11:2 2020

Algebraic Statistics

ALGEBRAIC ANALYSIS OF ROTATION DATA

Michael F. Adamer, András C. Lőrincz,
Anna-Laura Sattelberger and Bernd Sturmfels

msp

Algebraic Statistics
Vol. 11, No. 2, 2020
https://doi.org/10.2140/astat.2020.11.189

ALGEBRAIC ANALYSIS OF ROTATION DATA

MICHAEL F. ADAMER, ANDRÁS C. LŐRINCZ,
ANNA-LAURA SATTELBERGER AND BERND STURMFELS

We develop algebraic tools for statistical inference from samples of rotation matrices. This rests on the theory of D-modules in algebraic analysis. Noncommutative Gröbner bases are used to design numerical algorithms for maximum likelihood estimation, building on the holonomic gradient method of Sei, Shibata, Takemura, Ohara, and Takayama. We study the Fisher model for sampling from rotation matrices, and we apply our algorithms to data from the applied sciences. On the theoretical side, we generalize the underlying equivariant D-modules from SO(3) to arbitrary Lie groups. For compact groups, our D-ideals encode the normalizing constant of the Fisher model.

1. Introduction

Many of the multivariate functions that arise in statistical inference are holonomic. Being holonomic roughly means that the function is annihilated by a system of linear partial differential operators with polynomial coefficients whose solution space is finite-dimensional. Such a system of PDEs can be written as a left ideal in the Weyl algebra, or D-ideal, for short. This representation allows for the application of algebraic geometry and algebraic analysis, including the use of computational tools, such as Gröbner bases in the Weyl algebra [28; 30].

This important connection between statistics and algebraic analysis was first observed by a group of scholars in Japan, and it led to their development of the *holonomic gradient method* (HGM) and the *holonomic gradient descent* (HGD). We refer to [10; 16; 31] and to further references given therein. The point of departure for the present article is the work of Sei et al. [29], who developed HGD for data sampled from the rotation group SO(n), and the article of Koyama [16] who undertook a study of the associated equivariant D-module.

The statistical model we examine in this article is the Fisher distribution on the group of rotations, defined in (1) and (2). The aim of maximum likelihood estimation (MLE) is to learn the model parameters Θ that best explain a given data set. In our case, the MLE problem is difficult because there is no simple formula for evaluating the normalizing constant of the distribution. This is where algebraic analysis comes in. The normalizing constant is a holonomic function of the model parameters, and we can use its holonomic D-ideal to derive an efficient numerical scheme for solving the maximum likelihood estimation problem.

ORCID: Adamer 0000-0002-8996-7167 / Sattelberger 0000-0002-2308-4070.
MSC2020: 14F10, 33F10, 62F10, 62H11, 62R01, 90C90.
Keywords: algebraic analysis, Fisher model, maximum likelihood estimation, directional statistics, holonomic gradient method, Fourier transform.

The present article addresses diverse audiences and it offers multiple points of entry. First, we give an introduction to the use of D-module methods in statistics. We focus on data in the group of rotations in 3-space, and we advance both the theory and the practice of this application. Readers with an interest in the applied sciences may start in Section 5, as that displays a panorama of occurrences of rotation data in the real world. Experts in representation theory can jump straight to Section 6, where our approach is developed for arbitrary Lie groups. For such readers, the particular group SO(3) is merely an example.

Our presentation is organized as follows. Section 2 is purely expository. Here, we introduce the Fisher model, and we express its log-likelihood function in terms of the sufficient statistics of the given data. These are obtained from the singular value decomposition of the sample mean. In Section 3, we turn to algebraic analysis. We review the holonomic D-ideal in [29] that annihilates the normalizing constant of the Fisher distribution, and we derive its associated Pfaffian system. Passing to $n \geq 3$, we next study the D-ideals on $\mathrm{SO}(n)$ given in [16]. First new results can be found in Theorem 3.4 and in Propositions 3.5 and 3.6.

Section 4 is concerned with numerical algorithms for maximum likelihood estimation. We develop and compare holonomic gradient ascent (HGA), holonomic BFGS (H-BFGS), and a holonomic Newton method. We implemented these methods in the language R. Section 5 highlights how samples of rotation matrices arise in the sciences and engineering. Topics range from materials science and geology to astronomy and biomechanics. We apply holonomic methods to data from the literature, and we discuss both successes and challenges.

The D-ideal of the normalizing constant is of independent interest from the perspective of representation theory, as it generalizes naturally to other Lie groups. The development of that theory is our main new mathematical contribution. This work is presented in Section 6.

2. The Fisher model for random rotations

In this section, we introduce the Fisher model on the rotation group, building on [29]. The group $\mathrm{SO}(3)$ consists of all real 3×3 matrices Y that satisfy $Y^{\mathrm{t}}Y = \mathrm{Id}_3$ and $\det(Y) = 1$. This is a smooth algebraic variety of dimension 3 in the 9-dimensional space $\mathbb{R}^{3\times 3}$. See [5] for a study of rotation groups from the perspective of combinatorics and algebraic geometry.

The Haar measure on $\mathrm{SO}(3)$ is the unique probability measure μ that is invariant under the group action. The *Fisher model* is a family of probability distributions on $\mathrm{SO}(3)$ that is parametrized by 3×3 matrices Θ. For a fixed Θ, the density of the *Fisher distribution* equals

$$f_\Theta(Y) \;=\; \frac{1}{c(\Theta)} \cdot \exp(\mathrm{tr}(\Theta^{\mathrm{t}} \cdot Y)) \qquad \text{for all} \quad Y \in \mathrm{SO}(3). \tag{1}$$

This is the density with respect to the Haar measure μ. The denominator is the *normalizing constant*. It is chosen such that $\int_{\mathrm{SO}(3)} f_\Theta(Y)\mu(dY) = 1$. This requirement is equivalent to

$$c(\Theta) \;=\; \int_{\mathrm{SO}(3)} \exp(\mathrm{tr}(\Theta^{\mathrm{t}} \cdot Y))\mu(dY). \tag{2}$$

This function is the Fourier–Laplace transform of the Haar measure μ; see Remark 6.6. The Fisher model is an exponential family. It is one of the simplest statistical models on SO(3). The task at hand is the accurate numerical evaluation of the integral (2) for given Θ in $\mathbb{R}^{3\times 3}$. We begin with the observation that, since integration is against the Haar measure, the function (2) is invariant under multiplying Θ on the left or right by a rotation matrix:

$$c(Q \cdot \Theta \cdot R) \;=\; c(\Theta) \qquad \text{for all} \;\; Q, R \in \mathrm{SO}(3).$$

In order to evaluate (2), we can therefore restrict to the case of diagonal matrices. Namely, given any 3×3 matrix Θ, we first compute its *sign-preserving singular value decomposition*

$$\Theta \;=\; Q \cdot \mathrm{diag}(x_1, x_2, x_3) \cdot R.$$

Sign-preserving means that $Q,\ R \in \mathrm{SO}(3)$ and $|x_1| \geq x_2 \geq x_3 \geq 0$. For nonsingular Θ this implies that $x_1 > 0$ whenever $\det(\Theta) > 0$ and $x_1 < 0$ otherwise.

