Holonomic Gradient Method for Multivariate Normal Probabilities for Polyhedral and Spherical Regions

小山, 民雄

(Degree)
博士（理学）
(Date of Degree)
2014-03-25
(Date of Publication)
2015-03-01
(Resource Type)
doctoral thesis
(Report Number)
甲第6118号
(URL)
https://hdl.handle.net/20.500.14094/D1006118

博 士 論 文

Holonomic Gradient Method for Multivariate Normal Probabilities for Polyhedral and Spherical Regions

平成 26 年 1 月

神戸大学大学院理学研究科

小山 民雄

Contents

Chapter 1

A Holonomic system for the Fisher–Bingham Integral

1.1 Introduction

The Fisher–Bingham distribution is a probability distribution on the n-dimensional sphere $S^n(r)$ with the radius r defined by

$$\frac{1}{F(x,y,r)} \exp(t^T xt + yt)|dt(r)|. \tag{1.1}$$

Here, the variable x is an $(n+1) \times (n+1)$ symmetric matrix whose (i,j) component is x_{ij} when $i = j$ and $x_{ij}/2$ when $i \neq j$. The variable y (resp. t) is a row (resp. column) vector of length $n+1$, and the measure $|dt(r)|$ is the standard measure on $S^n(r)$. The function $F(x,y,r)$ is the normalizing constant defined by

$$F(x,y,r) = \int_{S^n(r)} \exp\left(\sum_{1 \leq i \leq j \leq n+1} x_{ij} t_i t_j + \sum_{i=1}^{n+1} y_i t_i \right) |dt(r)|. \tag{1.2}$$

The integral (1.2) is referred to as the Fisher–Bingham integral on the sphere $S^n(r)$.

The Fisher–Bingham distribution is used in directional statistics. Kent studied estimations, hypothesis testings and confidence regions with respect to the Fisher–Bingham distribution on the 2-dimensional sphere [17], and in the book by Mardia and Jupp on directional statistics [24, chapter 9], a definition of the Fisher–Bingham distribution having the same form as (1.1) and a relation with another probability distribution on the sphere are explained.

We are interested in estimating the value of parameters x and y which maximizes the likelihood function

$$L(x,y) = \frac{1}{F(x,y,1)^N} \prod_{i=1}^{N} \exp(t_i^T x t_i + y t_i)$$

for given data $t_1, \cdots, t_N \in S^n$. This problem is equivalent to estimating the value of parameters x and y which minimizes the function

$$F(x, y, 1) \exp \left(- \sum_{1 \leq i \leq j \leq n} S_{ij} x_{ij} - S_i y_i \right)$$

for given $\{S_{ij}\}_{1 \leq i \leq j \leq n}$, $\{S_i\}_{1 \leq i \leq n} \subset \mathbf{R}$. There are several approaches to solving this problem. Among them, the holonomic gradient descent proposed in [28] enables us to estimate the value by utilizing linear partial differential operators with polynomial coefficients which annihilate the Fisher–Bingham integral (1.2) and generate a holonomic ideal. Let D_d be the ring of differential operators $D_d = \mathbf{C}\langle z_1, \ldots, z_d, \partial_1, \ldots, \partial_d \rangle$. A left ideal I of D_d is called a holonomic ideal when the characteristic ideal $\mathrm{in}_{(0,e)}(I)$ generated by the principal symbols of I in $\mathbf{C}[z_1, \ldots, z_d, \xi_1, \ldots, \xi_d]$ has the Krull dimension d. See, e.g., [37, p 31, Definition 1.4.8], [31], and their references for details.

In [28], it is shown that the Fisher–Bingham integral $F(x, y, r)$ is annihilated by the following linear partial differential operators.

$$\begin{aligned}
&\partial_{x_{ij}} - \partial_{y_i} \partial_{y_j} \quad (i \leq j), \\
&\sum_{i=1}^{n+1} \partial_{x_{ii}} - r^2, \\
&x_{ij} \partial_{x_{ii}} + 2(x_{jj} - x_{ii}) \partial_{x_{ij}} - x_{ij} \partial_{x_{jj}} \\
&\quad + \sum_{k \neq i,j} \left(x_{kj} \partial_{x_{ik}} - x_{ik} \partial_{x_{jk}} \right) + y_j \partial_{y_i} - y_i \partial_{y_j} \quad (i < j, x_{k\ell} = x_{\ell k}), \\
&r \partial_r - 2 \sum_{i \leq j} x_{ij} \partial_{x_{ij}} - \sum_i y_i \partial_{y_i} - n.
\end{aligned} \tag{1.3}$$

They also show that (1.3) generates a holonomic ideal in the cases $n = 1$ and $n = 2$ by a calculation of their characteristic ideals on a computer, and conjecture that it holds for any n. We will prove this conjecture.

It is a fundamental problem to prove that a given system is holonomic. For example, the theory of A-hypergeometric systems is built on the fact that any A-hypergeometric system is holonomic. Moreover, many algorithms and theorems assume the holonomy property. Therefore, our main theorem is expected to be the foundation to study the Fisher-Bingham integral through differential equations.

In order to state the main result of this paper precisely, let us explain the notion of the integration ideal. For a holonomic ideal I in D_d, the left ideal $(I + \partial_{d'+1} D_d + \cdots + \partial_d D_d) \cap D_{d'}$ in $D_{d'}$ is called the integration ideal and it is known that the integration ideal is a holonomic ideal in $D_{d'}$ (see, e.g., [6, Chapter 1], [37, §5.5]). We note that Oaku gave an algorithm for computing the integration ideal in [32].

In the present paper, we show that the operators (1.3) generate the integration ideal of a holonomic ideal which annihilates the distribution

$$\exp\left(\sum_{1\leq i\leq j\leq n+1} x_{ij}t_it_j + \sum_{i=1}^{n+1} y_it_i\right)|dt(r)|$$

which has parameters x_{ij}, y_k, r. Here, $\{z_1, \ldots, z_{d'}\} = \{x_{ij}, y_k, r | 1 \leq i \leq j \leq n+1, 1 \leq k \leq n+1\}$ and $\{z_{d'+1}, \ldots, z_d\} = \{t_1, \ldots, t_{n+1}\}$. As its corollary, we show that (1.3) generates a holonomic ideal for any n and prove the conjecture in [28].

In section 1.2, we consider the holonomic ideal annihilating the measure $|dt(r)|$ on $S^n(r)$. In section 1.3, we give generators of the holonomic ideal which annihilates the integrand of the Fisher–Bingham integral. In section 1.4, we compute the integration ideal of the holonomic ideal which is given in section 1.3, and prove the main theorem of this paper.

1.2 The standard measure $|dt(r)|$ on $S^n(r)$

The Riemannian metric on the n-dimensional sphere with radius r is constructed by the pullback of the standard metric on the $(n+1)$-dimensional Euclidean space $\mathbf{R}^{n+1}$ along the embedding map. The metric defines a measure on $S^n(r)$, which is denoted by $|dt(r)|$. We define a distribution μ_r with a parameter $r > 0$ as

$$\langle \mu_r, \varphi(t)\rangle := \int_{S^n(r)} \varphi(t)|dt(r)|.$$

Here, $\varphi(t)$ is a test function.

Let $D = \mathbf{C}\langle x, y, r, t, \partial_x, \partial_y, \partial_r, \partial_t\rangle$ be the ring of differential operators with polynomial coefficients. For a given distribution F, we denote by $\mathrm{Ann}(F)$ the set of the operators in D which annihilate F.

Lemma 1.2.1. *A left ideal I in D generated by the following differential operators is a subset of* $\mathrm{Ann}(\mu_r)$.

$$\begin{aligned} &\partial_{x_{ij}}\ (1 \leq i \leq j \leq n+1), \quad \partial_{y_i}\ (1 \leq i \leq n+1), \quad t_1^2 + \cdots + t_{n+1}^2 - r^2, \\ &t_i\partial_{t_j} - t_j\partial_{t_i}\ (1 \leq i < j \leq n+1), \quad r\partial_r + \sum_{i=1}^{n+1} t_i\partial_{t_i} + 1 \end{aligned} \tag{1.4}$$

For computing the integration ideal, the following proposition is important.

Proposition 1.2.2. *The left ideal I in D is a holonomic ideal.*

This proposition may be well known, however we could not find a proof in the literature with the radius r as a variable. Therefore, we present a proof here.

Proof. By the fundamental theorem of algebraic analysis (see, e.g., [37, p30.Theorem1.4.5]), it is enough to show that the dimension of the characteristic ideal $\mathrm{in}_{(0,e)}(I)$ is not more than the number of variables $N := n(n+1)/2 + 2n + 1$.

We can find the operators $r^2\partial_{t_k} + t_k r\partial_r - t_k$ $(1 \le k \le n+1)$ in I as follows.

$$
\begin{aligned}
& t_{n+1}(t_{n+1}\partial_{t_{n+1}} + \cdots + t_1\partial_{t_1} + r\partial_r + 1) - \partial_{t_{n+1}}(t_{n+1}^2 + \cdots + t_1^2 - r^2) \\
= \quad & -\sum_{i=1}^{n} t_i(t_i\partial_{t_{n+1}} - t_{n+1}\partial_{t_i}) + r^2\partial_{t_{n+1}} + t_{n+1}r\partial_r - t_{n+1}, \\
& t_k(t_{n+1}\partial_{t_{n+1}} + \cdots + t_1\partial_{t_1} + r\partial_r + 1) - t_{n+1}(t_k\partial_{t_{n+1}} - t_{n+1}\partial_{t_k}) \\
= \quad & -\sum_{i=1}^{k-1} t_i(t_i\partial_{t_k} - t_k\partial_{t_i}) + \sum_{i=k+1}^{n} t_i(t_k\partial_{t_i} - t_i\partial_{t_k}) + \partial_{t_k}(t_{n+1}^2 + \cdots + t_1^2 - r^2) \\
& + r^2\partial_{t_k} + t_k r\partial_r - t_k \quad (1 \le k \le n)
\end{aligned}
$$

Then, the characteristic ideal $\mathrm{in}_{(0,e)}(I)$ contains the polynomials

$$
\begin{aligned}
& \xi_{x_{ij}}\ (1 \le i \le j \le n+1), \quad \xi_{y_i}\ (1 \le i \le n+1), \quad t_{n+1}^2 + \cdots + t_1^2 - r^2, \\
& t_i\xi_{t_j} - t_j\xi_{t_i}\ (1 \le i < j \le n+1), \quad r^2\xi_{t_i} + t_i r\xi_r (1 \le i \le n+1).
\end{aligned}
$$

Let I' be the ideal in the polynomial ring $\mathbf{C}[x, y, r, t, \xi_x, \xi_y, \xi_r, \xi_t]$ generated by these polynomials. Then, we have $I' \subset \mathrm{in}_{(0,e)}(I)$. Since $\dim I' \ge \dim \mathrm{in}_{(0,e)}(I)$, it is enough to show that $\dim I' \le N$.

Consider the graded lexicographic order satisfying

$$
\xi_{t_{n+1}} \succ \cdots \succ \xi_{t_1} \succ \xi_x \succ \xi_y \succ \xi_r \succ t_{n+1} \succ \cdots \succ t_1 \succ x \succ y \succ r.
$$

Since the degree of the Hilbert polynomial of an ideal in the polynomial ring equals that of the initial ideal with respect to the graded order of the ideal (see, e.g., [8, p448, Proposition 4]), the dimension of I' is equal to that of the initial ideal $\mathrm{LT}_{\prec}(I')$ with respect to this order. The initial ideal $\mathrm{LT}_{\prec}(I')$ contains the monomials $\xi_{x_{ij}}, \xi_{y_i}, t_i\xi_{t_j}, r^2\xi_{t_i}, t_{n+1}^2$. Let I'' be the ideal generated by these monomials. Analogously, we can show that it suffices to prove that the dimension of I'' is not more than N.

Computing the irreducible decomposition of the algebraic variety defined by

I'' as

$$
\begin{aligned}
& V(\xi_{x_{kl}}, \xi_{y_k}, t_i\xi_{t_j}, r^2\xi_{t_k}, t_{n+1}^2; 1 \le k \le l \le n+1, 1 \le i < j \le n+1) \\
= \quad & V(\xi_{x_{ij}}, \xi_{y_i}, t_{n+1}; 1 \le i \le j \le n+1) \cap \bigcap_{1\le i<j\le n+1} V(t_i\xi_{t_j}) \cap \bigcap_{i=1}^{n+1} V(r^2\xi_{t_i}) \\
= \quad & V(\xi_{x_{ij}}, \xi_{y_i}, t_{n+1}; 1 \le i \le j \le n+1) \cap \bigcup_{i=1}^{n+1} V(t_1, \dots, t_{i-1}, \xi_{t_{i+1}}, \dots, \xi_{t_{n+1}}) \\
& \cap \big(V(r) \cup V(\xi_{t_1}, \dots, \xi_{t_{n+1}})\big) \\
= \quad & \left(\bigcup_{k=1}^{n+1} V(\xi_{x_{ij}}, \xi_{y_i}, t_{n+1}, t_1, \dots, t_{k-1}, \xi_{t_{k+1}}, \dots, \xi_{t_{n+1}})\right) \\
& \cap \big(V(r) \cup V(\xi_{t_1}, \dots, \xi_{t_{n+1}})\big) \\
= \quad & \left(\bigcup_{i=1}^{n+1} V(\xi_{x_{kl}}, \xi_{y_l}, r, t_{n+1}, t_1, \dots, t_{i-1}, \xi_{t_{i+1}}, \dots, \xi_{t_{n+1}}; 1 \le k \le l \le n+1)\right) \\
& \cup \left(\bigcup_{i=1}^{n+1} V(\xi_{x_{kl}}, \xi_{y_k}, t_{n+1}, t_1, \dots, t_{i-1}, \xi_{t_1}, \dots, \xi_{t_{n+1}}; 1 \le k \le l \le n+1)\right),
\end{aligned}
$$

we conclude that the dimension of I'' is exactly N. $\square$

1.3 Holonomic ideal annihilating $\exp(g)\mu_r$

Let $g(x, y, t)$ be the polynomial $\sum_{1\le i\le j\le n+1} x_{ij}t_it_j + \sum_{i=1}^{n+1} y_it_i$. We can get a holonomic ideal annihilating the distribution $\exp(g(x, y, t))\mu_r$ by the following lemma.

Lemma 1.3.1. *[41] Consider the ring of differential operators with polynomial coefficients $\mathbf{C}\langle x_1, \dots, x_n, \partial_1, \dots, \partial_n\rangle$. Let u be a distribution and suppose that $I \subset \mathrm{Ann}(u)$ is a holonomic ideal. Let f be a polynomial and $f_i := \partial f/\partial x_i$. Then, the left ideal J generated by*

$$\{P(x_1, \dots, x_n; \partial_{x_1} - f_1, \dots, \partial_{x_n} - f_n) | P(x_1, \dots, x_n; \partial_{x_1}, \dots, \partial_{x_n}) \in I\}$$

is a holonomic ideal such that $J \subset \mathrm{Ann}(e^f u)$

For a proof of this lemma, we refer to [41]. It follows from this lemma that the left ideal J in D generated by the following differential operators is a holonomic ideal and included in $\mathrm{Ann}(\exp(g)\mu_r)$.

$$
\begin{aligned}
&\partial_{x_{ij}} - t_i t_j \quad (1 \le i \le j \le n+1), \\
&\partial_{y_i} - t_i \quad (1 \le i \le n+1), \\
&t_i(\partial_{t_j} - \sum_{k=1}^{n+1} x_{jk}t_k - x_{jj}t_j - y_j) - t_j(\partial_{t_i} - \sum_{k=1}^{n+1} x_{ik}t_k - x_{ii}t_i - y_i) \\
&\quad (1 \le i < j \le n+1), \\
&t_1^2 + \cdots + t_{n+1}^2 - r^2, \\
&r\partial_r + 1 + \sum_{i=1}^{n+1} t_i \left(\partial_{t_i} - \sum_{k=1}^{n+1} x_{ik}t_k - x_{ii}t_i - y_i\right)
\end{aligned}
\tag{1.5}
$$

In fact, we will show that the ideal J is generated by the differential operators

$$
\begin{aligned}
&t_i - \partial_{y_i}\,(1 \le i \le n+1), \quad \partial_{x_{ij}} - \partial_{y_i}\partial_{y_j}\,(1 \le i \le j \le n+1), \\
&\sum_{i=1}^{n+1} \partial_{x_{ii}} - r^2, \\
&x_{ij}\partial_{x_{ii}} + 2(x_{jj} - x_{ii})\partial_{x_{ij}} - x_{ij}\partial_{x_{jj}} \\
&\quad + \sum_{k \ne i,j} \left(x_{kj}\partial_{x_{ik}} - x_{ik}\partial_{x_{jk}}\right) + y_j\partial_{y_i} - y_i\partial_{y_j} + \partial_{t_i}\partial_{y_j} - \partial_{t_j}\partial_{y_i} \\
&\quad (1 \le i < j \le n+1, x_{k\ell} = x_{\ell k}), \\
&r\partial_r - 2\sum_{i \le j} x_{ij}\partial_{x_{ij}} - \sum_{i=1}^{n+1} y_i\partial_{y_i} - n + \sum_{i=1}^{n+1} \partial_{t_i}\partial_{y_i}
\end{aligned}
\tag{1.6}
$$

To prove this statement, we prepare the following lemma.

Lemma 1.3.2. *For any $\alpha \in \mathbf{Z}_{\ge 0}^{n+1}$, we have*

$$
t^\alpha \equiv \partial_y^\alpha \quad \mod D\{t_i - \partial_{y_i}; 1 \le i \le n+1\} \tag{1.7}
$$

Proof. When $\alpha = e_i$, the identity (1.7) obviously holds. Let us assume that (1.7) holds for the case of $\alpha - e_i$. Then, we have

$$
\begin{aligned}
t^\alpha &= t_i t^{(\alpha - e_i)} \\
&\equiv t_i \partial_y^{(\alpha - e_i)} \quad \mod D\{t_i - \partial_{y_i}; 1 \le i \le n+1\} \\
&= \partial_y^{(\alpha - e_i)} t_i = \partial_y^{(\alpha - e_i)}(t_i - \partial_{y_i}) + \partial_y^{(\alpha - e_i)}\partial_{y_i} \\
&\equiv \partial_y^{(\alpha - e_i)}\partial_{y_i} \quad \mod D\{t_i - \partial_{y_i}; 1 \le i \le n+1\} \\
&= \partial_y^\alpha
\end{aligned}
$$

Hence, (1.7) holds for α. Therefore, the identity (1.7) holds for any α. □

Finally, we prove the following lemma.

Lemma 1.3.3. *The differential operators* (1.6) *generate* J.

Proof. Let K be the left ideal generated by (1.6). First, let us show $J \subset K$. The identity

$$\partial_{x_{ij}} - t_i t_j \equiv \partial_{x_{ij}} - \partial_{y_i}\partial_{y_j} \mod D\{t_i - \partial_{y_i}; 1 \le i \le n+1\} \tag{1.8}$$

gives the inclusion $\partial_{x_{ij}} - t_i t_j \in K$.

The inclusion $\partial_{y_i} - t_i \in K$ is obvious. The inclusion $t_i(\partial_{t_j} - \sum_{k=1}^{n+1} x_{jk}t_k - x_{jj}t_j - y_j) - t_j(\partial_{t_i} - \sum_{k=1}^{n+1} x_{ik}t_k - x_{ii}t_i - y_i) \in K$ follows from

$$\begin{aligned}
& t_i(\partial_{t_j} - \sum_{k=1}^{n+1} x_{jk}t_k - x_{jj}t_j - y_j) - t_j(\partial_{t_i} - \sum_{k=1}^{n+1} x_{ik}t_k - x_{ii}t_i - y_i) \\
= \; & \sum_{k=1}^{n+1} x_{ik}t_k t_j - \sum_{k=1}^{n+1} x_{jk}t_k t_i - x_{jj}t_i t_j + x_{ii}t_i t_j - y_j t_i + y_i t_j + t_i\partial_{t_j} - t_j\partial_{t_i} \\
= \; & \sum_{k=1}^{n+1}(x_{ik}t_j - x_{jk}t_i)t_k + (x_{ii} - x_{jj})t_i t_j + y_i t_j - y_j t_i + t_i\partial_{t_j} - t_j\partial_{t_i} \\
\equiv \; & \sum_{k=1}^{n+1}(x_{ik}\partial_{y_j} - x_{jk}\partial_{y_i})\partial_{y_k} + (x_{ii} - x_{jj})\partial_{y_i}\partial_{y_j} \\
& + y_i\partial_{y_j} - y_j\partial_{y_i} + \partial_{y_i}\partial_{t_j} - \partial_{y_j}\partial_{t_i} \mod D\{t_i - \partial_{y_i}; 1 \le i \le n+1\} \\
\equiv \; & \sum_{k=1}^{n+1}(x_{ik}\partial_{x_{jk}} - x_{jk}\partial_{x_{ik}}) + (x_{ii} - x_{jj})\partial_{x_{ij}} \\
& + y_i\partial_{y_j} - y_j\partial_{y_i} + \partial_{y_i}\partial_{t_j} - \partial_{y_j}\partial_{t_i} \mod D\{\partial_{x_{ij}} - \partial_{y_i}\partial_{y_j}; 1 \le i \le j \le n+1\} \\
= \; & x_{ij}\partial_{x_{jj}} + \sum_{k \ne i,j}(x_{ik}\partial_{x_{jk}} - x_{jk}\partial_{x_{ik}}) - x_{ij}\partial_{x_{ii}} + 2(x_{ii} - x_{jj})\partial_{x_{ij}} \\
& + y_i\partial_{y_j} - y_j\partial_{y_i} + \partial_{y_i}\partial_{t_j} - \partial_{y_j}\partial_{t_i}.
\end{aligned}$$

Since

$$\begin{aligned}
& t_1^2 + \cdots + t_{n+1}^2 - r^2 \\
\equiv \; & \partial_{y_1}^2 + \cdots + \partial_{y_{n+1}}^2 - r^2 \mod D\{t_i - \partial_{y_i}; 1 \le i \le n+1\} \\
= \; & \sum_{i=1}^{n+1} \partial_{x_{ii}} - r^2 \mod D\{\partial_{x_{ij}} - \partial_{y_i}\partial_{y_j}; 1 \le i \le j \le n+1\},
\end{aligned}$$

we have $\sum_{i=1}^{n+1} \partial_{x_{ii}} - r^2 \in K$.

The inclusion $r\partial_r + 1 + \sum_{i=1}^{n+1} t_i \left(\partial_{t_i} - \sum_{k=1}^{n+1} x_{ik}t_k - x_{ii}t_i - y_i\right) \in K$ follows from

$$\begin{aligned}
& r\partial_r + 1 + \sum_{i=1}^{n+1} t_i \left(\partial_{t_i} - \sum_{k=1}^{n+1} x_{ik} t_k - x_{ii} t_i - y_i \right) \\
= \;& r\partial_r + 1 + \sum_{i=1}^{n+1} t_i \partial_{t_i} - \sum_{i=1}^{n+1}\sum_{k=1}^{n+1} x_{ik} t_i t_k - \sum_{i=1}^{n+1} x_{ii} t_i^2 - \sum_{i=1}^{n+1} y_i t_i \\
= \;& r\partial_r - n + \sum_{i=1}^{n+1} \partial_{t_i} t_i - \sum_{i=1}^{n+1}\sum_{k=1}^{n+1} x_{ik} t_i t_k - \sum_{i=1}^{n+1} x_{ii} t_i^2 - \sum_{i=1}^{n+1} y_i t_i \\
\equiv \;& r\partial_r - n + \sum_{i=1}^{n+1} \partial_{t_i} \partial_{y_i} - \sum_{i=1}^{n+1}\sum_{k=1}^{n+1} x_{ik} \partial_{x_{ik}} - \sum_{i=1}^{n+1} x_{ii} \partial_{x_{ii}} - \sum_{i=1}^{n+1} y_i \partial_{y_i} \\
& \bmod D\{t_i - \partial_{y_i}, \partial_{x_{ij}} - \partial_{y_i}\partial_{y_j}; 1 \le i \le j \le n+1\} \\
= \;& r\partial_r - n + \sum_{i=1}^{n+1} \partial_{t_i} \partial_{y_i} - 2\sum_{i \le j} x_{ij} \partial_{x_{ij}} - \sum_{i=1}^{n+1} y_i \partial_{y_i}.
\end{aligned}$$

Therefore, we have $J \subset K$.

Next, let us show the opposite inclusion $K \subset J$. The inclusion $t_i - \partial_{y_i} \in J$ is obvious. The inclusion $\partial_{x_{ij}} - \partial_{y_i}\partial_{y_j} \in J$ follows from identity (1.8). Other generators of K are also in J because of the above equivalence relation. □

1.4 The Fisher–Bingham Integral

Let D' be the ring of differential operators with polynomial coefficients $\mathbf{C}\langle x, y, r, \partial_x, \partial_y, \partial_r \rangle$. The left ideal $J' := D' \cap (J + \{\partial_{t_1}, \ldots, \partial_{t_{n+1}}\} \cdot D)$ in D' is the integration ideal of J. The Fisher–Bingham integral (1.2) can be written as

$$F(x, y, r) = \langle e^{g(x,y,t)} \mu_r, 1 \rangle = \int_{\mathbf{R}^{n+1}} \exp(g(x, y, t)) \mu_r dt.$$

Hence, the operators in J' annihilate $F(x, y, r)$. It is known that the integration ideal of a holonomic ideal is also a holonomic ideal (see, e.g., [6, 2, chapter1]). Therefore, if we obtain a set of generators of J', then this set generates a holonomic ideal. In this section, we compute a set of generators of J'. As the first step, we prove the following lemma.

Lemma 1.4.1. *Let P be an arbitrary differential operator in* (1.6)*; then we have*

$$t^\alpha P \equiv \partial_y^\alpha P \quad \bmod D\{t_i - \partial_{y_i}; 1 \le i \le n+1\}.$$

Proof. For simplicity,we put

$$Q_{ij} = x_{ij}\partial_{x_{ii}} + 2(x_{jj} - x_{ii})\partial_{x_{ij}} - x_{ij}\partial_{x_{jj}} + \sum_{k \ne i,j} \left(x_{kj}\partial_{x_{ik}} - x_{ik}\partial_{x_{jk}} \right)$$

and $R = r\partial_r - 2\sum_{i\le j} x_{ij}\partial_{x_{ij}} - n$. The following identities prove the lemma.

$$
\begin{aligned}
& t^\alpha \left(Q_{ij} + y_j\partial_{y_i} - y_i\partial_{y_j} + \partial_{t_i}\partial_{y_j} - \partial_{t_j}\partial_{y_i}\right) \\
= \;& \left(Q_{ij} + y_j\partial_{y_i} - y_i\partial_{y_j} + \partial_{t_i}\partial_{y_j} - \partial_{t_j}\partial_{y_i}\right) t^\alpha - \alpha_i\partial_{y_j} t^{(\alpha-e_i)} + \alpha_j\partial_{y_i} t^{(\alpha-e_j)} \\
\equiv \;& \left(Q_{ij} + y_j\partial_{y_i} - y_i\partial_{y_j} + \partial_{t_i}\partial_{y_j} - \partial_{t_j}\partial_{y_i}\right) \partial_y^\alpha \\
& -\alpha_i\partial_{y_j} y^{(\alpha-e_i)} + \alpha_j\partial_{y_i} y^{(\alpha-e_j)} \quad \bmod D\{t_i - \partial_{y_i}; 1\le i\le n+1\} \\
= \;& \partial_y^\alpha \left(Q_{ij} + y_j\partial_{y_i} - y_i\partial_{y_j} + \partial_{t_i}\partial_{y_j} - \partial_{t_j}\partial_{y_i}\right), \\
& t^\alpha \left(R - \sum_{i=1}^{n+1} y_i\partial_{y_i} + \sum_{i=1}^{n+1} \partial_{t_i}\partial_{y_i}\right) \\
= \;& \left(R - \sum_{i=1}^{n+1} y_i\partial_{y_i} + \sum_{i=1}^{n+1} \partial_{t_i}\partial_{y_i}\right) t^\alpha - \sum_{i=1}^{n+1} \alpha_i\partial_{y_i} t^{(\alpha-e_i)} \\
\equiv \;& \left(R - \sum_{i=1}^{n+1} y_i\partial_{y_i} + \sum_{i=1}^{n+1} \partial_{t_i}\partial_{y_i}\right) \partial_y^\alpha - \sum_{i=1}^{n+1} \alpha_i\partial_{y_i} \partial_y^{(\alpha-e_i)} \quad \bmod D\{t_i - \partial_{y_i}; 1\le i\le n+1\} \\
= \;& \partial_y^\alpha \left(R - \sum_{i=1}^{n+1} y_i\partial_{y_i} + \sum_{i=1}^{n+1} \partial_{t_i}\partial_{y_i}\right).
\end{aligned}
$$

□

Theorem 1.4.2. *The integration ideal J' is generated by the differential operators in* (1.3).

Proof. Let F and F' be the sets consisting of the differential operators (1.6) and (1.3) respectively. The inclusion $D' \cdot F' \subset J'$ is obvious. We need to show the opposite inclusion $D' \cdot F' \supset J'$. If a differential operator P is contained in J', then P can be written as

$$P = \sum_i Q_i P_i + \sum_j \partial_{t_j} R_j \quad (P_i \in F,\, Q_i \in D, R_j \in D),$$

from the definition of J'. Without loss of generality, we can assume that no term of Q_i contains ∂_t. Note that

$$t^\alpha P_i \equiv \partial_y^\alpha P_i \quad \bmod D\{t_k - \partial_{y_k}; 1\le k\le n+1\};$$

then, P can be written as

$$P = \sum_i Q_i' P_i + \sum_j \partial_{t_j} R_j + \sum_k S_k(t_k - \partial_{y_k}) \quad (P_i \in F,\, Q_i' \in D',\, R_j \in D, S_k \in D).$$

Since all differential operators in F except $t_i - \partial_{y_i}$ have the form $P' + \sum_i \partial_{t_i} U_i'$ $(P' \in F',\, U_i' \in D')$, P can be written as

$$P = \sum_i Q_i' P_i' + \sum_j \partial_{t_j} R_j + \sum_k S_k(t_k - \partial_{y_k}) \quad (P_i \in F,\, Q_i' \in D',\, R_j \in D, S_k \in D).$$

Moving some terms to the left-hand side, we obtain

$$P-\sum_i Q_i'P_i'-\sum_k S_k(t_k-\partial_{y_k}) = \sum_j \partial_{t_j} R_j \quad (P_i' \in F',\, Q_i' \in D',\, R_j \in D, S_k \in D)$$

Without loss of generality, if we assume that no term of S_k contains ∂_t, then the left-hand side of the identity does not contain ∂_t. Expanding both sides and comparing the coefficients, we get $\sum_j \partial_{t_j} R_j = 0$, in other words, we obtain

$$P - \sum_i Q_i'P_i' = \sum_k S_k(t_k - \partial_{y_k}) \quad (P_i' \in F',\, Q_i' \in D',\, S_k \in D).$$

The right-hand side of this identity is included in the left ideal $D \cdot \{t_i - \partial_{y_i} | 1 \leq i \leq n+1\}$ in D. Let the weight of t_i be 1 and that of other variables be 0, and consider a term order $\prec$ with this weight. The Gröbner basis of $D \cdot \{t_i - \partial_{y_i} | 1 \leq i \leq n+1\}$ with this order is $\{t_i - \partial_{y_i} | 1 \leq i \leq n+1\}$, and the initial ideal is generated by $\{t_i | 1 \leq i \leq n+1\}$. Hence, the leading term of $P - \sum Q_k' P_k' \in D'$ with respect to the order $\prec$ must divide some t_i. However, the differential operator in D' which satisfies this condition is only 0. Then, we have $P \in D'F'$. □

Corollary 1.4.3. *The integration ideal J' is a holonomic ideal.*

Notes. The result of this chapter will be generalized in section 4.2.

Chapter 2

The Holonomic Rank of the Fisher-Bingham System of Differential Equations

2.1 Introduction

Let $x = (x_{ij})$ and $y = (y_i)$ be parameters such that $x_{ij} = x_{ji}$ for $i \neq j$. Let Z be a function, which is the normalization constant of the Fisher-Bingham distribution, defined as

$$Z(x, y, r) = \int_{S^n(r)} \exp\left(\sum_{1 \leq i \leq j \leq n+1} x_{ij} t_i t_j + \sum_{i=1}^{n+1} y_i t_i \right) |dt| \qquad (2.1)$$

where $S^n(r) = \{(t_1, \ldots, t_{n+1}) \mid \sum_{i=1}^{n+1} t_i^2 = r^2, r > 0\}$ is the n-dimensional sphere and $|dt|$ denotes the Haar measure on the sphere.

Let D be the Weyl algebra

$$D = \mathbf{C}\langle x_{ij}, y_k, r, \partial_{ij}, \partial_k, \partial_r \mid 1 \leq i \leq j \leq n+1, 1 \leq k \leq n+1 \rangle$$

where $\partial_{ij} = \partial/\partial x_{ij}$, $\partial_k = \partial/\partial y_k$ and $\partial_r = \partial/\partial r$. It is shown in [18] and [28] that the normalization constant (2.1) of the Fisher-Bingham distribution is a holonomic function in x, y, r and consequently it is annihilated by the following holonomic ideal I in D generated by the following operators in D:

$$\begin{aligned}
&\partial_{ij} - \partial_i \partial_j, \\
&\sum_{i=1}^{n+1} \partial_i^2 - r^2, \\
&x_{ij}\partial_i^2 + 2(x_{jj} - x_{ii})\partial_i\partial_j - x_{ij}\partial_j^2 \\
&\qquad\qquad + \sum_{s \neq i,j} (x_{sj}\partial_i\partial_s - x_{is}\partial_j\partial_s) + y_j\partial_i - y_i\partial_j, \\
&r\partial_r - 2\sum_{i \leq j} x_{ij}\partial_i\partial_j - \sum_i y_i\partial_i - n.
\end{aligned}$$

We call the system of differential equations defined by I *the Fisher-Bingham system*.

