$  となるので, その基底たちをとり  $F = (1, \partial, \partial^2)^T$  とおく.  $I$  の Pfaff 系は,

$$\partial \begin{pmatrix} 1 \\ \partial \\ \partial^2 \end{pmatrix} \equiv \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -4 & -3x & -2x^2 \end{pmatrix} \begin{pmatrix} 1 \\ \partial \\ \partial^2 \end{pmatrix} \pmod{I}$$

$R$  が 1 変数の場合, Pfaff 系への書き換えは, 高階の常微分方程式を 1 階の微分方程式系に書き換えることにほかならない.  $f$  が  $P \cdot f = (\partial^3 + 2x^2\partial^2 + 3x\partial + 4) \cdot f = 0$  を満たす関数としたら,

$$\partial \begin{pmatrix} f \\ \partial \cdot f \\ \partial^2 \cdot f \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -4 & -3x & -2x^2 \end{pmatrix} \begin{pmatrix} f \\ \partial \cdot f \\ \partial^2 \cdot f \end{pmatrix}$$

を満たす.

Pfaff 系をグレブナー基底を用いて計算するアルゴリズムがある. アルゴリズムの詳細は [5] を参照.

**定理 5 ([5])**

左イデアル  $I \subset R$  が 0 次元イデアルである時,  $I$  のグレブナー基底を使い, Pfaff 系をアルゴリズム的に計算できる.

**定理 6 (Fisher-Bingham 積分の満たすホロノミック系)**

微分作用素環  $R = \mathbb{C}(x, y, r)\langle \partial_x, \partial_y, \partial_r \rangle$  の左イデアルで次の元たちで生成されるものを  $I$  とおく.

$$\begin{aligned} A_{ij} &= \partial_{ij} - \partial_i \partial_j \quad (1 \leq i \leq j \leq n+1), \\ B &= \sum_{i=1}^{n+1} \partial_i^2 - r^2, \\ C_{ij} &= x_{ij} \partial_i^2 + 2(x_{jj} - x_{ii}) \partial_i \partial_j - x_{ij} \partial_j^2 \\ &\quad + \sum_{1 \leq k \leq n+1, k \neq i, j} (x_{kj} \partial_{ik} - x_{ik} \partial_{jk}) + y_j \partial_i - y_i \partial_j \quad (1 \leq i < j \leq n+1), \\ E &= r \partial_r - 2 \sum_{1 \leq i \leq j \leq n+1} x_{ij} \partial_{ij} - \sum_{i=1}^{n+1} y_i \partial_i - n. \end{aligned}$$

ただし,  $\partial_{ij} = \frac{\partial}{\partial x_{ij}}, \partial_i = \frac{\partial}{\partial y_i}$  とする. この左イデアル  $I$  について, 次のことがわかる.

1.  $I$  は Fisher-Bingham 積分  $Z(x, y, r)$  を消去する. すなわち,  $I$  は Fisher-Bingham 積分の満たす微分方程式系 ([5]).
2.  $I$  はホロノミックイデアルである ([1]).
3.  $R/I$  は  $\mathbb{C}(x, y, r)$  ベクトル空間として

$$1, \partial_1, \dots, \partial_{n+1}, \partial_1^2, \dots, \partial_n^2$$

で張られる. 特に  $I$  のホロノミックランクは  $2n+2$  以下である ([5]).

4.  $I$  のホロノミックランクは  $2n+2$  ([3]).

5.  $R/I$  の  $\mathbb{C}(x, y, r)$  ベクトル空間として基底は

$$1, \partial_1, \dots, \partial_{n+1}, \partial_1^2, \dots, \partial_n^2$$

である ([3]).

この微分方程式系  $I$  より, Fisher-Bingham 積分の満たす Pfaff 系を, グレブナー基底を使ったアルゴリズム的な方法 (定理 5) で求めることができるが, 計算量の問題により高々  $n=2$  次元までしか Pfaff 系を計算できない.  $n=1$  次元の場合に Pfaff 系の計算を, グレブナー基底を使った方法と, 使わない方法の 2 つに分けて説明する.

