

# ホロノミック勾配法 (HGM, Holonomic Gradient Method)

高山信毅 (神戸大学)

参考文献等 [hgm OpenXM search](#).

(ある点  $x = a$  の近傍でなめらかな) 多変数関数  
 $f(x) = f(x_1, \dots, x_n)$ . 全方向で多項式係数の常微分方程式

$$\sum_{k=0}^{r_i} a_{ik}(x) \partial_i^k \bullet f = 0, \quad (i = 1, \dots, n) \quad (1)$$

を満す.  $a_{ik} : x = (x_1, \dots, x_n)$  の多項式.  $\partial_i = \frac{\partial}{\partial x_i}$ ,  $\partial_i^k \bullet f = \frac{\partial^k f}{\partial x_i^k}$ .  
このような性質を満す多変数関数 (の解析接続) を, **holonomic 解析関数**と呼ぶ.

例:  $f = \exp(-x_2 x_1^2)$  は 2 変数の holonomic 解析関数. 実際,  
 $\partial_1 \bullet f = -2x_1 x_2 f$ ,  $\partial_2 \bullet f = -x_1^2 f$ .

(超関数を含む一般の) Holonomic 関数  $f$  は, 全方向で常微分方程式 (1) をみたしさらにもう少し沢山の多項式係数偏微分方程式系を満す関数.

例:  $H(x) = 1 (x \geq 0), H(x) = 0 (x < 0)$  を Heaviside 関数とする.  
 $x\partial_x \bullet H = 0. \exp(-x_2x_1^2)H(x_1)$  は holonomic 関数.

Holonomic 関数  $f$  を零化する線形偏微分作用素全部の集合は **holonomic ideal (system)**.

holonomic 系を特殊関数の代わりにしよう (佐藤幹夫の哲学, 1970 年代).  
精密に定義するには, イデアル, Hilbert 関数, 等の概念.



## Holonomic 関数の性質

- ①  $f, g$  が holonomic 関数なら,  $f \pm g, fg$  (定義できれば) も holonomic 関数.
- ②  $f(\theta, t)$  を  $n + m$  個の変数  $\theta = (\theta_1, \dots, \theta_n), t = (t_1, \dots, t_m)$  についての holonomic 関数. パラメータ付き積分 (例:  $f$  が (非正規化) 分布密度関数なら  $Z$  は正規化定数)

$$Z(\theta) = \int_{\mathbf{R}^m} f(\theta, t) dt_1 \cdots dt_m$$

は (積分についてのある technical な条件のもと)  $n$  変数の holonomic 関数.

証明は  $D$ -加群の理論. 微分方程式系を導出するアルゴリズムもあり (Zeilberger 1990, Oaku 1997, ...).

例:  $Z(\theta) = \int_{\mathbf{R}} \exp(-\theta t^2) H(t) dt$  は holonomic 関数. 実際,  $(2\theta \partial_\theta + 1) \bullet Z = 0$ .

練習:  $\int_{\mathbf{R}} \exp(-y(x-1)^2) H(x) dx$  では?

[ [(Dy+(x-1)^2) (x\*Dx+2\*x\*(x-1)\*y)] [(x)] [[(x) (y)] []] 0]  
integration ::

[ 2 , [ -2\*e\_ \*y-2\*y\*Dy-2\*y-1 , -2\*y\*Dy^2-2\*y\*Dy-3\*Dy-1 ] ]

練習 以下の中で holonomic 関数はどれ?

- ①  $\exp(f(x))$  where  $f$  is a rational function,
- ②  $\frac{1}{\sin x}$  [Hint] Use Th: Any solution of the ordinary differential equation  $(a_m(x)\partial^m + \cdots + a_0(x)) \bullet f = 0$ ,  $a_i \in \mathbf{C}[x]$ , is holomorphic out of the singular locus  $\{x \mid a_m(x) = 0\}$ .
- ③  $\Gamma(x)$ , [Hint]  $\Gamma(x)$  has poles at  $x = -n$ ,  $n \in \mathbf{N}_0$ .
- ④  $2^x$ ,
- ⑤  $H(x)$  (Heaviside function),
- ⑥  $x^a$  where  $a$  is a constant,
- ⑦  $|x|$ ,
- ⑧  $\int_{-\infty}^{+\infty} \exp(-xt^6 - t) dt$ ,  $x > 0$ .