For a left ideal J in D, the holonomic rank of J is defined as the dimension of the $K = \mathbf{C}(x_{ij}, y_k, r \mid 1 \leq i \leq j \leq n+1, 1 \leq k \leq n+1)$ vector space $K\langle \partial_{ij}, \partial_k, \partial_r \rangle / K\langle \partial_{ij}, \partial_k, \partial_r \rangle J$. The rank is denoted by $\mathrm{rank}\,(J)$. When J is a holonomic ideal, the rank is finite. The holonomic rank agrees with the dimension of the holomorphic solutions of the associated system of linear partial differential equations at generic points and with the size of the Pfaffian equation associated to J. As to general facts on the holonomic rank, we refer to, e.g., the chapters 1 and 2 of [37]. The holonomic rank is a fundamental invariant of the D-module D/J and there are several attractive studies on holonomic ranks. For example, Miller, Matusevich and Walther studied holonomic ranks of A-hypergeometric systems by introducing a new homological method [25].

We are interested in the holonomic rank of the Fisher-Bingham system I. We prove the following theorem in this paper.

Theorem 2.1.1.

$$\mathrm{rank}\,(I) = 2n + 2$$

In [28], we proposed a new method in the statistical inference which is called the holonomic gradient descent. The method utilizes a holonomic system of linear partial differential equations associated to the normalization constant. The complexity of the method depends on the holonomic rank and correctness of the method are proved by utilizing the holonomic rank. In the case of the Fisher-Bingham distribution, which is the most fundamental distribution in the directional statistics, Theorem 2.1.1 is applied in [19], which gives a generalization of the result in [28] shown with a help of a computer program. Our method to prove the theorem is the Gröbner deformation to the direction $(-w, w)$, which is discussed in [37] for A-hypergeometric systems, and a determination of Gröbner bases *by hand* with adding several *slack variables* which do not change the holonomic rank.

2.2 The Rank of the Diagonal System

When the matrix x is diagonal, the normalization constant Z satisfies a system of linear partial differential equations for the variables x_{ii}, y_k, r. Let $\tilde{I}$ be the left ideal in D generated by

$$\begin{aligned}
&A_i = \partial_{ii} - \partial_i^2 \quad (1 \le i \le n+1),\\
&B = \sum_{i=1}^{n+1} \partial_i^2 - r^2,\\
&C_{ij} = 2(x_{ii} - x_{jj})\partial_i\partial_j + y_i\partial_j - y_j\partial_i \quad (1 \le i < j \le n+1),\\
&E = r\partial_r - 2\sum_{i=1}^{n+1} x_{ii}\partial_i^2 - \sum_{i=1}^{n+1} y_i\partial_i - n
\end{aligned}$$

and ∂_{ij}, $i \neq j$. The ideal $\tilde{I}$ annihilates the function Z restricted to the diagonal of x [19].

Theorem 2.2.1. *The holonomic rank of $\tilde{I}$ is $2n+2$.*

Our proof of the theorem reduces to the proof of the following proposition.

Proposition 2.2.2. *Let R be the ring of differential operators with rational function coefficients*

$$R = \mathbf{C}(x_{11}, \dots, x_{n+1n+1}, y_1, \dots, y_{n+1}, r)\langle \partial_{11}, \dots, \partial_{n+1n+1}, \partial_1, \dots, \partial_{n+1}, \partial_r \rangle.$$

Let $R\tilde{I}$ be the left ideal of R generated by A_i, B, C_{ij}, E. Let $<$ be the term order on R which is the block order with $\partial_r \gg \{\partial_{ii}\} \gg \{\partial_j\}$. The order of the block $\{\partial_{ii}\}$ is the graded lexicographic order with $\partial_{11} > \cdots > \partial_{n+1n+1}$ and that of the block $\{\partial_i\}$ is the graded lexicographic order with $\partial_1 > \cdots > \partial_{n+1}$. A Gröbner basis of $R\tilde{I}$ with respect to the term order $<$ is

$$\begin{aligned}
&A_i = \partial_{ii} - \partial_i^2 \quad (i = 1, \dots, n+1),\\
&B = \sum_{i=1}^{n+1} \partial_i^2 - r^2,\\
&C_{ij} = 2(x_{ii} - x_{jj})\partial_i\partial_j + y_i\partial_j - y_j\partial_i \quad (1 \le i < j \le n+1),\\
&(\textit{We put } a_{ij} = 2(x_{ii} - x_{jj}), F_{ij} = y_i\partial_j - y_j\partial_i, C_{ij} = a_{ij}\partial_i\partial_j + F_{ij}),\\
&D_k = \partial_k B - \partial_1 a_{1k}^{-1} C_{1k} - \cdots - \partial_{k-1} a_{k-1k}^{-1} C_{k-1k} \quad (k = 1, \dots, n+1),\\
&E = r\partial_r - 2\sum_{i=1}^{n+1} x_{ii}\partial_i^2 - \sum_{i=1}^{n+1} y_i\partial_i - n.
\end{aligned}$$

The initial monomials of the Gröbner basis are

$$\begin{aligned}
&\mathrm{in}_<(A_i) = \mathrm{in}_<(\partial_{ii}), \mathrm{in}_<(B) = \mathrm{in}_<(\partial_1)^2, \mathrm{in}_<(C_{ij}) = \mathrm{in}_<(\partial_i)\mathrm{in}_<(\partial_j),\\
&\mathrm{in}_<(D_k) = \mathrm{in}_<(\partial_k)^3, \mathrm{in}_<(E) = \mathrm{in}_<(\partial_r).
\end{aligned}$$

Here, the initial monomial $\mathrm{in}_<(c(x,y,r)\partial_r^\gamma \prod \partial_{ii}^{\alpha_{ii}} \prod \partial_k^{\beta_k})$ is defined as the element $c(x,y,r)\iota^\gamma \prod \xi_{ii}^{\alpha_{ii}} \prod \eta_k^{\beta_k}$ in the polynomial ring with rational function coefficients $\mathbf{C}(x,y,r)[\xi_{11},\ldots,\xi_{n+1n+1},\eta_1,\ldots,\eta_{n+1},\iota]$ (see, e.g., [37, Chapter 1]).

By the proposition, the standard monomials of the quotient ring $R/R\tilde{I}$ are $1,\partial_1,\partial_2,\partial_2^2,\ldots,\partial_{n+1},\partial_{n+1}^2$. Therefore, the holonomic rank of $\tilde{I}$ is $2n+2$ (Theorem 2.2.1). Let us prove the proposition.

Since D_k is expressed by $B, C_{1k},\ldots,C_{k-1k}$, the operator D_k is the element in $R\tilde{I}$. In order to prove Proposition 2.2.2, we will show that any S-pair for A_i, B, C_{ij}, D_k, E is reduced to 0 by A_i, B, C_{ij}, D_k, E. The following lemmas are proved by straight forward calculations.

Lemma 2.2.3. *Let P and Q be elements in R. If the initial monomials are coprime, i.e.,* $\gcd(\mathrm{in}_<(P),\mathrm{in}_<(Q)) = 1$*, then the S-pair $S(P,Q)$ is reduced to $[P,Q]$ by P and Q (we denote the reduction by $S(P,Q) \xrightarrow[P,Q]{}{}^* [P,Q]$), where $[P,Q]$ is the commutator of P and Q. In particular, when $[P,Q]=0$, the S-pair $S(P,Q)$ is reduced to* 0*.*

Lemma 2.2.4. *We have*

$$[A_p, C_{ij}] = 0,$$
$$[B, C_{ij}] = 0,$$
$$[C_{ij}, C_{jk}] = C_{ik}, [C_{ij}, C_{ik}] = -C_{jk}, [C_{ik}, C_{jk}] = -C_{ij} \quad (i<j<k),$$
$$[C_{ij}, C_{pq}] = 0 \quad (\{i,j\}\cap\{p,q\} = \emptyset).$$

Lemma 2.2.5. *We have*

$$[D_i, A_j] = \begin{cases} 0 & (i<j) \\ 2a_{ji}^{-2}\partial_j C_{ji} & (i>j), \\ \sum_{l=1}^{i-1} 2a_{li}^{-2}\partial_l C_{li} & (i=j) \end{cases}$$

$$[D_i, B] = 0,$$

$$[D_i, D_j] = -B\partial_j + \sum_{l<i} a_{li}^{-1}a_{ij}^{-1}\partial_l(\partial_i C_{lj} + \partial_l C_{ij}) + \sum_{l<i} a_{li}^{-1}a_{lj}^{-1}(-2\partial_l C_{ij}).$$

When $i,j \neq k$, we obtain

$$[C_{ij}, D_k] = \begin{cases} 0 & (k-1<i) \\ 0 & (j \le k-1) \\ -a_{ik}^{-1}[C_{ij}, \partial_i C_{ik}] = a_{ik}^{-1}(\partial_i C_{jk} + \partial_j C_{ik}) & (i \le k-1 < j) \end{cases}.$$

Lemma 2.2.6. *We have*

$$[A_i, E] = 0,$$
$$[B, E] = -2B,$$
$$[C_{ij}, E] = 0,$$
$$[D_i, E] = -3D_i - 2\sum_{k=1}^{i-1} a_{ki}^{-1}\partial_k C_{ki}.$$

Proof of Proposition 2.2.2. We prove that any S-pair for A_i, B, C_{ij}, D_k, E is reduced to 0 by A_i, B, C_{ij}, D_k, E.

S-pairs of A_i and A_j, B, C_{ij}. The initial monomials are coprime, and the elements commute. By Lemma 2.2.3, we obtain

$$\begin{aligned}
S(A_i, A_j) &\longrightarrow^* 0,\\
S(A_i, B) &\longrightarrow^* 0,\\
S(A_i, C_{jk}) &\longrightarrow^* 0.
\end{aligned}$$

S-pair of B and C_{ij}. When $i > 1$, the initial monomials $\mathrm{in}_<(B) = \mathrm{in}_<(\partial_1)^2$, $\mathrm{in}_<(C_{ij}) = \mathrm{in}_<(\partial_i)\mathrm{in}_<(\partial_j)$ are coprime. Operators B and C_{ij} commute by Lemma 2.2.4. By Lemma 2.2.3, we have $S(B, C_{ij}) \longrightarrow^* 0$.

When $i = 1$, the initial monomials $\mathrm{in}_<(B) = \mathrm{in}_<(\partial_1)^2$, $\mathrm{in}_<(C_{1j}) = \mathrm{in}_<(\partial_1)\mathrm{in}_<(\partial_j)$ are not coprime. We obtain the following reduction sequence of the S-pair:

$$\begin{aligned}
&S(B, C_{1j}) = a_{1j}\partial_j B - \partial_1 C_{1j}\\
&= a_{1j}(\partial_j\partial_2^2 + \cdots + \partial_j^3 + \cdots + \partial_j\partial_{n+1}^2 - \partial_j r^2) - \partial_1 F_{1j}\\
&= a_{1j}((\underline{\partial_j\partial_2^2} + \cdots + \partial_j^3 + \cdots + \partial_j\partial_{n+1}^2 - \partial_j r^2) - \partial_1 a_{1j}^{-1} F_{1j})\\
&\xrightarrow[C_{2j}]{}^* a_{1j}((\underline{\partial_j\partial_3^2} + \cdots + \partial_j^3 + \cdots + \partial_j\partial_{n+1}^2 - \partial_j r^2) - \partial_1 a_{1j}^{-1} F_{1j} - \partial_2 a_{2j}^{-1} F_{2j})\\
&\xrightarrow[C_{3j}]{}^* \cdots \xrightarrow[C_{j-1j}]{}^* a_{1j} D_j \xrightarrow[D_j]{} 0.
\end{aligned}$$

S-pair of C_{ij} and C_{kl} ($\{i,j\} \cap \{k,l\} = \emptyset$). The initial monomials $\mathrm{in}_<(C_{ij}) = \mathrm{in}_<(\partial_i)\mathrm{in}_<(\partial_j), \mathrm{in}_<(C_{kl}) = \mathrm{in}_<(\partial_k)\mathrm{in}_<(\partial_l)$ are coprime. Operators C_{ij} and C_{kl} commute by Lemma 2.2.4. By Lemma 2.2.3, we obtain

$$S(C_{ij}, C_{kl}) \xrightarrow[C_{ij}, C_{kl}]{}^* 0.$$

S-pair of C_{ij} and C_{jk} ($i < j < k$). We have the following reduction sequence of the S-pair:

$$\begin{aligned}
S(C_{ij}, C_{jk}) &= a_{jk}\partial_k C_{ij} - a_{ij}\partial_i C_{jk} = -a_{ik} y_j \partial_i\partial_k + a_{jk} y_i \partial_j\partial_k + a_{ij} y_k \partial_i\partial_j\\
&\xrightarrow[C_{ij}, C_{jk}, C_{kl}]{}^* y_i(-F_{jk}) + y_k(-F_{ij}) - y_j(-F_{ik}) = 0.
\end{aligned}$$

The S-pair of C_{ij} and C_{ik} and that of C_{ik} and C_{jk} are also reduced to 0.

S-pair of D_i and A_j. The initial monomials $\mathrm{in}_<(D_i) = \mathrm{in}_<(\partial_i)^3, \mathrm{in}_<(A_j) = \mathrm{in}_<(\partial_{jj})$ are coprime. When $i < j$, operators D_i and A_j commute. By Lemma 2.2.3, we have

$$S(D_i, A_j) \xrightarrow[D_i, A_j]{}^* 0.$$

When $i > j$, by Lemmas 2.2.3 and 2.2.5, we have

$$S(D_i, A_j) \xrightarrow[D_i, A_j]{}^* [D_i, A_j] = 2a_{ji}^{-2}\partial_j C_{ji} \xrightarrow[C_{ji}]{}^* 0.$$

When $i = j$, by Lemmas 2.2.3 and 2.2.5, we have

$$S(D_i, A_i) \xrightarrow[D_i,A_j]{}^* [D_i, A_i] = \sum_{l=1}^{i-1} 2a_{li}^{-2}\partial_l C_{li} \xrightarrow[C_{1i},\dots,C_{i-1i}]{}^* 0.$$

S-pair of D_i and B. When $i = 1$, the S-pair is $S(D_1, B) = \partial_1 B - \partial_1 B = 0$. When $i > 1$, the initial monomials $\mathrm{in}_<(D_i) = \mathrm{in}_<(\partial_i)^3$, $\mathrm{in}_<(B) = \mathrm{in}_<(\partial_1)^2$ are coprime. By Lemmas 2.2.3 and 2.2.5, we have

$$S(D_i, B) \xrightarrow[D_i,B]{}^* [D_i, B] = 0.$$

S-pair of D_i and D_j. The initial monomials $\mathrm{in}_<(D_i) = \mathrm{in}_<(\partial_i)^3$, $\mathrm{in}_<(D_j) = \mathrm{in}_<(\partial_j)^3$ are coprime. By Lemmas 2.2.3 and 2.2.5, we have

$$S(D_i, D_j) \xrightarrow[D_i,D_j]{}^* [D_i, D_j] \xrightarrow[B,C_{1j},\dots,C_{ij}]{}^* 0.$$

S-pair of C_{ij} and D_k. The initial monomials are $\mathrm{in}_<(C_{ij}) = \mathrm{in}_<(\partial_i)\mathrm{in}_<(\partial_j)$, $\mathrm{in}_<(D_k) = \mathrm{in}_<(\partial_k)^3$. When $i \neq k$ and $j \neq k$, the initial monomials are coprime. By Lemmas 2.2.3 and 2.2.5, we have

$$S(C_{ij}, D_k) \xrightarrow[C_{ij},D_k]{}^* [C_{ij}, D_k] \xrightarrow[C_{ik},C_{jk}]{}^* 0.$$

When $i = k$, the initial monomials $\mathrm{in}_<(C_{ij}) = \mathrm{in}_<(\partial_i)\mathrm{in}_<(\partial_j), \mathrm{in}_<(D_i) = \mathrm{in}_<(\partial_i)^3$ are not coprime. This case needs a care of an order of applying reductions. We will reduce the S-pair by D_j and then reduce remainders by C_{ij}'s.

$$\begin{aligned}
S(C_{ij}, D_i) &= \partial_i^2 C_{ij} - a_{ij}\partial_j D_i \\
&= \partial_i^2 F_{ij} - a_{ij}\underline{\partial_i\partial_j^3} - a_{ij}\partial_j\Big(\sum_{l=i+1,l\neq j}^{n+1} \partial_i\partial_l^2 - \partial_i r^2 - \sum_{l=1}^{i-1}\partial_l a_{li}^{-1}F_{li}\Big) \\
&\xrightarrow[D_j]{}^* \partial_i^2 F_{ij} + a_{ij}\partial_i\Big(\sum_{l=j+1}^{n+1} \partial_j\partial_l^2 - \partial_j r^2 - \sum_{l=1}^{j-1}\partial_l a_{lj}^{-1}F_{lj}\Big) \\
&\qquad\qquad - a_{ij}\partial_j\Big(\sum_{l=i+1,l\neq j}^{n+1} \partial_i\partial_l^2 - \partial_i r^2 - \sum_{l=1}^{i-1}\partial_l a_{li}^{-1}F_{li}\Big) \\
&= -a_{ij}\partial_i(\partial_{i+1}a_{i+1j}^{-1}C_{i+1j} + \cdots + \partial_{j-1}a_{j-1j}^{-1}C_{j-1j}) \\
&\qquad\qquad - a_{ij}\partial_i\Big(\sum_{l=1}^{i-1} a_{lj}^{-1}F_{lj}\Big) + a_{ij}\partial_j\Big(\sum_{l=1}^{i-1}\partial_l a_{li}^{-1}F_{li}\Big) \\
&\xrightarrow[C_{i+1j},\dots,C_{j-1j}]{}^* -a_{ij}\partial_i\Big(\sum_{l=1}^{i-1} a_{lj}^{-1}F_{lj}\Big) + a_{ij}\partial_j\Big(\sum_{l=1}^{i-1}\partial_l a_{li}^{-1}F_{li}\Big) \\
&= -a_{ij}\sum_{l=1}^{i-1}(a_{lj}^{-1}\partial_l\partial_i F_{lj} - a_{li}^{-1}\partial_l\partial_j F_{li}).
\end{aligned}$$

Since $a_{lj}^{-1}\partial_l\partial_i F_{lj} - a_{li}^{-1}\partial_l\partial_j F_{li} \xrightarrow[C_{lj},C_{li},C_{ij}]{}{}^* 0$, the S-pair $S(C_{ij}, D_i)$ is reduced to 0.

When $j = k$, the initial monomials $\mathrm{in}_<(C_{ij}) = \mathrm{in}_<(\partial_i)\mathrm{in}_<(\partial_j), \mathrm{in}_<(D_j) = \mathrm{in}_<(\partial_j)^3$ are not coprime. This case also needs a care of an order of applying reductions.

$$\begin{aligned}
S(C_{ij}, D_j) &= \partial_j^2 C_{ij} - a_{ij}\partial_i D_j \\
&= F_{ij}(\partial_i^2 + \partial_j^2) - a_{ij}\partial_i\Big(\sum_{l=j+1}^{n+1} \partial_j\partial_l^2 - \partial_j r^2 - \sum_{l=1,l\neq i}^{j-1} \partial_l a_{lj}^{-1} F_{lj}\Big) \\
&\xrightarrow[B]{}{}^* F_{ij}\Big(-\sum_{l=1,l\neq i,j}^{n+1} \partial_l^2 + r^2\Big) - a_{ij}\partial_i\Big(\sum_{l=j+1}^{n+1} \partial_j\partial_l^2 - \partial_j r^2 - \sum_{l=1,l\neq i}^{j-1} \partial_l a_{lj}^{-1} F_{lj}\Big) \\
&= F_{ij}\Big(-\sum_{l=1,l\neq i,j}^{n+1} \partial_l^2\Big) + r^2(a_{ij}\partial_i\partial_j + F_{ij}) - a_{ij}\partial_i\Big(\sum_{l=j+1}^{n+1} \partial_j\partial_l^2 - \sum_{l=1,l\neq i}^{j-1} \partial_l a_{lj}^{-1} F_{lj}\Big) \\
&\xrightarrow[C_{ij}]{}{}^* F_{ij}\Big(-\sum_{l=1,l\neq i,j}^{n+1} \partial_l^2\Big) - a_{ij}\partial_i\Big(\sum_{l=j+1}^{n+1} \partial_j\partial_l^2 - \sum_{l=1,l\neq i}^{j-1} \partial_l a_{lj}^{-1} F_{lj}\Big) \\
&= \Big(\sum_{l=j+1}^{n+1} \partial_l^2\Big)(-a_{ij}\partial_i\partial_j - F_{ij}) - F_{ij}\sum_{l=1,l\neq i}^{j-1} \partial_l^2 + a_{ij}\partial_i \sum_{l=1,l\neq i}^{j-1} \partial_l a_{lj}^{-1} F_{lj} \\
&\xrightarrow[C_{ij}]{}{}^* -F_{ij}\sum_{l=1,l\neq i}^{j-1} \partial_l^2 + a_{ij}\partial_i \sum_{l=1,l\neq i}^{j-1} \partial_l a_{lj}^{-1} F_{lj} \\
&= \sum_{l=1,l\neq i}^{j-1} \partial_l a_{ij}(-a_{ij}^{-1}\partial_l F_{ij} + a_{lj}^{-1}\partial_i F_{lj}).
\end{aligned}$$

Since $-a_{ij}^{-1}\partial_l F_{ij} + a_{lj}^{-1}\partial_i F_{lj} \xrightarrow[C_{lj},C_{li},C_{ij}]{}{}^* 0$, the S-pair $S(C_{ij}, D_j)$ is reduced to 0.

S-pair of E and A_i. The initial monomials $\mathrm{in}_<(E) = \mathrm{in}_<(\partial_r), \mathrm{in}_<(A_i) = \mathrm{in}_<(\partial_{ii})$ are coprime. By Lemmas 2.2.3 and 2.2.6, we have

$$S(A_i, E) \xrightarrow[A_i,E]{}{}^* [A_i, E] = 0.$$

S-pair of E and B. The initial monomials $\mathrm{in}_<(E) = \mathrm{in}_<(\partial_r), \mathrm{in}_<(B) = \mathrm{in}_<(\partial_1)^2$ are coprime. By Lemmas 2.2.3 and 2.2.6, we have

$$S(B, E) \xrightarrow[B,E]{}{}^* [B, E] = -2B \xrightarrow[B]{}{}^* 0.$$

S-pair of E and C_{ij}. The initial monomials $\mathrm{in}_<(E) = \mathrm{in}_<(\partial_r), \mathrm{in}_<(C_{ij}) = \mathrm{in}_<(\partial_i)\mathrm{in}_<(\partial_j)$ are coprime. By Lemmas 2.2.3 and 2.2.6, we have

$$S(C_{ij}, E) \xrightarrow[C_{ij},E]{}{}^* [C_{ij}, E] = 0.$$

S-pair of D_i and E. The initial monomials $\mathrm{in}_<(E) = \mathrm{in}_<(\partial_r), \mathrm{in}_<(D_i) = \mathrm{in}_<(\partial_i)^3$ are coprime. By Lemmas 2.2.3 and 2.2.6, we have

$$S(D_i, E) \xrightarrow[D_i, E]{}^{*} [D_i, E] = -3D_i - 2\sum_{k=1}^{i-1} a_{ki}^{-1} \partial_k C_{ki} \xrightarrow[D_i, C_{1i}, \dots, C_{i-1i}]{}^{*} 0.$$

We have proved that any S-pair is reduced to 0. By Buchberger's criterion, the set $\{A_i, B, C_{ij}, D_k, E\}$ is a Gröbner basis of $R\tilde{I}$. □

2.3 Gröbner Deformation of the Fisher-Bingham System

Consider the system of differential equations $I \cdot f = 0$. Intuitively speaking, we want to prove that the system I can be deformed to the diagonal system $\tilde{I}$ without increasing the holonomic rank. This can be done by a Gröbner basis computation with a weight vector $(-w, w)$ [37, Theorem 2.2.1]. However a straightforward calculation does not seem to be easy. We need to use some technical tricks to determine a suitable Gröbner deformation. Since these tricks may look too technical for the general n, we explain them in the case of $n = 1$ in Section 2.4 to clarify our idea without technical details of this section. Readers are expected to refer to the Section 2.4 when technicalities get complicated.

We will introduce new indeterminates to make our Gröbner basis computation possible by hand with employing the idea of the proof of [37, Theorem 3.1.3]. Let a_{pq}, b_i, c_i, d $(1 \le p \le q \le n+1,\ 1 \le i \le n+1)$ be constants, which we call slack variables when they are regarded as indeterminates. We put $g = \left(r^{d^3} \prod_{p \le q} x_{pq}^{a_{pq}^3} \right) f$ and make a change of the independent variables y_i by $y_i + b_i c_i$. Then, the system of differential equations for the function g is $I' \cdot g = 0$ where I' is the left ideal in D generated by the set of operators $G' = \{A'_{pq}, B, C'_{ij}, E'\}$ where

$$\begin{aligned}
A'_{pq} &= x_{pq}\partial_{pq} - x_{pq}\partial_p\partial_q - a_{pq}^3, \\
B &= \sum_{i=1}^{n+1} \partial_i^2 - r^2, \\
C'_{ij} &= x_{ij}\partial_i^2 + 2(x_{jj} - x_{ii})\partial_i\partial_j - x_{ij}\partial_j^2 \\
&\qquad + \sum_{s \ne i,j} (x_{sj}\partial_i\partial_s - x_{is}\partial_j\partial_s) + (y_j + b_j c_j)\partial_i - (y_i + b_i c_i)\partial_j, \\
E' &= r\partial_r - 2\sum_{i \le j} x_{ij}\partial_i\partial_j - \sum_{i=1}^{n+1} (y_i + b_i c_i)\partial_i - n - d^3.
\end{aligned}$$

The key fact is that the holonomic rank of I for f agrees with the holonomic rank of I' for g for any constants a_{pq}, b_i, c_i, d.

Let us make the same change of the variables for the diagonal system; let $\tilde{I}'$ be the left ideal of D generated by the set of operators $\tilde{G}' = \{\tilde{A}'_{ii}, B, \tilde{C}'_{ij}, \tilde{E}', x_{ij}\partial_{ij} - a^3_{ij}\ (i \neq j)\}$ where

$$\begin{aligned}
\tilde{A}'_{ii} &= x_{ii}\partial_i^2 - x_{ii}\partial_{ii} - a^3_{ii} \quad (1 \le i \le n+1),\\
B &= \sum_{i=1}^{n+1} \partial_i^2 - r^2,\\
\tilde{C}'_{ij} &= 2(x_{jj} - x_{ii})\partial_i\partial_j + (y_j + b_jc_j)\partial_i - (y_i + b_ic_i)\partial_j \quad (1 \le i < j \le n+1),\\
\tilde{E}' &= r\partial_r - 2\sum_{i=1}^{n+1} x_{ii}\partial_i^2 - \sum_{i=1}^{n+1}(y_i + b_ic_i)\partial_i - n - d^3.
\end{aligned}$$

The holonomic ranks of $\tilde{I}$ and $\tilde{I}'$ agree.

Define the weight vector w by $w_{ij} = 1$, $(i \neq j)$, $w_{ii} = 0$, $w_k = 0$, and $w_r = 0$. Here, w_{ij} stands for ∂_{ij}, w_k stands for ∂_k, and w_r stands for ∂_r. The initial form $\mathrm{in}_{(-w,w)}(\ell)$, $\ell \in D$, is the sum of the highest $(-w,w)$-degree terms in ℓ and $\mathrm{in}_{(-w,w)}(I')$ is the left ideal generated by $\ell \in I'$ where the weight $-w$ stands for space variables x_{ij}, y_k, r corresponding to differential operators ∂_{ij}, ∂_k, ∂_r respectively (see, e.g., [37, Chapter 1]).

Theorem 2.3.1. *For generic complex numbers a_{pq}, b_i, c_i, d, we have*

$$\mathrm{in}_{(-w,w)}(I') = \tilde{I}'. \tag{2.2}$$

In order to prove the theorem, we regard a_{pq}, b_i, c_i, d as ring variables with the weight 0 and consider the following homogenized system of I':

$$\begin{aligned}
A'^h_{pq} &= hx_{pq}\partial_{pq} - x_{pq}\partial_p\partial_q - a^3_{pq},\\
B &= \sum_{i=1}^{n+1} \partial_i^2 - r^2,\\
C'^h_{ij} &= x_{ij}\partial_i^2 + 2(x_{jj} - x_{ii})\partial_i\partial_j - x_{ij}\partial_j^2\\
&\qquad + \sum_{s \neq i,j} (x_{sj}\partial_i\partial_s - x_{is}\partial_j\partial_s) + (hy_j + b_jc_j)\partial_i - (hy_i + b_ic_i)\partial_j,\\
E'^h &= hr\partial_r - 2\sum_{i \le j} x_{ij}\partial_i\partial_j - \sum_{i=1}^{n+1}(hy_i + b_ic_i)\partial_i - nh^3 - d^3.
\end{aligned}$$

Let I'^h be a left $D^h[a,b,c,d]$-ideal generated by the set of operators $G'^h := \{A'^h_{pq}, B, C'^h_{ij}, E'^h\}$, where $D^h[a,b,c,d]$ is the homogenized Weyl algebra of the ring $D[a,b,c,d] = \mathbf{C}[a,b,c,d]\langle x_{ij}, y_k, r, \partial_{ij}, \partial_k, \partial_r\rangle$ with the homogenization variable h (see, e.g., [34, Section 9]). We introduce a new term order $<^h_{(-w,w,0)}$ over $D^h[a,b,c,d]$, which compares the total degree first, $(-w,w,0)$-degree second, otherwise we apply the following block order as a tie breaker: $d \gg r \gg \{a_{pq} \mid i \le j\} \gg \{b_k\} \gg \{c_k\} \gg \{y_k\} \gg \partial_r \gg \{\partial_{ij} \mid i < j\} \gg \{\partial_{ii}\} \gg \{\partial_k\} \gg \{x_{ij} \mid$

$i < j\} \gg \{x_{ii}\} \gg h$. Here, the block $\{b_k\}$ has a lexicographic order so that $b_1 > b_2 > \cdots > b_{n+1}$. The use of this tie breaker is a key of our calculation. Although, the initial monomial $\mathrm{in}_{<^h_{(-w,w,0)}}(\ell)$ is an element of the associated commutative ring, we denote it by the associated element in $D^h[a,b,c,d]$ as long as no confusion arises. For example, we denote $\mathrm{in}_{<^h_{(-w,w,0)}}(\partial_{ij})$ by ∂_{ij} instead of ξ_{ij}.

Proposition 2.3.2. *A Gröbner basis of I'^h with respect to $<^h_{(-w,w,0)}$ is $G'^h = \{A'^h_{pq}, B, C'^h_{ij}, E'^h\}$.*

We need four lemmas for proving the proposition. These can be obtained by a straightforward calculation.

Lemma 2.3.3. *The initial monomials of the generators of I'^h are*

1. $\mathrm{in}_{<^h_{(-w,w,0)}}(A'^h_{pq}) = -a^3_{pq}$,
2. $\mathrm{in}_{<^h_{(-w,w,0)}}(B) = -r^2$,
3. $\mathrm{in}_{<^h_{(-w,w,0)}}(C'^h_{ij}) = -b_i c_i \partial_j$,
4. $\mathrm{in}_{<^h_{(-w,w0)}}(E'^h) = -d^3$.

In particular, they are pairwise coprime expect in the case that the pair $\mathrm{in}_{<^h_{(-w,w,0)}}(C'^h_{ij})$ and $\mathrm{in}_{<^h_{(-w,w,0)}}(C'^h_{ik})$ and the pair $\mathrm{in}_{<^h_{(-w,w,0)}}(C'^h_{ik})$ and $\mathrm{in}_{<^h_{(-w,w,0)}}(C'^h_{jk})$.

Lemma 2.3.4. *The commutators of two generators of I'^h are*

1. $[A'^h_{pq}, A'^h_{rs}] = [A'^h_{pq}, B] = [A'^h_{pq}, C'^h_{ij}] = [A'^h_{pq}, E'^h] = 0$ *(for any p, q, r, s, i, j)*,
2. $[B, C'^h_{ij}] = 0$, $[B, E'^h] = -2hB$,
3. $[C'^h_{ij}, C'^h_{kl}] = 0 \quad (\{i,j\} \cap \{k,l\} = \emptyset)$,
4. $[C'^h_{ij}, C'^h_{jk}] = C'^h_{ki} := -C'^h_{ik}$, $[C'^h_{ij}, C'^h_{ik}] = C'^h_{jk}$, $[C'^h_{ik}, C'^h_{jk}] = C'^h_{ij} \quad (i < j < k)$,
5. $[C'^h_{ij}, E'^h] = 0$.

Lemma 2.3.5. *The following holds for S-pairs of generators of I'^h:*

1. $S(A'^h_{pq}, A'^h_{rs}), S(A'^h_{pq}, B), S(A'^h_{pq}, C'^h_{ij}), S(A'^h_{pq}, E'^h) \longrightarrow^* 0$ *(for any p, q, r, s, i, j)*,
2. $S(B, C'^h_{ij}) \longrightarrow^* 0$, $S(B, E'^h) \longrightarrow^* 0$,
3. $S(C'^h_{ij}, C'^h_{kl}) \longrightarrow^* 0 \quad (\{i,j\} \cap \{k,l\} = \emptyset)$,
4. $S(C'^h_{ij}, C'^h_{jk}) \longrightarrow^* 0 \quad (i < j < k)$,
5. $S(C'^h_{ij}, E'^h) \longrightarrow^* 0$.

Proof. Lemma 2.2.3 holds in the homogenized Weyl algebra $D^h[a,b,c,d]$ too. Therefore, these follow from Lemmas 2.3.3 and 2.3.4. $\square$

Lemma 2.3.6. *We put $\hat{C}'^h_{ij} := b_jc_j\partial_i - b_ic_i\partial_j$ and $\check{C}'^h_{ij} := C'^h_{ij} - \hat{C}'^h_{ij}$. Then the following cyclic relations hold:*

1. $\partial_k\hat{C}'^h_{ij} + \partial_i\hat{C}'^h_{jk} + \partial_j\hat{C}'^h_{ki} = \hat{C}'^h_{ij}\partial_k + \hat{C}'^h_{jk}\partial_i + \hat{C}'^h_{ki}\partial_j = 0,$

2. $\partial_k\check{C}'^h_{ij} + \partial_i\check{C}'^h_{jk} + \partial_j\check{C}'^h_{ki} = \check{C}'^h_{ij}\partial_k + \check{C}'^h_{jk}\partial_i + \check{C}'^h_{ki}\partial_j = 0,$

3. $\partial_kC'^h_{ij} + \partial_iC'^h_{jk} + \partial_jC'^h_{ki} = C'^h_{ij}\partial_k + C'^h_{jk}\partial_i + C'^h_{ki}\partial_j = 0,$

4. $b_kc_k\hat{C}'^h_{ij} + b_ic_i\hat{C}'^h_{jk} + b_jc_j\hat{C}'^h_{ki} = \hat{C}'^h_{ij}b_kc_k + \hat{C}'^h_{jk}b_ic_i + \hat{C}'^h_{ki}b_jc_j = 0.$

Proof of Proposition 2.3.2. By Lemma 2.3.5, we only need to check that the S-pairs $S(C'^h_{ij}, C'^h_{ik})$ and $S(C'^h_{ik}, C'^h_{jk})$ are reduced to zero.