### 例 3 (1 次元 Fisher-Bingham 積分の満たす Pfaff 系の計算)

1 次元 Fisher-Bingham 積分  $Z(x_{11}, x_{12}, x_{22}, y_1, y_2, r)$  の満たすホロノミック系は,

$$A_{11} = \partial_{11} - \partial_1^2, A_{12} = \partial_{12} - \partial_1 \partial_2, A_{22} = \partial_{22} - \partial_2^2$$

$$B = \partial_1^2 + \partial_2^2 - r^2$$

$$C_{12} = x_{12} \partial_1^2 - 2(x_{22} - x_{11}) \partial_1 \partial_2 - x_{12} \partial_2^2 + y_2 \partial_1 - y_1 \partial_2$$

$$E = r \partial_r - 2(x_{11} \partial_1^2 + x_{12} \partial_1 \partial_2 + x_{22} \partial_2^2) - (y_1 \partial_1 + y_2 \partial_2) - 1$$

であった (定理 6). これらの生成する  $R$  のイデアルを  $I$  とおく.

#### 1. $I$ のグレブナー基底 $G$ がわかっている場合の Pfaff 系の計算方法

$I$  の単項式順序  $<$  についてのグレブナー基底  $G$  が計算できたとする. グレブナー基底  $G$  より得られる  $R/I$  の  $\mathbb{C}(x, y, r)$  ベクトル空間としての基底が

$$1, \partial_1, \partial_2, \partial_1^2$$

であったとする.  $y_1$  の Pfaff 系を求めるには,

$$\partial_1 \begin{pmatrix} 1 \\ \partial_1 \\ \partial_2 \\ \partial_1^2 \end{pmatrix} \equiv \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} \begin{pmatrix} 1 \\ \partial_1 \\ \partial_2 \\ \partial_1^2 \end{pmatrix} \pmod{I}$$

が成り立つような行列  $A = (a_{ij})$  を求めればよい. ここで  $a_{ij} \in \mathbb{C}(x, y, r)$  である. 1, 2 行目は自明な関係式なので, 3, 4 行目の書き換えだけが問題である.  $\partial_1 \partial_2$  を  $G$  で割った余りと  $\partial_1^3$  を  $G$  で割った余りから, 3, 4 行目の関係式が得られる.

2.  $I$  のグレブナー基底を計算せず Pfaff 系を導く方法

$R/I$  の有理関数  $\mathbb{C}(x, y, r)$  ベクトル空間としての基底は

$$1, \partial_1, \partial_2, \partial_1^2$$

とわかっているので (定理 6),  $y_1$  の Pfaff 系を求めるには,

$$\partial_1 \begin{pmatrix} 1 \\ \partial_1 \\ \partial_2 \\ \partial_1^2 \end{pmatrix} \equiv \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} \begin{pmatrix} 1 \\ \partial_1 \\ \partial_2 \\ \partial_1^2 \end{pmatrix} \pmod{I}$$

が成り立つ行列  $A = (a_{ij})$  を求めればよい. ここで  $a_{ij} \in \mathbb{C}(x, y, r)$  である. 1, 2 行目は自明な関係式なので, 3, 4 行目の書き換えが問題になる.

$$\begin{aligned} \partial_1 \partial_2 &\equiv a_{31} + a_{32} \partial_1 + a_{33} \partial_2 + a_{34} \partial_1^2 \\ \partial_1^3 &\equiv a_{41} + a_{42} \partial_1 + a_{43} \partial_2 + a_{44} \partial_1^2 \end{aligned}$$

なる有理関数  $a_{ij}$  を計算することを考える.  $I$  の元

$$\begin{aligned} C_{12} &= x_{12} \partial_1^2 - 2(x_{22} - x_{11}) \partial_1 \partial_2 - x_{12} \partial_2^2 + y_2 \partial_1 - y_1 \partial_2 \\ B &= \partial_1^2 + \partial_2^2 - r^2 \end{aligned}$$