The normalizing constant $c(\Theta)$ is the following function of the three singular values:

$$\tilde{c}(x_1, x_2, x_3) \;:=\; c(\mathrm{diag}(x_1, x_2, x_3)) \;=\; \int_{\mathrm{SO}(3)} \exp(x_1 y_{11} + x_2 y_{22} + x_3 y_{33}) \mu(dY). \tag{3}$$

The statistical problem we address in this paper is parameter estimation for the Fisher model. Suppose we are given a finite sample $\{Y_1, Y_2, \ldots, Y_N\}$ from the rotation group SO(3). We refer to Figure 1 for a concrete example. Our aim is to find the parameter matrix Θ whose Fisher distribution f_Θ best

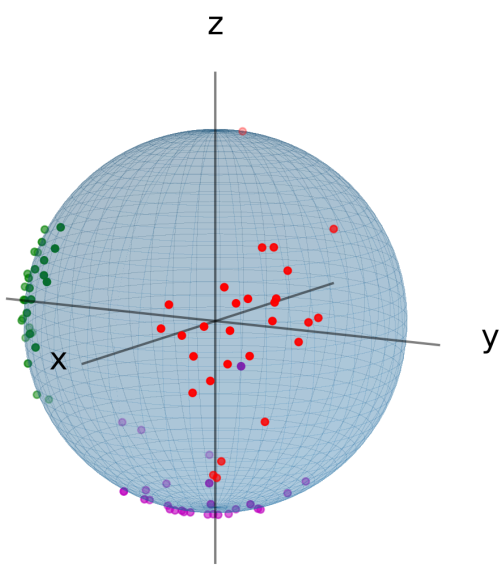

Figure 1. A dataset of 28 rotations from a study in vectorcardiography [7], a method in medical imaging. Each point represents the rotation of the unit standard vector on the x-axis (depicted in red color), the y-axis (green), and the z-axis (purple). This sample from the group SO(3) will be analyzed in Section 5.1.

explains the data. We work in the classical framework of likelihood inference, i.e. we seek to compute the maximum likelihood estimate (MLE) for the given data $\{Y_1, Y_2, \ldots, Y_N\}$. By definition, the MLE is the 3×3 parameter matrix $\hat{\Theta}$ which maximizes the log-likelihood function. Thus, we must solve an optimization problem.

From our data we obtain the *sample mean* $\bar{Y} = \frac{1}{N}\sum_{k=1}^{N} Y_k$. Of course, the sample mean $\bar{Y}$ is generally not a rotation matrix anymore. We next compute the sign-preserving singular value decomposition of the sample mean, i.e., we determine $Q, R \in \mathrm{SO}(3)$ such that

$$\bar{Y} \;=\; Q \cdot \mathrm{diag}(g_1, g_2, g_2) \cdot R.$$

The signed singular values g_1, g_2, g_3 together with Q and R are sufficient statistics for the Fisher model. The sample $\{Y_1, \ldots, Y_N\}$ enters the log-likelihood function only via g_1, g_2, g_3.

Lemma 2.1 [29, Lemma 2]. The log-likelihood function for the given sample from $\mathrm{SO}(3)$ is

$$\ell : \mathbb{R}^3 \longrightarrow \mathbb{R}, \quad x \;\mapsto\; x_1 g_1 + x_2 g_2 + x_3 g_3 - \log(\tilde{c}(x_1, x_2, x_3)). \tag{4}$$

If $(\hat{x}_1, \hat{x}_2, \hat{x}_3)$ is the maximizer of the function ℓ, then the matrix $\hat{\Theta} = Q\,\mathrm{diag}(\hat{x}_1, \hat{x}_2, \hat{x}_3)R$ is the MLE of the Fisher model (1) of the sample $\{Y_1, \ldots, Y_N\}$ from the rotation group $\mathrm{SO}(3)$.

Lemma 2.1 says that we need to maximize the function (4) in order to compute the MLE in the Fisher model. We note that a local maximum is already a global one since (4) is a strictly concave function. The maximum is attained at a unique point in $\mathbb{R}^3$. We shall compute this point using tools from algebraic analysis that are discussed in the next section.

Remark 2.2. The singular values of the sample mean $\bar{Y}$ are bounded from above and below, namely $1 \geq |g_1| \geq g_2 \geq g_3 \geq 0$. If g_3 is close to 1, i.e., the average of the rotation matrices is almost a rotation matrix, then the data is typically concentrated about a preferred rotation. In this case the normalizing constant becomes very large and MLE on $\mathrm{SO}(3)$ is numerically intractable; see also Remark 4.3. However, due to the small spread of the data around a point in $\mathrm{SO}(3)$, a matrix valued Gaussian model on $\mathbb{R}^3$ is an accurate approximation.

3. Holonomic representation

We shall represent the normalizing constant $\tilde{c}$ by a system of linear differential equations it satisfies. This is known as the holonomic representation of this function. We work in the *Weyl algebra* D and in the *rational Weyl algebra* R with complex coefficients:

$$D \;=\; \mathbb{C}[x_1, x_2, x_3]\langle \partial_1, \partial_2, \partial_3 \rangle \qquad \text{and} \qquad R \;=\; \mathbb{C}(x_1, x_2, x_3)\langle \partial_1, \partial_2, \partial_3 \rangle.$$

We refer to [28; 30] for basics on these two noncommutative algebras of linear partial differential operators with polynomial and rational function coefficients, respectively. In order to stress the number of variables, we sometimes write D_3 instead of D and R_3 instead of R. By a *D-ideal* we mean a left ideal in D, and by an *R-ideal* a left ideal in R. The use of these algebras in statistical inference was pioneered by

Takemura, Takayama, and their collaborators [10; 16; 17; 29; 31]. We begin with an exposition of their results from [29].