The former is $S(C'^h_{ij}, C'^h_{ik}) = \partial_kC'^h_{ij} - \partial_jC'^h_{ik} = \partial_kC'^h_{ij} + \partial_jC'^h_{ki}$, because the initial monomials are $\mathrm{in}_{<^h_{(-w,w,0)}}(C'^h_{ij}) = -b_ic_i\partial_j$ and $\mathrm{in}_{<^h_{(-w,w,0)}}(C'^h_{ik}) = -b_ic_i\partial_k$. The S-pair is equal to $-\partial_iC'^h_{jk}$ by the 3rd formula of Lemma 2.3.6. This implies that it is reduced to zero.

The latter is $S(C'^h_{ik}, C'^h_{jk}) = b_jc_jC'^h_{ik} - b_ic_iC'^h_{jk}$, because the initial monomials are $\mathrm{in}_{<^h_{(-w,w,0)}}(C'^h_{ik}) = -b_ic_i\partial_k$ and $\mathrm{in}_{<^h_{(-w,w,0)}}(C'^h_{jk}) = -b_jc_j\partial_k$.

Firstly, we show that it can be expressed as

$$\begin{aligned} S(C'^h_{ik}, C'^h_{jk}) = &\left(\sum_{s=1}^{n+1}(\delta_{ks}+1)x_{ks}\partial_s + hy_k + b_kc_k\right)C'^h_{ij} \\ &+ \left(\sum_{s=1}^{n+1}(\delta_{is}+1)x_{is}\partial_s + hy_i\right)C'^h_{jk} \\ &+ \left(\sum_{s=1}^{n+1}(\delta_{js}+1)x_{js}\partial_s + hy_j\right)C'^h_{ki}. \end{aligned} \tag{2.3}$$

Here, the symbol δ is the Kronecker delta. This expression of the S-pair is a key of our proof. Since each monomial of the left hand side (LHS in short) has at least one variable in b_i, b_j or b_k, we divide the right hand side (RHS in short) into two parts: S_1 contains these variables, S_2 does not contain them. Then,

$$\begin{aligned} S_2 = &\left(\sum_{s=1}^{n+1}(\delta_{ks}+1)x_{ks}\partial_s + hy_k\right)\check{C}'^h_{ij} + \left(\sum_{s=1}^{n+1}(\delta_{is}+1)x_{is}\partial_s + hy_i\right)\check{C}'^h_{jk} \\ &+ \left(\sum_{s=1}^{n+1}(\delta_{js}+1)x_{js}\partial_s + hy_j\right)\check{C}'^h_{ki} \end{aligned}$$

is equal to zero by a straightforward calculation and we have

$$\begin{aligned}
S_1 &= b_k c_k C'^h_{ij} + \left(\sum_{s=1}^{n+1}(\delta_{ks}+1)x_{ks}\partial_s + hy_k\right)\hat{C}'^h_{ij} \\
&\quad + \left(\sum_{s=1}^{n+1}(\delta_{is}+1)x_{is}\partial_s + hy_i\right)\hat{C}'^h_{jk} + \left(\sum_{s=1}^{n+1}(\delta_{js}+1)x_{js}\partial_s + hy_j\right)\hat{C}'^h_{ki} \\
&= b_k c_k C'^h_{ij} \\
&\quad + b_j c_j \left(\sum_{s=1}^{n+1}(\delta_{ks}+1)x_{ks}\partial_s\partial_i + hy_k\partial_i\right) - b_i c_i \left(\sum_{s=1}^{n+1}(\delta_{ks}+1)x_{ks}\partial_s\partial_j + hy_k\partial_j\right) \\
&\quad + b_k c_k \left(\sum_{s=1}^{n+1}(\delta_{is}+1)x_{is}\partial_s\partial_j + hy_i\partial_j\right) - b_j c_j \left(\sum_{s=1}^{n+1}(\delta_{is}+1)x_{is}\partial_s\partial_k + hy_i\partial_k\right) \\
&\quad + b_i c_i \left(\sum_{s=1}^{n+1}(\delta_{js}+1)x_{js}\partial_s\partial_k + hy_j\partial_k\right) - b_k c_k \left(\sum_{s=1}^{n+1}(\delta_{js}+1)x_{js}\partial_s\partial_i + hy_j\partial_i\right) \\
&= b_j c_j C'^h_{ik} - b_i c_i C'^h_{jk} = S(C'^h_{ik}, C'^h_{jk}).
\end{aligned}$$

Finally, we show that the expression (2.3) is a standard representation. We note that the RHS of (2.3) has three terms, which appear in the first, in the second and in the third lines of (2.3) respectively. We may show that the initial monomial of the LHS is no less than the initial monomials of the three terms of the RHS with respect to the term order $<^h_{(-w,w,0)}$. The initial monomial of the LHS $b_i c_i b_k c_k \partial_j$ coincides with that of the 1st term of the RHS. Moreover, all monomials appearing in the 2nd and the 3rd term of the RHS have the same total degree 5 and $(-w,w,0)$-degree 0, and they have a degree at most one with respect to b_i, b_j and b_k. This implies that they are less than $b_i c_i b_k c_k \partial_j$. Thus, we have shown that the expression is a standard representation and the S-pair is reduced to zero.

Since all S-pairs are reduced to zero, the conclusion follows from Buchberger's criterion. □

Proof of Theorem 2.3.1. Let $I'(a,b,c,d)$ be the left ideal generated by G' in the ring $D[a,b,c,d]$. It follows from Proposition 2.3.2 and [37, Theorem 1.2.4] that $G'^h|_{h=1} = G'$ is a Gröbner basis of $I'(a,b,c,d)$ with respect to $<_{(-w,w,0)}$. Therefore, we have

$$\mathrm{in}_{<_{(-w,w,0)}}(I'(a,b,c,d)) = \langle \mathrm{in}_{<_{(-w,w,0)}}(G') \rangle = \langle \tilde{G}' \rangle \quad \text{in } D[a,b,c,d].$$

We may replace $D[a,b,c,d]$ by $D(a,b,c,d) = \mathbf{C}(a,b,c,d)\langle x_{ij}, y_k, r, \partial_{ij}, \partial_k, \partial_r \rangle$. Here, $\mathbf{C}(a,b,c,d)$ is the rational function field with variables $a = (a_{pq}), b = (b_i), c = (c_i)$ and d. Then, the following holds:

$$\mathrm{in}_{(-w,w)}(I'(a,b,c,d)) = \langle \tilde{G}' \rangle \quad \text{in } D(a,b,c,d).$$

This implies the conclusion. □

Proof of the Main theorem 2.1.1. It follows from (2.2) that $\mathrm{rank}\,(I) \geq \mathrm{rank}\,(\mathrm{in}_{(-w,w)}(I'))$ by Theorem 2.2.1 in [37]. Therefore, we have $\mathrm{rank}\,(I) \geq 2n+2$ by Theorem 2.2.1. The opposite inequality follows from Theorem 3 of [28]. □

2.4 A Proof in the case of $n = 1$

In order to clarify ideas of the proof in Section 2.3, we present a proof of our theorem in the case of $n = 1$.

The 1-dimensional Fisher-Bingham system of differential equations $I \subset \mathbf{C}\langle x_{11}, x_{12}, x_{22}, y_1, y_2, r, \partial_{11}, \partial_{12}, \partial_{22}, \partial_1, \partial_2, \partial_r\rangle$ is

$$\begin{aligned}
I = \langle & A_{11} = \partial_{11} - \partial_1^2,\ A_{12} = \partial_{12} - \partial_1\partial_2, \\
& A_{22} = \partial_{22} - \partial_2^2, \\
& B = \partial_1^2 + \partial_2^2 - r^2, \\
& C_{12} = x_{12}\partial_1^2 + 2(x_{22} - x_{11})\partial_1\partial_2 - x_{12}\partial_2^2 + y_2\partial_1 - y_1\partial_2, \\
& E = r\partial_r - 2(x_{11}\partial_1^2 + x_{12}\partial_1\partial_2 + x_{22}\partial_2^2) - (y_1\partial_1 + y_2\partial_2) - 1\rangle.
\end{aligned}$$

The upper bound of $\mathrm{rank}(I)$ is $2 \cdot 1 + 2 = 4$ as given in [28, Theorem 3]. We will show the lower bound of the $\mathrm{rank}(I)$ is 4 by using the following general inequality [37, Theorem 2.2.1]:

$$\mathrm{rank}(I) \geq \mathrm{in}_{(-w,w)}(I).$$

Firstly, we make some change of variables, because it seems to be difficult to calculate $\mathrm{in}_{(-w,w)}(I)$ directly. Let $a_{11}, a_{12}, a_{22}, b_1, b_2, c_1, c_2, d$ be constants, which will be used as slack variables. We put $g = r^{d^3} x_{11}^{a_{11}^3} x_{12}^{a_{12}^3} x_{22}^{a_{22}^3} f$ where the function f is a solution of the system of differential equations $I \cdot f = 0$. Moreover, we make a change of variables y_1, y_2 by $y_1 + b_1c_1, y_2 + b_2c_2$ respectively. Then, the system of differential equations for g is given by $I' \cdot g = 0$, where

$$\begin{aligned}
I' = \langle & A'_{11} = x_{11}\partial_{11} - x_{11}\partial_1^2 - a_{11}^3,\ A'_{12} = x_{12}\partial_{12} - x_{12}\partial_1\partial_2 - a_{12}^3, \\
& A'_{22} = x_{22}\partial_{22} - x_{22}\partial_2^2 - a_{22}^3, \\
& B = \partial_1^2 + \partial_2^2 - r^2, \\
& C'_{12} = x_{12}\partial_1^2 + 2(x_{22} - x_{11})\partial_1\partial_2 - x_{12}\partial_2^2 + (y_2 + b_2c_2)\partial_1 - (y_1 + b_1c_1)\partial_2, \\
& E' = r\partial_r - 2(x_{11}\partial_1^2 + x_{12}\partial_1\partial_2 + x_{22}\partial_2^2) \\
& \qquad\qquad - ((y_1 + b_1c_1)\partial_1 + (y_2 + b_2c_2)\partial_2) - 1 - d^3\rangle.
\end{aligned}$$

We note that $\mathrm{rank}(I) = \mathrm{rank}(I')$ holds for any set of values of the constants.

Secondly, we calculate $\mathrm{in}_{(-w,w)}(I')$ for the weight vector $w = (0, 1, 0, 0, 0, 0)$ where each weight stands for the variables $\partial_{11}, \partial_{12}, \partial_{22}, \partial_1, \partial_2, \partial_r$ respectively. In order to perform Buchberger's algorithm with respect to the $(-w, w)$-weight order, we need to consider the homogenized system I'^h for I':

$$\begin{aligned}
I'^h = \langle & A'^h_{11} = hx_{11}\partial_{11} - x_{11}\partial_1^2 - \underline{a_{11}^3},\ A'^h_{12} = hx_{12}\partial_{12} - x_{12}\partial_1\partial_2 - \underline{a_{12}^3}, \\
& A'^h_{22} = hx_{22}\partial_{22} - x_{22}\partial_2^2 - \underline{a_{22}^3}, \\
& B = \partial_1^2 + \partial_2^2 - \underline{r^2}, \\
& C'^h_{12} = x_{12}\partial_1^2 + 2(x_{22} - x_{11})\partial_1\partial_2 - x_{12}\partial_2^2 + (y_2 h + b_2 c_2)\partial_1 - (y_1 h + \underline{b_1 c_1)\partial_2}, \\
& E'^h = hr\partial_r - 2(x_{11}\partial_1^2 + x_{12}\partial_1\partial_2 + x_{22}\partial_2^2) \\
& \qquad\qquad - ((hy_1 + b_1c_1)\partial_1 + (hy_2 + b_2c_2)\partial_2) - h^3 - \underline{d^3}\rangle.
\end{aligned}$$

We denote by $<^h_{(-w,w,0)}$ an order in the homogenized Weyl algebra which compares the total degree first, $(-w, w, 0)$-degree second, otherwise we apply the following block order as a tie breaker: $d \gg r \gg \{a_{11}, a_{12}, a_{22}\} \gg \{b_1 > b_2\} \gg \{c_1, c_2\} \gg \{y_1, y_2\} \gg \partial_r \gg \{\partial_{12}\} \gg \{\partial_{11}, \partial_{22}\} \gg \{\partial_1, \partial_2\} \gg \{x_{12}\} \gg \{x_{11}, x_{22}\} \gg h$. Here, the symbol $>$ represents the lexicographic order. The underlined parts in I'^h are initial terms with respect to $<^h_{(-w,w,0)}$. They are pairwise coprime and their commutators are equal to zero except $[B, E'^h] = -2hB$. From Lemma 2.2.3, we conclude that the set $\{A'^h_{11}, A'^h_{12}, A'^h_{22}, B, C'^h_{12}, E'^h\}$ is a Gröbner basis of I'^h in $D^h[a_{11}, a_{12}, a_{22}, b_1, b_2, c_1, c_2, d]$ with respect to $<^h_{(-w,w,0)}$. In other words, the transformation of dependent and independent variables gives us the Gröbner basis without adding new elements. Dehomogenizing I'^h, we obtain

$$\begin{aligned}
\mathrm{in}_{(-w,w)}(I') = \langle & \tilde{A}'_{11} = x_{11}\partial_{11} - x_{11}\partial_1^2 - a_{11}^3,\ \tilde{A}'_{12} = x_{12}\partial_{12} - a_{12}^3, \\
& \tilde{A}'_{22} = x_{22}\partial_{22} - x_{22}\partial_2^2 - a_{22}^3, \\
& B = \partial_1^2 + \partial_2^2 - r^2, \\
& \tilde{C}'_{12} = 2(x_{22} - x_{11})\partial_1\partial_2 + (y_2 + b_2c_2)\partial_1 - (y_1 + b_1c_1)\partial_2, \\
& \tilde{E}' = r\partial_r - 2(x_{11}\partial_1^2 + x_{22}\partial_2^2) \\
& \qquad\qquad - ((y_1 + b_1c_1)\partial_1 + (y_2 + b_2c_2)\partial_2) - 1 - d^3\rangle.
\end{aligned}$$

In this calculation, we regard $a_{11}, a_{12}, a_{22}, b_1, b_2, c_1, c_2, d$ as ring variables with the weight 0. As in the proof of Theorem 2.3.1, the equation above holds when $a_{11}, \dots, d$ are specialized to generic complex numbers. The holonomic rank of the $(-w, w)$-initial ideal $\tilde{I}' := \mathrm{in}_{(-w,w)}(I')$ coincides with that of the diagonal system transformed by the same change of variables for I' from I. The holonomic rank of $\tilde{I}'$ agrees with that of

$$\begin{aligned}
\tilde{I} = \langle & \tilde{A}_{11} = \underline{\partial_{11}} - \partial_1^2,\ \tilde{A}_{12} = \underline{\partial_{12}}, \\
& \tilde{A}_{22} = \underline{\partial_{22}} - \partial_2^2, \\
& B = \underline{\partial_1^2} + \partial_2^2 - r^2, \\
& \tilde{C}_{12} = 2(x_{22} - x_{11})\underline{\partial_1\partial_2} + y_2\partial_1 - y_1\partial_2, \\
& \tilde{E} = r\underline{\partial_r} - 2(x_{11}\partial_1^2 + x_{22}\partial_2^2) - (y_1\partial_1 + y_2\partial_2) - 1\rangle.
\end{aligned}$$

Hence, we obtain the following inequality:

$$\mathrm{rank}(I) = \mathrm{rank}(I') \geq \mathrm{rank}(\mathrm{in}_{(-w,w)}(I')) = \mathrm{rank}(\tilde{I}') = \mathrm{rank}(\tilde{I}).$$

Finally, we show that $\mathrm{rank}(\tilde{I}) = 4$. Proposition 2.2.2 tells us that the set

$$\begin{aligned} &\{\tilde{A}_{11}, \tilde{A}_{12}, \tilde{A}_{22}, B, \tilde{C}_{12}, \tilde{E}, \\ &\quad D_2 = 2(x_{22} - x_{11}) \left(\underline{\partial_2^3} - r^2 \partial_2 - \frac{y_2 \partial_1^2 - y_1 \partial_1 \partial_2 - \partial_2}{2(x_{22} - x_{11})} \right) \Bigg\}. \end{aligned}$$

is a Gröbner basis of $R\tilde{I}$ with respect to the block order $\{\partial_r\} \gg \{\partial_{11} > \partial_{22}\} \gg \{\partial_1 > \partial_2\}$, where the tie breaker $>$ represents the graded lexicographic order. The set of standard monomials is $\{1, \partial_1, \partial_2, \partial_2^2\}$. It means that the holonomic rank of $\tilde{I}$ is 4.

Chapter 3

Holonomic Gradient Descent for the Fisher-Bingham Distribution

3.1 Introduction

Let $x = (x_{ij})$ and $y = (y_i)$ be a symmetric matrix parameter of size $(d+1) \times (d+1)$ and a vector parameter of size $d+1$, respectively. We are interested in the Fisher-Bingham probability distribution

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

on the d-dimensional sphere $S^d(r) = \{(t_1, \ldots, t_{d+1}) \mid \sum_{i=1}^{d+1} t_i^2 = r^2, r > 0\}$ and the maximum likelihood estimate of the parameters x and y of this probability distribution. Here, the function Z is the normalizing constant defined as

$$Z(x, y, r) = \int_{S^d(r)} \exp\left(\sum_{1 \le i \le j \le d+1} x_{ij} t_i t_j + \sum_{i=1}^{d+1} y_i t_i \right) |dt| \qquad (3.1)$$

and $|dt|$ denotes the standard measure on the sphere with the radius r such that $\int_{S^d(r)} |dt| = r^d \frac{2\pi^{(d+1)/2}}{\Gamma((d+1)/2)}$.

Solving the MLE problem involves finding the maximum of the function in x, y

$$\prod_{k=1}^{N} \mu(T^{(k)}; x, y, 1)$$

for given data vectors $T^{(k)}$, $k = 1, \dots, N$ in the t-space $S^d(1)$. In order to compute the MLE, we need approximate values for the normalizing constant Z and its derivatives. In the case of $d = 2$, the normalizing constant is expressed in terms of the Bessel function and for $d = 2$ there are several approaches for computing MLEs in directional statistics [17, 24, 43]. However, there are few studies on approximating the normalizing constant for the case of $d > 2$ and applications to the MLE. Among these, [22] proposed a method to evaluate the normalizing constant by utilizing the Laplace approximation of the integral for $d > 2$ and [21] gave a series approximation of the normalizing constant.

In this paper, we propose a different method for evaluating it and present applications to the MLE. Our method is based on the holonomic gradient descent (HGD) proposed in [28], which utilizes a holonomic system of linear differential equations satisfied by the normalizing constant and gives the MLE accurately. The HGD consists of four steps. The first step is to derive a holonomic system of linear partial differential equations for the normalizing constant. The second step is to translate the holonomic system into a Pfaffian system, which is roughly speaking a set of ordinary differential equations with respect to the parameters x_{ij} and y_i for the normalizing constant. These two steps can be performed by a *symbolic* computation (the Gröbner basis method) if the size of the problem is moderate. The remaining steps utilize *numerical* computation. The third step is to evaluate the normalizing constant and its derivatives at an initial point. To do this, we can use numerical integration for a rough evaluation or a series expansion for a more accurate evaluation. The last step is to extend the evaluated values to other points needed for the MLE by using the Pfaffian system and a numerical solver of ordinary differential equations.

It is shown in [18] and [28] that the normalizing constant of the Fisher-Bingham distribution is a holonomic function in x, y, r and further more it is annihilated by a holonomic ideal for which an explicit expression is given. This is the first step when applying the HGD. For the second step, we need to translate the ideal into a Pfaffian system. This is performed on a computer for $d \leq 2$ in [28]; however, this is not possible for $d > 2$ on current computers using Gröbner basis algorithms due to the high computational complexity.

In this paper, we overcome the difficulty of high complexity and complete the remaining steps for a general dimension: we give an accelerated version of the HGD as a general method, we derive the Pfaffian system of the Fisher-Bingham distribution for general *d by hand*, and we derive a series expansion of the normalizing constant with an error estimation. We demonstrate that the accelerated version of HGD on our Pfaffian system works well up to $d = 7$ to a specified accuracy for a certain class of problems. We also propose a general method for evaluating the numerical errors and apply it to the Fisher-Bingham distribution.

3.2 Holonomic Gradient Descent with Pfaffian System of Factored Form

The HGD introduced in [28] is a general algorithm for solving MLE problems for holonomic unnormalized distributions accurately. We herein propose an accelerated version of the HGD. The modification is small, but it allows a drastic performance improvement as we will see in the case of the Fisher-Bingham distribution.

In this section, we maintain a general setting to explain our accelerated method. A function $f(y_1, \ldots, y_n)$ is called a *holonomic function* when it satisfies an ordinary differential equation with polynomial coefficients for each variable y_i. In other words, the function f is a holonomic function when the function is annihilated by an ordinary differential operator of the form

$$\sum_{j=0}^{r_i} a_{ij}(y) \left(\frac{\partial}{\partial y_i}\right)^j, \quad a_{ij}(y) \in \mathbf{C}[y_1, \ldots, y_n].$$

By virtue of this ordinary differential operators, the function f can be regarded as a solution of a Pfaffian system discussed below. An important property of holonomic functions is that the integral $\int_R f(y_1, \ldots, y_n) dy_n$ is a holonomic function with respect to $y_1, \ldots, y_{n-1}$ under a suitable condition on the integration domain R. In the landmark paper by [44], he introduced the notion of holonomic functions and applied it to mechanically prove special function identities with systems of linear partial differential equations and this important property. It has been proved that the normalizing constant Z of the Fisher-Bingham distribution is a holonomic function in x, y, r ([28]). Let $f(t; \theta)$ be a holonomic unnormalized probability distribution with respect to t and θ where $\theta = (\theta_1, \ldots, \theta_m)$ is a parameter vector and let $Z(\theta) = \int_U f(t; \theta) dt$ be the normalizing constant. Let I be a holonomic ideal in the ring of differential operators in θ that annihilates Z. The operator $\partial/\partial\theta_i$ is denoted by ∂_{θ_i}. In order to apply the HGD to MLE problems, we need an explicit expression for the Pfaffian system associated with the holonomic ideal I as an input to a numerical solver. The Pfaffian system is used to numerically evaluate the likelihood function and its gradient or its Hessian. Let us review the definition of the Pfaffian system (see, e.g., [28] for details). Let $\mathrm{rank}(I)$ be the holonomic rank of the ideal I and let F be a vector of the standard monomials of a Gröbner basis of I. The length of this vector is $\mathrm{rank}(I)$. We denote the elements of F by ∂^α where $\alpha \in S$. We assume that the first element of F is $\partial^0 = 1$. The Pfaffian system is a set of differential operators which annihilate the vector-valued function $F(Z) = (\partial^\alpha Z \,|\, \alpha \in S)^T$ that are of the form $\partial_{\theta_i} - P_i$ where P_i are $\mathrm{rank}(I) \times \mathrm{rank}(I)$ matrices with rational function entries which satisfy

$$\nabla \circ \nabla = 0, \quad \nabla = d - \sum P_i d\theta_i.$$

In this section, the symbol d in the definition of ∇ indicates the exterior derivative with respect to the variables θ. In some of the literatures, the definition of the Pfaffian system does not include the integrability condition $\nabla \circ \nabla = 0$, but we will call the integrable Pfaffian system of equations simply the Pfaffian system for short. Our Pfaffian system can be regarded as the integrable connection ∇.

The Pfaffian system can be obtained by an algorithmic method explained, e.g., in [28]. However, it requires heavy computation. For example, the computation for the Fisher-Bingham distribution could be performed within a reasonable time using current computer technology only up to the case of a 2-dimensional sphere. Moreover, the Pfaffian system obtained with this method requires heavy numerical computation in the HGD, because in general the entries of P_i are huge rational functions. The latter drawback is removed by using our accelerated version of HGD introduced below.

Algorithm 3.2.1.

1. Construct a Pfaffian system of the form
$$\partial_{\theta_i} - R_i^{-1}(\theta)Q_i(\theta) \tag{3.2}$$
where Q_i and R_i are $\mathrm{rank}(I) \times \mathrm{rank}(I)$ matrices with *polynomial* function entries.
2. Evaluate the normalizing constant $F(Z)$ at an initial parameter θ^0.
3. $k = 0$
4. Evaluate the gradient of the likelihood function by $F(Z)$ at θ^k (see [28]). If the gradient is 0, then stop. Determine the value of the new parameter θ^{k+1} by standard procedures of gradient descent. θ^{k+1} must be sufficiently close to θ^k.
5. Evaluate the approximate value of $F(Z)$ at θ^{k+1}. It is, for instance, approximately equal to
$$F(Z)(\theta^k) + \sum_{i=1}^{d} R_i(\theta^k)^{-1}Q_i(\theta^k)F(Z)(\theta^k) \cdot (\theta_i^{k+1} - \theta_i^k). \tag{3.3}$$
6. $k \to k+1$ and go to 4.

We call the Pfaffian system of the form (3.2) the Pfaffian system *of factored form*. The main difference between the HGD in [28] and our proposed method is (3.3). In the original version, the factored matrix $R_i^{-1}Q_i$ is expressed as a single matrix with entries of (huge) rational functions, but in our proposed method, we express it as the two matrices Q_i and R_i and the inverse of R_i is calculated numerically in each iteration step. We note that the approximation in (3.3) should be replaced with a more accurate and efficient numerical scheme such as the Runge-Kutta method.

Remark 3.2.2. When we have a Gröbner basis of I, the Pfaffian system of the form (3.2) can be obtained by the computation of normal forms by using the Gröbner basis and then solving linear equations in the ring of polynomials.

This procedure is general, but it can require a huge amount of computational resources. When we apply this method to problems, we need to find shortcuts based on individual problems in order to solve the problems efficiently. We will do this for the Fisher-Bingham distribution in the next section.

Remark 3.2.3. The matrix $R_i^{-1}Q_i$ may have the form

$$\sum_j R_{ij}^{-1}Q_{ij}T_{ij}^{-1}S_{ij}\cdots \tag{3.4}$$

where R_{ij}, Q_{ij}, S_{ij}, T_{ij}, $\ldots$ are $\mathrm{rank}(I)\times\mathrm{rank}(I)$ matrices in polynomial function entries. This form is referred to as the multi-factored form or sometimes simply factored form.

In the following, we will sometimes call the accelerated version of HGD simply the HGD.

3.3 Pfaffian System for the Normalizing Constant

It is shown in [28] and [18] that the normalizing constant Z in (3.1) of the Fisher-Bingham distribution is a holonomic function in x, y, r and furthermore that it is annihilated by a holonomic ideal I. The holonomic ideal I is generated by the following operators in the ring of differential operators.

$$\partial_{ij} - \partial_i\partial_j \quad (1 \le i \le j \le d+1), \tag{3.5}$$

$$\sum_{i=1}^{d+1}\partial_i^2 - r^2, \tag{3.6}$$

$$x_{ij}\partial_i^2 + 2(x_{jj}-x_{ii})\partial_i\partial_j - x_{ij}\partial_j^2 + \sum_{1\le k\le d+1, k\ne i,j}(x_{kj}\partial_i\partial_k - x_{ik}\partial_j\partial_k) + y_j\partial_i - y_i\partial_j \quad (1 \le i < j \le d+1), \tag{3.7}$$

$$r\partial_r - 2\sum_{1\le i\le j\le d+1} x_{ij}\partial_i\partial_j - \sum_{i=1}^{d+1} y_i\partial_i - d \tag{3.8}$$

Here, ∂_{ij}, ∂_i, and ∂_r stand for $\frac{\partial}{\partial x_{ij}}$, $\frac{\partial}{\partial y_i}$, and $\frac{\partial}{\partial r}$ respectively. Note that we assume $x_{ij} = x_{ji}$.

We want to translate these into a Pfaffian system of the form (3.2) or (3.4) which is used in the accelerated HGD explained in the previous section.

Before proceeding to the discussion of the general d-dimensional case, we illustrate our method in the case of $d = 1$ and $r = 1$. Let I_1 be the left ideal generated by

$$\partial_{11} - \partial_1^2, \quad \partial_{12} - \partial_1\partial_2, \quad \partial_{22} - \partial_2^2, \tag{3.9}$$

$$\partial_1^2 + \partial_2^2 - 1, \tag{3.10}$$

$$x_{12}\partial_1^2 + 2(x_{22}-x_{11})\partial_1\partial_2 - x_{12}\partial_2^2 + y_2\partial_1 - y_1\partial_2 \tag{3.11}$$

in the ring of differential operators. The holonomic rank of I_1 is 4. Let F be a vector of operators $(1, \partial_1, \partial_2, \partial_1^2)^T$. We want to find a matrix P whose entries are rational functions such that $\partial_1 F \equiv PF$ holds modulo the left ideal I_1. Here, $s \equiv t$ means that each element of $s - t$ belongs to I_1. Since $\partial_1 F = (\partial_1, \partial_1^2, \partial_1\partial_2, \partial_1^3)^T$, we need to express $\partial_1\partial_2$ and ∂_1^3 in terms of F modulo I_1. Eliminating ∂_2^2 from (3.11) by (3.10), we obtain

$$\begin{aligned} 2(x_{22} - x_{11})\partial_1\partial_2 &\equiv -x_{12}\partial_1^2 + x_{12}\underline{\partial_2^2} - y_2\partial_1 + y_1\partial_2 \\ &\equiv -x_{12}\partial_1^2 + x_{12}(1 - \partial_1^2) - y_2\partial_1 + y_1\partial_2 \\ &\qquad \text{(the underlined term is reduced by (3.10))} \\ &= (x_{12}, -y_2, y_1, -2x_{12})F. \end{aligned} \tag{3.12}$$

Thus, we have expressed $\partial_1\partial_2$ in terms of F. We now try to express ∂_1^3 in terms of F. From $\partial_1 \times$ (3.11), we obtain

$$\begin{aligned} &x_{12}\partial_1^3 + 2(x_{22} - x_{11})\partial_1^2\partial_2 \equiv x_{12}\partial_1\underline{\partial_2^2} - y_2\partial_1^2 + y_1\underline{\partial_1\partial_2} + \partial_2 \\ &\equiv x_{12}\partial_1(1 - \partial_1^2) - y_2\partial_1^2 + \frac{y_1}{2(x_{22} - x_{11})}(x_{12} - y_2\partial_1 + y_1\partial_2 - 2x_{12}\partial_1^2) + \partial_2 \\ &\qquad \text{(the underlined terms are reduced by (3.10) and (3.12))}, \end{aligned}$$

and consequently we have $2x_{12}\partial_1^3 + 2(x_{22} - x_{11})\partial_1^2\partial_2 \equiv (a, b, c, d)F$, where

$$\begin{aligned} a &= \frac{y_1 x_{12}}{2(x_{22} - x_{11})}, \quad b = x_{12} - \frac{y_1 y_2}{2(x_{22} - x_{11})}, \\ c &= 1 + \frac{y_1^2}{2(x_{22} - x_{11})}, \quad d = -y_2 - \frac{x_{12} y_1}{x_{22} - x_{11}}. \end{aligned}$$

By a similar computation for $\partial_2 \times$ (3.11), we have

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

where

$$\begin{aligned} a' &= -y_1 + \frac{x_{12} y_2}{2(x_{22} - x_{11})}, \quad b' = 1 + 2(x_{22} - x_{11}) - \frac{y_2^2}{2(x_{22} - x_{11})}, \\ c' &= -x_{12} + \frac{y_1 y_2}{2(x_{22} - x_{11})}, \quad d' = y_1 - \frac{x_{12} y_2}{x_{22} - x_{11}}. \end{aligned}$$

Therefore, we have

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

Multiplying the both sides by the inverse matrix $\begin{pmatrix} 2x_{12} & 2(x_{22} - x_{11}) \\ 2(x_{22} - x_{11}) & -2x_{12} \end{pmatrix}^{-1}$, we can express ∂_1^3 in terms of F. Thus we have obtained a factored form of P in

$\partial_1 F \equiv PF$. The identity $\partial_1 F \equiv PF$ gives a Pfaffian equation for the direction y_1. In other words, the differential equation

$$\frac{\partial F(Z)}{\partial y_1} = PF(Z), \quad F(Z) = \left(Z, \frac{\partial Z}{\partial y_1}, \frac{\partial Z}{\partial y_2}, \frac{\partial^2 Z}{\partial y_1^2} \right)^T$$

holds. This is an ordinary differential equation for the vector-valued function $F(Z)$ with respect to the variable y_1. It is easy to see that P is of the form (3.4). Ordinary differential equations for the other directions $\partial_2, \partial_{11}, \partial_{12}, \partial_{22}$ can be obtained analogously.

For the general d, let F be the vector of operators

$$(1, \partial_1, \ldots, \partial_{d+1}, \partial_1^2, \ldots, \partial_d^2)^T. \tag{3.13}$$

Theorem 3.3.1. *There exists a* $(2d+2) \times (2d+2)$ *matrix* H_i *which has a factored form (3.4) and satisfies the relation* $\partial_i F \equiv H_i F \bmod I$.

An expression of H_i as a factored form and a proof of this theorem, which is technical, will be given in the Appendix.