**Holonomic 分布:** ホロノミック関数で (非正規化) 分布密度関数を書ける分布.  $f(\theta, t)/Z(\theta)$  が  $t$  空間での分布密度関数.

Holonomic 関数の差分版. 差分作用素  $E_i : E_i \bullet f(y_i) = f(y_i + 1)$ .

Mellin 変換:  $E_i \mapsto x_i, y_i \mapsto -x_i \partial_i - 1 = -\partial_i x_i$ .

$$F(x_i) \mapsto G(y_i) = \int_0^\infty F(x_i) x_i^{y_i} dx_i$$

$f(y_1, \dots, y_n)$  が差分版 holonomic 関数 (holonomic 数列) とは,  $f$  の満す線形差分作用素達をそれぞれ Mellin 変換したとき, holonomic ideal をなすこと.

この時  $f$  は全方向に多項式係数の線形漸化式を満す  $+\alpha$ .

例:  $Z(n; p) = \sum_{u_1+u_2=n, u_i \geq 0} \frac{p_1^{u_1} p_2^{u_2}}{u_1! u_2!}$  とおく.

$n!Z(n; p) = \frac{t^{-n-1}}{1 - p_1 t - p_2 t}$  を  $t$  で展開したときの  $t^{-1}$  の係数

$$Z(n; p) = \frac{1}{n! 2\pi i} \int_C f(t, p_1, p_2) dt, \quad f = \frac{t^{-n-1}}{1 - p_1 t - p_2 t}$$

$E_n \bullet f(n, p) = t f(n, p), \dots, \Rightarrow [(n+1)E_n - (p_1 + p_2)] \bullet Z(n; p) = 0$   
( $Z$  は二項展開).

例: 離散 A-超幾何分布 (予稿 7 節) は  $\beta$  について差分 holonomic 関数.

## Pfaffian 方程式とは?

$$\left[ a_2(x) \frac{d^2}{dx^2} + a_1(x) \frac{d}{dx} + a_0(x) \right] \bullet f = 0$$

$F = \begin{pmatrix} f \\ \frac{df}{dx} \end{pmatrix}$  とおくと,  $\frac{d}{dx} \bullet F = \begin{pmatrix} 0 & 1 \\ -a_0/a_2 & -a_1/a_2 \end{pmatrix} F$  を満たす. これの多変数化.

$\delta_i = \partial_i, i = 1, \dots, n. \alpha^1 = \mathbf{0} \in \mathbf{N}_0^n, \alpha^i \in \mathbf{N}_0^n, i = 2, \dots, r$  をうまく与えると,  $c$  基底ベクトル

$$F = \begin{pmatrix} \delta^{\alpha^1} \bullet Z \\ \delta^{\alpha^2} \bullet Z \\ \vdots \\ \delta^{\alpha^r} \bullet Z \end{pmatrix}$$

に対して,

$$\partial_i \bullet F = P_i F, i = 1, \dots, n \quad \text{微分 Pfaffian 系}$$

となる  $r \times r$  行列  $P_i(x)$  が存在. この方程式を **Pfaffian (方程式) 系** とよぶ.

注: holonomic ideal から Pfaffian 系への変換はグレブナー基底で原理的には可能. yang.rr など.

例: Von Mises 分布.  $\exp(\theta_1 \cos(t) + \theta_2 \sin(t))/Z(\theta)$ ,  
 $Z(\theta) = \int_0^{2\pi} \exp(\theta_1 \cos(t) + \theta_2 \sin(t)) dt$ . (以下の式は数学セミナー  
の記事 [20] より, 予稿 3 節 Fisher-Bingham 分布).

$$\frac{\partial Z}{\partial \theta_1} = \begin{pmatrix} 0 & 1 \\ \frac{\theta_1^2}{\theta_1^2 + \theta_2^2} & \frac{\theta_2^2 - \theta_1^2}{\theta_1(\theta_1^2 + \theta_2^2)} \end{pmatrix} Z,$$
$$\frac{\partial Z}{\partial \theta_2} = \begin{pmatrix} 0 & \frac{\theta_2}{\theta_1} \\ \frac{\theta_1 \theta_2}{\theta_1^2 + \theta_2^2} & \frac{-2\theta_2}{\theta_1^2 + \theta_2^2} \end{pmatrix} Z$$

ここで  $Z = (Z, \partial_1 \bullet Z)^T$ ,  $\partial_i = \partial/\partial\theta_i$  である.