The normalizing constant $\tilde{c}$ is closely related to the hypergeometric function ${}_0F_1$ of a matrix argument. In [29], annihilating differential operators of $\tilde{c}$ are derived from

$$H_i \;=\; \partial_i^2 - 1 + \sum_{j\neq i} \frac{1}{x_i^2 - x_j^2}(x_i\partial_i - x_j\partial_j) \qquad \text{for } i = 1,2,3. \tag{5}$$

These in turn can be obtained from Muirhead's differential operators in [21, Theorem 7.5.6] by a change of variables. In the notation of [28], we have $H_i \bullet \tilde{c} = 0$ for $i = 1,2,3$. Written in the more familiar form of linear PDEs, this says

$$\frac{\partial^2\tilde{c}}{\partial x_i^2} + \sum_{j\neq i} \frac{1}{x_i^2 - x_j^2}\left(x_i\frac{\partial\tilde{c}}{\partial x_i} - x_j\frac{\partial\tilde{c}}{\partial x_j}\right) \;=\; \tilde{c} \qquad \text{for } i = 1,2,3.$$

Note that the operators H_i are elements in the rational Weyl algebra R. Clearing the denominators, we obtain elements G_i in the Weyl algebra D that annihilate $\tilde{c}$, namely

$$G_i \;=\; \prod_{j\neq i}(x_i^2 - x_j^2)\cdot H_i. \tag{6}$$

By [29, Theorem 1], the following three additional differential operators in D annihilate $\tilde{c}$:

$$L_{ij} \;:=\; (x_i^2 - x_j^2)\partial_i\partial_j \;-\; (x_j\partial_i - x_i\partial_j) \;-\; (x_i^2 - x_j^2)\partial_k. \tag{7}$$

Here the indices are chosen to satisfy $1 \le i < j \le 3$ and $\{i,j,k\} = \{1,2,3\}$.

Let us consider the D-ideal that is generated by the six operators in (6) and (7):

$$I \;:=\; \langle G_1, G_2, G_3, L_{12}, L_{13}, L_{23}\rangle. \tag{8}$$

In the rational Weyl algebra, we have $RI = \langle H_1, H_2, H_3, L_{12}, L_{13}, L_{23}\rangle$ as R-ideals. We enter the D-ideal I into the computer algebra system Singular:Plural as follows:

```
ring r = 0,(x1,x2,x3,d1,d2,d3),dp;
def D = Weyl(r); setring D;
poly L12 = (x1^2-x2^2)*d1*d2 - (x2*d1-x1*d2)-(x1^2-x2^2)*d3;
poly L13 = (x1^2-x3^2)*d1*d3 - (x3*d1-x1*d3)-(x1^2-x3^2)*d2;
poly L23 = (x2^2-x3^2)*d2*d3 - (x3*d2-x2*d3)-(x2^2-x3^2)*d1;
poly G1 = (x1^2-x2^2)*(x1^2-x3^2)*d1^2 + (x1^2-x3^2)*(x1*d1-x2*d2)
            + (x1^2-x2^2)*(x1*d1-x3*d3) - (x1^2-x2^2)*(x1^2-x3^2);
poly G2 = (x2^2-x1^2)*(x2^2-x3^2)*d2^2 + (x2^2-x3^2)*(x2*d2-x1*d1)
            + (x2^2-x1^2)*(x2*d2-x3*d3) - (x2^2-x1^2)*(x2^2-x3^2);
poly G3 = (x3^2-x1^2)*(x3^2-x2^2)*d3^2 + (x3^2-x2^2)*(x3*d3-x1*d1)
            + (x3^2-x1^2)*(x3*d3-x2*d2) - (x3^2-x1^2)*(x3^2-x2^2);
ideal I = L12,L13,L23,G1,G2,G3;
```

We can now perform various symbolic computations in the Weyl algebra D. We used the libraries `dmodloc` [1] and `dmod` [18], due to Andres, Levandovskyy, and Martín-Morales. In particular, the following two lines confirm that I is holonomic and its holonomic rank is 4:

```
isHolonomic(I);
holonomicRank(I);
```

The rank statement means algebraically that $\dim_{\mathbb{C}(x_1,x_2,x_3)}(R/RI) = 4$. In terms of analysis, it means that the set of holomorphic solutions to I on a small open ball $\mathcal{U} \subset \mathbb{C}^3$ is a 4-dimensional vector space. Here $\mathcal{U}$ is chosen to be disjoint from the singular locus

$$\mathrm{Sing}(I) \;=\; \bigl\{ x \in \mathbb{C}^3 \,:\, (x_1^2 - x_2^2)(x_1^2 - x_3^2)(x_2^2 - x_3^2) \,=\, 0 \bigr\}. \tag{9}$$

We note that the normalizing constant $\tilde{c} = \tilde{c}(x_1, x_2, x_3)$ is a real analytic function on $\mathbb{R}^3 \backslash \mathrm{Sing}(I)$ that extends to a holomorphic function on all of complex affine space $\mathbb{C}^3$.

Using Gröbner bases in the rational Weyl algebra R, we find that the initial ideal of RI for the degree reverse lexicographic order is generated by the symbols of our six operators:

$$\mathrm{in}(RI) \;=\; \langle\, \partial_1\partial_2\,,\; \partial_1\partial_3\,,\; \partial_2\partial_3\,,\; \partial_1^2\,,\; \partial_2^2\,,\; \partial_3^2 \,\rangle.$$

The set of standard monomials equals $S = \{1, \partial_1, \partial_2, \partial_3\}$. This is a $\mathbb{C}(x_1, x_2, x_3)$-basis for the vector space R/RI. In this situation, we can associate a *Pfaffian system* to the D-ideal I. For the general theory, we refer the reader to [30] and specifically to [28, (23)].

The Pfaffian system is a system of first-order linear differential equations associated to the holonomic function $\tilde{c}$. It consists of three 4×4 matrices P_1, P_2, P_3 whose entries are rational functions in x_1, x_2, x_3. We introduce the column vector $C = (\,\tilde{c},\, \partial_1 \bullet \tilde{c},\, \partial_2 \bullet \tilde{c},\, \partial_3 \bullet \tilde{c}\,)^{\mathrm{t}}$.

Theorem 3.1 [29, Theorem 2]. The Pfaffian system associated to the normalizing constant $\tilde{c}$ of the Fisher distribution (1) consists of the following three vector equations:

$$\partial_i \bullet C \;=\; P_i \cdot C \qquad \text{for } i \,=\, 1, 2, 3, \tag{10}$$

where the matrices $P_1, P_2, P_3 \in \mathbb{C}(x_1, x_2, x_3)^{4\times 4}$ are

$$P_1 \;=\; \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & \frac{x_1(-2x_1^2+x_2^2+x_3^2)}{(x_1^2-x_3^2)(x_1^2-x_2^2)} & \frac{x_2}{x_1^2-x_2^2} & \frac{x_3}{x_1^2-x_3^2} \\ 0 & \frac{x_2}{x_1^2-x_2^2} & \frac{-x_1}{x_1^2-x_2^2} & 1 \\ 0 & \frac{x_3}{x_1^2-x_3^2} & 1 & \frac{-x_1}{x_1^2-x_3^2} \end{pmatrix}, \quad P_2 \;=\; \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & \frac{-x_2}{x_2^2-x_1^2} & \frac{x_1}{x_2^2-x_1^2} & 1 \\ 1 & \frac{x_1}{x_2^2-x_1^2} & \frac{x_2(x_1^2-2x_2^2+x_3^2)}{(x_2^2-x_1^2)(x_2^2-x_3^2)} & \frac{x_3}{x_2^2-x_3^2} \\ 0 & 1 & \frac{x_3}{x_2^2-x_3^2} & \frac{-x_2}{x_2^2-x_3^2} \end{pmatrix},$$

$$\text{and} \quad P_3 \;=\; \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & \frac{-x_3}{x_3^2-x_1^2} & 1 & \frac{x_1}{x_3^2-x_1^2} \\ 0 & 1 & \frac{-x_3}{x_3^2-x_2^2} & \frac{x_2}{x_3^2-x_2^2} \\ 1 & \frac{x_1}{x_3^2-x_1^2} & \frac{x_2}{x_3^2-x_2^2} & \frac{x_3(x_1^2+x_2^2-2x_3^2)}{(x_3^2-x_1^2)(x_3^2-x_2^2)} \end{pmatrix}.$$