The relation between $\partial_{ij} F$ and F can be easily obtained by Theorem 3.3.1. In fact, since $\partial_{ij} F \equiv \partial_i \partial_j F \equiv \partial_j \partial_i F$ by (3.5), we have

$$\partial_{ij} F \equiv \partial_j \partial_i F \equiv \partial_j (H_i F) \equiv \frac{\partial H_i}{\partial y_j} F + H_i (\partial_j F) \equiv \left(\frac{\partial H_i}{\partial y_j} + H_i H_j \right) F. \tag{3.14}$$

We denote by H_{ij} the matrix $\frac{\partial H_i}{\partial y_j} + H_i H_j$. The matrix such that $\partial_r F \equiv H_r F$ can be obtained easily by utilizing (3.8). Thus, we have obtained the relations

$$\partial_i F \equiv H_i F, \quad \partial_{ij} F \equiv H_{ij} F, \quad \partial_r F \equiv H_r F. \tag{3.15}$$

In [20], we prove that the holonomic rank of I is equal to $2d+2$. Therefore, the Pfaffian equations are expressed in terms of $(2d+2) \times (2d+2)$ matrices. The matrices in (3.15) are exactly these matrices. The integrability conditions of Pfaffian equations imply $\frac{\partial H_i}{\partial y_j} + H_i H_j = \frac{\partial H_j}{\partial y_i} + H_j H_i$.

In [28], the differential equations satisfied by the likelihood function for $d = 1$ and $d = 2$ are derived by a heavy Gröbner basis computation and we could not obtain them for $d \geq 3$. It is known that the Gröbner basis computation has the double-exponential complexity with respect to the number of variables (see, e.g., [5]) and we usually have to avoid deriving Gröbner bases by a computer for large problems. Instead, we can sometimes derive Gröbner bases by hand and apply them to interesting applications. By virtue of Theorem 3.3.1 for the general dimension, we can describe the differential equation satisfied by the likelihood function with matrices in factored form of which factors are relatively small matrices with polynomial entries. If we calculate the inverse matrices in the factored forms by symbolic computation, we would obtain the same result with the Gröbner basis method. In order to apply for the HGD, we do not need to calculate these inverse matrices with polynomial entries symbolically, instead we need only calculate inverse matrices numerically when variables are restricted to real number values in each step of the Runge-Kutta method. This will become a key ingredient of our algorithm, which will be discussed in section 3.6.

Remark 3.3.2. The matrices H_i, H_{ij}, H_r have simple forms when x is a diagonal matrix. In [40], the MLE of the Fisher distribution on $SO(3)$ is obtained by the HGD with differential equations for the normalizing constant with diagonalized arguments. It is a natural question to ask whether a simplification analogous to the diagonal x case is possible. Unfortunately, an analog of Lemma 2 of [40] does not hold except for the case of $y = 0$. It is possible to evaluate the gradient of Z from values of Z for diagonal x by Proposition 1 given later, however this requires computation of a transformation matrix to diagonalize the matrix x at each step of the gradient descent. On the other hand, we do not need to do this computation for the diagonal form in the HGD by the Pfaffian system for the full parameters x, y, r.

3.4 Series Expansion for the Normalizing Constant

Let us define the function $\tilde{Z}$ by the integral

$$\tilde{Z}(\tilde{x}, \tilde{y}, \tilde{r}) = \int_{S^d(r)} \exp\left(\sum_{i=1}^{d+1} (\tilde{x}_i t_i^2 + \tilde{y}_i t_i)\right) |dt|. \tag{3.16}$$

The function satisfies the invariance relation

$$\tilde{Z}(\tilde{x}, \tilde{y}, 1) = \tilde{Z}(r^{-2}\tilde{x}, r^{-1}\tilde{y}, r). \tag{3.17}$$

This function is the restriction of the normalizing constant Z to the diagonalized $\tilde{x}$. Since the normalizing constant is invariant under the action of the orthogonal group $O(d+1)$, we can express $F(Z)$ in terms of $F(\tilde{Z})$. The following proposition can be obtained by a straight forward calculation.

Proposition 3.4.1. *Suppose that the real symmetric matrix x is diagonalized by an orthogonal matrix $P = (p_{ij})$, and set $\tilde{x} = P^T x P$, $\tilde{y} = P^T y$, $\tilde{r} = r$. Then, we have*

$$\begin{aligned}
Z(x, y, r) &= \tilde{Z}(\tilde{x}, \tilde{y}, \tilde{r}) \\
\frac{\partial Z}{\partial y_i}(x, y, r) &= \sum_{k=1}^{d+1} p_{ik} \frac{\partial \tilde{Z}}{\partial \tilde{y}_k}(\tilde{x}, \tilde{y}, \tilde{r}) \\
\frac{\partial^2 Z}{\partial y_i^2}(x, y, r) &= \sum_{k=1}^{d+1} p_{ik}^2 \frac{\partial^2 \tilde{Z}}{\partial \tilde{y}_k^2}(\tilde{x}, \tilde{y}, \tilde{r}) \\
&\quad - \sum_{1 \le k < \ell \le d+1} \frac{p_{ik} p_{i\ell}}{\tilde{x}_k - \tilde{x}_\ell} \left(\tilde{y}_k \frac{\partial \tilde{Z}}{\partial \tilde{y}_\ell}(\tilde{x}, \tilde{y}, \tilde{r}) - \tilde{y}_\ell \frac{\partial \tilde{Z}}{\partial \tilde{y}_k}(\tilde{x}, \tilde{y}, \tilde{r}) \right).
\end{aligned}$$

[21] give a series approximation of the Fisher-Bingham distribution and consequently that of the normalizing constant $\tilde{Z}(\tilde{x}, \tilde{y}, 1)$. Their expression is easily rescaled to the case including the parameter r. We will give an error estimate of this series approximation. We will omit the tilde symbol ('$\tilde{}$') for x and y in the following where this should cause no confusion.

Theorem 3.4.2. *1.* ([21]) *The restricted normalizing constant has the following series expansion*

$$\tilde{Z}(x, y, r) = S_d \cdot \sum_{\alpha,\beta \in \mathbf{N}_0^{d+1}} r^{d+2|\alpha+\beta|} \frac{(d-1)!! \prod_{i=1}^{d+1} (2\alpha_i + 2\beta_i - 1)!!}{(d-1+2|\alpha|+2|\beta|)!! \alpha! (2\beta)!} x^\alpha y^{2\beta}. \tag{3.18}$$

Here, $S_d = \int_{S^d(1)} |dt|$ denotes the surface area of the d-sphere of radius 1, and $\mathbf{N}_0 = \{0, 1, 2, \ldots\}$. For a multi-index $\alpha \in \mathbf{N}_0^{d+1}$, we put $\alpha! = \prod_{i=1}^{d+1} \alpha_i!$, $\alpha!! = \prod_{i=1}^{d+1} \alpha_i!!$, and $|\alpha| = \sum_{i=1}^{d+1} \alpha_i$.

2. The truncation error of the series is estimated as

$$\begin{aligned} &\left| S_d \cdot \sum_{|\alpha+\beta| \geq N} r^{d+2|\alpha+\beta|} \frac{(d-1)!! \prod_{i=1}^{d+1} (2\alpha_i + 2\beta_i - 1)!!}{(d-1+2|\alpha|+2|\beta|)!! \alpha! (2\beta)!} x^\alpha y^{2\beta} \right| \\ \leq\ & S_d \cdot \frac{r^d}{N!} \left(r^2 \sum_i (|x_i| + |y_i|^2) \right)^N \frac{N+1}{N+1 - r^2 \sum_i (|x_i| + |y_i|^2)} \end{aligned} \tag{3.19}$$

when N is sufficiently large.

We note that the series (3.18) converges slowly when $r^2 \sum_i (|x_i| + |y_i|^2) > 1$ and converges relatively rapidly when $r^2 \sum_i (|x_i| + |y_i|^2) \leq 1$. The derivatives of $\tilde{Z}$ are expressed as derivatives of the right-hand side of (3.18).

Proof of 2. We have the following estimates

$$\begin{aligned} &\left| \sum_{|\alpha+\beta| \geq N} r^{d+2|\alpha+\beta|} \frac{(d-1)!! \prod_{i=1}^{d+1} (2\alpha_i + 2\beta_i - 1)!!}{(d-1+2|\alpha|+2|\beta|)!! \alpha! (2\beta)!} x^\alpha y^{2\beta} \right| \\ \leq\ & \sum_{|\alpha+\beta| \geq N} \left| r^{d+2|\alpha+\beta|} \frac{(d-1)!! \prod_{i=1}^{d+1} (2\alpha_i + 2\beta_i - 1)!!}{(d-1+2|\alpha|+2|\beta|)!! \alpha! (2\beta)!} x^\alpha y^{2\beta} \right| \\ \leq\ & \sum_{|\alpha+\beta| \geq N} \left| r^{d+2|\alpha+\beta|} \frac{1}{\alpha! (2\beta)!} x^\alpha y^{2\beta} \right| \\ \leq\ & \sum_{|\alpha+\beta| \geq N} \left| r^{d+2|\alpha+\beta|} \frac{1}{\alpha! \beta!} x^\alpha y^{2\beta} \right| \leq r^d \sum_{n \geq N} \frac{r^{2n}}{n!} \sum_{|\alpha+\beta|=n} \frac{n!}{\alpha! \beta!} |x|^\alpha |y|^{2\beta} \\ \leq\ & r^d \sum_{n \geq N} \frac{r^{2n}}{n!} \left(\sum_{i=1}^{d+1} (|x_i| + |y_i^2|) \right)^n \leq r^d \sum_{n \geq N} \frac{1}{n!} \left(r^2 \sum_{i=1}^{d+1} (|x_i| + |y_i^2|) \right)^n . \end{aligned}$$

Set $L(x,y,r) = r^2 \sum_{i=1}^{d+1}(|x_i| + |y_i^2|)$. Assume that N is sufficiently large so that $L/(N+1) < 1$. We have the estimate

$$r^d \sum_{n\geq N} \frac{1}{n!} L(x,y,r)^n \leq \frac{r^d}{N!} L(x,y,r)^N \frac{N+1}{N+1-L(x,y,r)} \tag{3.20}$$

by the estimate

$$\begin{aligned}
\sum_{n\geq N} \frac{1}{n!} L^n &= \frac{1}{N!} L^N \sum_{n\geq N} \frac{1}{(N+1)_{n-N}} L^{n-N} \\
&= \frac{1}{N!} L^N \sum_{n=0}^{\infty} \frac{1}{(N+1)_n} L^n \\
&\leq \frac{1}{N!} L^N \sum_{n=0}^{\infty} \frac{1}{(N+1)^n} L^n \\
&= \frac{1}{N!} L^N \frac{N+1}{N+1-L}.
\end{aligned}$$

□

3.5 Numerical Evaluation of the Normalizing Constant

In order to efficiently evaluate $\tilde{Z}$ numerically, we use the holonomic gradient method (HGM) (see, e.g., [15]). The HGM is a method for evaluating the normalizing constant by utilizing a system of differential equations. In the case of the Fisher-Bingham distribution, we numerically evaluate the series (3.4.2) in the domain $r^2 \sum_i (|x_i| + |y_i|^2) \leq 1$ and extend the numerical evaluation outside this domain by using a differential equation with respect to r. To use this method, we prepare the following theorem.

Theorem 3.5.1. *1. The function $\tilde{Z}$ is annihilated by the left ideal $\tilde{I}$ generated by*

$$\begin{aligned}
&A_i = \partial_{y_i}^2 - \partial_{x_i} \quad (1 \leq i \leq d+1), \\
&B = \partial_{y_1}^2 + \cdots + \partial_{y_{d+1}}^2 - r^2, \\
&C_{ij} = 2(x_i - x_j)\partial_{y_i}\partial_{y_j} + y_i\partial_{y_j} - y_j\partial_{y_i} \quad (1 \leq i < j \leq d+1), \\
&E = r\partial_r - 2\sum_{i=1}^{d+1} x_i \partial_{y_i}^2 - \sum_{i=1}^{d+1} y_i \partial_{y_i} - d.
\end{aligned}$$

2. Set $\tilde{F} = (\partial_{y_1}, \ldots, \partial_{y_{d+1}}, \partial_{y_1}^2, \ldots, \partial_{y_{d+1}}^2)^T$. Then, we have $\partial_r \tilde{F} \equiv P^{(r)} \tilde{F} \bmod \tilde{I}$.

Here, the matrix $P^{(r)} = (p_{ij}^{(r)})$ *is defined by*

$$\begin{aligned} rp_{ij}^{(r)} &= (2x_i r^2 + 1)\delta_{ij} + \sum_{k=1}^{d+1} y_i \delta_{j(k+d+1)} \quad (1 \le i \le d+1), \\ rp_{(i+d+1)j}^{(r)} &= y_i r^2 \delta_{ij} + (2x_i r^2 + 2)\delta_{j(i+d+1)} + \sum_{k \ne i} \delta_{j(k+d+1)} \quad (1 \le i \le d+1) \end{aligned}$$

for $1 \le j \le 2d+2$.

The proof of this theorem is analogous to that for the non-diagonal x case.

Example 3.5.2. In the case of $d = 1$, the matrix $P^{(r)}$ is

$$\frac{1}{r}\begin{pmatrix} 2r^2x_1+1 & 0 & y_1 & y_1 \\ 0 & 2r^2x_2+1 & y_2 & y_2 \\ r^2y_1 & 0 & 2r^2x_1+2 & 1 \\ 0 & r^2y_2 & 1 & 2r^2x_2+2 \end{pmatrix}.$$

We note that the largest eigenvalue of $P^{(r)}$ is $O(r)$. Our implementation of the HGM numerically solves the ordinary differential equation

$$\frac{\partial G}{\partial r} = (P^{(r)} - r\lambda E)G \tag{3.21}$$

instead of solving $\partial_r \tilde{F}(\tilde{Z}) = P^{(r)}\tilde{F}(\tilde{Z})$, where λ is the largest eigenvalue of $\lim_{r\to+\infty} P^{(r)}/r$ and the vector-valued function G is defined by $\tilde{F}(\tilde{Z}) = \exp(\lambda r^2/2)G$. This scalar scaling is necessary, because the adaptive Runge-Kutta method requires an absolute error bound of the solution to automatically make meshes finer and when a solution grows exponentially, meshes become too small to maintain an absolute error bound.

Now let us discuss on the accuracy of the HGM. The truncation error of the series approximation is estimated in Theorem 3.4.2. We want to estimate the numerical error caused by applying the HGM. In other words, we want to estimate how much the truncation error is magnified by solving the ordinary differential equation numerically. We propose a practical method to do this. Note that this method can be applied to any HGM, but we will explain it in the case of the Fisher-Bingham distribution. In this method, we assume that initial values are governed by a probability measure. This assumption is natural in, e.g., molecular modeling and some classes of non-linear ordinary differential equations are studied under this assumption (see, e.g., [7]). In our case, the ordinary differential equation is linear and the problem is much easier. Since we have not found a reference relevant to our case, we include below a discussion on the behavior of solutions of a linear equation under random initial data. For a given initial value vector Z_0 at $r = r_0$, we denote by RZ_0 the output obtained at $r = r_1$ by solving the ordinary differential equation where we may suppose that R is a constant matrix, because the Runge-Kutta solver can be regarded

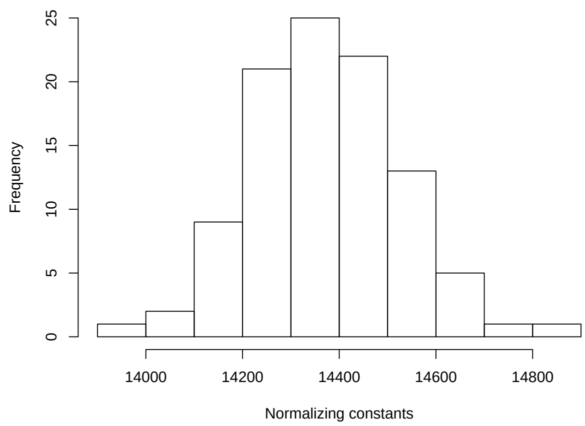

Figure 3.1: Histogram of normalizing constants by the HGM with random initial values

as a linear map from the input to the output under the assumption that round-off errors and cancellation errors by floating point arithmetic are sufficiently small and that the automatic mesh refinement process is fixed. When Z_0 is regarded as a random vector distributed as a multivariate normal distribution, the output RZ_0 is also a random vector distributed as a multivariate normal distribution. In our implementation, we perturb Z_0 with random numbers of which the standard deviation is $\varepsilon/2$ where ε is a truncation error and solve the ordinary differential equation for these perturbed initial values. We evaluate the mean and the standard deviation of the first component of RZ_0 and these give an evaluation of the normalizing constant and its statistical error bound. This bound is more practical than that by interval arithmetic.

For example, Figure 3.1 shows a histogram of the normalizing constant which is generated by the above procedure with $\varepsilon/2 = 0.1$, $r_0 = 1$, $r_1 = \sum |x_i| + \sum y_i^2$, $d = 3$, $x = \mathrm{diag}(1.2, 2.5, 3.2, 3.6)$, $y = (2.3, 5.3, 4.2, 0.1)$. The series is evaluated at $x/r_1^2, y/r_1, r = 1$ and extended to $r = r_1$. In this case, the standard derivation of the normalizing constant is evaluated as 156.6288 and the confidence interval with probability 0.95 is $[14065.6, 14679.6]$

[22] gave a saddle point approximation of the normalizing constant for the Fisher-Bingham distribution. Our method evaluates the normalizing constant with an error bound. Table 3.1 shows values by the HGM and by the third order saddle point approximation by Kume and Wood. Here, $d = 4$ and $(x_{ij}) = \mathrm{diag}(x_{11}, 2x_{11}, 3x_{11}, 4x_{11}, 5x_{11})$, $0.5 \le x_{11} \le 10$, $(y_k) = (0.5y_0, 0.4y_0, 0.3y_0, 0.2y_0, 0.1y_0)$, $y_0 = 3$. The absolute error bound to solve (3.21) by the adaptive Runge-Kutta method is set to $10^{-6}\sum_{i=1}^{2d+2} G_i/(2d+2)$ and ε is 10^{-5}. The values of the stan-

dard deviation imply that the values by the HGM have at least 6-digit accuracy with 95 % confidence.

x_{11}	HGM		Kume-Wood, 3
	HGM	standard deviation	
0.5	189.243	1.737976e-04	$\underline{189}$.763
1.0	985.529	9.102497e-04	$\underline{9}$94.043
1.5	5856.78	5.424156e-03	$\underline{58}$08.16
2.0	39075.8	3.624707e-02	$\underline{3}$7602.6
2.5	287231	2.667160e-01	$\underline{2}$71557
3.0	2.28420e+06	2.122623e+00	$\underline{2}$.15158e+06
3.5	1.93448e+07	1.798630e+01	$\underline{1}$.82924e+07
4.0	1.72236e+08	1.602082e+02	$\underline{1}$.63939e+08
4.5	1.59584e+09	1.484901e+03	$\underline{1.5}$2931e+09
5.0	1.52663e+10	1.420891e+04	$\underline{1}$.4717e+10
5.5	1.49868e+11	1.395204e+05	$\underline{1.4}$5179e+11
6.0	1.50274e+12	1.399244e+06	$\underline{1}$.46123e+12
6.5	1.53345e+13	1.428082e+07	$\underline{1}$.49556e+13
7.0	1.58797e+14	1.479060e+08	$\underline{1.5}$5222e+14
7.5	1.66504e+15	1.551038e+09	$\underline{1.6}$302e+15
8.0	1.76459e+16	1.643961e+10	$\underline{1.7}$299e+16
8.5	1.88748e+17	1.758618e+11	$\underline{1.8}$5223e+17
9.0	2.03531e+18	1.896519e+12	1.99905e+18
9.5	2.21040e+19	2.059834e+13	$\underline{2}$.1716e+19
10.0	2.41579e+20	2.251392e+14	$\underline{2}$.37462e+20

Table 3.1: Normalizing constants

3.6 Algorithm and Numerical Results

In [28, Algorithm 1 and Theorem 2], we give an algorithm to obtain the MLE for the Fisher-Bingham distribution. This algorithm is valid for general dimensions, but it cannot be used for more than two dimensions with the current level of computer technology because of the high computational complexity of the Gröbner basis computation. We replace the Gröbner basis computation part (steps 1,2,3 in [28, Algorithm 1]) with our derivation of the Pfaffian system of factored form given in Theorem 3.3.1 and replace the numerical integration part of (3.1) with the evaluation by the series (3.4.2) and extend values to slowly convergent domains of the series by the HGM. For efficiency, we calculate the inverse matrices in our expressions for H_{ij} and H_i numerically during the steps of the adaptive Runge-Kutta method as explained in Section 3.2 regarding the accelerated version of the HGD. This enables us to solve maximum likelihood estimation problems in more than two dimensions case with the HGD. More precisely, we have the following complexity result.

Theorem 3.6.1. *The complexity of the series expansion method, the HGM and the HGD for the Fisher-Bingham distribution on the d-dimensional sphere is*

$$O((2d+2)^{N+1}/N!) + (\text{complexity of solving the ODE with respect to } r)$$
$$+O((2d+2)^3) \times (\text{steps of the convergence of gradient descent}).$$

The first and the second terms are the complexity to evaluate the initial values $F(Z)$ up to degree N and the third term is the complexity of the HGD.

Proof. The number of terms of the truncated series of (3.18) is $\binom{2d+2+N}{2d+2} = \binom{2d+2+N}{N} = O((2d+2)^N/N!)$. The coefficients of the series can be evaluated by a recursive relation. We need $2d+1$ derivatives of $\tilde{Z}$. Thus, we obtain the first term.

Our HGD requires the computation of the inverses of $(2d+2) \times (2d+2)$ matrices in each step of the HGD by Theorem 3.3.1. This corresponds to the third term. □

We implemented our algorithm firstly in Maple and next in the C language using the GNU scientific library [14]. The prototype written in Maple is useful for debugging our C codes. Our C code is automatically generated by our code generation program `pfn_gen_c_2.rr`, which can be obtained from the URL in the Example 3.6.4, on [36].

In order to apply the HGD, we need to find a good starting point $\theta^0 = (x^0, y^0)$. We use the following method.

Algorithm 3.6.2.

1. Take a random point $\tilde{\theta}^0 = (x, y)$ satisfying $0 < x_{ij} < 1$ and $0 < y_k < 1$.

2. Apply the Nelder-Mead algorithm, which does not require the gradient and may also be replaced with other methods, to find an approximate optimal point θ^0 of the likelihood function from the starting point $\tilde{\theta}^0$ (see, e.g., [30]). Normalizing constants are evaluated by the HGM, which may be replaced by other methods.

3. Apply the HGD with the starting point θ^0. If the HGD stops normally, we are done. If the HGD stops at θ^1 because of a numerical instability, go to the step 2 with $\tilde{\theta}^0 =$ a point in a neighborhood of θ^1.

An alternative and heuristic way to avoid the retry in the last step is to abort the computation when a numerical instability occurs and then restart the algorithm with a new randomly chosen starting point. This procedure can be implemented in parallel.

We present some examples to illustrate the performance of our new algorithm and its implementation.

Example 3.6.3. The problem "Astronomical data" given in [28] is solved in 2.58 seconds on a 32-bit virtual machine, the host machine of which is an Intel Xeon E5410 (2.33GHz) processor based computer. In contrast, our implementation in [28] spends 17.3 seconds for the HGD and more than an hour for computing a Gröbner basis and deriving a Pfaffian system.

Example 3.6.4. Problem names beginning with sk_'s in Table 3.2 are problems on the k-dimensional sphere. These problems are generated by a random number generator according to the Fisher-Bingham distribution. We chose 8 random points in the parameter space as starting points for Algorithm 3.6.2. The HGD aborts 2 times in the 8 tries in the worse case on S^3. This rate increases to 7 aborts of 8 tries in the worse case on S^7. The timing data shown are those for the first successful HGD among the 8 starting points. The first step time in the table is that of the step of applying the Nelder-Mead algorithm with the HGM. These sample data and programs are obtainable from our web page. [1]

Problem	Time (1st step)	Time of the HGD and steps	Total
s3_e1	9.8	3.2(73)	13
s3_e3	10	3.1(66)	13
s4_e1	50	14(93)	64
s4_e2	50	28(183)	78
s4_e3	50	11(75)	61
s5_e1	220	80(142)	300
s5_e2	221	140(121)	361
s5_e3	222	66(117)	288
s6_e1	828	247(172)	1075
s7_e1	2679	571(183)	3250

Table 3.2: Performance of the accelerated HGD

3.7 Conclusion and Open Problems

We show that the HGD can solve some MLE problems up to dimension $d = 7$ by utilizing an explicit expression of the the Pfaffian system of factored form and the series expansion of the normalizing constant. However, there are two problems in applying our method efficiently to arbitrary data.

1. In examples, we find a starting point for applying the HGD by the Nelder-Mead algorithm with the HGM for evaluating the normalizing constant. This method seems to work well for our examples, but it is not very efficient. Finding a good starting point efficiently for arbitrary data is an open question.

[1] `http://www.math.kobe-u.ac.jp/OpenXM/Math/Fisher-Bingham-2`

2. There are domains in (x, y)-space where the normalizing constant cannot be evaluated to a given accuracy within a reasonable time by the HGM, because the normalizing constant is huge in these domains, which includes domains where $|y|$ is large.

Although, there still remain important open problems, our proposed method evaluates the normalizing constant and its derivatives to a specified accuracy for a sufficiently broad set of parameters and solves MLE problems. We can easily control the accuracy of evaluations of the normalizing constant and so it is possible to apply our method for evaluations to other approximation methods.

3.8 Proof of Theorem 3.3.1

We define two auxiliary vectors of operators to present the expression. We sort the set of the square free second order operators

$$\{\partial_i\partial_j | 1 \le i < j \le d+1\}$$

by the lexicographic order. This gives a vector of operators of the length $d(d+1)/2$:

$$F^{(2)} = (\partial_1\partial_2,\, \partial_1\partial_3, \dots, \partial_d\partial_{d+1})^T. \tag{3.22}$$

We sort the set of the third order operators

$$\{\partial_i\partial_j\partial_k | 1 \le i \le j \le k \le d+1,\, j \le d\}$$

by the lexicographic order. We denote by $F^{(3)}$ the sorted vector

$$F^{(3)} = (\partial_1\partial_1\partial_1,\, \partial_1\partial_1\partial_2, \dots, \partial_1\partial_1\partial_{d+1},\, \partial_1\partial_2\partial_2, \dots, \partial_d\partial_d\partial_{d+1})^T. \tag{3.23}$$

The length of this vector $d(d+1)(d+5)/6$ is denoted by m.

When two operators ℓ_1 and ℓ_2 are the same modulo the ideal I, we denote it by $\ell_1 \equiv \ell_2$. By examining the proof of Lemmas 2 and 3 of [28], we obtain the following two lemmas which give an expression of the second and the third order operators $F^{(2)}$ and $F^{(3)}$ in terms of F.

Lemma 3.8.1. *We have*

$$P^{(2)}F^{(2)} + Q^{(2)}F \equiv 0. \tag{3.24}$$

Here, $P^{(2)}$ is an invertible $d(d+1)/2 \times d(d+1)/2$ matrix and $Q^{(2)}$ is a $d(d+1)/2 \times (2d+2)$ matrix of which entries are as follows.

$$P^{(2)}_{ij,kl} = \begin{cases} 2(x_{jj} - x_{ii}) & (i = k, j = l) \\ x_{jl} & (i = k, j \neq l) \\ x_{jk} & (i = l, j \neq k) \\ -x_{ik} & (i \neq k, j = l) \\ -x_{il} & (i \neq l, j = k) \end{cases}$$

$$Q^{(2)}_{ij,k} = \begin{cases} y_j\delta_{k,i+1} - y_i\delta_{k,j+1} + x_{ij}\delta_{k,i+d+2} - x_{ij}\delta_{k,j+d+2} & (j \le d) \\ y_j\delta_{k,i+1} - y_i\delta_{k,j+1} + x_{ij}\delta_{k,i+d+2} & \\ \qquad -r^2 x_{i,d+1}\delta_{k1} + \sum_{\ell=1}^{d} x_{i,d+1}\delta_{k,\ell+d+2} & (j = d+1) \end{cases}$$

Here, δ is Kronecker's δ and $P^{(2)}_{ij,kl}$ is the matrix element of $P^{(2)}$ standing for $\partial_i\partial_j$ and $\partial_k\partial_l$ in $F^{(2)}$. We use this notation of the index of the matrix element in the sequel.

Lemma 3.8.2. *We have*

$$P^{(3)}F^{(3)} + Q^{(3)}F^{(2)} + R^{(3)}F \equiv 0. \tag{3.25}$$

Here, $P^{(3)}$, $Q^{(3)}$, and $R^{(3)}$ are an invertible $m \times m$ matrix, an $m \times d(d+1)/2$ matrix and an $m \times (2d+2)$ matrix of polynomial entries respectively. Entries are defined as follows.

$$P^{(3)}_{ijk,abc} = \begin{cases} (\delta_{k,d+1}+1)x_{jk}\delta_{ai}\delta_{bj}\delta_{cj} \\ \qquad +2(x_{kk}-x_{jj})\delta_{ai}\delta_{bj}\delta_{ck} + (\delta_{k,d+1}-1)x_{jk}\delta_{ai}\delta_{bk}\delta_{ck} \\ \qquad +\sum_{l\neq j,k}\left(x_{kl}\delta'_{abc;ijl} - x_{jl}\delta'_{abc;ikl} + x_{jk}\delta_{k,d+1}\delta'_{abc;ill}\right) \\ \qquad\qquad\qquad (i \le j < k \le d+1) \\ x_{ij}\delta_{ai}\delta_{bi}\delta_{cj} + 2(x_{jj}-x_{ii})\delta_{ai}\delta_{bj}\delta_{cj} - x_{ij}\delta_{aj}\delta_{bj}\delta_{cj} \\ \qquad +\sum_{l\neq i,j}\left(x_{jl}\delta'_{abc;ijl} - x_{il}\delta'_{abc;jjl}\right) \\ \qquad\qquad\qquad (i < j = k < d+1) \\ \sum_{s=1}^{d}(x_{s,d+1}\delta'_{abc;is,d+1} - 2(x_{d+1,d+1}-x_{ii})\delta'_{abc;iss} \\ \qquad +\sum_{l\neq i}x_{il}\delta'_{abc;lss}) \\ \qquad\qquad\qquad (i = j = k < d+1) \end{cases}$$

$$Q^{(3)}_{ijk,ab} = \begin{cases} (1-\delta_{ij})y_k\delta_{ai}\delta_{bj} - y_j\delta_{ai}\delta_{bk} & (i \le j < k \le d+1) \\ y_j\delta_{ai}\delta_{bj} & (i < j = k < d+1) \\ y_{d+1}\delta_{ai}\delta_{b,d+1} & (i = j = k < d+1) \end{cases}$$

$$R^{(3)}_{ijk,a} = \begin{cases} -x_{jk}r^2\delta_{k,d+1}\delta_{a,i+1} - \delta_{ij}\delta_{a,k+1} + y_k\delta_{ij}\delta_{a,i+d+2} \\ \qquad\qquad (i \le j < k \le d+1) \\ -y_i\delta_{a,j+d+2} + \delta_{a,i+1} \\ \qquad\qquad (i < j = k < d+1) \\ -y_ir^2\delta_{a1} + (2(x_{d+1,d+1}-x_{ii})r^2+1)\delta_{a,i+1} \\ \qquad -\sum_{l\neq i}x_{il}r^2\delta_{a,l+1} + \sum_{l<d+1}y_i\delta_{a,l+d+2} \\ \qquad\qquad (i = j = k < d+1) \end{cases}$$

where

$$\delta'_{abc;ijk} = \begin{cases} 1 & (\partial_a\partial_b\partial_c = \partial_i\partial_j\partial_k) \\ 0 & (\partial_a\partial_b\partial_c \neq \partial_i\partial_j\partial_k) \end{cases}.$$

Proof. We denote by C_{ij} the differential operator (3.7) in I; we put

$$C_{ij} = x_{ij}\partial_i^2 + 2(x_{jj}-x_{ii})\partial_i\partial_j - x_{ij}\partial_j^2 + \sum_{k\neq i,j}(x_{kj}\partial_i\partial_k - x_{ik}\partial_j\partial_k) + y_j\partial_i - y_i\partial_j.$$

Define a differential operator G_{ijk} $(i \le j \le k \le d+1, j \le d)$ by

$$G_{ijk} = \begin{cases} \partial_i C_{jk} & (i \le j < k \le d+1), \\ \partial_j C_{ij} & (i < j = k \le d), \\ \partial_{d+1} C_{i,d+1} & (i = j = k \le d). \end{cases}$$

We expand G_{ijk} in the ring of differential operators and express it in terms of the elements of F, $F^{(2)}$, and $F^{(3)}$. For example, when $i < j < k < d+1$, we have

$$\begin{aligned} G_{ijk} &= \partial_i C_{jk} \\ &= \partial_i \big(x_{jk}\partial_j^2 + 2(x_{kk} - x_{jj})\partial_j\partial_k - x_{jk}\partial_k^2 \\ &\quad + \sum_{l \ne j,k} (x_{lk}\partial_j\partial_l - x_{jl}\partial_k\partial_l) + y_k\partial_j - y_j\partial_k \big) \\ &= x_{jk}\partial_i\partial_j^2 + 2(x_{kk} - x_{jj})\partial_i\partial_j\partial_k - x_{jk}\partial_i\partial_k^2 \\ &\quad + \sum_{l \ne j,k} (x_{kl}\partial_i\partial_j\partial_l - x_{jl}\partial_i\partial_k\partial_l) + y_k\partial_i\partial_j - y_j\partial_i\partial_k \end{aligned}$$

which yields $P^{(3)}_{ijk,ijj}, P^{(3)}_{ijk,ijk}, \ldots, Q^{(3)}_{ijk,ij}, Q^{(3)}_{ijk,ik}$. Analogous expansions and rewritings for the other cases give the conclusion. $\square$

We denote by $\mathrm{Mat}(k, l, S)$ the space of the $k \times l$ matrices with entries in the set S. Let $\mathbf{Q}[x, y, r]$ denote the ring of polynomials with coefficients in $\mathbf{Q}$.

Lemma 3.8.3. *The vector F satisfies the identity*

$$A\partial_i F \equiv BF + CF^{(2)} + EF^{(3)}. \tag{3.26}$$

Here, $A = (a_{pj}) \in \mathrm{Mat}(2d+2, 2d+2, \mathbf{Q}[x,y,r]), B = (b_{pj}) \in \mathrm{Mat}(2d+2, 2d+2, \mathbf{Q}[x,y,r]), C = (c_{p,jk}) \in \mathrm{Mat}(2d+2, d(d+1)/2, \mathbf{Q}[x,y,r]), E = (e_{p,jk\ell}) \in \mathrm{Mat}(2d+2, m, \mathbf{Q}[x,y,r])$ and A is invertible in the space of the matrices with entries in the field of rational functions $\mathbf{Q}(x,y,r)$. Explicit expressions of these matrices are given in (3.27), (3.28), (3.29), (3.30), (3.31), (3.32), (3.33), (3.34). Note that A, B, C, E depend on the index i.