を使い, 書き換えを行う.  $C_{12}$  より,

$$\begin{aligned} 2(x_{22} - x_{11}) \partial_1 \partial_2 &\equiv x_{12} \partial_1^2 - x_{12} \partial_2^2 + y_2 \partial_1 - y_1 \partial_2 \\ &\equiv x_{12} \partial_1^2 - x_{12} (-\partial_1^2 + r^2) + y_2 \partial_1 - y_1 \partial_2 \quad (B \text{ で書き換え}) \\ &\equiv -x_{12} r^2 + y_2 \partial_1 - y_1 \partial_2 + 2x_{12} \partial_1^2 \end{aligned} \tag{1}$$

この式より  $a_{31}, a_{32}, a_{33}, a_{34}$  が定まる. 次のような 2 つの  $I$  の元を考える.

$$\begin{aligned} \partial_1 C_{12} &= x_{12} \partial_1^3 + 2(x_{22} - x_{11}) \partial_1^2 \partial_2 - x_{12} \partial_1 \partial_2^2 + y_2 \partial_1^2 - y_1 \partial_1 \partial_2 - \partial_2 \\ \partial_2 C_{12} &= x_{12} \partial_1^2 \partial_2 + 2(x_{22} - x_{11}) \partial_1 \partial_2^2 - x_{12} \partial_2^3 + y_2 \partial_1 \partial_2 + \partial_1 - y_1 \partial_2^2 \end{aligned}$$

$\partial_1 C_{12}$  より,

$$\begin{aligned} x_{12} \partial_1^3 + 2(x_{22} - x_{11}) \partial_1^2 \partial_2 &\equiv x_{12} \partial_1 \partial_2^2 - y_2 \partial_1^2 + y_1 \partial_1 \partial_2 - \partial_2 \\ &\equiv x_{12} \partial_1 (-\partial_1^2 + r^2) - y_2 \partial_1^2 + y_1 \partial_1 \partial_2 - \partial_2 \quad (B \text{ で書き換え}) \\ 2x_{12} \partial_1^3 + 2(x_{22} - x_{11}) \partial_1^2 \partial_2 &\equiv a + b \partial_1 + c \partial_2 + d \partial_1^2 \quad ((1) \text{ で書き換え}) \end{aligned}$$

ここで  $a, b, c, d$  は具体的に定まる有理関数である. 同様の計算を  $\partial_2 C_{12}$  に行えば,

$$-2(x_{22} - x_{11}) \partial_1^3 + 2x_{12} \partial_1^2 \partial_2 \equiv a' + b' \partial_1 + c' \partial_2 + d' \partial_1^2$$

が得られる. ここで  $a', b', c', d'$  は具体的に定まる有理関数である. こうして得られた 2 式を行列表示すれば,

$$\begin{pmatrix} 2x_{12} & 2(x_{22} - x_{11}) \\ -2(x_{22} - x_{11}) & 2x_{12} \end{pmatrix} \begin{pmatrix} \partial_1^3 \\ \partial_1^2 \partial_2 \end{pmatrix} \equiv \begin{pmatrix} a & b & c & d \\ a' & b' & c' & d' \end{pmatrix} \begin{pmatrix} 1 \\ \partial_1 \\ \partial_2 \\ \partial_1^2 \end{pmatrix}$$

となり, 左辺に現れる行列の逆行列を計算してやれば,  $\partial_1^3 \equiv a_{41} + a_{42}\partial_1 + a_{43}\partial_2 + a_{44}\partial_1^2$  が得られる. このような計算法により, 1次元 Fisher-Bingham 積分の満たす  $y_1$  についての Pfaff 系は

$$\partial_1 F \equiv \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ \frac{-r^2 x_{12}}{2(x_{11}-x_{22})} & \frac{y_2}{2(x_{11}-x_{22})} & \frac{-y_1}{2(x_{11}-x_{22})} & \frac{x_{12}}{x_{11}-x_{22}} \\ a'' & b'' & c'' & d'' \end{pmatrix} F$$