We reproduced this Pfaffian system from the operators $G_1, G_2, G_3, L_{12}, L_{13}, L_{23}$ with the Mathematica package `HolonomicFunctions` [15]. This was done by running Gröbner basis computations in the rational Weyl algebra R with the degree reverse lexicographic order. See [28, Example 3.4] for an illustration on how this is done.

The Pfaffian system (10) allows us to recover the ith partial derivative of the normalizing constant as the first coordinate of the column vector $P_i \cdot C$. In symbols we have $\partial_i \bullet \tilde{c} = (P_i \cdot C)_1$. We make extensive use of this fact when computing the MLE in Section 4. In the same vein, we can recover the Hessian of $\tilde{c}$ from the Pfaffian system of $\tilde{c}$ as follows:

$$\begin{aligned} \partial_1^2 \bullet \tilde{c} &= (P_1 \cdot C)_2, \quad \partial_1\partial_2 \bullet \tilde{c} = (P_2 \cdot C)_2, \quad \partial_1\partial_3 \bullet \tilde{c} = (P_3 \cdot C)_2, \\ \partial_2^2 \bullet \tilde{c} &= (P_2 \cdot C)_3, \quad \partial_2\partial_3 \bullet \tilde{c} = (P_3 \cdot C)_3, \quad \partial_3^2 \bullet \tilde{c} = (P_3 \cdot C)_4. \end{aligned} \tag{11}$$

This allows for the use of second order optimization algorithms, see Section 4.

An object of interest—from the algebraic analysis perspective—is the Weyl closure of the D-ideal I. By definition, the *Weyl closure* is the following D-ideal which clearly contains I:

$$W(I) := RI \cap D.$$

In general, it is a challenging problem to compute the Weyl closure of a D-ideal. This computation is reminiscent of finding the radical of a polynomial ideal, which, according to Hilbert's Nullstellensatz, consists of all polynomials that vanish on the complex solutions to the given polynomials. The Weyl closure plays a similar role for holonomic functions. It turns out that computing $W(I)$ is fairly benign for the D-ideal I studied in this section.

Lemma 3.2. *Let I be the holonomic D-ideal in (8). Then the Weyl closure $W(I)$ is generated by I and the one additional operator $x_1\partial_1\partial_3 + x_2\partial_2\partial_3 + x_3\partial_3^2 - x_2\partial_1 - x_1\partial_2 - x_3 + 2\partial_3$.*

Proof. We used the Singular library `dmodloc` [1] to compute the Weyl closure of I. We found that I is not Weyl-closed, i.e., $I \subsetneq W(I)$. Moreover, by Gröbner basis reductions in the Weyl algebra, we find that adding the claimed operator results in a Weyl-closed ideal. □

Following [16; 29], we now consider the Fisher distribution on $\mathrm{SO}(n)$. The normalizing constant $c(\Theta)$ is defined as in (2), with the integral taken over $\mathrm{SO}(n)$ with its Haar measure. Let D_{n^2} be the Weyl algebra whose variables are the entries of the $n \times n$ matrix $\Theta = (t_{ij})$. The corresponding $n \times n$ matrix of differential operators in D_{n^2} is denoted by $\partial = (\partial_{ij})$. The following result was established by Koyama [16], based on earlier work of Sei et al. [29]. We shall prove a more general statement for arbitrary compact Lie groups in Section 6.

Theorem 3.3. *The annihilator of $c(\Theta)$ is the D-ideal J generated by the following operators*:

$$\begin{aligned} d &= 1 - \det(\partial), \quad g_{ij} = \delta_{ij} - \textstyle\sum_{k=1}^n \partial_{ik}\partial_{jk} && \text{for } 1 \le i \le j \le n, \\ P_{ij} &= \textstyle\sum_{k=1}^n \big(t_{ik}\partial_{jk} - t_{jk}\partial_{ik}\big) && \text{for } 1 \le i < j \le n. \end{aligned}$$

Above we omitted half of the equations given in [16, (12)], which is justified by the results in [25, Section 8.7.3]. Also, the operators P_{ij} are induced from left matrix multiplication (as in (26)) rather than right multiplication as in [16, (11)].

A problem that was left open in [16; 29], even for $n = 3$, is the determination of the holonomic rank of J. We now address this by introducing dimensionality reduction via invariant theory. The same approach makes sense for the general definition of J seen in (30).

Let J' be the D-ideal generated by the operators P_{ij}, g_{ij}. This is the analogue of J for the orthogonal group $\mathrm{O}(n)$ in its standard representation in $\mathrm{GL}_n(\mathbb{C})$ (see Section 6). Since $\mathrm{O}(n)$ has two connected components, the corresponding module in Theorem 6.2 is a direct sum of two simple holonomic D_{n^2}-modules. By symmetry, we obtain

$$\operatorname{rank}(J') \;=\; 2 \cdot \operatorname{rank}(J). \tag{12}$$

The ring of $\mathrm{O}(n)$-invariant polynomials on $\mathbb{C}^{n\times n}$ is generated by the $\binom{n+1}{2}$ entries $\{y_{kl}\}_{1\le k\le l\le n}$ of the symmetric matrix $Y = \Theta^{\mathrm{t}} \cdot \Theta$ (see [25, Section 11.2.1]). These matrix entries y_{kl} are algebraically independent quadratic forms in the n^2 unknowns t_{ij}.

We now work in the Weyl algebra $D_{\binom{n+1}{2}}$ with the convention $y_{kl} = y_{lk}$ and $\partial_{kl} = \partial_{lk}$. Let K denote the left ideal in that Weyl algebra which is generated by the operators

$$h_{ij} \;=\; 2^{\delta_{ij}}\, n \cdot \partial_{ij} \;-\; \delta_{ij} \;+\; \sum_{k,\,l=1}^{n} 2^{\delta_{ki}+\delta_{lj}}\, y_{kl} \cdot \partial_{ik}\partial_{jl} \qquad \text{for } 1 \le i \le j \le n. \tag{13}$$

Theorem 3.4. *A holomorphic function is a solution to J' if and only if it is of the form $\Theta \mapsto \phi(y_{ij}(\Theta))$, where ϕ is a solution to K. In particular,* $\operatorname{rank}(K) = 2 \cdot \operatorname{rank}(J)$.