Notation: $c_{p,jk}$ means the element at the p-th row of C and the column of C standing for $\partial_j\partial_k = \partial_k\partial_j$. $e_{p,jk\ell}$ is defined analogously.

Proof. The both sides of (3.26) is a column vector of the length $2d+2$. We will determine the rows of A, B, C, E from generators of I. Note that the index i is fixed over the proof.

The first rows. The first element of the vector $\partial_i F$ is ∂_i, then we have

$$a_{11} = 1, \quad b_{1,i+1} = 1 \tag{3.27}$$

and the other elements of the first rows of A, B, C, E are 0.

The $(j+1)$-th rows $(1 \le j \le d, i \ne j)$. Using the differential operator (3.7) in I, we have

$$x_{ij}\partial_i^2 + 2(x_{jj} - x_{ii})\partial_i\partial_j + \sum_{k\ne i,j} x_{kj}\partial_i\partial_k \equiv x_{ij}\partial_j^2 + \sum_{k\ne i,j} x_{ik}\partial_j\partial_k + y_i\partial_j - y_j\partial_i.$$

Therefore, we may put as

$$\begin{aligned}
&a_{j+1,i+1} = x_{ij}, \quad a_{j+1,j+1} = 2(x_{jj} - x_{ii}),\\
&a_{j+1,k+1} = x_{kj} \quad (1 \le k \le d+1, k \ne i, k \ne j),\\
&b_{j+1,j+1} = y_i, \quad b_{j+1,i+1} = -y_j, \quad b_{j+1,j+d+2} = x_{ij},\\
&c_{j+1,jk} = x_{ik} \quad (1 \le k \le d+1, k \ne i, k \ne j).
\end{aligned} \tag{3.28}$$

Notation: when an index is out of bound, ignore the setting. For example, we set $b_{j+1,j+d+2} = x_{ij}$ when $j+d+2 \le 2d+2$. The other elements of the $(j+1)$-th rows of A, B, C, E are 0.

The $(i+1)$-th rows. The $(i+1)$-th element of the vector $\partial_i F$ is ∂_i^2. When $i \le d$, we put

$$a_{i+1,i+1} = 1, \quad b_{i+1,i+d+2} = 1 \tag{3.29}$$

and the other elements of the $(i+1)$-th rows are 0. When $i = d+1$, we consider the operator (3.6) in the ideal I. Then, we have

$$\partial_{d+1}^2 \equiv r^2 - \sum_{k=1}^{d} \partial_k^2$$

and hence we put

$$\begin{aligned}
&a_{d+2,d+2} = 1,\\
&b_{d+2,1} = r^2, \quad b_{d+2,k+d+2} = -1 \quad (1 \le k \le d).
\end{aligned} \tag{3.30}$$

The other elements of the $(i+1)$-th rows of A, B, C, E are 0.

The $(d+2)$-th rows. When $i = d+1$, it is reduced to the case of the $(i+1)$-th rows. We assume that $i \le d$. Using the operators (3.7) and (3.6) in I, we have

$$\begin{aligned}
& x_{i,d+1}\partial_i^2 + 2(x_{d+1,d+1} - x_{ii})\partial_i\partial_{d+1} + \sum_{k\ne i,d+1} x_{k,d+1}\partial_i\partial_k\\
\equiv\ & x_{i,d+1}\partial_{d+1}^2 + \sum_{k\ne i,d+1} x_{ik}\partial_{d+1}\partial_k + y_i\partial_{d+1} - y_{d+1}\partial_i\\
\equiv\ & x_{i,d+1}r^2 - \sum_{k=1}^{d} x_{i,d+1}\partial_k^2 + \sum_{k\ne i,d+1} x_{ik}\partial_{d+1}\partial_k + y_i\partial_{d+1} - y_{d+1}\partial_i.
\end{aligned}$$

Hence, we put

$$\begin{aligned}
&a_{d+2,i+1} = x_{i,d+1}, \quad a_{d+2,d+2} = 2(x_{d+1,d+1} - x_{ii}),\\
&a_{d+2,k+1} = x_{k,d+1} \quad (1 \le k \le d, k \ne i),\\
&b_{d+2,1} = x_{i,d+1}r^2, \quad b_{d+2,d+2} = y_i, \quad b_{d+2,i+1} = -y_{d+1},\\
&b_{d+2,l+d+2} = -x_{i,d+1}, \quad (1 \le l \le d),\\
&c_{d+2,k(d+1)} = x_{ik} \quad (1 \le k \le d, k \ne i).
\end{aligned} \tag{3.31}$$

The other elements of the $(d+2)$-th rows of A, B, C, E are 0.

The $(j+d+2)$*-th rows* $(1 \le j \le d, i \ne j)$. Using the operator (3.7) multiplied by ∂_j from the left hand side, we have

$$-2(x_{jj}-x_{ii})\partial_i\partial_j^2 \equiv x_{ij}\partial_i^2\partial_j - x_{ij}\partial_j^3 + \sum_{k\ne i,j}\left(x_{kj}\partial_i\partial_j\partial_k - x_{ik}\partial_j^2\partial_k\right) + y_j\partial_i\partial_j - y_i\partial_j^2 + \partial_i.$$

When $i \le d$, we put

$$\begin{aligned}
&a_{j+d+2,j+d+2} = -2(x_{jj}-x_{ii}), \\
&b_{j+d+2,i+1} = 1, \quad b_{j+d+2,j+d+2} = -y_i, \\
&c_{j+d+2,ij} = y_j, \\
&e_{j+d+2,iij} = x_{ij}, \quad e_{j+d+2,jjj} = -x_{ij}, \\
&e_{j+d+2,ijk} = x_{kj}, \quad e_{j+d+2,jjk} = -x_{ik} \quad (1 \le k \le d+1, k \ne i, k \ne j).
\end{aligned} \tag{3.32}$$

The other elements in the $(j+d+2)$-th rows of A, B, C, E are 0.

When $i = d+1$, We use the operator (3.6) and obtain

$$\begin{aligned}
& -2(x_{jj}-x_{d+1,d+1})\partial_{d+1}\partial_j^2 \\
\equiv\ & x_{d+1,j}r^2\partial_j - 2x_{d+1,j}\partial_j^3 + y_j\partial_{d+1}\partial_j - y_{d+1}\partial_j^2 + \partial_{d+1} \\
+ & \sum_{k\ne d+1,j}\left(x_{kj}\partial_{d+1}\partial_j\partial_k - x_{d+1,k}\partial_j^2\partial_k - x_{d+1,j}\partial_j\partial_k^2\right).
\end{aligned}$$

Therefore, we may put as

$$\begin{aligned}
&a_{j+d+2,j+d+2} = -2(x_{jj}-x_{d+1,d+1}), \\
&b_{j+d+2,j+1} = x_{d+1,j}r^2, \quad b_{j+d+2,d+2} = 1, \quad b_{j+d+2,j+d+2} = -y_{d+1}, \\
&c_{j+d+2,j(d+1)} = y_j, \\
&e_{j+d+2,jjj} = -2x_{ij}, \\
&e_{j+d+2,ijk} = x_{kj}, \quad e_{j+d+2,jjk} = -x_{ik}, \quad e_{j+d+2,jkk} = -x_{ij} \quad (1 \le k \le d, k \ne j).
\end{aligned} \tag{3.33}$$

The other elements of the $(j+d+2)$-th rows of A, B, C, E are 0.

The $(i+d+2)$*-th rows*. We may assume that $i \le d$. Since the $(i+d+2)$-th element of the vector $\partial_i F$ is ∂_i^3, we put

$$a_{i+d+2,i+d+2} = 1, \quad e_{i+d+2,iii} = 1. \tag{3.34}$$

The other elements of the $(i+d+2)$-th rows of A, B, C, E are 0. □

From Lemmas 3.8.1, 3.8.2, and 3.8.3, we have Theorem 3.3.1, which gives a differential equation satisfied by the normalizing constant with respect to the variable y_i. As we remarked in Lemma 3.8.3, we note that A, B, C, E depend on the index i and we omit to denote the dependency.

Proof of Theorem 3.3.1.

$$\begin{aligned}
\partial_i F &\equiv A^{-1}(BF + CF^{(2)} + EF^{(3)}) \quad \text{by Lemma 3.8.3} \\
&\equiv A^{-1}\left(BF + CF^{(2)} - E(P^{(3)})^{-1}\left(Q^{(3)}F^{(2)} + R^{(3)}F\right)\right) \\
&\qquad \text{by Lemma 3.8.2} \\
&\equiv A^{-1}\left(BF - C(P^{(2)})^{-1}Q^{(2)}F - E(P^{(3)})^{-1}\left(-Q^{(3)}(P^{(2)})^{-1}Q^{(2)}F + R^{(3)}F\right)\right) \\
&\qquad \text{by Lemma 3.8.1} \\
&\equiv A^{-1}\left(B - C(P^{(2)})^{-1}Q^{(2)} + E(P^{(3)})^{-1}\left(Q^{(3)}(P^{(2)})^{-1}Q^{(2)} - R^{(3)}\right)\right) F.
\end{aligned}$$

□

Example 3.8.4. In the case of $d = 1$ and for the y_1 direction, these matrices are as follows.

$$\begin{aligned}
F &= \begin{pmatrix} 1 & \partial_1 & \partial_2 & \partial_1^2 \end{pmatrix}^T, \\
F^{(2)} &= \begin{pmatrix} \partial_1\partial_2 \end{pmatrix}, \quad F^{(3)} = \begin{pmatrix} \partial_1^3 & \partial_1^2\partial_2 \end{pmatrix}^T,
\end{aligned}$$

$$\begin{aligned}
A &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & x_{12} & -2x_{11} + 2x_{22} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ r^2 x_{12} & -y_2 & y_1 & -x_{12} \\ 0 & 0 & 0 & 0 \end{pmatrix}, \\
C &= \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad E = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{pmatrix},
\end{aligned}$$

$$P^{(2)} = \begin{pmatrix} -2x_{11} + 2x_{22} \end{pmatrix}, \quad Q^{(2)} = \begin{pmatrix} -r^2 x_{12} & y_2 & -y_1 & 2x_{12} \end{pmatrix},$$

$$\begin{aligned}
P^{(3)} &= \begin{pmatrix} 2x_{11} - 2x_{22} & 2x_{12} \\ 2x_{12} & -2x_{11} + 2x_{22} \end{pmatrix}, \quad Q^{(3)} = \begin{pmatrix} y_2 \\ -y_1 \end{pmatrix}, \\
R^{(3)} &= \begin{pmatrix} -r^2 y_1 & -2r^2 x_{11} + 2x_{22} r^2 + 1 & -r^2 x_{12} & y_1 \\ 0 & -r^2 x_{12} & -1 & y_2 \end{pmatrix}.
\end{aligned}$$

Chapter 4

Orhtant probabilities

4.1 Introduction

The holonomic gradient method (HGM) introduced by
Nakayama et al. [28] is a new method of numerical calculation which utilizes algebraic properties of differential equations. This method has many applications in statistics. For example, an application to the evaluation of the exact distribution function of the largest root of a Wishart matrix is introduced in [15] and an application to the maximum likelihood estimation for the Fisher-Bingham distribution on the d-dimensional sphere is introduced in [19]. These applications greatly expand the scope of the field of algebraic statistics.

In this paper, we utilize the holonomic gradient method for an accurate evaluation of the orthant probability

$$\Phi(\Sigma, \mu) = \mathbf{P}(X_1 \geq 0, \ldots, X_d \geq 0), \tag{4.1}$$

where the d-dimensional random vector $X = (X_1, \ldots, X_d) \in \mathbf{R}^d$ is normally distributed with mean μ and covariance matrix Σ, i.e., $X \sim N(\mu, \Sigma)$. Since evaluation of the orthant probability has important applications in statistical practice, there are many studies about it. Genz introduced a method to calculate the orthant probability utilizing the quasi Monte-Carlo method in [11]. Miwa, Hayter and Kuriki proposed a recursive integration algorithm to evaluate the orthant probability in [26].

When the mean vector μ is equal to zero, the orthant probability can be interpreted as the area of a $d-1$ dimensional spherical simplex (cf. [1]). In [38], Schläfli gave a classical differential recurrence formula for $\Phi(\Sigma, 0)$. Plackett generalized Schläfli's result and gave a recurrence formula for $\Phi(\Sigma, \mu)$ in [35]. He evaluated orthant probabilities by a recursive integration utilizing his formula. Gassmann implemented the Plackett's method in the case of higher dimensions in [10]. Their method is a recursive integration based on a differential recurrence formula, whereas the holonomic gradient method utilizes differential equations. Plackett's recurrence formula is not suitable for the holonomic gradient method and in this paper we give a new recurrence formula, which is more natural from the viewpoint of holonomic gradient method.

Let $y = \Sigma^{-1}\mu$ and $x = -\frac{1}{2}\Sigma^{-1}$. We denote the i-th element of y by y_i and the (i, j) element of x by x_{ij}. By this transformation of the parameters, the orthant probability in (4.1) can be written as

$$(\pi)^{-d/2} \det(-x)^{1/2} \exp\left(-\frac{1}{2} y^t x^{-1} y\right) g(x, y),$$

where

$$g(x, y) = \int_0^\infty \dots \int_0^\infty \exp\left(\sum_{i,j=1}^d x_{ij} t_i t_j + \sum_{i=1}^d y_i t_i\right) dt \qquad (dt = dt_1 \dots dt_d). \tag{4.2}$$

In order to evaluate the orthant probability, it is enough to evaluate $g(x, y)$.

To apply the holonomic gradient method, we need an explicit form of a Pfaffian system ([19, Section 2]) associated with $g(x, y)$. The Pfaffian system can be obtained from a holonomic system for $g(x, y)$ (see [28]). For a given d, we can obtain a holonomic system for $g(x, y)$ by applying the algorithms introduced in [33] with computer algebra systems. However, for general d, these algorithms can not be applied and we need theoretical considerations. In this paper, we give a holonomic system for the function $g(x, y)$ and construct a Pfaffian system from the holonomic system.

Note that the integral $g(x, y)$ satisfies an incomplete A-hypergeometric system ([29]) when the integration domain is not the orthant but a polytope.

The organization of this paper is as follows. In Section 4.2, we describe a holonomic system for an integral

$$g(x, y) = \int_{\mathbf{R}^d} f(t) \exp\left(\sum_{i,j=1}^d x_{ij} t_i t_j + \sum_{i=1}^d y_i t_i\right) dt, \tag{4.3}$$

where $f(t)$ is a holonomic function or a holonomic distribution. In Section 4.3, we construct a holonomic system for the function in (4.2) by utilizing the result in Section 4.2. Then we construct a Pfaffian system from the holonomic system. Finally in Section 4.4 we describe numerical experiments of the holonomic gradient method.

4.2 Holonomic system associated with the expectation under multivariate normal distributions

In this section, we consider the holonomic ideal which annihilates the integral (4.3), which is the expectation under multivariate normal distributions (for the definition of the holonomic ideal, see [37]). We develop a general theory, where

$f(t)$ in (4.3) is a smooth function or a distribution in the sense of Schwartz ([39]). This theory is a generalization of the result introduced in [18]. In Section 4.3 we will specialize $f(t)$ to be the indicator function of the positive orthant. Note that the results of this section can be applied to various problems of the multivariate normal distribution theory, other than the orthant probability.

We denote the ring of differential operators in x with polynomial coefficient by $D = \mathbf{C}\langle x_1, \ldots, x_n, \partial_{x_1}, \ldots, \partial_{x_n}\rangle$. The operator $\partial/\partial x_i$ is denoted by ∂_{x_i}. We frequently use the following rings:

$$\begin{aligned} D_{xyt} &:= \mathbf{C}\langle x_{ij}, y_k, t_k, \partial_{x_{ij}}, \partial_{y_k}, \partial_{t_k} : 1 \le i \le j \le d, 1 \le k \le d\rangle, \\ D_{xy} &:= \mathbf{C}\langle x_{ij}, y_k, \partial_{x_{ij}}, \partial_{y_k} : 1 \le i \le j \le d, 1 \le k \le d\rangle, \\ D_x &:= \mathbf{C}\langle x_{ij}, \partial_{x_{ij}} : 1 \le i \le j \le d\rangle, \\ D_t &:= \mathbf{C}\langle t_i, \partial_{t_i} : 1 \le i \le d\rangle. \end{aligned} \tag{4.4}$$

We use the following notation.

$$y = \Sigma^{-1}\mu, \quad x = -\frac{1}{2}\Sigma^{-1}, \quad h(x,y,t) = \sum_{i,j=1}^{d} x_{ij}t_i t_j + \sum_{i=1}^{d} y_i t_i. \tag{4.5}$$

4.2.1 The case of a smooth function

We first consider the case when $f(t)$ is a smooth function. In this case, the integral (4.3) converges if the matrix $-x$ is positive definite and $f(t)$ is of exponential growth.

To state the main theorem of this section, we show a general lemma. The notation f, g, x, y in the following lemma is generic and not related to (4.5).

Lemma 4.2.1. *Let D_x (resp. D_{xy}) be the ring of differential operators with polynomial coefficient $\mathbf{C}\langle x_i, \partial_{x_i} : 1 \le i \le n\rangle$ (resp. $\mathbf{C}\langle x_i, y_j, \partial_{x_i}, \partial_{y_j} : 1 \le i \le n, 1 \le j \le m\rangle$). Suppose that a holonomic ideal I annihilates a function $f(x,y)$ on $\mathbf{R}^n \times \mathbf{R}^m$ and the function $f(x,y)$ is rapidly decreasing with respect to the variable y for any x in an open set $O \subset \mathbf{R}^n$. Then the integration ideal of I with respect to the variable y annihilates*

$$g(x) := \int_{\mathbf{R}^m} f(x,y)dy \tag{4.6}$$

defined on the open set O.

Proof. Since the function $f(x,y)$ is rapidly decreasing, the integral of (4.6) converges. Let P be in the integration ideal $J = \left(I + \sum_{j=1}^m \partial_{y_j} D_{x,y}\right) \cap D_x$, then we have $P \bullet g = \int P \bullet f dy$ by the Lebesgue convergence theorem. Since the differential operator P can be written as

$$P = P_0 + \sum_{i=1}^{m} \partial_{y_i} P_i \in D_x \quad (P_0 \in I,\, P_i \in D_{xy}),$$

we have $\int P \bullet f dy = \sum \int \partial_{y_i} P_i \bullet f dy$. Since $P_i \bullet f$ is rapidly decreasing, we have $\int \partial_{y_i} P_i \bullet f dy = 0$. □

We now go back to $g(x,y)$ in (4.3) and use the notation in (4.4) and (4.5). Consider a $\mathbf{C}$-algebra morphism φ from D_t to D_{xy} defined by

$$\varphi : D_t \to D_{xy} \qquad \left(t_i \mapsto \partial_{y_i},\ \partial_{t_i} \mapsto -y_i - 2\sum_{k=1}^{d} x_{ik}\partial_{y_k} \right).$$

Here, we assume $x_{ij} = x_{ji}$, $\partial_{x_{ij}} = \partial_{x_{ji}}$. Since $[\varphi(\partial_{t_i}), \varphi(\partial_{t_j})] := \varphi(\partial_{t_i})\varphi(t_j) - \varphi(t_j)\varphi(\partial_{t_i}) = \delta_{ij}$, where δ_{ij} is Kronecker's delta, φ is well-defined as a morphism of $\mathbf{C}$-algebra.

Now we state the main theorem in this section.

Theorem 4.2.2. *Suppose a function f on $\mathbf{R}^d$ is smooth and of exponential growth. If differential operators $P_1, \ldots, P_s \in D_t$ annihilate f, then the following differential operators annihilate the integral $g(x,y)$ in* (4.3)*.*

$$\varphi(P_k) \qquad (1 \le k \le s), \tag{4.7}$$
$$\partial_{x_{ij}} - 2\partial_{y_i}\partial_{y_j} \qquad (1 \le i < j \le d), \tag{4.8}$$
$$\partial_{x_{ii}} - \partial_{y_i}^2 \qquad (1 \le i \le d). \tag{4.9}$$

Moreover, the differential operators (4.7), (4.8), (4.9) *generate a holonomic ideal in D_{xy} if the differential operators $P_1, \ldots, P_s$ generate a holonomic ideal in D_t.*

Proof. For a differential operator

$$P = \sum_{\alpha,\beta} c_{\alpha,\beta} t_1^{\alpha_1} \cdots t_d^{\alpha_d} \partial_{t_1}^{\beta_1} \cdots \partial_{t_d}^{\beta_d} \in D_t$$

and $p_i, q_i \in D_{xyt}$ $(i = 1, \ldots, d)$, we put

$$P(p_i; q_i) = \sum_{\alpha,\beta} c_{\alpha,\beta} p_1^{\alpha_1} \cdots p_d^{\alpha_d} q_1^{\beta_1} \cdots q_d^{\beta_d}. \tag{4.10}$$

By the assumption, the differential operators

$$P_\ell\ (\ell = 1, \ldots, s), \quad \partial_{x_{ij}}\ (1 \le i \le j \le d), \quad \partial_{y_i}\ (1 \le i \le d)$$

annihilate f and generate a holonomic ideal in D_{xyt}. By the lemma introduced in [41, Section 3.3], the differential operators

$$P_\ell(t_i; \partial_{t_i} - \frac{\partial h}{\partial t_i})\ (\ell = 1, \ldots, s), \quad \partial_{x_{ij}} - \frac{\partial h}{\partial x_{ij}}\ (1 \le i \le j \le d), \quad \partial_{y_i} - \frac{\partial h}{\partial y_i}\ (1 \le i \le d)$$

annihilate $\exp(h)f$ and generate holonomic ideal I in D_{xyt}. As we will show in Lemma 4.2.4 below, the differential operators (4.7), (4.8), (4.9) generate the integration ideal J of I with respect to the variables t_i $(i = 1, \ldots, d)$.

Since the function $\exp(h) f$ is rapidly decreasing, the integration ideal J annihilates the integral $g(x,y)$ by Lemma 4.2.1. Hence, the differential operators (4.7),(4.8),
(4.9) annihilate $g(x,y)$.

Since the integration ideal of a holonomic ideal is also holonomic (see, e.g., [6, Chap 1]), the ideal J is holonomic when the ideal I is holonomic. □

Now, we show that the integration ideal J is generated by the differential operators (4.7), (4.8), (4.9) in the following two lemmas. We denote $P \equiv Q$ when $P - Q \in \sum_{i=1}^{d} D_{xyt}(\partial_{y_i} - t_i)$.

Lemma 4.2.3. *Let $p_i = \partial_{t_i} - y_i - 2\sum_{k=i}^{d} x_{ik} t_i$ and $q_i = \partial_{t_i} - y_i - 2\sum_{k=i}^{d} x_{ik}\partial_{y_k}$. For a differential operator $P \in D_t$ and a multi index $\alpha \in \mathbf{N}_0^d = \{0, 1, 2 \ldots\}^d$, the following equivalence relations hold.*

$$P(t_i; p_i) \equiv P(\partial_{y_i}; q_i), \tag{4.11}$$
$$t^\alpha P(\partial_{y_i}; q_i) \equiv \partial_y^\alpha P(\partial_{y_i}; q_i). \tag{4.12}$$

Here, we use the notation in (4.10).

Proof. By the straightforward calculation, we have the following relations for $1 \le i, j \le d$.

$$[p_i, q_j] = 0, \tag{4.13}$$
$$[p_i, p_j] = [q_i, q_j] = 0, \tag{4.14}$$
$$[q_i, t_j] = [q_i, \partial_{y_j}] = \delta_{ij}. \tag{4.15}$$

In order to prove (4.11), it is sufficient to show

$$t_1^{\alpha_1} \cdots t_d^{\alpha_d} p_1^{\beta_1} \cdots p_d^{\beta_d} \equiv \partial_{y_1}^{\alpha_1} \cdots \partial_{y_d}^{\alpha_d} q_1^{\beta_1} \cdots q_d^{\beta_d} \quad (\alpha, \beta \in \mathbf{N}_0^d). \tag{4.16}$$

By the induction on β_i, we have a relation

$$t_j q_i^{\beta_i} \equiv \partial_{y_j} q_i^{\beta_i} \quad (1 \le i, j \le d, \quad \beta_i \in \mathbf{N}_0). \tag{4.17}$$

When $\beta_i = 0$, the relation (4.17) holds clearly. Suppose that the relation (4.17) holds for β_i. Then we have

$$\begin{aligned} t_j q_i^{\beta_i+1} &= t_j q_i q_i^{\beta_i} = (q_i t_j - \delta_{ij}) q_i^{\beta_i} \\ &\equiv (q_i \partial_{y_j} - \delta_{ij}) q_i^{\beta_i} = \partial_{y_j} q_i q_i^{\beta_i} = \partial_{y_j} q_i^{\beta_i+1}. \end{aligned}$$

By induction, the relation (4.17) holds for any β_i.

By the relations (4.13) and (4.17), we have

$$p_1^{\beta_1} \cdots p_d^{\beta_d} \equiv q_1^{\beta_1} \cdots q_d^{\beta_d} \quad (\beta_i \in \mathbf{N}_0). \tag{4.18}$$

Now we prove the relation (4.16) by induction on the multi index $\alpha \in \mathbf{N}_0^d$. When $\alpha = 0$, the relation (4.16) holds because of (4.18). Suppose the relation (4.16) holds for α, then we have

$$\begin{aligned}
t_i t_1^{\alpha_1} \cdots t_d^{\alpha_d} p_1^{\beta_1} \cdots p_d^{\beta_d} &\equiv t_i \partial_{y_1}^{\alpha_1} \cdots \partial_{y_d}^{\alpha_d} q_1^{\beta_1} \cdots q_d^{\beta_d} \\
&= \partial_{y_1}^{\alpha_1} \cdots \partial_{y_d}^{\alpha_d} (\prod_{j \neq i} q_j^{\beta_j}) t_i q_i^{\beta_i} \quad \text{(by (4.14) and (4.15))} \\
&\equiv \partial_{y_1}^{\alpha_1} \cdots \partial_{y_d}^{\alpha_d} (\prod_{j \neq i} q_j^{\beta_j}) \partial_{y_i} q_i^{\beta_i} \quad \text{(by (4.17))} \\
&\equiv \partial_{y_i} \partial_{y_1}^{\alpha_1} \cdots \partial_{y_d}^{\alpha_d} q_1^{\beta_1} \cdots q_d^{\beta_d} \quad \text{(by (4.14) and (4.15))}.
\end{aligned}$$

By induction, the relation (4.11) holds for any α.

Finally, the relation (4.12) holds by (4.14), (4.15) and (4.17). □

Lemma 4.2.4. *The integration ideal J is generated by the differential operators* (4.7), (4.8), (4.9).

Proof. Note that the ideal I is generated by differential operators

$$\begin{aligned}
&P_\ell(t_i; \partial_{t_i} - y_i - 2\sum_{k=i}^{d} x_{ik} t_i) \quad (\ell = 1, \ldots, s), \\
&\partial_{x_{ij}} - 2t_i t_j \quad (1 \leq i < j \leq d), \\
&\partial_{x_{ii}} - t_i^2, \ \partial_{y_i} - t_i \quad (1 \leq i \leq d).
\end{aligned}$$

By (4.11), the ideal I is generated by

$$P_\ell(\partial_{y_i}; \partial_{t_i} - y_i - 2\sum_{k=i}^{d} x_{ik} \partial_{y_k}) \quad (\ell = 1, \ldots, s), \tag{4.19}$$

$$\partial_{x_{ij}} - 2\partial_{y_i}\partial_{y_j} \quad (1 \leq i < j \leq d), \tag{4.20}$$

$$\partial_{x_{ii}} - \partial_{y_i}^2 \quad (1 \leq i \leq d), \tag{4.21}$$

$$\partial_{y_i} - t_i \quad (1 \leq i \leq d). \tag{4.22}$$

We denote by $\tilde{J}$ the left ideal generated by (4.7), (4.8), (4.9). Clearly we have $\tilde{J} \subset J$. If P is a differential operator in J, then P can be written as

$$P = Q + \sum_{i=1}^{d} \partial_{t_i} R_i + \sum_{i=1}^{d} S_i (\partial_{y_i} - t_i) \quad (Q \in I, \ R_i, S_i \in D_{xyt}) \tag{4.23}$$

by the definition of integration ideal. The differential operator Q is written as a linear combination of the differential operators (4.19)–(4.21) with D_{xyt} coefficients. By the second term of the right-hand side of (4.23), we can assume without loss of generality that the variables ∂_{t_i} do not appear in these coefficients. By the relation (4.12) in Lemma 4.2.3, we can assume that the coefficient in Q is an element in D_{xy}. Since any differential operator in (4.19)–(4.21) is an element of $\tilde{J} + \sum_{i=1}^{d} \partial_{t_i} \cdot D_{xyt}$, we can assume $Q \in J$.

Consider the equation

$$P - Q - \sum_{i=1}^{d} S_i(\partial_{y_i} - t_i) = \sum_{i=1}^{d} \partial_{t_i} R_i.$$

We can assume that the variables $\partial_{t_1}, \ldots, \partial_{t_d}$ do not appear in $S_1, \ldots, S_d$. For example, if $t_1\partial_{t_1}$ is a term of S_2 then we can replace S_2 and R_1 to $S_2 - t_1\partial_{t_1} - 1$ and $R_1 - t_1(\partial_{y_2} - t_2)$ respectively. The term $t_1\partial_{t_1}$ in S_2 is removed. In the same way, we can remove all terms which include ∂_{t_i}.

Expanding both sides and comparing the coefficients of ∂_{t_i}, we have

$$P - Q = \sum S_i(\partial_{y_i} - t_i).$$

The right-hand side of this equation is an element of the left ideal $J' := \sum_{i=1}^{d} D_{xyt} \cdot (\partial_{y_i} - t_i)$. Let the weight of t_i be 1 and that of other variables 0, and consider a term order $\prec$ with this weight. The set $\{t_i - \partial_{y_i} | 1 \leq i \leq d\}$ is a Gröbner basis of J' with the order, so that the initial term of $P - Q$ has to divide some t_i. Since $P - Q$ is in D_{xy}, we have $P - Q = 0$. Thus $P \in J$. □

4.2.2 The case of a non-smooth function

Next we consider the case when $f(t)$ is not smooth. In this case we consider $f(t)$ as a distribution in the sense of Schwartz ([39]).

Let Ω be a domain defined by

$$\{(x, y) | {-x} \text{ is positive definite}\}.$$

For a tempered distribution f on $\mathbf{R}^d$, we can define a function on Ω as

$$g(x, y) = \langle f, \exp(h(t, y, x)) \rangle. \tag{4.24}$$

Since $\exp(h(t, y, x))$ is rapidly decreasing with respect to the variable t when $-x$ is positive definite, the right-hand side of (4.24) is finite.

A holonomic system for (4.24) is given as follows.

Theorem 4.2.5. *If differential operators $P_1, \ldots, P_s \in D_t$ annihilate a tempered distribution f on $\mathbf{R}^d$, then the differential operators* (4.7), (4.8), (4.9) *annihilate the function $g(x, y)$ in* (4.24). *Moreover, the differential operators* (4.7), (4.8), (4.9) *generate a holonomic ideal in D_{xy} if the differential operators $P_1, \ldots, P_s$ generate a holonomic ideal in D_t.*

Proof. We only need to prove that the differential operators (4.7), (4.8), (4.9) annihilate $g(x, y)$. Let $P_\ell = \sum c_{\alpha\beta} t^\alpha \partial_t^\beta$ and $P_\ell^* = \sum c_{\alpha\beta} t^\beta \partial_t^\alpha$. We have

$$P_\ell(\partial_{y_i}; -y_i - 2\sum_{k=i}^{d} x_{ik}\partial_{y_k})\langle f, \exp(h(t, y, x))\rangle$$

$$
\begin{aligned}
&= \langle f, P_\ell(\partial_{y_i}; -y_i - 2\sum_{k=i}^{d} x_{ik}\partial_{y_k}) \exp(h(t,y,x))\rangle \\
&= \langle f, P_\ell(\partial_{y_i}; -\partial_{t_i}) \exp(h(t,y,x))\rangle \\
&= \langle f, P_\ell^*(-\partial_{t_i}; \partial_{y_i}) \exp(h(t,y,x))\rangle \\
&= \langle f, P_\ell^*(-\partial_{t_i}; t_i) \exp(h(t,y,x))\rangle \\
&= \langle P_\ell(t_i; \partial_{t_i}) f, \exp(h(t,y,x))\rangle \\
&= 0.
\end{aligned}
$$

□

4.3 Holonomic system associated with the orthant probability

In this section we specialize $f(t)$ of the last section to the indicator function of the positive orthant and we will construct a Pfaffian system associated with the integral (4.2) for the orthant probability.

4.3.1 Generators of the holonomic ideal

At first, we obtain generators of a holonomic ideal which annihilates (4.2) by Theorem 4.2.5. Let E be the positive orthant in $\mathbf{R}^d$ defined by

$$\left\{ t = (t_1, \dots, t_d) \in \mathbf{R}^d | t_i \geq 0, (i = 1, \dots, d) \right\},$$

and $\mathbf{1}_E$ be the indicator function of E. A holonomic ideal which annihilates $\mathbf{1}_E$ is given as follows.

Lemma 4.3.1. *The indicator function $\mathbf{1}_E$ is annihilated by the following differential operators as a distribution.*

$$t_1\partial_{t_1}, \dots, t_d\partial_{t_d} \tag{4.25}$$

The differential operators (4.25) *generate a holonomic ideal J in the ring D_t.*

Proof. At first, we show that the differential operators $t_i\partial_{t_i}$ annihilates the function $\mathbf{1}_E$. It suffices to prove for $i = 1$. Let $\varphi(t)$ be a rapidly decreasing function on $\mathbf{R}^d$. Then, we have

$$-\int_0^\infty \partial_{t_1}(t_1\varphi(t))dt_1 = 0$$

for any $t_j \in \mathbf{R}$, $2 \leq j \leq d$. Integrating both sides with respect to the variables $t_2, \dots, t_d$, we have $\langle t_1\partial_{t_1}\mathbf{1}_E, \varphi\rangle = 0$. Therefore, the distribution $t_1\partial_{t_1}\mathbf{1}_E$ is equal to 0.