$$\begin{aligned} a'' &= \frac{r^2((x_{11}-x_{22})x_{12}v_2 + (-x_{12}^2 + 2x_{11}^2 - 4x_{22}x_{11} + 2x_{22}^2)v_1)}{4(x_{11}-x_{22})(x_{12}^2 + x_{11}^2 - 2x_{22}x_{11} + x_{22}^2)} \\ b'' &= \frac{-((x_{11}-x_{22})v_2^2 - x_{12}v_1v_2 + (-2r^2x_{11} + 2x_{22}r^2)x_{12}^2 - 4r^2x_{11}^3 + (12x_{22}r^2 + 2)x_{11}^2 + (-12x_{22}^2r^2 - 4x_{22})x_{11} + 4x_{22}^3r^2 + 2x_{22}^2)}{4(x_{11}-x_{22})(x_{12}^2 + x_{11}^2 - 2x_{22}x_{11} + x_{22}^2)} \\ c'' &= \frac{(x_{11}-x_{22})v_1v_2 - x_{12}v_1^2 + (2r^2x_{11}^2 + (-4x_{22}r^2 + 2)x_{11} + 2x_{22}r^2 - 2x_{22})x_{12}}{4(x_{11}-x_{22})(x_{12}^2 + x_{11}^2 - 2x_{22}x_{11} + x_{22}^2)} \\ d'' &= \frac{-((2x_{11}-2x_{22})x_{12}v_2 + (-x_{12}^2 + x_{11}^2 - 2x_{22}x_{11} + x_{22}^2)v_1)}{2(x_{11}-x_{22})(x_{12}^2 + x_{11}^2 - 2x_{22}x_{11} + x_{22}^2)} \end{aligned}$$

となることがわかる. 1次元の場合でもこのように大きな Pfaff 系になる.

例 3 のグレブナー基底を計算せずに Pfaff 系を求める方法は, 高次元の場合にも同様に適用できる. こうして次の結果が得られる.

#### 定理 7 ([2])

Fisher-Bingham 積分  $Z(x, y, r)$  の満たすホロノミック系のなすイデアルを  $I$  とする.  $I$  の各変数について次のような形で Pfaff 系を明示的に得ることができる. たとえば,  $y_i$  については,

$$A\partial_i F \equiv PF \pmod{I}$$

ここで,  $A$  は多項式を成分に持つ行列,  $P$  は有理関数を成分に持つ行列,  $A$  は正則行列である. ( $A, P$  の具体形については [2] を参照)

## 4 Fisher-Bingham 積分の評価

勾配降下法を行うには, ある初期点を選び, そこでの Fisher-Bingham 積分  $Z$  とその偏微分の値を知る必要がある. 例えば  $n=1$  次元の場合には,

$$F \cdot Z = (Z, \partial_1 \cdot Z, \partial_2 \cdot Z, \partial_1^2 \cdot Z)^T$$

のある点での値が必要となる.

Fisher-Bingham 積分

$$Z(x, y, r) = \int_{S^n(r)} \exp \left( \sum_{1 \leq i \leq j \leq n+1} x_{ij} t_i t_j + \sum_{i=1}^{n+1} y_i t_i \right) |dt|.$$

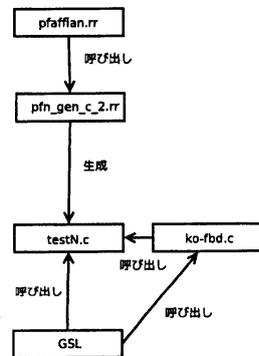
のある点での値を計算するには, 被積分関数の  $\exp$  をべき級数展開して, 項別積分をすればよい. 各項の積分の評価は単項式を球面上で積分した結果の公式を使う. こうして得られたべき級数は  $r$  が大きい所では収束が遅いので,  $r$  の小さい所でべき級数を計算し, ホロノミック勾配法 (Holonomic Gradient Method) で  $r$  が大きい所に持っていく ([6] の手法). 詳細については, [2] を参照.