Proof. The Lie algebra operators P_{ij} express left invariance under $\mathrm{SO}(n)$. The fact that every solution to J' is expressible in Y follows from Luna's Theorem [19] (see also [11, Section 6.4]). We note that the determinant $\det(\Theta)$ is an $\mathrm{SO}(n)$-invariant that we may omit, due to the relation $\det(\Theta)^2 = \det Y$. The D-ideal K is the invariant version of J'. The operator h_{ij} is derived from g_{ij} by the chain rule. The result therefore follows from (12). □

As an application of Theorem 3.4, we answer a question left open in [29, Proposition 2].

Proposition 3.5. *For $n = 3$, we have* $\operatorname{rank}(J) = 4$.

Proof. We used the computer algebra system Macaulay2 [9]. Unlike for $\operatorname{rank}(J)$, the calculations for $\operatorname{rank}(K)$ finished, and we found $\operatorname{rank}(K) = 8$. We conclude by Theorem 3.4. □

For arbitrary $n \ge 2$, we define I to be the D-ideal generated by the n operators G_i analogous to (6) and the $\binom{n}{2}$ operators L_{ij} analogous to (7). We saw this D-ideal in (8) for $n = 3$. We now explain how the D_{n^2}-ideal J and the D_n-ideal I are connected. We use the construction of the *restriction ideal*. For the general definition see [28, (13)]. In our case, the construction works as follows. We set $x_i = t_{ii}$ for $i = 1, \ldots, n$ and we write D_n for the corresponding Weyl algebra. Then

$$J_{\mathrm{diag}} \;:=\; (J \,+\, \{t_{ij} \,:\, 1 \le i \ne j \le n\} \cdot D_{n^2}) \;\cap\; D_n \tag{14}$$

is the D_n-ideal obtained by restricting the annihilator of $c(\Theta)$ to the diagonal entries of the matrix Θ. Note that the second summand in (14) is a *right* ideal in the Weyl algebra D_{n^2}.

If $f(\Theta)$ is a function in the n^2 variables t_{ij} that is annihilated by J, then the restriction ideal J_{diag} annihilates the function $f(\text{diag}(x_1, \ldots, x_n))$ in n variables. Therefore, J_{diag} annihilates the restricted normalizing constant $\tilde{c}(x_1, \ldots, x_n)$. The result to be presented next guarantees that the Pfaffian system in Theorem 3.1 is indeed of minimal size.

Proposition 3.6. *The following inclusions hold among holonomic D_n-ideals representing $\tilde{c}$:*

$$I \subseteq J_{\text{diag}} \subseteq W(J_{\text{diag}}) \subseteq \text{ann}_{D_n}(\tilde{c}).$$

Equality holds for $n \leq 3$ in the rightmost inclusion.

Sketch of proof. The proof of [29, Theorem 1] shows that I is contained in J_{diag}. Note that for $n = 3$ the middle inclusion is strict by Lemma 3.2. We have $W(J_{\text{diag}}) \subseteq \text{ann}_{D_n}(\tilde{c})$ because the annihilator of a smooth function such as $\tilde{c}$ is Weyl-closed, by an argument spelled out in [8].

The equality on the right for $n = 3$ is shown by proving $W(I) = \text{ann}_{D_3}(\tilde{c})$. We use the following argument and computations. The Fourier transform $W(I)^{\mathcal{F}}$ is the D-ideal obtained by switching ∂_i and x_i (up to sign). We find that its holonomic rank is 1. We next compute the *holonomic dual* of the module $D_3/W(I)^{\mathcal{F}}$. This is another D_3-module, as defined in [12, Section 2.6]. There is a built-in command for the holonomic dual in Macaulay2 [9]. Another computation, using localization techniques, verifies that both $D_3/W(I)^{\mathcal{F}}$ and its holonomic dual are torsion-free as $\mathbb{C}[x_1, x_2, x_3]$-modules. These facts imply that $D_3/W(I)^{\mathcal{F}}$ is a simple D-module, and hence so is $D_3/W(I)$. From this we conclude that $W(I) = \text{ann}_{D_3}(\tilde{c})$. □

We conjecture that the inclusion on the right is an equality for all positive integers n. Using results from Section 6, we can argue that $W(J_{\text{diag}})^{\mathcal{F}}$ is regular holonomic for any n. It appears that its singular locus is a hyperplane arrangement. The special combinatorial structure encountered in this arrangement gives strong evidence for the conjecture above.

4. Maximum likelihood estimation

We now proceed to computing the maximum of the log-likelihood function of Lemma 2.1 for given datasets. Since the objective function (4) is strictly concave, a local maximum is the global maximizer and attained at a unique point $\hat{x} = (\hat{x}_1, \hat{x}_2, \hat{x}_3) \in \mathbb{R}^3$. In order to compute $\hat{x}$, we run a number of algorithms, each using the holonomic gradient method. This is based on the results presented in the previous section, especially on Theorem 3.1 and equation (11). These are used to compute the function values, gradients, and Hessians in each iteration. The code for the numerical computations of this section can be found at https://github.com/MikeAdamer/hgm_MLE.

A critical step in running any local optimization method is finding a suitable starting point. As mentioned in Section 3, solutions to the D-ideal I are analytic outside the singular locus $\text{Sing}(I)$. Starting points need to be chosen in $\mathbb{R}^3 \backslash \text{Sing}(I)$. For the Fisher model on SO(3), the singular locus $\text{Sing}(I)$ is the arrangement (9) of six planes through the origin in $\mathbb{R}^3$. This partitions $\mathbb{R}^3$ into 24 distinct chambers. For the algorithms described below, we choose starting points in each of the 24 connected components of $\mathbb{R}^3/\text{Sing}(I)$, and we evaluate the vector C at these points. This initialization can be done

either via the series expansion method of [29, Section 3.2] or, equivalently, using the package `hgm` [31] in the statistical software R, which uses a series expansion in conjunction with HGM.

In this section, we present three optimization methods based on algebraic analysis, building on the methods given in [22]. The holonomic part of the algorithms stems from the basic structure of most optimization schemes. In essence, there are always two steps: a gradient evaluation step and a gradient descent step.

In this paper, we show how HGD is used in the first step for exact evaluation of the gradients. For further details on the second step in a number of optimization schemes we refer to [23]. The simplest algorithm is *holonomic gradient ascent* (HGA). This is a straightforward adaptation of the HGD method in [29]. Second, we introduce a holonomic version of the Broyden–Fletcher–Goldfarb–Shanno (BFGS) method [23, Chapter 6, §1]. BFGS is a quasi-Newton method that requires the gradient and the function value as inputs. Both can be calculated directly using (10). This turns BFGS into *holonomic BFGS* (H-BFGS). Our third algorithm is a *holonomic Newton method*. This second-order method exploits the fact that the Hessian is easy to calculate from (11) and that the objective function is strictly concave.