例,  $2 \times 2$  分割表, 予稿 7 節 離散  $A$  超幾何分布

$$A = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}, Z(\beta; x) = \sum_{Au=\beta, u \in \mathbf{N}_0^n} \frac{x^u}{u!}$$

$$F = (Z, E_3 \bullet Z)^T, D = x_1 x_4 - x_2 x_3, B = \beta_1 - \beta_2 - \beta_3,$$

$$E_1 - \frac{1}{x_4} \begin{pmatrix} \frac{-\beta_1 x_1 x_4 + \beta_3 D}{(B-1)x_3} & \frac{-x_2 D}{(B-1)x_3} \\ \beta_3 & -x_2 \end{pmatrix}$$

$$E_2 - \frac{1}{x_1} \begin{pmatrix} -B & -x_2 \\ \frac{\beta_3 B x_3}{D} & \frac{\beta_2 x_1 x_4 + \beta_3 x_2 x_3}{D} \end{pmatrix}$$

$$Z(\beta = 0, x) = 1.$$

$$E_1 \bullet F(\beta_1, \beta_2, \beta_3; x) = F(\beta_1 + 1, \beta_2, \beta_3; x)$$



# HGM (Holonomic Gradient Method) の 3 ステップ

- 1 正規化定数  $Z$  に対する holonomic 系を求める. さらに holonomic 系を Pfaffian に変換する. 理論的方法または  $D$ -加群のアルゴリズムを用いる.
- 2 正規化定数とその偏微分または偏差分を必要な rank まである点で計算する. [c 基底ベクトルの初期値を求める.]
- 3 2 で求めた値を Pfaffian で必要な点まで延長する. [初期値問題を解く.]

注: rank  $r$  は step 3 の計算量を決める.

注: 最尤推定, e.g., [20] では,

$$\sum_{T \in \text{観測値の } N \text{ 個の集合}} \log f(\theta, T) - N \log Z(\theta) = \log \prod_T \frac{f(\theta, T)}{Z}$$

の最大化を  $\theta$  について行う.

c 基底ベクトル  $F$  の数値がわかれば,  $Z$  の Hessian やより高い階数の微分の値もわかる.

$$\frac{\partial^2 F}{\partial x_i \partial x_j} = \frac{\partial P_i(\beta, \rho)}{\partial x_j} F + P_i \frac{\partial F}{\partial x_j} = \left( \frac{\partial P_i(\beta, \rho)}{\partial x_j} + P_i P_j \right) F. \quad (2)$$

## Fisher-Bingham 分布 (予稿3節)

$d$ 次元球面  $S^d(r) = \{(u_1, \dots, u_{d+1}) \mid \sum_{i=1}^{d+1} u_i^2 = r^2, r > 0\}$  の上でパラメータ  $x_{ij}, y_i$  を持つ分布

$$\mu(u; x, y, r)\delta := \frac{1}{Z(x, y, r)} \exp \left( \sum_{1 \leq i < j \leq d+1} x_{ij} u_i u_j + \sum_{i=1}^{d+1} y_i u_i \right) \delta$$

を Fisher-Bingham 分布と呼ぶ (1980年代から始まる A.T.A. Wood 等の研究が先駆). ここで  $\delta = \delta(r, u)$  は

$\int_{\mathbf{R}^{d+1}} \delta(r, u) = r^d \frac{2\pi^{(d+1)/2}}{\Gamma((d+1)/2)}$  をみたす球面に台をもつデルタ関数,  $Z$  は次で定義される正規化定数

$$Z(x, y, r) = \int_{\mathbf{R}^{d+1}} \exp \left( \sum_{1 \leq i < j \leq d+1} x_{ij} u_i u_j + \sum_{i=1}^{d+1} y_i u_i \right) \delta(r, u). \quad (3)$$

(清, 竹村が発端. その後, 小山民雄ら)