Next, we show that the left ideal J is holonomic. By the Buchberger's criterion, we can show that the set of the differential operators in (4.25) is a

Gröbner basis of J with the weight $w = (0, 1)$. The characteristic variety (see, e.g. [31], [37]) of J is $\mathrm{ch}(J) = \{(t, \xi) \in \mathbf{C}^{2d} | t_i \xi_i = 0, 1 \leq i \leq d\}$. The variety $\mathrm{ch}(J)$ can be decomposed as follows,

$$\bigcup_{J \subset \{1,\ldots,d\}} \{(t, \xi) | t_i = 0, \xi_j = 0, i \in J, j \notin J\}.$$

Since the Krull dimension of the each component is d, we have $\dim(\mathrm{ch}(J)) = d$ and the ideal J is holonomic. □

The function $g(x, y)$ in (4.2) can be written as $g(x, y) = \langle \mathbf{1}_E(t), \exp(h(x, y, t)) \rangle$. This is a case of Theorem 4.2.5 in which the distribution f is $\mathbf{1}_E$. A holonomic ideal which annihilates $\mathbf{1}_E$ is given in Lemma 4.3.1. Hence we have the following theorem for $g(x, y)$ in (4.2).

Theorem 4.3.2. *Differential operators*

$$2 \sum_{k=1}^{d} x_{ik} \partial_{y_i} \partial_{y_k} + y_i \partial_{y_i} + 1 \quad (i = 1, \ldots, d, \quad x_{ij} = x_{ji}), \tag{4.26}$$

$$\partial_{x_{ij}} - 2 \partial_{y_i} \partial_{y_j} \quad (1 \leq i < j \leq d), \tag{4.27}$$

$$\partial_{x_{ii}} - \partial_{y_i}^2 \quad (1 \leq i \leq d) \tag{4.28}$$

annihilate the function $g(x, y)$ in (4.2), and generate a holonomic ideal I in the ring D_{xy}.

4.3.2 Differential recurrence formula

Next, we give a Pfaffian system associated with (4.2). In order to write the Pfaffian system, we define new notation. For $J \subset [d] = \{1, \ldots, d\}$, we put

$$h_J(x, y, x) = \sum_{i \in J} \sum_{j \in J} x_{ij} t_i t_j + \sum_{k \in J} y_k t_k,$$

$$g_J(x, y) = \int_0^\infty \cdots \int_0^\infty \exp(h_J(x, y, t))\, dt_J,$$

where $dt_J = \prod_{j \in J} dt_j$. When the set J is empty, we set $g_\emptyset = 1$. For example, the functions are written as follows when $d = 2$.

$$\begin{aligned}
g_{\{1,2\}}(x, y) &= \int_0^\infty \exp\left(t_1^2 x_{11} + 2 t_1 t_2 x_{12} + t_2^2 x_{22} + y_1 t_1 + y_2 t_2\right) dt_1 dt_2 \\
&= g(x, y), \\
g_{\{1\}}(x, y) &= \int_0^\infty \exp\left(t_1^2 x_{11} + y_1 t_1\right) dt_1, \\
g_{\{2\}}(x, y) &= \int_0^\infty \exp\left(t_2^2 x_{22} + y_2 t_2\right) dt_2, \\
g_\emptyset &= 1.
\end{aligned}$$

Let J be a subset of $[d]$ and put

$$Q_j = -\left(y_j + 2\sum_{k=1}^{d} x_{jk}\partial_{y_k}\right) \quad (j=1,\ldots,d), \tag{4.29}$$

$$Q_{j,J} = -\left(y_j + 2\sum_{k\in J} x_{jk}\partial_{y_k}\right) \quad (j\in J). \tag{4.30}$$

Lemma 4.3.3. *The following equation holds.*

$$Q_j g_J = Q_{j,J} g_J = g_{J\setminus\{j\}} \quad (j\in J). \tag{4.31}$$

Proof. Since g_J is constant with respect to y_ℓ $(\ell \notin J)$, we have $Q_j g_J = Q_{j,J} g_J$. We assume that $J=[d]$ without loss of generality. Applying Q_j to the integrand of $g_{[d]}$, we have

$$\begin{aligned} & -\left(y_j + 2\sum_{k=1}^{d} x_{jk}\partial_{y_k}\right)\exp\left(h(x,y,t)\right) \\ = & -\left(y_j + 2\sum_{k=1}^{d} x_{jk}t_k\right)\exp\left(h(x,y,t)\right) \\ = & -\partial_{t_j}\exp\left(h(x,y,t)\right). \end{aligned}$$

Integrating both sides of the equation from 0 to ∞ with respect to t_j, we have

$$Q_j\int_0^\infty \exp(h(x,y,t))dt_j = \exp\left(h_{[d]\setminus\{j\}}(x,y,t)\right).$$

Integrating both sides of the equation with respect to the remaining variables, we have the equation (4.31). □

Remark. By Lemma 4.3.3, we have $Q_iQ_jg = g_{[d]\setminus\{i,j\}}$ for $i\neq j$ When the vector y is equal to zero, this equation can be written as

$$\left(2x_{ij} + 4\sum_{k=1}^{d} x_{ik}x_{jk}\partial_{x_{kk}} + 2\sum_{1\le k\neq \ell\le d} x_{ik}x_{j\ell}\partial_{x_{k\ell}}\right) g = g_{[d]\setminus\{i,j\}}.$$

By the transformation of parameters with $\Sigma = (\sigma_{k\ell}) = -\frac{1}{2}x^{-1}$, we have

$$\frac{\partial}{\partial\sigma^J_{k\ell}}\left(|\sigma|^{-1/2}g\right) = |\sigma|^{-1/2}g_{[d]\setminus\{i,j\}}.$$

It can be easily checked that this equation corresponds to the classical Schläfli's formula.

For $J=\{j_1,\ldots,j_s\}\subset[d]$, $(j_1<j_2<\cdots<j_s)$, we denote

$$\begin{aligned} x_J &= (x_{j_kj_\ell})_{1\le k,\ell\le s}, \quad y_J = (y_{j_1},\ldots,y_{j_s})^T, \\ \Sigma_J &= -\frac{1}{2}x_J^{-1} = (\sigma^J_{ij}), \quad \mu_J = \Sigma_J y_J = (\mu^J_{j_1},\ldots,\mu^J_{j_s}). \end{aligned}$$

The following differential recurrence formula holds.

Theorem 4.3.4. *For any $J \subset [d]$.*

$$\partial_{y_i} g_J = \begin{cases} \mu_i^J g_J + \sum_{j\in J} \sigma_{ij}^J g_{J\setminus\{j\}} & i \in J \\ 0 & i \notin J \end{cases} \tag{4.32}$$

$$\partial_{x_{ij}} g_J = \begin{cases} 2\partial_{y_i}\partial_{y_j} g_J & \{i,j\} \subset J,\ i<j \\ \partial_{y_i}^2 g_J & \{i\} \subset J,\ i=j \\ 0 & \textit{else}. \end{cases} \tag{4.33}$$

Proof. For $i \notin J$, we have $\partial_{y_i} g_J = 0$ since the g_J is constant with respect to y_i. For $i \in J$ Lemma 4.3.3 implies

$$\begin{aligned} \mu_i^J g_J + \sum_{j\in J} \sigma_{ij}^J g_{J\setminus\{j\}} &= \mu_i^J g_J - \sum_{j\in J} \sigma_{ij}^J \left(y_j^J + 2\sum_{k\in J} x_{jk}^J \partial_{y_k} \right) g_J \\ &= \left(\mu_i^J - \sum_{j\in J} \sigma_{ij}^J y_j^J - 2\sum_{k,j\in J} \sigma_{ij}^J x_{jk}^J \partial_{y_k} \right) g_J. \end{aligned}$$

We have $\sum_{j\in J} \sigma_{ij}^J y_j^J = \mu_i^J$ by the relation $\mu^J = \Sigma^J y^J$ and we also have

$$-2\sum_{k,j\in J} \sigma_{ij}^J x_{jk}^J \partial_{y_k} = \sum_{k\in J} \delta_{ik}\partial_{y_k} = \partial_{y_i},$$

since $-2\Sigma^J x^J$ is the identity matrix of size $|J|$. Hence, the right-hand side of (4.32) equals $\partial_{y_i} g_J$.

For $\{i,j\} \not\subset J$, $\partial_{x_{ij}} g_J = 0$ since the g_J is constant with respect to x_{ij}. For $\{i,j\} \subset J$

$$\partial_{x_{i,j}} \exp\left(h_J(x,y,t)\right) = (2-\delta_{ij})\partial_{y_i}\partial_{y_j} \exp\left(h_J(x,y,t)\right).$$

Integrating both sides we have the relation (4.33). □

By theorem 4.3.4, we have differential operators $\partial_{x_{ij}} - A_{ij}, \partial_{y_i} - A_i$ which annihilate the vector value function $G(x,y) = (g_J(x,y))_{j\subset[d]}$. Here, A_{ij} and A_i are $2^d \times 2^d$ matrices with rational function entries. In the next subsection, we prove that the system of these differential operators is a Pfaffian system in the meaning of [19, Section 2]. Note that the Pfaffian system have no singular point on

$$\{(x,y)|-x \text{ is positive definite }\}. \tag{4.34}$$

4.3.3 Pfaffian system and the holonomic rank

In this section, we give the holonomic rank of the ideal I generated by (4.26)–(4.28), and show that the differential recurrence formula (4.32) and (4.33) in the subsection 4.3.2 give a Pfaffian system associated with (4.2).

In this subsection, we denote the ring of differential operators in the variables x, y by R. The holonomic rank of I is the dimension of R/RI as a vector space over the field of rational functions $\mathbf{C}(x, y)$ (see, e.g., [28]).

At first, we give a lower bound of the holonomic rank. Since the holonomic rank equals to the dimension of holomorphic solutions of $I \bullet f = 0$ at generic points, we can obtain a lower bound of the holonomic rank by constructing linearly independent functions annihilated by I.

Lemma 4.3.5. *The holonomic rank of the ideal I in Theorem* 4.3.2 *is not less than* 2^d*, i.e.,* $\operatorname{rank} I \geq 2^d$.

Proof. For a vector $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_d) \in \{\pm 1\}^d$, let E_ε be an orthant

$$\left\{t = (t_1, \ldots, t_d) \in \mathbf{R}^d | \varepsilon_i t_i > 0 \ (i = 1, \ldots, d)\right\}.$$

It is enough to show that the following 2^d functions are linearly independent and annihilated by I;

$$g_\varepsilon(x, y) = \int_{E_\varepsilon} \exp\left(h(x, y, t)\right) dt.$$

By an analogous way as in the proof of Lemma 4.3.1, we can show that the indicator function of E_ε is annihilated by the differential operators in (4.25) for any $\varepsilon \in \{\pm 1\}^d$. Analogously to the proof of Theorem 4.2.5, we can show that the differential operators (4.26)–(4.28) annihilate $g_\varepsilon(x, y)$.

Let c_ε be a real number for $\varepsilon \in \{\pm 1\}^d$, and suppose $\sum_\varepsilon c_\varepsilon g_\varepsilon = 0$. Multiplying both sides of the equation by $(2\pi)^{-d/2}(\det \Sigma) n^{-1/2} \exp(-\frac{1}{2}\mu^t \Sigma^{-1} \mu)$, we have

$$\sum_{\varepsilon \in \{\pm 1\}^d} c_\varepsilon \mathbf{P}(E_\varepsilon | \mu, \Sigma) = 0.$$

Here, $\mathbf{P}(E_\varepsilon | \mu, \Sigma)$ is the probability of the event E_ε under the multivariate normal distribution $N(\mu, \Sigma)$. Substituting $\mu = t\varepsilon, (\varepsilon \in \{\pm 1\}^d)$ and taking a limit $t \to +\infty$, we have

$$c_\varepsilon = 0$$

Hence, the functions $g_\varepsilon(x, y)$ are linearly independent. $\square$

In order to obtain an upper bound of the holonomic rank of I, we construct bases of R/RI as a linear space over $\mathbf{C}(x, y)$. The bases correspond to the functions g_J.

For $J \subset [d]$, we put a differential operator

$$P_J = \prod_{j' \in [d] \setminus J} Q_{j'}, \tag{4.35}$$

where Q_j is the differential operator in (4.29). Note that the differential operators Q_j commute with each other. By Lemma 4.3.3, we have

$$P_J g = g_J. \tag{4.36}$$

Equation (4.36) means that the differential operator P_J corresponds to the function g_J. For example, When $d = 2$ and $J = \emptyset$, the equation (4.36) is written as follows.

$$(y_1 + 2x_{11}\partial_{y_1} + 2x_{12}\partial_{y_2})\,(y_2 + 2x_{21}\partial_{y_1} + 2x_{22}\partial_{y_2})\,g_{\{1,2\}} = 1.$$

Since the differential operator Q_j commutes with $\partial_{x_{ij}} - 2\partial_{y_i}\partial_{y_j}$ $(1 \le i < j \le d)$ and $\partial_{x_{ii}} - \partial_{y_i}^2$ $(1 \le i \le d)$, we have the following lemma.

Lemma 4.3.6. *The following formulas hold in R/RI.*

$$\partial_{x_{ij}} P_J = 2\partial_{y_i}\partial_{y_j} P_J \quad (1 \le i < j \le d, J \subset [d])\,, \tag{4.37}$$
$$\partial_{x_{ii}} P_J = \partial_{y_i}^2 P_J \quad (1 \le i \le d, J \subset [d])\,. \tag{4.38}$$

Proof. For $1 \le i < j \le d$ and $1 \le k \le d$, we have

$$\partial_{x_{ij}} Q_k = -\partial_{x_{ij}} \left(y_k + 2\sum_{\ell=1}^{d} x_{k\ell}\partial_{y_\ell} \right) = Q_k \partial_{x_{ij}} - 2\delta_{ik}\partial_{y_j} - 2\delta_{jk}\partial_{y_i},$$
$$2\partial_{y_i}\partial_{y_j} Q_k = -2\partial_{y_i}\partial_{y_j} \left(y_k + 2\sum_{\ell=1}^{d} x_{k\ell}\partial_{y_\ell} \right) = Q_k 2\partial_{y_i}\partial_{y_j} - 2\delta_{ik}\partial_{y_j} - 2\delta_{jk}\partial_{y_i}.$$

For $1 \le i \le d$ and $1 \le k \le d$, we have

$$\partial_{x_{ii}} Q_k = -\partial_{x_{ii}} \left(y_k + 2\sum_{\ell=1}^{d} x_{k\ell}\partial_{y_\ell} \right) = Q_k \partial_{x_{ii}} - 2\delta_{ik}\partial_{y_i},$$
$$\partial_{y_i}^2 Q_k = -\partial_{y_i}^2 \left(y_k + 2\sum_{\ell=1}^{d} x_{k\ell}\partial_{y_\ell} \right) = Q_k \partial_{y_i}^2 - 2\delta_{ik}\partial_{y_i}.$$

Therefore, the differential operator Q_j commutes with $\partial_{x_{ij}} - 2\partial_{y_i}\partial_{y_j}$ $(1 \le i < j \le d)$ and $\partial_{x_{ii}} - \partial_{y_i}^2$ $(1 \le i \le d)$. Then we have

$$(\partial_{x_{ij}} - 2\partial_{y_i}\partial_{y_j})P_J = P_J(\partial_{x_{ij}} - 2\partial_{y_i}\partial_{y_j}) = 0$$

in R/RI. Similarly we have $(\partial_{x_{ii}} - \partial_{y_i}^2)P_J = 0$. □

The following lemma corresponds to Lemma 4.3.3.

Lemma 4.3.7. *With the same notations as in Lemma* 4.3.3*,*

$$Q_j P_J = Q_{j,J} P_J = P_{J\setminus\{j\}} \quad (j \in J,\ J \subset [d]). \tag{4.39}$$

holds in R/RI.

Proof. By definition of P_J, it is clear that $Q_j P_J = P_{J\setminus\{j\}}$. Since $\partial_{y_\ell} Q_\ell \in I$, $\partial_\ell P_J$ is in I if $\ell \notin J$. Hence we have

$$Q_j P_J = -\left(y_j + 2\sum_{k=1}^{d} x_{jk}\partial_{y_k} \right) \prod_{\ell \in [d]\setminus J} Q_\ell = Q_{j,J} P_J$$

in R/RI. Note that Q_j and ∂_{y_k} commute if $j \neq k$. □

Theorem 4.3.8. *The following relations hold in R/RI for any $J \subset [d]$.*

$$\partial_{y_i} P_J \equiv \begin{cases} \mu_i^J P_J + \sum_{j \in J} \sigma_{ij}^J P_{J \setminus \{j\}} & i \in J \\ 0 & i \notin J, \end{cases} \tag{4.40}$$

$$\partial_{x_{ij}} P_J \equiv \begin{cases} 2\partial_{y_i}\partial_{y_j} P_J & \{i,j\} \subset J,\, i < j \\ \partial_{y_i}^2 P_J & \{i\} \subset J,\, i = j \\ 0 & else. \end{cases} \tag{4.41}$$

Proof. For $i \in J$, we can show the equation (4.40) by Lemma 4.3.7 and an analogous calculation as in the proof of (4.32). For $i \notin J$, the equation (4.40) is shown as in the proof of Lemma 4.3.7. By Lemma 4.3.6, we have the equation (4.41). □

Corollary 4.3.9. *The set of the differential operators $P_J\,(J \subset [d])$ in (4.35) spans the quotient space R/RI as a vector space over $\mathbf{C}(x, y)$.*

This corollary together with Lemma 4.3.5 establishes the following theorem.

Theorem 4.3.10.

$$\operatorname{rank} I = 2^d$$

4.4 Numerical experiments

In this section we present numerical experiments of our holonomic gradient method for orthant probabilities. Our experiments show that the holonomic gradient method is very accurate and fast compared to existing methods.

By theorem 4.3.4, we have an explicit formula for $\frac{\partial}{\partial t} G(x(t), y(t))$ when $x(t)$ and $y(t)$ are smooth functions. In order to evaluate $G(x, y)$ at (x_1, y_1), we put

$$x(t) = (1-t)x_0 + tx_1,\ y(t) = ty_1 \quad 0 \le t \le 1. \tag{4.42}$$

Here, x_0 is the diagonal matrix whose (i,i)-entry equals to that of x_1. The initial value can be written as

$$g_J(x(0), y(0)) = \prod_{j \in J} \left(-\frac{\pi}{4} x_{jj}(0) \right)^{\frac{1}{2}}. \tag{4.43}$$

Note that $\frac{\partial}{\partial t} G(x(t), y(t))$ does not have singular points on $[0, 1]$ since the Pfaffian system for $G(x, y)$ does not have singular point on (4.34).

The accuracy of the holonomic gradient method can be checked by looking at the summation $\sum_{\varepsilon \in \{\pm 1\}^d} \mathbf{P}(X \in E_\varepsilon) = 1$. Table 4.1 shows errors $|1 - \sum_{\varepsilon \in \{\pm 1\}^d} \mathbf{P}(X \in E_\varepsilon)|$ for sample data.

For a correlation matrix $R = \{\rho_{ij}\}$ with $\rho_{ij} = \rho, i \neq j$, the orthant probability can be written as a one-dimensional integral. Table 4.2 shows the values of the orthant probability calculated by the one-dimensional integral and HGM for

Table 4.1: Errors

No.	dim	error
1	2	1.760124e-08
2	3	5.473549e-08
3	4	3.373671e-08
4	5	2.265284e-09
5	6	1.120033e-08
6	7	7.330036e-09
7	8	8.705609e-09
8	9	2.288549e-09
9	10	5.024879e-10

Table 4.2: Comparing of one-dimensional integral

rho	Dunnett	HGM	error
0.0	0.00097656249807343551	0.000976562500000	1.926564e-12
0.1	0.0065864743711607976	0.006586475124405	7.532442e-10
0.25	0.026603192349344017	0.026603192572268	2.229240e-10
0.5	0.090909089922689604	0.090909086078297	3.844393e-09

certain values of ρ. The differences of the result of HGM and the one-dimensional integral are less than 10^{-06}.

Table 4.3 shows averages of computational times of holonomic gradient method and Miwa's method for 100 sample data. The mean vectors of the sample data are all zero, and the covariance matrices are randomly generated. Table 4.3 shows that our holonomic gradient method evaluates orthant probabilities faster than Miwa's method, as the dimension becomes larger.

However, it should be noted that the holonomic gradient method can be very slow when the mean vector is far away from zero. When the mean vector is far away from zero and some eigenvalue of the covariance matrix is very small, the value of the integral (4.2) becomes very large since the parameter y becomes large. In such a case, the Runge-Kutta method takes a long time.

Table 4.3: Averages of computational times

dim	Miwa	HGM	dim	Miwa	HGM
5	0.002	0.016	9	6.078	1.050
6	0.011	0.056	10	60.171	2.371
7	0.080	0.154	11	671.370	5.411
8	0.664	0.390	12	-	13.48

Chapter 5

Probability content of polyhedra

5.1 Introduction

A convex polyhedron P is the intersection of half-spaces of the d-dimensional Euclidean space $\mathbf{R}^d$. We are interested in numerical calculation of the probability $\mathbf{P}(X \in P)$, where X is a random vector distributed as a d-dimensional normal distribution with mean μ and covariance matrix Σ. When P is the orthant $\{x \in \mathbf{R}^d : x_i \geq 0, 1 \leq i \leq d\}$ in $\mathbf{R}^d$, the probability is called the orthant probability and can be regarded as a function of μ and Σ. By the inclusion–exclusion identity given in [9] and [27], the probability content of the convex polyhedron can be written as a linear combination of the orthant probabilities. For this reason, the literature includes many discussions of methods for evaluating the orthant probabilities. For example, [26] proposed a method based on recursive integration, and [11] and [13] proposed the use of the randomized quasi-Monte Carlo procedure. For details, see [12]. In [23], the probability content of a general convex polyhedron was evaluated by calculating the orthant probabilities.

The motivation of our study is to evaluate the probability content of a convex polyhedron by a completely different and novel approach that uses the holonomic gradient method (HGM) proposed in [28]. The HGM, which is based on the theory and algorithms of D-modules, is a method for numerically calculating definite integrals. It can be applied to a broad class of problems. In fact, various applications of HGM have been proposed, for example, [15], [40], and [16]. In order to apply HGM, we need to regard the probability $\mathbf{P}(X \in P)$ as a function and provide an explicit Pfaffian equation for it. In the case where P is the orthant and $\mu = 0$, the orthant probability is a function of Σ and Schläfli gave a recurrence formula for it in [38]. In [35], Plackett generalized Schläfli's result for the case where P is the orthant and $\mu \neq 0$. In [42], we provided a holonomic system and a Pfaffian equation that is associated with the orthant probability. Our Pfaffian equation corresponds with the reduction formula in [35] and [10]. In this paper, we generalize our previous results [42] and give a recurrence formula as a Pfaffian equation for the case of a general convex polyhedron.

Let $\Sigma = AA^\top$ be a Cholesky decomposition of the covariance matrix Σ, and let Y be a random vector that is d-variate normally distributed with mean vector 0 and for which the covariance matrix is the identity matrix. Then, we have $\mathbf{P}(X \in P) = \mathbf{P}(AY + \mu \in P)$, and the set $\{y \in \mathbf{R}^d : Ay + \mu \in P\}$ is also a convex polyhedron. Hence, it is enough to consider the case in which the mean vector of X is 0 and the covariance matrix of X is the identity matrix. Under this assumption, the probability $\mathbf{P}(X \in P)$ can be written as

$$\mathbf{P}(X \in P) = \frac{1}{(2\pi)^{d/2}} \int_{x \in P} \exp(-\frac{1}{2}\sum_{i=1}^{d} x_i^2) dx_1 \cdots dx_d.$$

The polyhedron P can be written as

$$P = \{x \in \mathbf{R}^d : \sum_{i=1}^{d} \tilde{a}_{ij} x_i + \tilde{b}_j \geq 0,\ 1 \leq j \leq n\}, \tag{5.1}$$

where $\tilde{a}_{ij}$ and $\tilde{b}_j$ are real numbers. We wish to study this integral with the HGM, and as a first step, we will assume that the convex polyhedron is in the "general position;" the precise definition of "general position" will be given in Section 5.3.

Let a be a $d \times n$ matrix, and let b be a vector with length n. We are interested in the analytic properties of the function

$$\varphi(a, b) = \int_{\mathbf{R}^d} \exp(-\frac{1}{2}\sum_{i=1}^{d} x_i^2) \prod_{j=1}^{n} H\left(\sum_{i=1}^{d} a_{ij} x_i + b_j\right) dx_1 \cdots dx_d, \tag{5.2}$$

which is defined on a neighborhood of $\tilde{a} = (\tilde{a}_{ij}), \tilde{b} = (\tilde{b}_j)$. Here, we denote by H the Heaviside function. Note that this function is an interesting specialization of the one studied by Aomoto in [2] and Aomoto, Kita, Orlik, and Terao in [4]. For the meaning of the specialization, see Remark 5.6.3, below. In this paper, we provide a holonomic system and a Pfaffian equation associated with this function. The Pfaffian equation is required by the HGM for $\varphi(a, b)$. In order to explicitly provide the holonomic system, we decompose the function by the inclusion–exclusion identity associated with the polyhedron P. We also show that the holonomic rank of the system is equal to the number of nonempty faces of the polyhedron P. This Pfaffian equation is a generalization of the recursion formula given by Plackett [35]. In addition, the singular locus of the Pfaffian equation is compatible with that of the Schläfli function given in [1].

This paper is constructed as follows. In Section 5.2, we will give a brief explanation of holonomic modules and Pfaffian equations. In Section 5.3, we provide an analytic continuation of the function $\varphi(a, b)$. In section 5.4, we prove an existence of an open neiborhood of a given point in general position. In Section 5.5, we give a system of linear partial differential equations with polynomial coefficients for the function $\varphi(a,b)$ and show that the system induces a holonomic module. In Section 5.6, we show that the holonomic rank of the module is equal to the number of nonempty faces of P and explicitly provide the Pfaffian equation associated with the module.

5.2 Holonomic module and Pfaffian equation

Before starting the main discussion, we briefly review holonomic modules and Pfaffian equations. For a comprehensive presentation, see [28] and the references cited therein. We denote by $D_x = \mathbf{C}\langle x_i, \partial_{x_i} : 1 \le i \le n\rangle$ the ring of differential operators of n variables $x_1, \cdots, x_n$ with polynomial coefficients. Here, we put $\partial_{x_i} = \partial/\partial x_i$. Let us consider a system of linear partial differential equations

$$\sum_{j=1}^{m} P_{kj} g_j = 0 \quad (P_{kj} \in D_x,\ 1 \le k \le m') \tag{5.3}$$

for unknown functions $g_1, \dots, g_m$. Let $(D_x)^m$ be the free D_x-module with the basis $\{g_1, \dots, g_m\}$, and let N be a D_x-submodule of $(D_x)^m$ generated by $P_k = \sum_{j=1}^{m} P_{kj} g_j \in (D_x)^m\,(1 \le k \le m')$. We denote by M the quotient module $(D_x)^m/N$.

The set consisting of the holomorphic functions on a domain $U \subset \mathbf{C}^n$ forms a left D_x-module $\mathcal{O}(U)$. For a morphism $\varphi : M \to \mathcal{O}(U)$ of left D_x-modules, the functions $\varphi(g_1), \dots, \varphi(g_m)$ satisfy system (5.3). For this reason, we call a vector-valued function $(g'_1, \dots, g'_m)$ on U a solution of M when there is a morphism of D_x-modules $\varphi : M \to \mathcal{O}(U)$ such that $\varphi(g_j) = g'_j\,(1 \le j \le m)$.

By the theory of the Gröbner basis in Weyl algebra, the *characteristic variety* $\mathrm{Char}(M)$ of M can be computed explicitly. For details, see [31]. According to the Bernstein inequality, the Krull dimension of the characteristic variety is not less than n (see, e.g., [6]). When the dimension of $\mathrm{Char}(M)$ is equal to n, the D_x-module M is said to be *holonomic*. When the system of differential equations (5.3) induces a holonomic D_x-module, we call (5.3) a holonomic system.

We denote by R_x the ring of differential operators of n variables $x_1, \cdots, x_n$ with rational function coefficients. The left D_x-module $\mathbf{C}(x) \otimes_{\mathbf{C}[x]} M$ is a left R_x-module, where $\mathbf{C}(x)$ is the field of rational functions. When the module M is holonomic, $\mathbf{C}(x) \otimes_{\mathbf{C}[x]} M$ as a linear space over $\mathbf{C}(x)$ has finite dimension. This value is called the holonomic rank of M, and we denote it by $\mathrm{rank}\, M$. Let r be the holonomic rank of M, and let $\{f_1, \dots, f_r\}$ be a basis of $\mathbf{C}(x) \otimes_{\mathbf{C}[x]} M$ as a linear space over $\mathbf{C}(x)$. Then, there exist rational functions $c_{ijk} \in \mathbf{C}(x)(1 \le i, j, k \le r)$ such that

$$\partial_{x_i} f_j = \sum_{k=1}^{r} c_{ijk} f_k \quad (1 \le i, j \le r) \tag{5.4}$$

in $\mathbf{C}(x) \otimes_{\mathbf{C}[x]} M$. Moreover, the matrices $c_i = (c_{ijk})_{j,k=1}^r$ $(1 \le i \le r)$ satisfy the integrability condition

$$\frac{\partial c_i}{\partial x_j} + c_i c_j = \frac{\partial c_j}{\partial x_i} + c_j c_i \quad (1 \le i, j \le r).$$

We call equation (5.4) a Pfaffian equation associated with the holonomic module M. The union of the zero sets of the denominators of the elements of c_i's is called the singular locus of the Pfaffian equation. Note that a Pfaffian equation associated with M depends on the choice of the basis of $\mathbf{C}(x) \otimes_{\mathbf{C}[x]} M$, and it is not unique.

5.3 Integral representation of the probability content of a polyhedron

In this section, we show that the function $\varphi(a, b)$ in (5.2) can be regarded as a real analytic function. Since the Heaviside function $H(x)$ is the hyperfunction defined by $-\frac{1}{2\pi\sqrt{-1}} \log(-z)$, we can expect the function $\varphi(a, b)$ to be expressed in terms of a logarithmic function. However, we cannot find a suitable d-simplex for the d-form obtained by replacing $H\left(\sum_{i=1}^d a_{ij} x_i + b_j\right)$ in the integrand of (5.2) with $-\frac{1}{2\pi\sqrt{-1}} \log\left(-\left(\sum_{i=1}^d a_{ij} x_i + b_j\right)\right)$. In order to overcome this difficulty, we use a decomposition of $\varphi(a, b)$, which will be given in (5.5), and show that $\varphi(a, b)$ can be written as a linear combination of complex integrals. This implies that $\varphi(a, b)$ is a real analytic function.

First, let us review some notions of polyhedra. In the remainder of this paper, we will assume that d and n are positive integers. A subset $H \subset \mathbf{R}^d$ is called a *half-space* if H can be written as $H = \{x \in \mathbf{R}^d | \sum_{i=1}^d a_i x_i + a_0 \ge 0\}$ for some $a_i, a_0 \in \mathbf{R}$. A *polyhedron* is a finite intersection of half-spaces. An inequality $\sum_{i=1}^d a_i x_i + b \ge 0$ is called *valid* for P if all the points of P satisfy the inequality. We call a subset $S \subset \mathbf{R}^d$ an *affine subspace* if S can be written as an intersection of hyperplanes. For a subset $S \subset \mathbf{R}^d$, the affine hull of S is the smallest affine subspace which contains S, and we denote it by $\mathrm{aff}(S)$. The *dimension* of a polyhedron P is the dimension of $\mathrm{aff}(P)$. For details, see [45].

In order to describe the combinatorial structure of a polyhedron, we use the notion of the abstract simplicial complex [9]. Let $\mathcal{F}$ be a set consisting of subsets of $[n] := \{1, 2, \ldots, n\}$. We call $\mathcal{F}$ an abstract simplicial complex when $J \in \mathcal{F}$ and $J' \subset J$ implies $J' \in \mathcal{F}$. Let $\mathcal{F}$ and $\mathcal{F}'$ be two abstract simplicial complexes. We say that $\mathcal{F}$ is *equal* to $\mathcal{F}'$ and denote $\mathcal{F} = \mathcal{F}'$ when $\mathcal{F}$ and $\mathcal{F}'$ are equal as sets. We say that $\mathcal{F}$ and $\mathcal{F}'$ are equivalent and denote $\mathcal{F} \cong \mathcal{F}'$ when there is a bijection $\sigma : \mathcal{F} \to \mathcal{F}'$ such that $J_1 \subset J_2$ if and only if $\sigma(J_1) \subset \sigma(J_2)$.

Let $P \subset \mathbf{R}^d$ be a polyhedron, and let $F_1, \ldots, F_n$ be all the facets of P. For each facet F_j, there is a unique half-space $H_j \subset \mathbf{R}^d$ that satisfies $(\partial H_j) \cap P = F_j$ and $P \subset H_j$ (see, e.g., exercise 2.14(iv) of Lecture 2 in [45]). We call $\mathcal{H} = \{H_1, \ldots, H_n\}$ *the family of the bounding half-spaces for the polyhedron* P.

The *nerve* of $\{F_1, \dots, F_n\}$ is the abstract simplicial complex defined by

$$\mathcal{F} = \{J \subset [n] : F_J \neq \emptyset\}, \qquad \left(F_J := \bigcap_{j \in J} F_j \right).$$

We also call $\mathcal{F}$ the *abstract simplicial complex associated with the polyhedron* P. When $\mathcal{F}$ is an abstract simplicial complex associated with a polyhedron P, we have $\{j\} \in \mathcal{F}$ for any $j \in [n]$.