To get started, we need an expression for the gradient of the log-likelihood function ℓ and a holonomic gradient method (HGM) for evaluating that expression. By Lemma 2.1,

$$\nabla \ell(x) \;=\; \begin{pmatrix} g_1 \\ g_2 \\ g_3 \end{pmatrix} - \frac{1}{\tilde{c}(x)} \cdot \nabla \tilde{c}(x). \tag{15}$$

Note that $C(x) = (\tilde{c}(x), \nabla \tilde{c}(x))^{\mathrm{t}}$. Hence, our task to evaluate $\nabla \ell$ at any point amounts to evaluating the vector-valued function C at any point. This is where the HGM comes in.

In general, we approximate the function C at a point $x^{(n+1)}$ given its value at a previous point $x^{(n)}$. To this end, a path $x^{(n)} \to x^{(n)} + \delta^{(1)} \to x^{(n)} + \delta^{(2)} \to \cdots \to x^{(n)} + \delta^{(K)} \to x^{(n+1)}$ is chosen, where $\delta^{(1)}, \ldots, \delta^{(K)} \in \mathbb{R}^3$ with $\|\delta^{(m+1)} - \delta^{(m)}\|$ sufficiently small. The linear part of the Taylor series expansion of C at $x^{(n)}$ yields the following approximations:

$$\begin{aligned} C(x^{(n)} + \delta^{(m+1)}) \;&\approx\; C(x^{(n)} + \delta^{(m)}) + \sum_{i=1}^{3} (\delta_i^{(m+1)} - \delta_i^{(m)})\, (\partial_i \bullet C)(x^{(n)} + \delta^{(m)}) \\ &=\; C(x^{(n)} + \delta^{(m)}) + \sum_{i=1}^{3} (\delta_i^{(m+1)} - \delta_i^{(m)})\, P_i \cdot C(x^{(n)} + \delta^{(m)}). \end{aligned} \tag{16}$$

We choose a path consisting of points, separated by intervals of size Δt, on the line segment $x(t) = x^{(n)}(1-t) + x^{(n+1)}t$ with $t \in [0, 1]$. With this notation, (16) becomes

$$C(x((m+1)\Delta t)) \;\approx\; C(x(m\Delta t)) + \textstyle\sum_{i=1}^{3} (x_i^{(n+1)} - x_i^{(n)})\, \Delta t \cdot\; P_i \cdot C(x(m\Delta t)). \tag{17}$$

If we take the limit $\Delta t \to 0$, then the equation above becomes the differential equation

$$\frac{dC(t)}{dt} \;=\; \sum_{i=1}^{3} \frac{\partial x_i}{\partial t} \frac{\partial C}{\partial x_i} \;=\; \sum_{i=1}^{3} \big(x_i^{(n+1)} - x_i^{(n)}\big) P_i \cdot C.$$

This ordinary differential equation can be solved using any numerical ODE solver, e.g., an Euler scheme or Runge–Kutta scheme. This leads to the following algorithm.

Algorithm 1: Holonomic gradient method.

Input: $x^{(n)}$, $x^{(n+1)}$, $C(x^{(n)})$, a Pfaffian system P_1, P_2, P_3
Output: $C(x^{(n+1)})$
1 Set $x(t) = x^{(n)}(1-t) + x^{(n+1)}t$.
2 Let $\frac{dC(t)}{dt} = \sum_{i=1}^{3} \frac{\partial x_i}{\partial t}\frac{\partial C}{\partial x_i} = \sum_{i=1}^{3}(x_i^{(n+1)} - x_i^{(n)})P_i \cdot C$.
3 Numerically integrate line 2 from $t = 0$ to $t = 1$.

We employ Algorithm 1 as a subroutine for the holonomic gradient ascent algorithm, which will be described next. HGA is analogous to other gradient ascent/descent methods, however, with the special feature that the gradients are calculated via the HGM algorithm. A description of the algorithm, adapted for data from SO(3), is outlined below.

Algorithm 2: Holonomic gradient ascent.

Input: Matrices Q and R, singular values g_1, g_2, g_3 and a starting point $x^{(0)} \in \mathbb{R}^3$
Result: A maximum likelihood estimate for the data in the Fisher model (1)
1 Choose a learning rate γ_n.
2 Choose a threshold δ.
3 Evaluate C at the starting point $x^{(0)}$.
4 Evaluate $\nabla\ell$ at the starting point $x^{(0)}$.
5 Set $n = 0$.
6 **while** $\max|\nabla\ell(x^{(n)})| < \delta$ **do**
7 $\quad x^{(n+1)} = x^{(n)} + \gamma_n \nabla\ell(x^{(n)})$.
8 $\quad$ Calculate $C(x^{(n+1)})$ via HGM using Algorithm 1.
9 $\quad$ Calculate $\nabla\ell(x^{(n+1)})$ from $C(x^{(n+1)})$.
10 $\quad$ Set $n = n+1$.
11 **end**
12 Output the vector $x^{(n)} \in \mathbb{R}^3$ as our approximation for $(\hat{x}_1, \hat{x}_2, \hat{x}_3)$.
13 Output the rotation matrix $\hat{\Theta} = Q \cdot x^{(n)} \cdot R$ as our approximation for the MLE.

The given data is a list of rotation matrices $Y_1, \ldots, Y_N$ in SO(3). As explained in Section 2, we encode these in the singular values g_1, g_2, g_3 of the sample mean $\bar{Y} = \frac{1}{N}\sum_{k=1}^{N} Y_k$. Thus, the input for HGA consists primarily of just three numbers g_1, g_2, g_3. They are used in the evaluation in the first terms of $\nabla\ell$, as seen in (15). The second term is evaluated by matrix multiplication with P_1, P_2, P_3, as seen in (10). Part of the input are also the matrices Q and R that diagonalize the sample mean $\bar{Y}$. They are needed in the last step to recover $\hat{\Theta}$ from $\hat{x}_1, \hat{x}_2, \hat{x}_3$ as in Lemma 2.1. The HGA algorithm has two

parameters, namely the threshold δ which indicates a termination condition, and the learning rate γ_n. While δ can be chosen freely depending on the desired accuracy, choosing the learning rate can have significant effects on the convergence of the algorithm. In our computations we chose $\gamma_n = 10^{-2}$. This can clearly be improved. However, the standard technique of performing line searches to find a good γ_n is not recommended as evaluating C at a new point is costly.