# 例：神戸市の平均風向 (16 方位) のデータ

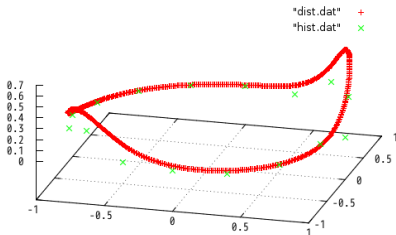
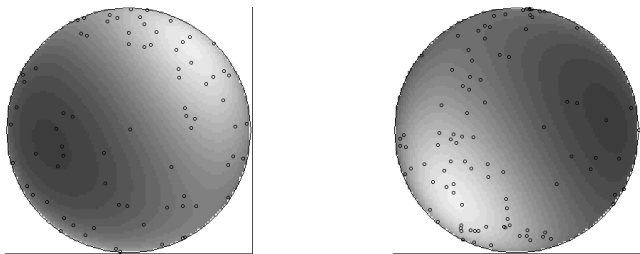


Figure: 頻度と密度関数

| 方位  | ラジアン     | 頻度   |
|-----|----------|------|
| 北   | 1.570760 | 0.00 |
| 北北東 | 1.178070 | 0.00 |
| 北東  | 0.785380 | 0.24 |
| 東北東 | 0.392690 | 0.24 |
| 東   | 0.000000 | 0.03 |
| 東南東 | 5.890349 | 0.12 |
| 南東  | 5.497659 | 0.03 |
| 南南東 | 5.104969 | 0.00 |
| 南   | 4.712279 | 0.02 |
| 南南西 | 4.319589 | 0.00 |
| 南西  | 3.926899 | 0.19 |
| 西南西 | 3.534209 | 0.07 |
| 西   | 3.141519 | 0.03 |
| 西北西 | 2.748829 | 0.00 |
| 北西  | 2.356139 | 0.00 |
| 北北西 | 1.963450 | 0.00 |

# 天球の星, $S^2$ での Fisher-Bingham 分布による fitting



出典: 中山, 西山, 野呂, 小原, 清, 高山, 竹村, Holonomic Gradient Descent and its Application to Fisher-Bingham  
Integral, *Advances in Applied Mathematics* 47 (2011), 639–658,

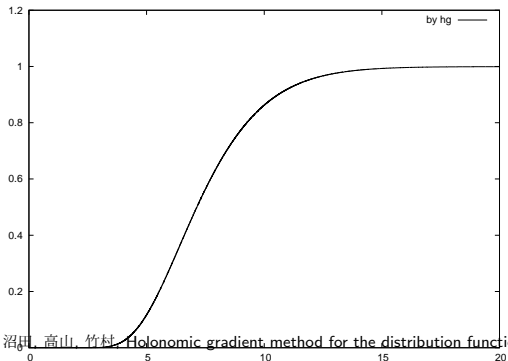
Q. ところで, 3.2 節の holonomic ideal (system) の中には常微分方程式はどこにもないですが.