Next, we introduce the notion of a polyhedron in the "general position." Since we need to consider information for points at infinity, we will use the idea of "homogenization." The *homogenization* $\hat{H}$ of a half-space $H = \{x \in \mathbf{R}^d | \sum a_i x_i + a_0 \geq 0\}$ is defined as

$$\hat{H} = \{(x_0, \dots, x_d) \in \mathbf{R}^{d+1} | \sum_{i=0}^{d} a_i x_i \geq 0\}.$$

For a family of half-spaces $\mathcal{H} = \{H_1, \dots, H_n\}$, we call $\hat{\mathcal{H}} = \{\hat{H}_0, \hat{H}_1, \dots, \hat{H}_n\}$ the *homogenization* of $\mathcal{H}$. Here, we put $\hat{H}_0 = \{x_0 \geq 0\}$. We say that a family of half-spaces $\mathcal{H} = \{H_1, \dots, H_n\}$ (or its homogenization $\hat{\mathcal{H}} = \{\hat{H}_0, \dots, \hat{H}_n\}$) is *in general position* when, for $J \subset [n+1]$,

$$\hat{F}_J := \left(\bigcap_{j \in J} \partial \hat{H}_j \right) \cap \left(\bigcap_{j=0}^{n} \hat{H}_j \right)$$

is a $d+1-|J|$-dimensional cone (i.e., the affine hull of the cone is $d+1-|J|$-dimensional affine space) or $\{0\}$. Here, we denote by $[n+1]$ the set $\{0, 1, \dots, n\}$. This is somewhat analogous to the "general position" for hyperplane arrangements in [3, Chap 2. Section 9], but we emphasize that they are different (see Example 5.3.1). The polyhedron P is *in general position* when the family $\mathcal{H}$ of the bounding half-spaces of P is in general position.

Example 5.3.1. Let $d = 2$. We define H_j $(1 \leq j \leq 4)$ by

$$H_1 := \left\{(x_1, x_2) \in \mathbf{R}^2 : x_1 \geq 0\right\}, \qquad H_2 := \left\{(x_1, x_2) \in \mathbf{R}^2 : x_1 \leq 1\right\},$$
$$H_3 := \left\{(x_1, x_2) \in \mathbf{R}^2 : x_2 \geq 0\right\}, \qquad H_4 := \left\{(x_1, x_2) \in \mathbf{R}^2 : x_2 \leq 1\right\}.$$

Then, the family of half-spaces $\mathcal{H} := \{H_1, H_2, H_3, H_4\}$ is in general position. However, the family of half-spaces $\mathcal{H}' := \{H_1, H_2, H_3\}$ is not in general position. In fact, the homogenization of $\mathcal{H}$ and $\mathcal{H}'$ can be written as

$$\hat{\mathcal{H}} = \left\{\hat{H}_0, \hat{H}_1, \hat{H}_2, \hat{H}_3, \hat{H}_4\right\}, \qquad \hat{\mathcal{H}}' = \left\{\hat{H}_0, \hat{H}_1, \hat{H}_2, \hat{H}_3\right\}$$

where

$$\hat{H}_0 := \left\{(x_0, x_1, x_2) \in \mathbf{R}^{2+1} : x_0 \geq 0\right\},$$
$$\hat{H}_1 := \left\{(x_0, x_1, x_2) \in \mathbf{R}^{2+1} : x_1 \geq 0\right\}, \quad \hat{H}_2 := \left\{(x_0, x_1, x_2) \in \mathbf{R}^{2+1} : x_1 \leq x_0\right\},$$
$$\hat{H}_3 := \left\{(x_0, x_1, x_2) \in \mathbf{R}^{2+1} : x_2 \geq 0\right\}, \quad \hat{H}_4 := \left\{(x_0, x_1, x_2) \in \mathbf{R}^{2+1} : x_2 \leq x_0\right\}.$$

Calculating $\hat{F}_J$ for each $J \subset \{0,1,2,3,4\}$, we can show that $\hat{\mathcal{H}}$ is in general position. For example, the set $\hat{H}_0 \cap \partial\hat{H}_1 \cap \hat{H}_2 \cap \partial\hat{H}_3 \cap \hat{H}_4$ is a $d+1-2$-dimensional cone, and the set $\hat{H}_0 \cap \partial\hat{H}_1 \cap \partial\hat{H}_2 \cap \hat{H}_3 \cap \hat{H}_4$ is equal to $\{0\}$. On the other hand, the family $\hat{\mathcal{H}}'$ is not in general position since the dimension of

$$\operatorname{aff}\left(\partial\hat{H}_0 \cap \partial\hat{H}_1 \cap \partial\hat{H}_2 \cap \hat{H}_3\right) = \{x \in \mathbf{R}^{d+1} : x_0 = x_1 = 0\}$$

is not equal to $d+1-3=0$.

Hence, the polyhedron $\bigcap_{j=1}^4 H_j$ in Figure 5.1(a) is in general position, but the polyhedron $\bigcap_{j=1}^3 H_j$ in Figure 5.1(b) is not in general position.

We note that the hyperplane arrangement $\{\partial H_1, \partial H_2, \partial H_3, \partial H_4\}$ is not in general position in the sense of [3], but $\mathcal{H}$ is in general position.

Example 5.3.2. Let $H_5 := \{(x_1,x_2) \in \mathbf{R}^2 : x_1 + x_2 \geq 0\}$. Using the notation in Example 5.3.1, the family of the half-space $\{H_j : 1 \leq j \leq 5\}$, which is shown in Figure 5.1(c), is not in general position. However, the polyhedron $\bigcap_{j=1}^5 H_j$ is in general position since the family of the bounding half-spaces is $\{H_j : 1 \leq j \leq 4\}$.

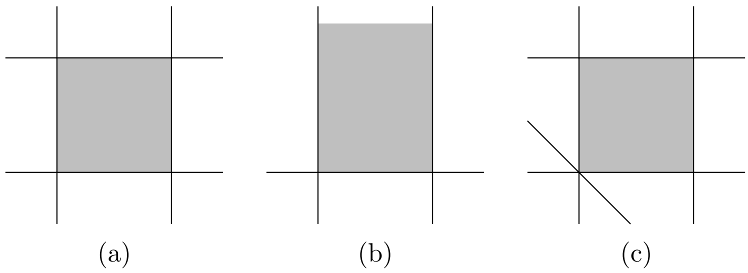

Figure 5.1: Examples of polyhedra

Remark 5.3.3. Note that our definition of general position is more restrictive than that of [27], and it is less restrictive than that of [23]. For example, the polyhedron $\bigcap_{j=1}^3 H_j$ in Example 5.3.1 is in general position by the definition in [27]; and the polyhedron $\bigcap_{j=1}^4 H_j$ in Example 5.3.1 is not in general position by the definition in [23].

Let $P \subset \mathbf{R}^d$ be a polyhedron. Suppose the family of bounding half-spaces for P is given by

$$\left\{x \in \mathbf{R}^d : \sum_{i=1}^d \tilde{a}_{ij} x_i + \tilde{b}_j \geq 0\right\} \quad (1 \leq j \leq n).$$

We denote by $\tilde{a}$ the $d \times n$ matrix $(\tilde{a}_{ij})$, and by $\tilde{b}$ the vector $(\tilde{b}_1, \ldots, \tilde{b}_n)$. Let F_j be the intersection of P and the hyperplane $\{x \in \mathbf{R}^d : \sum_{i=1}^d \tilde{a}_{ij} x_i + \tilde{b}_j = 0\}$. The sets $F_1, \dots, F_n$ are all of the facets of P. Let $\mathcal{F}$ be the nerve of $\{F_1, \dots, F_n\}$, which is the abstract simplicial complex of P.

Edelsbrunner showed the inclusion–exclusion identity for the indicator function of a polyhedron [9, Lemma 5.1.].

Proposition 5.3.4 (Edelsbrunner)**.** *If $\mathcal{F}$ is the abstract simplicial complex associated with a polyhedron P, then the indicator function of P can be written as*

$$\mathbf{1}_P = \sum_{J \in \mathcal{F}} \prod_{j \in J} \left(\mathbf{1}_{H_j} - 1\right).$$

Example 5.3.5. For the polyhedron $\bigcap_{j=1}^4 H_j$ in Example 5.3.1, the inclusion–exclusion identity can be written as follows:

$$\begin{aligned}\mathbf{1}_{\bigcap_{j=1}^4 H_j} &= 1 + (\mathbf{1}_{H_1} - 1) + (\mathbf{1}_{H_2} - 1) + (\mathbf{1}_{H_3} - 1) + (\mathbf{1}_{H_4} - 1)\\ &\quad + (\mathbf{1}_{H_1} - 1)(\mathbf{1}_{H_3} - 1) + (\mathbf{1}_{H_1} - 1)(\mathbf{1}_{H_4} - 1)\\ &\quad + (\mathbf{1}_{H_2} - 1)(\mathbf{1}_{H_3} - 1) + (\mathbf{1}_{H_2} - 1)(\mathbf{1}_{H_4} - 1).\end{aligned}$$

The first term of the right-hand side corresponds to the empty set.

With the Heaviside function H, the Edelsbrunner's identity can be written as

$$\prod_{j=1}^n H\left(\sum_{i=1}^d \tilde{a}_{ij} x_i + \tilde{b}_j\right) = \sum_{J \in \mathcal{F}} \prod_{j \in J} \left(H\left(\sum_{i=1}^d \tilde{a}_{ij} x_i + \tilde{b}_j\right) - 1\right).$$

Under the general position assumption, this identity can be generalized as follows.

Theorem 5.3.6. *In the notation above, if the polyhedron P is in general position, then there exists a neighborhood U of $(\tilde{a}, \tilde{b})$ such that the equation*

$$\prod_{j=1}^n H\left(\sum_{i=1}^d a_{ij} x_i + b_j\right) = \sum_{J \in \mathcal{F}} \prod_{j \in J} \left(H\left(\sum_{i=1}^d a_{ij} x_i + b_j\right) - 1\right)$$

holds for all $(a, b, x) \in U \times \mathbf{R}^d$.

Our proof of Theorem 5.3.6 is technical, and it is independent from the other parts of this paper; thus, it is given in Section 5.4.

Consider n polynomials $f_j(a, b, x) = \sum_{i=1}^d a_{ij} x_i + b_j$ $(1 \le j \le n)$ with variables a_{ij}, b_j, x_i $(1 \le i \le d, 1 \le j \le n)$, and let

$$\chi_F(a, b, x) = \prod_{j \in F} \left(H(f_j(a, b, x)) - 1\right)$$

for $F \in \mathcal{F}$. Note that $\chi_\emptyset(a, b, x) = 1$. We put

$$\varphi_F(a, b) = \int_{\mathbf{R}^d} \frac{1}{(2\pi)^{d/2}} \exp(-\frac{1}{2} \sum_{i=1}^d x_i^2) \chi_F(a, b, x) dx \quad (F \in \mathcal{F}).$$

By Theorem 5.3.6, the function $\varphi(a,b)$ in (5.2) can be decomposed as

$$\varphi(a,b) = \sum_{F\in\mathcal{F}} \varphi_F(a,b) \tag{5.5}$$

on a neighborhood of $(\tilde{a},\tilde{b})$ if the polyhedron P is in general position.

In order to give analytic continuations of the function $\varphi(a,b)$, it is enough to consider $\varphi_F(a,b)$. For $F\in\mathcal{F}$, let $\alpha_F(a) = (\alpha_{ij}(a))_{i,j\in F}$ be an $|F|\times|F|$ matrix, where $|F|$ is the number of the elements in F and

$$\alpha_{ij}(a) = \sum_{k=1}^{d} a_{ki}a_{kj} \quad (1\le i,j\le n).$$

This is a submatrix of the Gram matrix of a. The matrices $\alpha_F(a)$ are symmetric and positive semidefinite. Since the function $\varphi_F(a,b)$ can be written as

$$\int_{\mathbf{R}^d} \frac{1}{(2\pi)^{d/2}}(-1)^{|F|}\exp(-\frac{1}{2}\sum_{i=1}^{d} x_i^2)\prod_{j\in F} H(-f_j(a,b,x)),$$

we can expect that $\varphi_F(a,b)$ is written by a d-simplex and the d-form

$$\frac{1}{(2\pi)^{d/2}}\exp(-\frac{1}{2}\sum_{i=1}^{d} z_i^2)\prod_{j\in F}\left(\frac{\log(f_j(\tilde{a},\tilde{b},z))}{2\pi\sqrt{-1}}\right)dz.$$

In fact, we can find a suitable d-simplex and thus have the following lemma.

Lemma 5.3.7. *If the polyhedron P is in general position, then for $F\in\mathcal{F}$, the value of $\varphi_F(\tilde{a},\tilde{b})$ can be written as*

$$\sum_{\lambda\in\{\pm1\}^{|F|}}\int_{\gamma^\lambda} \frac{1}{(2\pi)^{d/2}}\exp(-\frac{1}{2}\sum_{i=1}^{d} z_i^2)\prod_{j\in F}\left(\frac{\log(f_j(\tilde{a},\tilde{b},z))}{2\pi\sqrt{-1}}\right)dz. \tag{5.6}$$

Here, for $\lambda\in\{\pm1\}^{|F|}$, γ^λ is a smooth map from $\mathbf{R}^d$ to $\mathbf{C}^d$. We suppose the multivalued function $\log$ *satisfies* $\log(1)=0$ *and the branch cut is* $\{z\in\mathbf{C} : \Re(z)\le 0\}$.

Proof. Let s be the number of elements in F. By the general position assumption, s is not greater than d. We denote by $j(1),\dots,j(s)$ all of the elements of F. Since the polyhedron P is in general position, the vectors $u_k = (\tilde{a}_{1j(k)},\dots,\tilde{a}_{dj(k)})^\top (1\le k\le s)$ are linearly independent . Consequently, the determinant $\alpha_F(\tilde{a})$ is not zero. Let $u_k (s<k\le d)$ be an orthonormal basis of the orthogonal complement of the subspace $\sum_{k=1}^{s}\mathbf{R}u_k\subset\mathbf{R}^d$. We denote the vector u_k by $u_k=(u_{1k},\dots,u_{dk})^\top$. The matrix $u=(u_{ij})$ is regular, and we set $u^{-1}=(u^{ij})$. Without loss of generality, we can assume $\det u>0$. Under this assumption, we have $\det u = \sqrt{|\det\alpha_F(\tilde{a})|}$. Let $\varepsilon(t)$ be a positive bounded function on $\mathbf{R}$, i.e., $\inf_{t\in\mathbf{R}} \varepsilon(t) > 0$ and $\sup_{t\in\mathbf{R}} \varepsilon(t) < \infty$. For a vector $(\lambda_1, \dots, \lambda_s) \in \{-1, 1\}^s$, we define $\gamma^\lambda : \mathbf{R}^d \to \mathbf{C}^d$ by

$$\gamma_j^\lambda(t) = \sum_{i=1}^d u^{ij} \left(\lambda_i t_i + \lambda_i \sqrt{-1}\varepsilon(\lambda_i t_i)\right) \quad (t \in \mathbf{R}^d).$$

Here, we put $\lambda_j = 1$ for $s < j \le d$.

By the coordinate transformation $w_j = \sum_{i=1}^d u_{ij} z_i$, the integral (5.6) can be written as

$$\frac{1}{(2\pi)^{d/2}} \frac{1}{|\alpha_F(\tilde{a})|^{1/2}} \sum_{\lambda\in\{\pm 1\}^s} \int_{\gamma'_\lambda} \exp(-\frac{1}{2}\sum_{i=1}^d (\sum_{j=1}^d u^{ji} w_j)^2) \prod_{k=1}^s \left(\frac{\log(w_k + \tilde{b}_k)}{2\pi\sqrt{-1}}\right) dw,$$

where $\gamma'_{\lambda j}(t) = \lambda_j t_j + \lambda_j \sqrt{-1}\varepsilon(\lambda_j t_j)$ $(t \in \mathbf{R}^d)$. Calculating this integral recursively, we have

$$\frac{1}{(2\pi)^{d/2}} \frac{1}{|\alpha_F(\tilde{a})|^{1/2}} \int_E \exp(-\frac{1}{2}\sum_{i=1}^d (\sum_{j=1}^d u^{ji} y_j)^2) dy,$$

where $E = \{y \in \mathbf{R}^d : y_j + \tilde{b}_j \le 0,\ j \in F\}$. By the coordinate transformation $x_i = \sum_{j=1}^d u^{ji} y_j$, the above integral is

$$\frac{1}{(2\pi)^{d/2}} \int_{f_j(\tilde{a},\tilde{b},x)\le 0, j\in F} \exp(-\frac{1}{2}\sum_{i=1}^d x_i^2) dx,$$

which is equal to $\varphi_F(\tilde{a}, \tilde{b})$. □

Moreover, we have the following proposition.

Proposition 5.3.8. *Let U_F be a domain $\{(a, b) : \det \alpha_F(a) \neq 0\}$. The function $\varphi_F(a, b)$ can be written as*

$$\sum_{\lambda\in\{\pm 1\}^{|F|}} \int_{\gamma_\lambda} \frac{1}{(2\pi)^{d/2}} \exp(-\frac{1}{2}\sum_{i=1}^d z_i^2) \prod_{j\in F} \left(-\frac{\log(f_j(a, b, z))}{2\pi\sqrt{-1}}\right) dz \qquad (5.7)$$

on a connected open neighborhood of $(\tilde{a}, \tilde{b})$ in U_F. Here, we suppose the multi-valued function $\log$ *satisfies* $\log(1) = 0$ *and the branch cut is* $\{z \in \mathbf{C} : \Re(z) \le 0\}$.

Proof. Since $\det(\alpha_F(a)) \neq 0$, by arguments similar to those in the proof of Lemma 5.3.7, $\varphi_F(a, b)$ can be written as

$$\sum_{\lambda\in\{\pm 1\}^{|F|}} \int_{\gamma'_\lambda} \frac{1}{(2\pi)^{d/2}} \exp(-\frac{1}{2}\sum_{i=1}^d z_i^2) \prod_{j\in F} \left(\frac{\log(f_j(a, b, z))}{2\pi\sqrt{-1}}\right) dz.$$

The matrix u' and the integral path γ'_λ can be constructed similarly.

We need to show that the above integral is equal to (5.7). There is a smooth path $u(t)$ ($t \in [0,1]$) in the general linear group of degree d over $\mathbf{C}$ such that $u(0) = u$ and $u(1) = u'$. The homotopy between γ_λ and γ'_λ is given by

$$\gamma_{\lambda j}(s)(t) = \sum_{k=1}^{d} u^{jk}(s)\left(\lambda_k t_k + \lambda_k \sqrt{-1}\varepsilon(\lambda_k t_k)\right) \quad (t \in \mathbf{R}^d).$$

Consequently, the value of the integral on the right-hand side of (5.7) does not change when we change the integral path with γ'_λ. □

Corollary 5.3.9. *The function $\varphi(a,b)$ is a real analytic function, and it has an analytic continuation along every path in $\bigcap_{F\in\mathcal{F}} U_F$.*

5.4 Proof of Theorem 5.3.6

In this section, we prove Theorem 5.3.6. For this purpose, we need to present some notation and some lemmas, most importantly, Theorem 5.4.6. The argument in the proof of Lemma 5.4.1 is also important, since analogous arguments will appear repeatedly in the proof of Theorem 5.4.6.

Let $a = (a_1, \ldots, a_n) \in \mathbf{R}^{(d+1)\times n}$ be a matrix where we denote by a_j the j-th column vector of a. Put $a_0 := (1, 0, \ldots, 0)^\top \in \mathbf{R}^{d+1}$. For a matrix a, we put

$$\begin{aligned}
H_j(a) &:= \left\{ x \in \mathbf{R}^d : \sum_{i=1}^{d} a_{ij}x_i + a_{0j} \geq 0 \right\} \quad (j \in [n]),\\
\mathcal{H}(a) &:= \{H_1(a), \ldots, H_n(a)\},\\
P(a) &:= \bigcap_{j=1}^{n} H_j(a),\\
F_j(a) &:= \partial H_j(a) \cap P(a) \quad (j \in [n]),\\
F_J(a) &:= \bigcap_{j\in J} F_j(a) \quad (J \subset [n]),\\
\mathcal{F}(a) &:= \{J \subset [n] : F_J(a) \neq \emptyset\}.
\end{aligned}$$

Note that $F_j(a)$ is not necessarily a facet of $P(a)$, and $\mathcal{F}(a)$ is not necessarily equivalent to the abstract simplicial complex associated with $P(a)$. For this difficulty, we need the notion of families of half-spaces in general position. In fact, in Lemma 5.4.5, we will show that the abstract simplicial complex associated with $P(a)$ is equivalent to $\mathcal{F}(a)$ under the general position assumption, which is required by the proof of Theorem 5.3.6.

In order to consider combinatorial structures at the point at infinity, we introduce the following notion: for the abstract simplicial complex $\mathcal{F}$ associated with a polyhedron, the *homogenization* of $\mathcal{F}$ is the abstract simplicial complex defined by

$$\hat{\mathcal{F}} := \left\{ J \subset [n+1] : \hat{F}_J \neq \{0\} \right\} \quad ([n+1] = \{0, 1, \ldots, n\}).$$

Since $\mathcal{F} \subset \hat{\mathcal{F}}$, we have $\{j\} \in \hat{\mathcal{F}}$ for $j \in [n]$. Note that $\{0\} \in \hat{\mathcal{F}}$ does not hold in general. The following are the "homogenization" of the notations given in the previous paragraph. For a matrix a, we put

$$\begin{aligned}
\hat{H}_j(a) &:= \left\{ x \in \mathbf{R}^{d+1} : \sum_{i=0}^{d} a_{ij} x_i \geq 0 \right\} \quad (j \in [n+1]), \\
\hat{\mathcal{H}}(a) &:= \left\{ \hat{H}_j(a) : j \in [n+1] \right\}, \\
\hat{P}(a) &:= \bigcap_{j=1}^{n+1} \hat{H}_j(a), \\
\hat{F}_j(a) &:= \partial \hat{H}_j(a) \cap \hat{P}(a) \quad (j \in [n+1]), \\
\hat{\mathcal{F}}(a) &:= \left\{ J \subset [n+1] : \hat{F}_J(a) \neq \{0\} \right\}, \\
\hat{F}_J(a) &:= \bigcap_{j \in J} \hat{F}_j(a) \quad (J \subset [n+1]).
\end{aligned}$$

For all $J \in [n+1]$, $\hat{F}_J(a)$ includes the zero vector. Analogous to the case of the abstract simplicial complex associated with polyhedra, we have $\mathcal{F}(a) \subset \hat{\mathcal{F}}(a)$. For a family of half-spaces in general position, we have the following lemma, which is required by the proof of Lemma 5.4.5.

Lemma 5.4.1. *Suppose $\mathcal{H}(a)$ is in general position. Let $J \subset [n]$. If $J \in \mathcal{F}(a)$, then the set $F_J(a)$ is a $d - |J|$-dimensional face of $P(a)$. If $J \notin \mathcal{F}(a)$, then the set $F_J(a)$ is empty. In particular, $P(a)$ is a d-dimensional polyhedron.*

Proof. If $J \notin \mathcal{F}(a)$, it is obvious that $F_J(a) = \emptyset$. Let us consider the case where $J \in \mathcal{F}(a)$. Put $J = \{j(0), \ldots, j(s)\}$, $J^c := [n+1] \backslash J$. Since $\mathcal{H}(a)$ is in general position, we have $\hat{F}_J(a) = \{0\}$ or

$$\hat{F}_J(a) \supsetneq \hat{F}_J(a) \cap \partial \hat{H}_k(a) \tag{5.8}$$

for any $k \in J^c$. In fact, the equality of (5.8) contradicts the assumption about the dimension when $\hat{F}_J(a) \neq \{0\}$. However, we have $\hat{F}_J(a) \neq \{0\}$, since $J \in \mathcal{F}(a)$ implies $J \in \hat{\mathcal{F}}(a)$. Hence, condition (5.8) holds for any $k \in J^c$. For $k \in J^c$, take $x^{(k)} \in \hat{F}_J(a) \backslash \partial \hat{H}_k(a)$. Then the affine combination

$$\frac{1}{n+1-|J|} \sum_{k \in J^c} x^{(k)}$$

is an element of

$$\hat{F}_J(a) \cap \left(\bigcap_{k \in J^c} \operatorname{int}(\hat{H}_k(a)) \right), \tag{5.9}$$

where $\operatorname{int}(S)$ denotes the interior of S. Since $0 \in J^c$, we have

$$\hat{F}_J(a) \cap \{x_0 = 1\} \cap \left(\bigcap_{k \in J^c} \operatorname{int}(\hat{H}_k(a)) \right) \neq \emptyset.$$

Let x be an element of this set, and let B be an open ball centered at x whose closure is included in $\bigcap_{k\in J^c} \operatorname{int}\left(\hat{H}_k(a)\right)$. Let x' be an arbitrary point in $\{x_0 = 1\}\cap\bigcap_{j\in J}\partial\hat{H}_j(a)$, and let x'' be an intersection point of ∂B and the line between x and x'. Since $x'' \in \hat{F}_J(a)$ and x' can be written as an affine combination of x and x'', we have $x' \in \operatorname{aff}(\hat{F}_J(a))$. By the arbitrariness of x', we have

$$\operatorname{aff}(F_J(a)) \cong \operatorname{aff}\left(\hat{F}_J(a)\cap\{x_0=1\}\right) = \bigcap_{j\in J}\partial\hat{H}_j(a)\cap\{x_0=1\}.$$

Analogously, we have

$$\operatorname{aff}(\hat{F}_J(a)) = \bigcap_{j\in J}\partial\hat{H}_j(a). \tag{5.10}$$

Since $\hat{\mathcal{H}}(a)$ is in general position, the left-hand side of (5.10) is a $d+1-|J|$-dimensional cone. Consequently, the vectors $a_{j(1)},\dots,a_{j(s)}$ are linearly independent. Moreover, the vectors $a_0, a_{j(1)},\dots,a_{j(s)}$ are linearly independent. In fact, if there are $c_j\,(j\in J)$ such that $a_0 = \sum_{j\in J} c_j a_j$, then we have $0 = \sum_{j\in J} c_j\left(\sum_{i=1}^d a_{ij}x_i + a_{0j}\right) = 1$ for $x \in F_J$. Hence, the dimension of $\operatorname{aff}(F_J)$ is equal to $d-|J|$. □

By the argument in the proof of Lemma 5.4.1, when $\mathcal{H}(a)$ is in general position, $J\in\hat{\mathcal{F}}(a)$ implies $F_J(a)\neq\emptyset$. Hence, we have the following corollary.

Corollary 5.4.2. *When $\mathcal{H}(a)$ is in general position, for $J\subset[n]$, $J\in\mathcal{F}(a)$ holds if and only if $J\in\hat{\mathcal{F}}(a)$ holds.*

Similarly, by the argument in the proof of Lemma 5.4.1, we have the following.

Corollary 5.4.3. *When $\mathcal{H}(a)$ is in general position and $J=\{j(1),\dots,j(s)\}\in\mathcal{F}(a)$, the column vectors $a_{j(1)},\dots,a_{j(s)}$ are linearly independent.*

We now review the Farkas lemma in [45].

Proposition 5.4.4 (Farkas lemma)**.** *The inequality $\sum_{i=1}^d c_i x_i + c_0 \geq 0$ is valid for $P(a)$ if and only if one of the following conditions holds:*

(i) There exist $\lambda_j\geq 0\,(j\in[n])$ such that

$$\sum_{j=1}^n a_{ij}\lambda_j = c_i, \quad \sum_{j=1}^n a_{0j}\lambda_j \leq c_0 \quad (1\leq i\leq d);$$

(ii) There exist $\lambda_j\geq 0\,(j\in[n])$ such that

$$\sum_{j=1}^n a_{ij}\lambda_j = 0, \quad \sum_{j=1}^n a_{0j}\lambda_j < 0 \quad (1\leq i\leq d).$$

Note that condition (ii) in the Farkas lemma implies $P(a) = \emptyset$.

Lemma 5.4.1 and Proposition 5.4.4 imply the following lemma, which is not required by the proof of Theorem 5.4.6 but is required by that of Theorem 5.3.6.

Lemma 5.4.5. *If $\mathcal{H}(a)$ is in general position, then the abstract simplicial complex associated with $P(a)$ is equivalent to $\mathcal{F}(a)$.*

Proof. Note that we assume $d \geq 1$ and $P(a)$ is a d-dimensional polyhedron. Let F be a facet of $P(a)$; then there is an inequality $\sum_{i=1}^d c_i x_i + c_0 \geq 0$ valid for $P(a)$ such that

$$F = P(a) \cap \left\{ x \in \mathbf{R}^d : \sum_{i=1}^d c_i x_i + c_0 = 0 \right\}.$$

Since $P(a)$ is not empty, condition (ii) in the Farkas lemma does not hold. By the Farkas lemma, there exist $\lambda_j \geq 0 \, (j \in [n])$ such that

$$\sum_j a_{ij}\lambda_j = c_i, \quad \sum_j a_{0j}\lambda_j \leq c_0 \quad (1 \leq i \leq d).$$

Moreover, there is a unique λ_j which is greater than 0. In fact, if there is not such a j, then we have $c_i = 0 \, (1 \leq i \leq d)$. This implies $F = P(a)$ or $F = \emptyset$. This is a contradiction.

If there are two indexes $j(1)$ and $j(2)$ such that $\lambda_{j(1)} > 0$, $\lambda_{j(2)} > 0$, and $j(1) \neq j(2)$, then we have $F \subset P(a) \cap \partial H_{j(1)}(a) \cap \partial H_{j(2)}(a)$. This contradicts the fact that F is $d - |J|$-dimensional.

Therefore, there is a unique j such that $a_{ij}\lambda_j = c_i$, $a_{0j}\lambda_j \leq c_0$. Since we have

$$0 \leq \lambda_j \left(\sum_{i=1}^d a_{ij} x_i + a_{0j} \right) \leq \sum_{i=1}^d c_i x_i + c_0 = 0$$

for $x \in F$, we have $c_0 = a_{0j}\lambda_j$. Consequently, $F = F_j(a)$.

If $F_j(a) \neq \emptyset$, then, by Lemma 5.4.1, $F_j(a)$ is a facet of $P(a)$.

Therefore, all the facets of $P(a)$ are given by $F_j(a) \, (\{j\} \in \mathcal{F}(a))$. Consequently, the abstract simplicial complex $\mathcal{F}(a)$ associated with $P(a)$ is equivalent to $\mathcal{F}(a)$. $\square$

Theorem 5.4.6. *Let $P \subset \mathbf{R}^d$ be a polyhedron in general position, let n be the number of facets of P, and let $\mathcal{F}$ be the abstract simplicial complex associated with P. Then, the set*

$$U = \left\{ a \in \mathbf{R}^{(d+1)\times n} : \mathcal{H}(a) \text{ is in general position and } \hat{\mathcal{F}}(a) = \hat{\mathcal{F}} \right\}$$

is an open set of $\mathbf{R}^{(d+1)\times n}$.

Proof. Our proof is given by writing U as an intersection of finite open sets $V_J\,(J \subset [n+1])$:

$$U = \left(\bigcap_{J \subset [n+1]} V_J \right). \tag{5.11}$$

STEP 1 . First, we define V_J. For $J = \{j(1), \ldots, j(s)\} \in \hat{\mathcal{F}}$, let V_J consist of $a \in \mathbf{R}^{(d+1)\times n}$ such that the vectors $a_{j(1)}, \ldots, a_{j(s)}$ are linearly independent and the set (5.9) includes a nonzero element. For $J \subset [n+1]$ such that $J \notin \hat{\mathcal{F}}$, let V_J be the intersection of

$$\left\{ a \in \mathbf{R}^{(d+1)\times n} : \hat{F}_J(a) = \{0\} \right\}$$

and

$$\bigcap_{j=1}^{n} \left\{ a \in \mathbf{R}^{(d+1)\times n} : a_j \neq 0 \right\}. \tag{5.12}$$

STEP 2 . Next, we show that the right-hand side of (5.11) includes the left-hand side. Let $J = \{j(1), \ldots, j(s)\} \subset [n+1]$.

Suppose $J \in \hat{\mathcal{F}}$. Let $a \in U$; then we have $\hat{\mathcal{F}}(a) = \hat{\mathcal{F}}$ by the definition of U. When $J^c = \emptyset$, the set (5.9) is equal to $\hat{F}_J(a)$ and includes a nonzero element. When $J^c \neq \emptyset$, by an argument analogous to that in the proof of Lemma 5.4.1, we can show that the set (5.9) includes a nonzero element. Consequently, equation (5.10) holds. Since the left-hand side of (5.10) is a $d+1-|J|$-dimensional affine subspace, the vectors $a_{j(1)}, \ldots, a_{j(s)}$ are linearly independent. Hence, we have $U \subset V_J\,(J \in \mathcal{F})$.

Suppose $J \notin \hat{\mathcal{F}}$. For $a \in U$, it is obvious that $\hat{F}_J(a) = \{0\}$. By the assumption for $\mathcal{F}$, we have $\{j\} \in \hat{\mathcal{F}}$ for arbitrary $j \in [n]$. By the argument in the previous paragraph, $a \in U$ implies $a_j \neq 0$. Hence, we have $U \subset V_J\,(J \notin \mathcal{F})$.

Therefore, the right-hand side of (5.11) includes the left-hand side.

STEP 3 . We show that the left-hand side of (5.11) includes the right-hand side. Suppose a matrix a is included in the right-hand side of (5.11). Let $J \subset [n]$. When $J \in \mathcal{F}$, $a \in V_J$ implies that the set (5.9) includes a nonzero element. Hence, we have $\hat{F}_J(a) \neq \{0\}$. When $J \notin \mathcal{F}$, $a \in V_J$ implies $\hat{F}_J(a) = \{0\}$. Consequently, we have $\hat{\mathcal{F}}(a) = \hat{\mathcal{F}}$.

Next, we show that $\mathcal{H}(a)$ is in general position. It is enough to show that $\hat{F}_J(a)$ is a $d+1-|J|$-dimensional cone or that it is equal to $\{0\}$ for $J = \{j(1), \ldots, j(s)\} \subset [n+1]$. When $J \in \hat{\mathcal{F}}$, set (5.9) is not empty and equation (5.10) holds. Since the vectors $a_{j(1)}, \ldots, a_{j(s)}$ are linearly independent, the right-hand side of (5.10) is a $d+1-|J|$-dimensional affine subspace. When $J \notin \hat{\mathcal{F}}$, we have $\hat{F}_J(a) = \{0\}$. Hence, $\mathcal{H}(a)$ is in general position. Therefore, equation (5.11) holds.

STEP 4 . Let $J \subset [n+1]$. We show that V_J is an open set.