To employ more advanced methods such as BFGS, and to avoid integrating along a path crossing the singular locus, we use [29, Corollary 1]. This states that the value of the column vector C at a point (x_1, x_2, x_3) can be obtained by fixing (x_1, x_2, x_3) and integrating the following ODE from $t = \epsilon \ll 1$ to $t = 1$. Here C is regarded as a function of the parameter t:

$$C'(t) \;=\; \begin{pmatrix} 0 & x_1 & x_2 & x_3 \\ x_1 & -2/t & x_3 & x_2 \\ x_2 & x_3 & -2/t & x_1 \\ x_3 & x_2 & x_1 & -2/t \end{pmatrix} \cdot C(t). \tag{18}$$

Using this approach for calculating C, we can employ BFGS optimization using HGM as a subroutine to calculate the gradients and function values required as inputs. This also prevents the accumulation of numerical errors as the initial conditions of the ODE are exact. The H-BFGS method achieves faster convergence rates than the simple HGA Algorithm 2.

A final very powerful algorithm for concave (or convex) functions is the Newton method which uses the Hessian matrix. Often, finding the Hessian matrix $\mathbf{H}[\ell(x)]$ of a function is a difficult task. However, using holonomic methods the Hessian is obtained for free via

$$\partial_i \partial_j \bullet \ell \;=\; \frac{1}{\tilde{c}^2} \left(\partial_i \bullet \tilde{c}\right) \left(\partial_j \bullet \tilde{c}\right) - \frac{1}{\tilde{c}} \partial_i \partial_j \bullet \tilde{c},$$

and the relations in (10) and (11). We found that the Newton method,

$$x^{(n+1)} \;=\; x^{(n)} - \mathbf{H}[\ell(x)]^{-1} \cdot \nabla \ell(x),$$

gives the fastest convergence. We refer to this approach as the *Holonomic Newton method*.

We implemented the H-BFGS method in a script in the software R. Interested readers may obtain our implementation at https://github.com/MikeAdamer/hgm_MLE. This code is custom-tailored for rotations in 3-space. The function C is evaluated at the starting point $x^{(0)}$ using the series expansion method that is described in [29, Section 3.2]. Here we truncate the series at order 41.

Example 4.1. We created a synthetic dataset consisting of $N = 500$ rotation matrices. These were sampled from the Fisher distribution with parameter matrix

$$\Theta \;=\; \begin{pmatrix} -1.178 & 0.2804 & 1.037 \\ -0.3825 & 0.9181 & 0.6016 \\ -0.0955 & 0.9037 & 1.695 \end{pmatrix}. \tag{19}$$

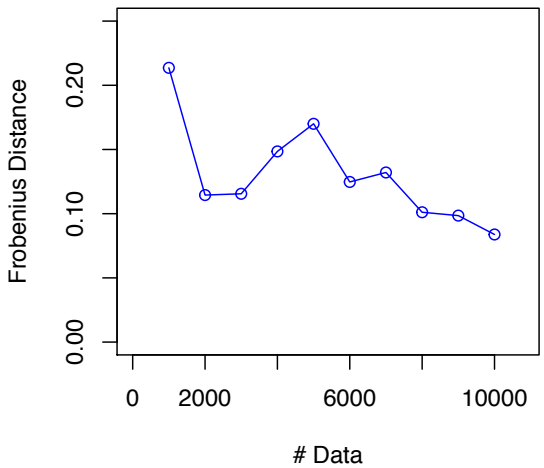

Figure 2. The Frobenius distance of the MLE parameters to the true parameter values. The convergence is relatively slow. This suggests that the MLE problem is not well conditioned.

The sample mean and its sign-preserving singular value decomposition are found to be

$$\bar{Y} = \begin{pmatrix} -0.2262 & 0.1021 & 0.2260 \\ -0.0233 & 0.0611 & 0.2779 \\ -0.0364 & 0.2802 & 0.3529 \end{pmatrix} = Q \cdot \begin{pmatrix} 0.5946 & 0.0000 & 0.0000 \\ 0.0000 & 0.1838 & 0.0000 \\ 0.0000 & 0.0000 & 0.1059 \end{pmatrix} \cdot R,$$

$$\text{with} \quad Q = \begin{pmatrix} -0.4977 & 0.8589 & 0.1211 \\ -0.4518 & -0.1376 & -0.8815 \\ -0.7404 & -0.4934 & 0.4565 \end{pmatrix}, \quad R = \begin{pmatrix} 0.2524 & -0.4808 & -0.8397 \\ -0.9419 & -0.3209 & -0.0993 \\ -0.2217 & 0.8160 & -0.5339 \end{pmatrix}.$$

Running H-BFGS on this input, the MLE is found to be

$$\hat{\Theta} = \begin{pmatrix} -0.8972 & 0.3446 & 0.9682 \\ -0.2392 & 0.7777 & 0.7856 \\ -0.0763 & 0.8664 & 1.616 \end{pmatrix} = Q \cdot \begin{pmatrix} 2.422 & 0.0000 & 0.0000 \\ 0.0000 & 0.7432 & 0.0000 \\ 0.0000 & 0.0000 & -0.3043 \end{pmatrix} \cdot R. \tag{20}$$

While the entries of the MLE $\hat{\Theta}$ have the correct sign and order of magnitude, the actual values are not very close to those in Θ. In order to isolate the effect of the sample size on the MLE, we extended the data to 10000 matrices. In the iterations we recorded the Frobenius distance (FD) from $\hat{\Theta}$ to Θ. Our findings are outlined in Figure 2.

Remark 4.2. The authors in [29] report that the HGD algorithm becomes numerically unstable when it is close to the singular locus of the Pfaffian system. They recommend picking a starting point in the same connected component of $\mathbb{R}^3 \setminus \mathrm{Sing}(I)$ where the MLE is suspected. In contrast, our computations suggest that the output of the HGA does not depend on the connected component which the starting point lies in, when a sufficiently stable numerical integration method (e.g. `lsode` from the R package `deSolve`) is

chosen in Algorithm 1. Therefore, the starting point of the optimization can be chosen arbitrarily, as long as it is close enough to the origin so that the series expansion converges.

Remark 4.3. The sample mean matrix $\bar{Y}$ lies in the convex hull of the rotation group. This convex body, denoted conv(SO(3)), was studied in [27, Section 4.4], and an explicit representation as a spectrahedron was given in [27, Proposition 4.1]. It follows from the theory of orbitopes [27] that the singular values of matrices in conv(SO(3)) are precisely the triples that satisfy $1 \geq |g_1| \geq g_2 \geq g_3 \geq 0$. These inequalities define two polytopes, which are responsible for the facial description of conv(SO(3)) found in [27, Theorem 4.11].

We can think of the MLE as a map from the interior of the orbitope conv(SO(3)) to $\mathbb{R}^3$. Using the singular value decomposition, we restricted this map to the open polytopes given by $1 > |g_1| > g_2 > g_3 > 0$. Note that the coordinates of the vector $\hat{x}$ goes off to infinity as the maximum of $\{g_1, g_2, g_3\}$ approaches 1. This follows from [14, (4.12)], where the analogue for $\mathrm{O}(n)$ was derived. This divergence can cause numerical problems.