A. これらに微分作用素を左から掛けたりして, 足したり, 引いたりすると常微分方程式を作れます.  $d = 1$  の時,

```
G=test1(); G[0];
((-8*x_1_1^3+24*x_2_2*x_1_1^2+(-8*x_1_2^2-24*x_2_2^2)*x_1_1+8*x_2_2*x_1_2^2+8*x_2_2^3)*x_1
+(-16*x_1_2*x_1_1^2+32*x_2_2*x_1_2*x_1_1-16*x_1_2^3-16*x_2_2^2*x_1_2)*x_2)*dx_2^5
+((-8*x_1_2*x_1_1+8*x_2_2*x_1_2)*x_1^2+(4*x_1_1^2-8*x_2_2*x_1_1-20*x_1_2^2+4*x_2_2^2)*x_2*x_1
+(16*x_1_2*x_1_1-16*x_2_2*x_1_2)*x_2^2+16*x_1_2*x_1_1^3-48*x_2_2*x_1_2*x_1_1^2
+(16*x_1_2^3+48*x_2_2^2*x_1_2)*x_1_1-16*x_2_2*x_1_2^3-16*x_2_2^3*x_1_2)*dx_2^4
+((-2*x_1_1+2*x_2_2)*x_1^3-8*x_1_2*x_2*x_1^2+((2*x_1_1-2*x_2_2)*x_2^2+8*x_1_1^3+(-24*x_2_2+24)*x_1_1^2
+(24*x_1_2^2+24*x_2_2^2-48*x_2_2)*x_1_1-24*x_2_2*x_1_2^2-8*x_2_2^3+24*x_2_2^2)*x_1-4*x_1_2*x_2^3
+(48*x_1_2*x_1_1+16*x_1_2^3-48*x_2_2*x_1_2)*x_2)*dx_2^3
+(-x_2*x_1_1^3+(8*x_1_2*x_1_1+(-8*x_2_2+2)*x_1_2)*x_1^2+(-x_2^3+(-4*x_1_1^2+(8*x_2_2-6)*x_1_1
+12*x_1_2^2-4*x_2_2^2+6*x_2_2)*x_2)*x_1+(-12*x_1_2*x_1_1+(12*x_2_2-20)*x_1_2)*x_2^2-16*x_1_2*x_1_1^3
+(48*x_2_2-16)*x_1_2*x_1_1^2+(-16*x_1_2^3+(-48*x_2_2+32*x_2_2)*x_1_2)*x_1_1+(16*x_2_2+16)*x_1_2^3
+(16*x_2_2^3-16*x_2_2^2)*x_1_2)*dx_2^2
+(x_1^3+2*x_1_2*x_2*x_1^2+((-2*x_1_1+2*x_2_2-1)*x_2^2-20*x_1_1^2+(-10*x_1_2^2+40*x_2_2-6)*x_1_1+(10*x_2_2
+(-x_1_2*x_1^2+(x_2^3+(6*x_1_1-x_1_2^2-6*x_2_2)*x_2)*x_1+(-4*x_1_2*x_1_1+(4*x_2_2+16)*x_1_2)*x_2^2+8*x_1_1
+(-4*x_2_2-8)*x_1_2^3+8*x_2_2^2*x_1_2)*dx_2^0)
[1957] vars(G[0]);
[x_1,x_2,x_1_1,x_1_2,x_2_2,dx_2]
```

# Wishart 分布の第一固有値, 行列引数超幾何系 ${}_1F_1$ (予稿 4 節)

累積分布関数  $\Pr[\ell_1 < x]$ ,  $m = 10, n = 12$ ,  
 $\Sigma^{-1} = \text{diag}(1, 2, \dots, 10)$



出典: 橋口, 沼田, 高山, 竹村, Holonomic gradient method for the distribution function of the largest root of a Wishart matrix, Journal of Multivariate Analysis, 117, (2013) 296-312.

# Orthant probability (予稿5節)

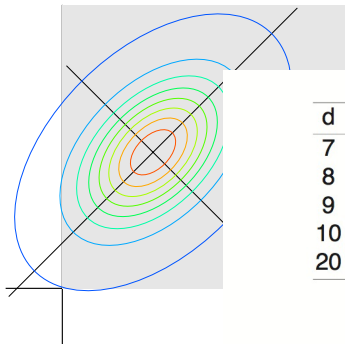
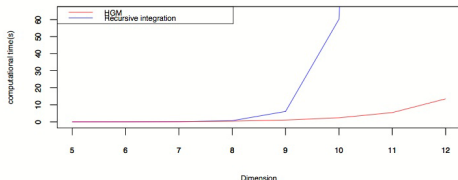


Table: Results of HGM

| d  | HGM          | Exact        |
|----|--------------|--------------|
| 7  | 0.125000000  | 0.125000000  |
| 8  | 0.111111111  | 0.111111111  |
| 9  | 0.100000000  | 0.100000000  |
| 10 | 0.090909091  | 0.090909091  |
| 20 | 0.0476190473 | 0.0476190476 |

Orthant  
Probability の  
HGM による計算  
(小山, 竹村,  
2012))



# 離散 $A$ 超幾何分布 (予稿 7 節, arxiv:1510.02269)

差分 HGM では厳密計算が可能. 近似公式の精度を調べるなど理論研究にも使える.