Suppose $J \in \mathcal{F}$. Let $a \in V_J$. Since the vectors $a_j (j \in J)$ are linearly independent, there is a subset $I \subset [n]$ such that $|I| = |J|$ and the $|I| \times |J|$

matrix $(a_{ij})_{i\in I, j\in J}$ has the inverse matrix $(c_{ji})_{i\in I,j\in J}$. Let $\sigma : I \to J$ be a bijection, then

$$y_k = \begin{cases} \sum_{i\in I} a_{i\sigma(k)} x_i & (k \in I) \\ x_k & (k \in I^c) \end{cases}$$

defines a coordinate transformation. By this coordinate transformation, the condition that set (5.9) includes a nonzero element can be written as follows: there exists $y \in \mathbf{R}^d$ such that

$$\begin{cases} y_{\sigma^{-1}(j)} + \sum_{i\in I^c} y_i a_{ij} = 0 & (j \in J), \\ \sum_{i\in I} \sum_{k\in J} y_{\sigma^{-1}(k)} c_{ki} a_{ij} + \sum_{i\in I^c} a_{ij} y_i > 0 & (j \in J^c), \\ y \neq 0. \end{cases} \tag{5.13}$$

Let V_{IJ} be the set consisting of a such that the matrix $(a_{ij})_{i\in I,j\in J}$ is invertible and there exists y satisfying condition (5.13). Then, we have

$$V_J = \bigcup_{\substack{I\subset[n],\\ |I|=|J|}} V_{IJ}.$$

In equation (5.13), the value of y_j $(j \in J)$ is determined uniquely by the value of the other variables. We can regard (5.13) as a condition for the variables a, $y_j (j \notin J)$. The set V_{IJ} is the projection of the following open subset of $\mathbf{R}^{(d+1)\times n} \times \mathbf{R}^{n-|J|}$: the set consisting of $(a, y_j)_{j\notin J} \in \mathbf{R}^{(d+1)\times n} \times \mathbf{R}^{n-|J|}$ that satisfies (5.13) and in which $(a_{ij})_{i\in I,j\in J}$ is invertible. This implies that V_{IJ} is open. Consequently, V_J is an open set.

Suppose $J \notin \mathcal{F}$. There is the isomorphism between the set (5.12) and $\mathbf{R}^n_{>0} \times (S^d)^n$. Here, $(S^d)^n$ denotes the n-th direct product of a d-dimensional sphere S^d. Since V_J is closed under the transformation which multiplies column vectors by positive numbers,

$$(a_{ij}) \mapsto (\lambda_j a_{ij}), \quad (\lambda_j \in \mathbf{R}_{>0}),$$

the image of V_J under the isomorphism is $\mathbf{R}^n_{>0} \times \big((S^d)^n \cap V_J\big)$. It is enough to show $(S^d)^n \cap V_J$ is an open set. Since the projection

$$(S^d)^n \times S^d \to (S^d)^n, \quad (a, x) \mapsto a$$

is a continuous map from a compact set to a Hausdorff space, it is a closed map. The set $(S^d)^n \backslash V_J$ is the projection of the following closed set of $(S^d)^n \times S^d$: the intersection of $(S^d)^n \times S^d$ and the subset of $\mathbf{R}^{(d+1)\times n} \times \mathbf{R}^{d+1}$ consisting of (a, x) satisfying

$$\begin{cases} \sum_{i=0}^d a_{ij} x_i = 0 & (j \in J) \\ \sum_{i=0}^d a_{ij} x_i \geq 0 & (j \in J^c). \end{cases}$$

Hence, $(S^d)^n \cap V_J$ is open. $\square$

Proof of Theorem 5.3.6. With the notation of Theorem 5.4.6, it is enough to show that for $a \in U$, the abstract simplicial complex associated with $P(a)$ is equivalent to $\mathcal{F}$. Let $a \in U$. By Lemma 5.4.5, the abstract simplicial complex associated with $P(a)$ is equivalent to $\mathcal{F}(a)$. By Corollary 5.4.2, $\hat{\mathcal{F}}(a) = \hat{\mathcal{F}}$ implies $\mathcal{F}(a) = \mathcal{F}$. □

5.5 Holonomic modules

In this section, we explicitly give a system of differential equations for the function

$$\chi_P(a, b, x) = \sum_{F \in \mathcal{F}} \chi_F(a, b, x), \tag{5.14}$$

and show that the system is holonomic. Since the equation

$$\varphi(a, b) = \int_{\mathbf{R}^d} \frac{1}{(2\pi)^{d/2}} \exp(-\frac{1}{2}\sum_{i=1}^{d} x_i^2)\chi_P(a, b, x)dx$$

holds on a neighborhood of $(\tilde{a}, \tilde{b})$, a holonomic system for $\varphi(a, b)$ can be given as the integration module of a holonomic module associated with

$$\exp(-\frac{1}{2}\sum_{i=1}^{d} x_i^2)\chi_P(a, b, x).$$

For more about the integration module, see [6] and [32].

For $J \subset [n]$, we define a hyperfunction χ^J by

$$\chi^J = \partial_b^J \chi_P \quad (\partial_b^J := \prod_{i \in J} \partial_{b_i}).$$

Note that $\chi^{\emptyset} = \chi_P$.

Lemma 5.5.1. *If $J \subset [n]$ is not an element of $\mathcal{F}$, then we have $\partial_b^J \chi_F = 0$ for $F \in \mathcal{F}$. Consequently, we have $\chi^J = 0$.*

Proof. Since $\mathcal{F}$ is an abstract simplicial complex, we have $J \not\subset F$. Take $k \in J \backslash F$, then we have $\partial_b^J \chi_F = 0$, since $\partial_{b_k} \prod_{j \in F} (H(f_j(a, b, x)) - 1) = 0$. □

We now provide a system of differential equations for the χ^J's. Let $g = (g^J)_{J \in \mathcal{F}}$ be a vector whose elements are functions indexed by the set $\mathcal{F}$. Let us consider the system defined by the following:

$$\partial_{x_i} g^J = \sum_{j=1}^{n} a_{ij}\partial_{b_j} g^J \qquad (1 \le i \le d,\ J \in \mathcal{F}), \tag{5.15}$$

$$\partial_{a_{ij}} g^J = x_i \partial_{b_j} g^J \qquad (1 \le i \le d,\ 1 \le j \le n,\ J \in \mathcal{F}), \tag{5.16}$$

$$\partial_{b_j} g^J = g^{J \cup \{j\}} \qquad (j \in J^c,\ J \in \mathcal{F}), \tag{5.17}$$

$$f_j g^J = 0 \qquad (j \in J,\ J \in \mathcal{F}), \tag{5.18}$$

where $g^{J \cup \{j\}} = 0$ for $J \cup \{j\} \notin \mathcal{F}$.

Lemma 5.5.2. *Let $g = (\partial_b^J \chi_F)_{J\in\mathcal{F}}$ for $F \in \mathcal{F}$. Then the function g satisfies equations* (5.15)*,* (5.16)*,* (5.17)*, and* (5.18)*.*

Proof. When $F = \emptyset$, it is obvious that equations (5.15), (5.16), (5.17), and (5.18) hold, since $\chi_\emptyset = 1$. Suppose F is not the empty set.

We first check equation (5.15). Since $\partial_{b_j}\chi_F(a,b,x) = 0$ for $j \notin F$, we have

$$\partial_{x_i}\chi_F(a,b,x) = \sum_{j\in F} a_{ij}\partial_{b_j}\chi_F(a,b,x) = \sum_{j=1}^{n} a_{ij}\partial_{b_j}\chi_F(a,b,x).$$

Here, we apply the chain rule for hyperfunctions. The equation above implies (5.15).

We now show equation (5.16). When $j \notin F$, both sides of (5.16) are equal to 0. When $j \in F$, we have $\partial_{a_{ij}} g^J = x_i \partial_{b_j} g^J$ by the chain rule.

Equation (5.17) holds by Lemma 5.5.1. For $J \subset F$, we have

$$g^J = \prod_{k\in J}\delta(f_k(a,b,x)) \prod_{k\in F\setminus J}\left(H(f_k(a,b,x)) - 1\right).$$

This implies (5.18). □

By (5.14), we have $\chi^J = \sum_{F\in\mathcal{F}} \partial_b^J \chi_F$. By this equation and Lemma 5.5.2, we have the following proposition.

Proposition 5.5.3. *The vector-valued function $g = (\chi^J)_{J\in\mathcal{F}}$ satisfies equations* (5.15)*,* (5.16)*,* (5.17)*, and* (5.18)*.*

Next, we show that the system defined by (5.15), (5.16), (5.17), and (5.18) for χ_P is a holonomic system. In the remainder of this paper, we will frequently use the following rings:

$$\begin{aligned} D_{abx} &:= \mathbf{C}\langle a_{ij}, b_j, x_i, \partial_{a_{ij}}, \partial_{b_j}, \partial_{x_i} : 1 \le i \le d,\ 1 \le j \le n\rangle, \\ D_{ab} &:= \mathbf{C}\langle a_{ij}, b_j, \partial_{a_{ij}}, \partial_{b_j} : 1 \le i \le d,\ 1 \le j \le n\rangle, \\ \mathbf{C}[a,b,x,\xi_a,\xi_b,\xi_x] &:= \mathbf{C}[a_{ij}, b_j, x_i, \xi_{a_{ij}}, \xi_{b_j}, \xi_{x_i} : 1 \le i \le d,\ 1 \le j \le n]. \end{aligned}$$

We also use the free modules $(D_{abx})^{|\mathcal{F}|}$, $(D_{ab})^{|\mathcal{F}|}$, and $\mathbf{C}[a,b,x]^{|\mathcal{F}|}$, whose basis is $\{g^J : J \in \mathcal{F}\}$.

Proposition 5.5.4. *Let N_χ be the sub left D_{abx}-module of $(D_{abx})^{|\mathcal{F}|}$ generated by*

$$\left(\partial_{x_i} - \sum_{j=1}^{n} a_{ij}\partial_{b_j}\right) g^J \quad (1 \le i \le d,\ J \in \mathcal{F}), \tag{5.19}$$

$$\left(\partial_{a_{ij}} - x_i\partial_{b_j}\right) g^J \quad (1 \le i \le d,\ 1 \le j \le n,\ J \in \mathcal{F}), \tag{5.20}$$

$$\partial_{b_j} g^J - g^{J\cup\{j\}} \quad (j \in J^c,\ J \in \mathcal{F}), \tag{5.21}$$

$$f_j g^J, \quad (j \in J, J \in \mathcal{F}). \tag{5.22}$$

Then, the quotient module $M_\chi = (D_{abx})^{|\mathcal{F}|}/N_\chi$ is holonomic.

Proof. The principal symbols of (5.19), (5.20), (5.21), and (5.22) are the following:

$$(\xi_{x_i} - \sum_{j=1}^{n} a_{ij}\xi_{b_j})g^J \quad (1 \le i \le d,\ J \in \mathcal{F}), \tag{5.23}$$

$$(\xi_{a_{ij}} - x_i\xi_{b_j})g^J \quad (1 \le i \le d,\ 1 \le j \le n,\ J \in \mathcal{F}), \tag{5.24}$$

$$f_j g^J,\ \xi_{b_k} g^J \quad (j \in J,\ k \in J^c,\ J \in \mathcal{F}). \tag{5.25}$$

For $J \in \mathcal{F}$, let V_J be an algebraic variety defined by

$$\xi_{x_i} - \sum_{j=1}^{n} a_{ij}\xi_{b_j},\ \xi_{a_{ij}} - x_i\xi_{b_j} \quad (1 \le i \le d,\ 1 \le j \le n), \tag{5.26}$$

$$f_j,\ \xi_{b_k} \quad (j \in J,\ k \in J^c). \tag{5.27}$$

By [31, Proposition 1], the union $\bigcup_{J\in\mathcal{F}} V_J$ includes $\mathrm{Char}(M)$. Since the rank of the Jacobian matrix of (5.26) and (5.27) is $nd + n + d$, the Krull dimension of V_J is equal to $nd + n + d$. Hence, the dimension of $\mathrm{Char}(M)$ is not greater than $nd + n + d$. □

Let N_q be a D_{abx}-submodule of $(D_{abx})^{|\mathcal{F}|}$ generated by (5.20), (5.21), (5.22), and

$$\left(x_i + \partial_{x_i} - \sum_{j=1}^{n} a_{ij}\partial_{b_j}\right) g^J \quad (1 \le i \le d,\ J \in \mathcal{F}). \tag{5.28}$$

Proposition 5.5.4 implies that the quotient module $M_q = (D_{abx})^{|\mathcal{F}|)}/N_q$ is a holonomic module. Moreover, the function

$$q^J(a,b,x) = \exp(-\frac{1}{2}\sum_{i=1}^{d} x_i^2)\chi^J \quad (J \in \mathcal{F})$$

is a solution of N_q (see [41]). By Lemma 5.5.2, the function

$$\exp(-\frac{1}{2}\sum_{i=1}^{d} x_i^2)\partial_b^J \chi_F \quad (J \in \mathcal{F})$$

is also a solution of N_q for $F \in \mathcal{F}$.

Calculating the integration module of N_q with respect to x, we have the following theorem.

Theorem 5.5.5. *Let N be the sub left D_{ab}-module of $(D_{ab})^{|\mathcal{F}|}$ generated by*

$$\left(\partial_{a_{ij}} - \sum_{k=1}^{n} a_{ik}\partial_{b_k}\partial_{b_j}\right) g^J \quad (1 \le i \le d,\ 1 \le j \le n,\ J \in \mathcal{F}), \tag{5.29}$$

$$\partial_{b_j} g^J - g^{J\cup\{j\}} \quad (j \in J^c,\ J \in \mathcal{F}), \tag{5.30}$$

$$(b_j + \sum_{k=1}^{n}\sum_{i=1}^{d} a_{ij}a_{ik}\partial_{b_k})g^J \quad (j \in J,\ J \in \mathcal{F}). \tag{5.31}$$

Then the quotient module $M = (D_{ab})^{|\mathcal{F}|}/N$ is isomorphic to the integration module $M_q/\sum_{i=1}\partial_{x_i}M_q$. Consequently, M is a holonomic module.

Proof. We denote by ι the canonical morphism from $(D_{ab})^{|\mathcal{F}|}$ to the integration module $M_q/\sum_{i=1}\partial_{x_i}M_q$. Let $P_i = x_i + \partial_{x_i} - \sum_{k=1}^n a_{ik}\partial_{b_k}$. Since we have

$$\left(\partial_{a_{ij}} - x_i\partial_{b_j}\right)g^J + \partial_{b_j}P_i g^J = \left(\partial_{x_i}\partial_{b_j} + \partial_{a_{ij}} - \sum_{k=1}^n a_{ik}\partial_{b_k}\partial_{b_j}\right)g^J, \tag{5.32}$$

$$f_j g^J - \sum_{i=1}^d a_{ij}P_i g^J = (b_j + \sum_{k=1}^n\sum_{i=1}^d a_{ij}a_{ik}\partial_{b_k})g^J, \tag{5.33}$$

the D_{abx}-module N_q is generated by (5.28), (5.21), and

$$\left(\partial_{x_i}\partial_{b_j} + \partial_{a_{ij}} - \sum_{k=1}^n a_{ik}\partial_{b_k}\partial_{b_j}\right)g^J \quad (1 \le i \le d,\ J \in \mathcal{F}),$$

$$(b_j + \sum_{k=1}^n\sum_{i=1}^d a_{ij}a_{ik}\partial_{b_k})g^J \quad (j \in J,\ J \in \mathcal{F}).$$

Consequently, we have $N \subset \ker\iota$.

Next, we show the opposed inclusion $\ker\iota \subset N$. Regarding $(D_{ab})^{|\mathcal{F}|}$ as a subset of $(D_{abx})^{|\mathcal{F}|}$, the left D_{ab}-module $N_q + \sum\partial_{x_i}(D_{abx})^{|\mathcal{F}|}$ includes $\ker\iota$. Since the module N_q is generated by (5.20), (5.21), (5.22), and (5.28), the module $N_q + \sum\partial_{x_i}(D_{abx})^{|\mathcal{F}|}$ can be written as

$$\sum_{\lambda\in\Lambda} D_{abx}Q_\lambda + \sum_{J\in\mathcal{F}}\sum_{i=1}^d D_{abx}P_i g^J + \sum_{i=1}^d \partial_{x_i}(D_{abx})^{|\mathcal{F}|}. \tag{5.34}$$

Here, we denote by $Q_\lambda\,(\lambda \in \Lambda)$ the differential operators in (5.20), (5.21), and (5.22). The left D_{ab}-module (5.34) is equal to

$$\sum_{\lambda\in\Lambda} D_{ab}Q_\lambda + \sum_{J\in\mathcal{F}}\sum_{i=1}^d D_{abx}P_i g^J + \sum_{i=1}^d \partial_{x_i}(D_{abx})^{|\mathcal{F}|}. \tag{5.35}$$

Note that the first term of (5.35) is different from that of (5.34). In fact, for any P_i and the differential operator in (5.20), (5.21), and (5.22), we have

$$P_i\left(\partial_{a_{k\ell}} - x_k\partial_{b_\ell}\right)g^J = \left(\partial_{a_{k\ell}} - x_k\partial_{b_\ell}\right)P_i g^J,$$
$$P_i\left(\partial_{b_k}g^J - g^{J\cup\{k\}}\right) = \partial_{b_k}P_i g^J - P_i g^{J\cup\{k\}},$$
$$P_i\left(\sum_{k=1}^d a_{k\ell}x_k + b_\ell\right)g^J = \left(\sum_{k=1}^d a_{k\ell}x_k + b_\ell\right)P_i g^J.$$

These equations imply that for any differential operator Q_λ in (5.20), (5.21), and (5.22), the operator P_iQ_λ is an element of $\sum_{i,J} D_{abx}P_i g^J$. Consequently, module (5.35) includes $x_i Q_\lambda = \sum_{k=1}^n a_{ik}\partial_{b_k} Q_\lambda + P_i Q_\lambda - \partial_{x_i} Q_\lambda$. By induction on the multi-index $\alpha \in \mathbf{N}_0^d$, module (5.35) includes $x^\alpha Q_\lambda$ for any $\alpha \in \mathbf{N}_0^d$. Hence, module (5.35) includes (5.34). The opposite inclusion is obvious.

We denote by Q'_λ ($\lambda \in \Lambda$) the differential operators in (5.29), (5.30), and (5.31). By (5.32) and (5.33), the left D_{ab}-module

$$\sum_{\lambda\in\Lambda} D_{ab}Q'_\lambda + \sum_{J\in\mathcal{F}}\sum_{i=1}^d D_{abx}P_i g^J + \sum_{i=1}^d \partial_{x_i}(D_{abx})^{|\mathcal{F}|}$$

is equal to module (5.35). Obviously, this module is equal to the left D_{ab}-module

$$\sum_{\lambda\in\Lambda} D_{ab}Q'_\lambda + \sum_{\alpha\in\mathbf{N}_0^d}\sum_{J\in\mathcal{F}}\sum_{i=1}^d x^\alpha D_{ab}P'_i g^J + \sum_{i=1}^d \partial_{x_i}(D_{abx})^{|\mathcal{F}|}, \tag{5.36}$$

where $P'_i := x_i - \sum_{k=1}^n a_{ik}\partial_{b_k}$. Note that the module $N_q + \sum \partial_{x_i}(D_{abx})^{|\mathcal{F}|}$ is equal to (5.36). Since $\ker\iota$ is a subset of the intersection of $(D_{ab})^{|\mathcal{F}|}$ and (5.36), we have

$$\ker\iota \subset \sum_{\lambda\in\Lambda} D_{ab}Q'_\lambda + \sum_{\alpha\in\mathbf{N}_0^d}\sum_{J\in\mathcal{F}}\sum_{i=1}^d x^\alpha D_{ab}P'_i g^J.$$

Let $P \in \ker\iota$. Then the element P can be written as $Q + R$, where $Q \in \sum_{\lambda\in\Lambda} D_{ab}Q'_\lambda$ and $R \in \sum_{\alpha\in\mathbf{N}_0^d}\sum_{J\in\mathcal{F}}\sum_{i=1}^d x^\alpha D_{ab}P'_i g^J$. The element $P-Q = R$ is an element of the D_{abx}-module $L := \sum_{J\in\mathcal{F}}\sum_{i=1}^d D_{abx}P'_i g^J$. Let $\prec$ be a lex order which satisfies $x_i \succ m$ for $1 \le i \le d$ and any monomial $m \in D_{ab}$. Since the Gröbner basis for L with respect to $\prec$ is given by $\left\{P'_i g^J : 1 \le i \le d,\ J \in \mathcal{F}\right\}$, the leading term of $P - Q$ must be divided by some $x_i g^J$ $(1 \le i \le d,\ J \in \mathcal{F})$. Since $P - Q \in (D_{ab})^{|\mathcal{F}|}$, we have $P - Q = 0$. Therefore, we have $\ker\iota \subset \sum D_{ab}Q'_\lambda$. □

5.6 Pfaffian equation and holonomic rank

We will evaluate the holonomic rank of M and derive a Pfaffian equation associated with M. The following lemma gives a lower bound of the holonomic rank.

Lemma 5.6.1. *The real analytic functions $\varphi_F(a,b)$, where F runs over the abstract simplicial complex $\mathcal{F}$ associated with P, are linearly independent solutions of M. Consequently, the holonomic rank of M is not less than the number of the nonempty faces of P.*

Proof. By Proposition 5.3.8 and Theorem 5.5.5, it is obvious that the function φ_F is a solution of M.

Suppose a matrix $\tilde{a} = (\tilde{a}_{ij})$ and a vector $\tilde{b} = (\tilde{b}_j)$ satisfy equation (5.1). Note that $\tilde{a}_{ij}$ and $\tilde{b}_j$ are real numbers. Let $U \subset \mathbf{C}^{d\times n}$ be a sufficiently small neighborhood of $\tilde{a}$. Note that the function φ_F is defined on $\{(a,b) : a \in U, b \in$

$\mathbf{C}^n\}$, by Proposition 5.3.8. We prove the linear independence of these functions. Suppose $\mathcal{F} = \{F_1, \dots, F_s\}$ and

$$|F_1| \le |F_2| \le \cdots \le |F_s|.$$

We denote φ_{F_j} by φ_j. Suppose $\sum_{i=1}^{s} c_j \varphi_j = 0$ for some complex numbers c_j $(1 \le i \le s)$. Take an arbitrary k such that $1 \le k \le s$, and suppose $c_1 = \cdots = c_{k-1} = 0$. It is enough to show that $c_k = 0$.

Note that $F_\ell \not\subset F_k$ for $\ell > k$, since $|F_\ell| \ge |F_k|$. Define $a(t)$ and $b(t)$ as

$$a_{ij}(t) = \tilde{a}_{ij}, \quad b_j(t) = \begin{cases} \tilde{b}_j + t & j \in F_k \\ \tilde{b}_j - t & j \notin F_k \end{cases}, \quad t \in [0, \infty).$$

Since there is an element of F_ℓ which is not included in F_k for $\ell > k$, we have $\lim_{t\to\infty} \prod_{j \in F_\ell} H(f_j(a(t), b(t), x)) = 0$ for all $x \in \mathbf{R}^d$. By the Lebesgue convergence theorem, we have

$$\lim_{t\to\infty} \varphi_\ell(a(t), b(t)) = 0 \quad (\ell > k).$$

Since we have $\lim_{t\to\infty} \prod_{j\in F_k} H(f_j(a(t), b(t), x)) = 1$ for all $x \in \mathbf{R}^d$, again by the Lebesgue convergence theorem, we have $\lim_{t\to\infty} \varphi_\ell(a(t), b(t)) = 1$. By the assumption of the induction, we have $\sum_{j=k}^{s} c_j \varphi_j(a(t), b(t)) = 0$. Taking the limit of both sides as $t \to \infty$, we have $c_k = 0$. $\square$

Theorem 5.6.2. *The holonomic rank of M is equal to the number of nonempty faces of P, i.e.,*

$$\mathrm{rank}(M) = |\mathcal{F}|.$$

In addition, a Pfaffian equation associated with M is given by

$$\partial_{a_{ij}} g^J = \sum_{k=1}^{n} a_{ik} \partial_{b_k} \partial_{b_j} g^J \quad (1 \le i \le d,\ 1 \le j \le n,\ J \in \mathcal{F}), \tag{5.37}$$

$$\partial_{b_j} g^J = g^{J \cup \{j\}} \quad (j \in J^c,\ J \in \mathcal{F}), \tag{5.38}$$

$$\partial_{b_j} g^J = -\sum_{k \in J} \alpha_J^{jk}(a) \left(b_k g^J + \sum_{\ell \in J^c} \alpha_{k\ell}(a) g^{J \cup \ell} \right) \quad (j \in J,\ J \in \mathcal{F}). \tag{5.39}$$

Here, $(\alpha_F^{ij}(a))_{i,j \in F}$ is the inverse matrix of $\alpha_F(a)$. Note that the right-hand side of (5.37) *can be reduced by* (5.38) *and* (5.39).

Proof. We have (5.37) and (5.38) by (5.29) and (5.30), respectively. By (5.30) and (5.31), the right-hand side of (5.39) can be written as

$$\begin{aligned} -\sum_{k \in J} \alpha_J^{jk} \left(b_k g^J + \sum_{\ell \in J^c} \alpha_{k\ell} g^{J \cup \ell} \right) &= \sum_{k \in J} \alpha_J^{jk} \left(\sum_{\ell=1}^{d} \alpha_{k\ell} \partial_{b_\ell} g^J - \sum_{\ell \in J^c} \alpha_{k\ell} \partial_{b_\ell} g^J \right) \\ &= \sum_{k \in J} \alpha_J^{jk} \sum_{\ell \in F} \alpha_{k\ell} \partial_{b_\ell} g^J = \partial_{b_j} g^J. \end{aligned}$$

Consequently, we have (5.39).

By (5.37), (5.38), and (5.39), the module $\mathbf{C}(a,b) \otimes_{\mathbf{C}[a,b]} M$ is spanned by g^J $(J \in \mathcal{F})$ as a linear space over $\mathbf{C}(a,b)$, and we have $\mathrm{rank}(M) \leq |\mathcal{F}|$. By this inequality and Lemma 5.6.1, we have $\mathrm{rank}(M) = |\mathcal{F}|$. $\square$

Remark 5.6.3. We note that the Heaviside function $H\left(\sum_{i=1}^{d} a_{ij}x_i + b_j\right)$ is equal to $\lim_{\lambda \to 0}\left(\sum_{i=1}^{d} a_{ij}x_i + b_j\right)_+^{\lambda}$ as a Schwartz distribution. Thus we may expect that our integral representation is a "specialization" of the integral considered in the cohomology groups in [4]. However, we have no rigorous understanding of this limiting procedure of the twisted cohomology. An interesting observation is that the holonomic rank of M can be smaller than the dimension of the twisted cohomology group $H^d(\mathcal{G}, \nabla)$, where $\mathcal{G} := \mathbf{C}^d - V\left(\prod_{j=1}^{n}\left(\sum_{i=1}^{d} a_{ij}x_i + b_j\right)\right)$ and ∇ is a connection defined in [4]. In our case, the dimension of the twisted cohomology is $\sum_{i=0}^{d}\binom{n}{i}$. The number of faces of the polyhedron in general position with n facets can be smaller than this value.

Acknowledgements. The author is deeply grateful to Professor Nobuki Takayama for useful discussions and constant encouragement, and wishes to thank Professor Akimichi Takemura for important suggestions.

Bibliography

[1] K. Aomoto. Analytic structure of Schläfli function. *Nagoya Math.J.*, 68:1–16, 1977.

[2] K. Aomoto. Configurations and invariant gauss-manin connections of integrals i. *Tokyo J. Math.*, 5(2):249–287, 1982.

[3] K. Aomoto and M. Kita. *Theory of Hypergeometric Functions*. Springer, New York, 2011.

[4] K. Aomoto, M. Kita, P. Orlik, and H. Terao. Twisted de rham cohomology groups of logarithmic forms. *Advances in Mathematics*, 128:119–152, 1997.

[5] D. Bayer and M. Stillman. On the complexity of computing syzygies. *Journal of Symbolic Computation*, 6:135–147, 1988.

[6] J.E. Björk. *Rings of differential operators*. North-Holland, New York, 1979.

[7] B. Cano, A. M. Stuart, E. Suli, and J. O. Warren. Stiff oscillatory systems, delta jumps and white noise. *Foundations of Computational Mathematics*, 1:69–99, 2001.

[8] D. O'Shea D. Cox, J. Little. *Ideals, Varieties, and Algorithms*. Springer, 1992.

[9] H. Edelsbrunner. The union of balls and its dual shape. *discrete & computational geometry*, 13:415–440, 1995.

[10] H. I. Gassmann. Multivariate normal probabilities: implementing an old idea of plackett's. *Journal of Computational and Graphical Statistics*, 12(3):731–752, 2003.

[11] A. Genz. Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics*, 1:141–150, 1992.

[12] A. Genz and F. Bretz. *Computation of Multivariate Normal and t Probabilities*. Lecture Notes in Statistics. Springer-Verlag, Heidelberg, 2009.

[13] A. Genz, F. Bretz, T. Miwa, X. Mi, F. Leisch, F. Scheipl, and T. Hothorn. *mvtnorm: Multivariate Normal and t Distributions*, 2013. R package version 0.9-9996.

[14] GSL. GNU scientific library. `http://www.gnu.org/software/gsl/`, 2012.

[15] H. Hashiguchi, Y. Numata, N. Takayama, and A. Takemura. Holonomic gradient method for the distribution function of the largest root of Wishart matrix. *Journal of Multivariate Analysis*, 117:296–312, 2013.

[16] T. Hibi, K. Nishiyama, and N. Takayama. Pfaffian systems of a-hypergeometric equations. `http://arxiv.org/abs/1212.6103`, 2012.

[17] J. T. Kent. The Fisher-Bingham distribution on the sphere. *Journal of Royal Statistical Society Ser. B*, 44:71–80, 1982.

[18] T. Koyama. A holonomic ideal which annihilates the Fisher-Bingham integral. *Funkcialaj Ekvacioj*, 56:51–61, 2013.

[19] T. Koyama, H. Nakayama, K. Nishiyama, and N. Takayama. Holonomic gradient descent for the fisher-bingham distribution on the d-dimensional sphere. arXiv:1201.3239, to apper in Computational Statistics, 2012.

[20] T. Koyama, H. Nakayama, K. Nishiyama, and N. Takayama. The holonomic rank of the Fisher-Bingham system of differential equations. `http://arxiv.org/abs/1205.6144`, 2012.

[21] K. Kume and S. G. Walker. On the Fisher-Bingham distribution. *Statistics and Computing*, 19:167–172, 2009.

[22] K. Kume and A. T. A. Wood. Saddlepoint approximations for the Bingham and Fisher-Bingham normalising constants. *Biometrika*, 92:465–476, 2005.

[23] S. Kuriki, T. Miwa, and A. J. Hayter. Abstract tubes associated with perturbed polyhedra with applications to mulitidimensional normal probability computations. In T. Hibi, editor, *Harmony of Gröbner Bases and the Modern Industrial Society*, pages 169–183, Singapore, 2012. World Scientific.

[24] K. V. Mardia and P. E. Jupp. *Directional Statistics*. John Wiley & Sons, 2000.

[25] E. Miller, L.Matusevich, and U.Walther. Homological methods for hypergeometric families. *Journal of the American Mathematical Society*, 18:919–941, 2005.

[26] A. Miwa, J. Hayter, and S. Kuriki. The evaluation of general non-centered orthant probabilities. *Journal of the Royal Statistical Society, Ser. B*, 65:223–234, 2003.

[27] Daniel Q. Naiman and Henry P. Wynn. Abstract tubes,improved inclusion-exclusion identities and inequalities and importance sampling. *The Annals of Statistics*, 25(5):1954–1983, 1997.

[28] H. Nakayama, K. Nishiyama, M. Noro, K. Ohara, T. Sei, N. Takayama, and A. Takemura. Holonomic gradient descent and its application to the Fisher-Bingham integral. *Advances in Applied Mathematics*, 47:639–658, 2011.

[29] K. Nishiyama and N. Takayama. Incomplete $\dashv$-hypergeometric systems. In T. Hibi, editor, *Harmony of Gröbner Bases and the Modern Industrial Society*, pages 193–212, Singapore, 2012. World Scientific.

[30] J. Nocedal and S. Wright. *Numerical Optimization*. Springer, 2007.

[31] T. Oaku. Computation of the characteristic variety and the singular locus of a system of differential equations with polynomial coefficients. *Japan Journal of Industrial and Applied Mathematics*, 11:485–497, 1994.

[32] T. Oaku. Algorithms for b-functions, restrictions, and algebraic local cohomology groups of d-modules. *Advances in Applied Mathematics*, 19:61–105, 1997.

[33] T. Oaku. Algorithms for integrals of holonomic functions over domains defined by polynomial inequalities. *Journal of Symbolic Computation*, 50:1–27, 2013.

[34] T. Oaku and N. Takayama. Algorithms for d-modules —restriction, tensor product, localization, and local cohomology groups. *Journal of Pure and Applied Algebra*, 156:267–308, 2001.

[35] R. L. Plackett. A reduction formula for normal multivariate integrals. *Biometrika*, 41(3/4):351–360, 1954.

[36] Risa/Asir. A computer algebra system. `http://www.math.kobe-u.ac.jp/Asir/`, 2012.

[37] M. Saito, B. Strumfels, and N. Takayama. *Gröbner Deformations of Hypergeometric Differential Equations*. Springer, 2000.

[38] L. Schläfli. On the multiple integral $\int^n dxdy \ldots dz$ whose limits are $p_1 = a_1 x + b_1 y + \cdots + h_1 z > 0, \quad p_2 > 0, \ldots, p_n > 0$ and $x^2 + y^2 + \cdots + z^2 < 1$. *Quart. J. Pure Appl. Math*, 2:269–301, 1858.

[39] L. Schwartz. *Théorie des distributions*. Hermann, Paris, 1966.

[40] T. Sei, H. Shibata, A. Takemura, K. Ohara, and N. Takayama. Properties and applications of fisher distribution on the rotation group. *Journal of Multivariate Analysis*, 116:440–455, 2013.

[41] N. Takayama T. Oaku, Y. Shiraki. Algorithms for d-modules and numerical analysis. *Z.M.Li, W.Sit(editors), Computer Mathematics, World scientific*, pages 23–39, 2003.

[42] T.Koyama and A.Takemura. Calculation of orthant probabilities by the holonomic gradient method. `http://arxiv.org/abs/1211.6822`, 2012.

[43] A. T. A. Wood. Some notes on the Fisher-Bingham family on the sphere. *Communications in Statistics — Theory and Methods*, 17:3881–3897, 1988.

[44] D. Zeilberger. A holonomic systems approach to speical function identities. *Journal of Computational and Applied Mathematics*, 32:321–368, 1990.

[45] G. M. Ziegler. *Lectures on Polytopes.* Springer–Verlag, New York, 1995.