## 5 ホロノミック勾配降下法の実装

Fisher-Bingham 分布に対するホロノミック勾配降下法の実装は、次のプログラムからなる。(プログラムと計算結果は <http://www.math.kobe-u.ac.jp/OpenXM/Math/Fisher-Bingham-2/> より入手できる.)

- ko-fbd.c — Fisher-Bingham 積分の数値評価 (初期値計算) を行う C プログラム
- pfaffian.rr — Fisher-Bingham 積分の Pfaff 系の計算 (有理関数を成分に持つ行列を返す) する Asir のプログラム
- pfn\_gen\_c\_2.rr — 勾配降下法を行う C プログラム (testN.c (N は次元を表す数値)) を生成する Asir のプログラム
- testN.c — 勾配降下法を行う C プログラム

数式処理部分は、数式処理ソフト Risa/Asir で計算させて、数値計算部分は C プログラムから GSL (Gnu Scientific Library) を呼び出すことで計算させている。これらプログラムの間の関係は次のようになる。



### 例 4 (自動生成された C のプログラム test1.c の例)

test1.c は、1 次元 Fisher-Bingham 分布の最尤推定を行うプログラムである。機械的に生成されるものなので、巨大なプログラムで決して読みやすいものではない。例えば、次はある点での  $y_1$  についての Pfaff 系に現れる行列の値を計算する部分である。

```

void pf1(double x11, double x12, double x22, double y1, double y2, double r)
{
  extern gsl_matrix *a1, *inv_a1, *b1, *c1, *e1;
  extern gsl_matrix *p2, *inv_p2, *q2;
  extern gsl_matrix *p3, *inv_p3, *q3, *r3;
  extern gsl_matrix *pfi_m;
  extern double g_s1;

  set_abca_1(x11, x12, x22, y1, y2, r);
  set_pqr(x11, x12, x22, y1, y2, r);
  invmat(a1, inv_a1);
  invmat(p2, inv_p2);
  invmat(p3, inv_p3);

  gsl_matrix *t1 = gsl_matrix_alloc(inv_p2->size1, q2->size2);
  gsl_blas_dgemm(CblasNoTrans, CblasNoTrans, 1.0, inv_p2, q2, 0.0, t1);
  gsl_matrix *t2 = gsl_matrix_alloc(c1->size1, t1->size2);
  gsl_blas_dgemm(CblasNoTrans, CblasNoTrans, 1.0, c1, t1, 0.0, t2);
  gsl_matrix *t3 = gsl_matrix_alloc(q3->size1, t1->size2);
  gsl_blas_dgemm(CblasNoTrans, CblasNoTrans, 1.0, q3, t1, 0.0, t3);
  gsl_matrix_sub(t3, r3);
  gsl_matrix *t4 = gsl_matrix_alloc(inv_p3->size1, t3->size2);
  gsl_blas_dgemm(CblasNoTrans, CblasNoTrans, 1.0, inv_p3, t3, 0.0, t4);
  gsl_matrix *t5 = gsl_matrix_alloc(e1->size1, t4->size2);
  gsl_blas_dgemm(CblasNoTrans, CblasNoTrans, 1.0, e1, t4, 0.0, t5);
  gsl_matrix_add(t5, b1);
  gsl_matrix_sub(t5, t2);
  gsl_blas_dgemm(CblasNoTrans, CblasNoTrans, 1.0, inv_a1, t5, 0.0, t2);
  gsl_matrix_memcpy(pfi_m, t2);
}

```

```

gsl_matrix_set_identity(t2);
gsl_matrix_scale(t2, -g_sl);
gsl_matrix_add(pf1_m, t2);

gsl_matrix_free(t1);
gsl_matrix_free(t2);
gsl_matrix_free(t3);
gsl_matrix_free(t4);
gsl_matrix_free(t5);
}