Theorem (T-栗木-竹村 (arxiv:1510.02269))

$k \rightarrow +\infty$  で,

$$Z(k\beta; x) \sim \frac{x^{km}}{\Gamma(km+1)} \frac{(2\pi k)^{n-d}}{(\det(\bar{A}M^{-1}\bar{A}^T))^{1/2}}$$

$M = \text{diag}(m)$ .

$2 \times 4$  分割表. 行和  $(4, 19)k$ , 列和  $(9, 5, 3, 6)k$ ,  
 $k\beta = k(19, 9, 5, 3, 6)$ .  $x = (1, 1/3, 12, 1/5001, 1, 1, 1, 1)$ .  $x_4$  は他より小さい.  $x_i$  の期待値 ( $x_i \partial_i \bullet Z/Z$ ) の近似 (by IPS)  $m = (2.79518, 0.652785, 0.551505, 0.000540425, 6.20482, 4.34722, 2.4485, 5.99946)$ .



期待値の近似 (by IPS)  $m =$

(2.79518, 0.652785, 0.551505, 0.000540425, 6.20482, 4.34722, 2.4485, 5.99946).

HGM による厳密値:

(2.83214, 0.627808, 0.539555, 0.000496547, 6.16786, 4.37219, 2.46044, 5.9995).

比:

(0.98695, 1.03978, 1.02215, 1.08837, 1.00599, 0.994289, 0.995147, 0.999993).

$k \rightarrow +\infty$  の時

| $k$ | $\log Z$    | Approx by Th 5 | error  |
|-----|-------------|----------------|--------|
| 9   | -568.0127   | -569.8179      | 1.8052 |
| 200 | -26598.4556 | -26598.9446    | 0.4890 |
| 300 | -42685.5415 | -42685.9149    | 0.3734 |

注: 近似値で確率を計算すると誤差は大きい. 例.

$u = (33, 1, 1, 1, 48, 44, 26, 53)$ ,  $k = 9$ .  $u$  を得る正確な確率は

$3.26465 \times 10^{-7}$ . 上記の  $\log Z$  の近似による確率は

$1.98529 \times 10^{-6}$ . 約 6 倍異なる.

厳密計算なのでこんな数が出力されます

N=200

A=[[1,0,0,1,0,1,0,1,0,1], [0,1,0,1,0,1,0,1,0,1], [0,0,1,-1,0,0,0,0,0,0], [0,0,0,0,1,-1,0,0,0,0],  
[0,0,0,0,0,0,1,-1,0,0], [0,0,0,0,0,0,0,0,1,-1]]

Beta=[452,412,-37,-23,-13,31]

at ([x1,x2,x3,x4,x5,x6,x7,x8,x9,x10]=[140/411,40/137,25/822,31/411,14/411,17/274,17/822,5/137,10/137,29/822])

oohg\_native=0, oohg\_curl=1

x3 の期待値を計算.

EV([x3])=[484018240471728953822203320553380653219481012643866487201043272204554116427335942534923953734366  
8636569983916892438594752962343521375555177302221592210472215250465284561475111662762276502434509742280774  
305750092193523229313167685161576286201466399466487213469381535663734384193880974741829514261324096233334  
344275350822035203131054916726819435165178778325389866000027699548897905993488167196392728277735383730885  
/1944222849842515553043842429125888595116006553306378943684005607207680083449525569604031294035766826584  
206368590575510231394395404443601780545808586417609373178438189812637405870280353563181965119049387640350  
941772514489533194749781746840208705674606008876031734288671532476200701856516011956451597268538379935874  
3209062720142982595156985628080863960988690611022042551157063876491557859146442800043022086834093773944354  
9573932056327206030262721912023810463723569352286063413912998077871191506911]

Time=84562.4

| N   | 級数の直接計算                 | Macaulay 型行列の方法    |            |
|-----|-------------------------|--------------------|------------|
| 0   | 6822s (1.89 時間)         | 61399s (約 17 時間)   | Intel Xeon |
| 100 | 138640s (1 日と約 14.5 時間) | 73126s(約 20.3 時間)  |            |
| 200 | 未計測                     | 84562s (約 23.5 時間) |            |

E5-4650 (2.7GHz) with 256G memory, the computer algebra system Risa/Asir (20140528). パッケージ [ot\\_hgm\\_ahg.rr](http://ot.hgm-ahg.rr).