```

このプログラムで行われる数値計算は、行列の逆行列の計算、連立 1 階の微分方程式系を数値的に解く (Runge-Kutta 法)、勾配降下法により最適値を求めるなどがあるが、これらは *GSL(Gnu Scientific Library)* を呼び出すことにより実行している。GSL を使う利点には次のようなものがある。

- 少ない命令で、ブラックボックス的に用いることができる。
- 数値計算の非専門家が自分でコードを書くよりも高速で精度が良い。
- アルゴリズムも簡単に選択できる。たとえば、微分方程式を数値的に解く *Runge-Kutta* 法があるが、差分スキームのとり方で幾つかバリエーションがあるが次のような 1 文で選択可能。

```
#define ODEIV_STEP_TYPE gsl_odeiv_step_rkf45 /* 刻み幅制御の Runge-Kutta Fehlberg 法 */
```

ホロノミック勾配降下法の勾配降下法の部分は、次のように行われる。



### 例 5 ( $n = 1$ 次元の場合の実行例 (神戸の風向データの例))

例 1 のデータに関して、プログラムを実行すると次のような結果が得られる。

```

[5.0484, 2.43327, 0.566476, 2.99465] <--- 初期点での値 (FB 積分を繰返開した値より出したもの)

1, 0.502267, 0.506488, 0.147743, 0.994617, 0.145672, 5.03304
points = [0.502267, 0.506488, 0.147743, 0.994617, 0.145672] <--- 初期点
values = [5.03304, 2.41783, 0.56737, 2.98617] <--- 初期点での値
grad : -0.346338 -0.996036 0.346338 0.819518 0.658977 <--- この点での gradient
norm(grad): 1.529002 <--- gradient のノルム
gradient 方向に移動 (Runge-Kutta)

2, 0.506771, 0.519465, 0.143229, 0.98385, 0.137015, 5.00272 <--- 次の点
points = [0.506771, 0.519465, 0.143229, 0.98385, 0.137015] <--- 次の点
values = [5.00272, 2.38703, 0.541216, 2.96941] <--- 次の点での値
grad : -0.343022 -0.987293 0.343022 0.798350 0.642210 <--- この点での gradient
norm(grad): 1.503287 <--- gradient のノルム

3, 0.515799, 0.545417, 0.134201, 0.982316, 0.119702, 4.94363
points = [0.515799, 0.545417, 0.134201, 0.982316, 0.119702]
values = [4.94363, 2.32576, 0.509122, 2.83671]
grad : -0.336588 -0.970034 0.336588 0.755845 0.608924
norm(grad): 1.452460

繰り返し ... 略

Minimum found
33, 1.1462, 3.58647, -0.496196, 0.951055, -0.882713, 3.43639 <--- 極小点 (近似)
points = [1.1462, 3.58647, -0.496196, 0.951055, -0.882713]
values = [3.43639, 1.09163, -0.0690454, 2.2753]
grad : -0.000018 -0.000321 0.000018 0.000357 0.000328 <--- gradient のノルムが十分小さくなったので終了
norm(grad): 0.000582
search_min :
[3.43639, 1.09163, -0.0690454, 2.2753] <--- 極小値 (近似)

```

通常の勾配降下法とホロノミック勾配降下法は次のような点異なる。

- 通常の勾配降下法であれば、目的関数が明示的に与えられており、任意の点でその値を評価し、gradientも容易に計算できる。
- ホロノミック勾配降下法では、目的関数が明示的に与えられておらず、代わりに目的関数の満たす微分方程式系 (Pfaff 系) とある点での目的関数の値が与えられている。他の点で目的関数の値を得るためには、Pfaff 系を数値的に解いてある点での目的関数の値から他の点での目的関数の値を近似的に計算する。ある点での gradient も Pfaff 系から計算できる。

方法	入力	目的関数の評価	勾配の計算
勾配降下法	目的関数 (明示的に与えられている)	容易	容易
ホロノミック勾配降下法	目的関数の満たす Pfaff 系 ある点での目的関数の値	Pfaff 系より計算	Pfaff 系より計算

さらに Fisher-Bingham 積分を目的関数とする場合、定理 7 より Pfaff 系は

$$A\partial_i F \equiv PF \pmod{I} \quad (A \text{ は各成分が多項式の行列}, P \text{ は各成分が有理関数の行列})$$

の形で与えられているので、 $\partial_i F = A^{-1}PF$  の形に直す必要がある。すなわち、 $A$  の逆行列を計算する必要がある。ただしこの逆行列の計算は前もって symbolic に計算するのではなく、Pfaff 系を解く必要が生じたごとに数値的に逆行列を計算している。

- symbolic に計算すると確かに最初に一回だけ逆行列を計算するだけで済むが、非常に時間のかかる計算であり、次元が  $n = 2$  までしか計算できない。また逆行列が得られたとしても非常に巨大なものであり、数値を代入して評価するにも時間がかかることになる。
- 今の場合、Pfaff 系を解く必要が生じたごとに数値的に逆行列  $A^{-1}$  を計算してこの計算の困難を回避している。ただし、複数回 (Runge-Kutta 法のステップ数分) 逆行列を数値計算する必要が生じる。こうすることで、次元が  $n = 7$  までの計算を実行することができた。

こうして Fisher-Bingham 積分の最尤推定に関して、あるデータに関して  $n = 7$  次元まで計算を実行させることができた。計算時間は以下ようになった。

次元	初期値計算	勾配降下法部分 (ステップ数)	Total
3	0.016	5.5 (100 steps)	5.5
4	0.128	18 (100 steps)	18
5	0.624	56 (100 steps)	57
6	2.9	168 (100 steps)	171
7	13	486 (100 steps)	499

(単位は秒、勾配降下法の終了条件は、ステップ数が 100 に到達したら終了するようにしている)

#### 定理 8 (計算量, [2])

得られたアルゴリズムに関して、 $n$  次元 Fisher-Bingham 分布の最尤推定にかかる計算量は、

$$\begin{aligned} & (\text{初期値計算にかかる計算量}) + (\text{勾配降下法部分の計算量}) \\ & = (\text{初期値計算にかかる計算量}) + O((2n+2)^3) \times (\text{Runge-Kutta 法のステップ数}) \end{aligned}$$

となる。ここで、 $2n+2$  は Pfaff 系に現れる行列のサイズであり、 $O((2n+2)^3)$  はサイズ  $(2n+2) \times (2n+2)$  の行列の逆行列の計算のコストから来るものである。

## 参 考 文 献

- [1] T. Koyama : A Holonomic Ideal Annihilating the Fisher-Bingham Integral, preprint, arXiv:1104.1411
- [2] T. Koyama, H. Nakayama, K. Nishiyama, N. Takayama : Holonomic Gradient Descent for the Fisher-Bingham Distribution on the  $n$ -dimensional Sphere, preprint, arXiv:1201.3239
- [3] T. Koyama, H. Nakayama, K. Nishiyama, N. Takayama : The Holonomic Rank of the Fisher-Bingham System of Differential Equations, preprint, arXiv:1205.6144
- [4] K. Kume, A. T. A. Wood : Saddlepoint approximations for the Bingham and Fisher-Bingham normalising constants, *Biometrika* **92** (2005), 465–476.
- [5] H. Nakayama, K. Nishiyama, M. Noro, K. Ohara, T. Sei, N. Takayama, A. Takemura : Holonomic Gradient Descent and its Application to the Fisher-Bingham Integral, *Advances in Applied Mathematics* **47** (2011), 639–658
- [6] Hashiguchi, Numata, N. Takayama, A. Takemura : Holonomic Gradient Method for the Distribution Function of the Largest Root of Wishart Matrix, preprint, arXiv:1201.0472
- [7] Risa/Asir, a computer algebra system,  
<http://www.math.kobe-u.ac.jp/Asir>
- [8] GNU Scientific Library,  
<http://www.gnu.org/software/gsl/>