In this section, we have turned the earlier results on D-ideals into numerical algorithms. This is just a first step. The success of any local method relies heavily on a clear understanding of the numerical analysis that is relevant for the problem at hand. A future study of condition numbers from the perspective of holonomic representations would be desirable.

5. Rotation data in the sciences

Rotation data arise in any field of science in which the orientation of an object in 3-space is important. Occurrences include a diverse number of research areas such as medical imaging, biomechanics, astronomy, geology, and materials science. In this section, we apply our methods to a prominent dataset of vectorcardiograms and to biomechanical data. We also review previous findings on rotation data in astronomy, geology, and materials science.

5.1. ***Medical imaging.*** One important occurrence of rotational data in the applied sciences stems from medical imaging, and more precisely from vectorcardiography. In that field, the electrical forces generated by the heart are studied and their magnitude and direction are recorded.

The dataset presented in [7] is a famous example of directional data. It contains the orientation of the vectorcardiogram (VC) loop of 98 children aged $2-19$. In particular, the orientation is measured using two different techniques. Both measurements are given in the form of two vectors. The first identifies the VC loop of greatest magnitude and the second is the normal direction to the loop. We add as a third vector the cross product of the magnitude and normal vector to form a right handed set and, therefore, a rotation matrix.

This dataset has been used to exemplify a range of methods in directional statistics, see, e.g., [24]. We applied the optimization methods from Section 4 to the same dataset. In other words, we computed the maximum of the log-likelihood function (4) for the orientations of the VC loop. In order to match our analysis with the results of [24], we only consider the 28 data points of the boys aged $2-10$. A colorful illustration of the action of these 28 rotation matrices on the coordinate axes is shown in Figure 1.

We now proceed to the MLE. The sample mean has the singular value decomposition

$$\bar{Y} = \begin{pmatrix} 0.6868 & 0.5756 & 0.1828 \\ 0.5511 & -0.7372 & -0.0045 \\ 0.1216 & 0.1417 & -0.8630 \end{pmatrix} = Q \cdot \begin{pmatrix} 0.9469 & 0.0000 & 0.0000 \\ 0.0000 & 0.8962 & 0.0000 \\ 0.0000 & 0.0000 & 0.8737 \end{pmatrix} \cdot R, \tag{21}$$

where

$$Q = \begin{pmatrix} 0.6112 & 0.7636 & 0.2079 \\ -0.7498 & 0.4748 & 0.4608 \\ 0.2532 & -0.4376 & 0.8628 \end{pmatrix}, \quad R = \begin{pmatrix} 0.03941 & 0.99324 & -0.1092 \\ 0.81778 & 0.03072 & 0.5747 \\ 0.57418 & -0.11194 & -0.8110 \end{pmatrix}. \tag{22}$$

By forming the matrix product QR we recover the result of [24]. The matrix QR, however, is only one part of the MLE as described in [14]. By using H-BFGS, we can find the full MLE of the Fisher model. We compared H-BFGS to other methods. For that, we estimated x_1, x_2, x_3 with a BFGS optimization of the log-likelihood using the series expansion of the normalizing constant. We then compare the resulting estimate to the output of H-BFGS. The H-BFGS algorithm finds the MLE

$$\begin{aligned} \hat{x}_1 &= 20.072407, \\ \hat{x}_2 &= 12.513841, \\ \hat{x}_3 &= -6.510704, \end{aligned}$$

which corresponds to a log-likelihood of $\hat{\ell} = 3.97299$. The runtime of the algorithm is highly dependent on the number of nonzero terms in the series expansion for $\tilde{c}$. In this calculation, the first 6000 nonzero terms are used and the runtime is about 4 seconds. To improve the runtime one could try to truncate the series at lower order. For further refinement of the MLE one can combine H-BFGS with H-Newton by using the output of H-BFGS as the starting point of H-Newton. The classical BFGS method is not convergent if only the first 6000 nonzero terms are used. Hence, we need to truncate the series expansion at higher order. If we use the first 48000 nonzero terms, then the series expansion BFGS method finds the MLE $\hat{x}_1 = 17.604156$, $\hat{x}_2 = 10.024591$, $\hat{x}_3 = -3.881811$, which gives $\hat{\ell} = 3.96330$. The computation takes about 20 seconds. Hence, the holonomic BFGS outperformed the classical method by finding a better likelihood value in much shorter time.

5.2. ***Biomechanics.*** Rotational data is ubiquitous in the biomedical sciences. A prominent experiment in this area is the human kinematics study of [26]. In this experiment, the rotations of four different upper body parts were tracked while the subject was drilling holes into six different locations of a vertical panel. In [3], this dataset was studied and maximum likelihood and Bayesian point estimates for the orientation of the wrist were obtained and credible regions constructed.

A further experiment concerns the heel orientation of primates. In the experiments, the rotation of the calcaneus bone (the heel) and the cuboid bone, which is horizontally adjacent to the heel and closer to the toes, was measured. A load was applied to three sedentary primates, a human, a chimpanzee, and a baboon and the rotation of their ankle was recorded. While the data is actually a time series, the

simplifying assumption of independent identically distributed data is made in its analysis [4]. We study this dataset which was kindly provided by Melissa Bingham. The sample mean for the human data equals

$$\bar{Y} = \begin{pmatrix} -0.1013 & -0.9127 & -0.3811 \\ 0.3275 & -0.3895 & 0.8535 \\ -0.9335 & -0.0358 & 0.3475 \end{pmatrix} = Q \cdot \begin{pmatrix} 0.9997 & 0.0000 & 0.0000 \\ 0.0000 & 0.9926 & 0.0000 \\ 0.0000 & 0.0000 & 0.9923 \end{pmatrix} \cdot R, \tag{23}$$

with

$$Q = \begin{pmatrix} 0.4771 & 0.8753 & -0.0791 \\ -0.4320 & 0.1552 & -0.8884 \\ -0.7654 & 0.4580 & 0.4521 \end{pmatrix} \text{ and } R = \begin{pmatrix} 0.5248 & -0.2399 & -0.8167 \\ -0.4690 & -0.8822 & -0.0422 \\ -0.7104 & 0.4051 & -0.5754 \end{pmatrix}.$$

We see on the right hand side in (23) that the singular values for this dataset only differ in the third significant figure and the smallest singular value is approximately 1. We found that the normalizing constant gets too large to be computed directly. Indeed, our simulations returned a value error when $\tilde{c} \approx 10^{308}$. This is a serious numerical issue, arising in any MLE algorithm that attempts to directly calculate $\tilde{c}$ when the sample mean is almost a rotation matrix. Singular values close to 1 imply that the samples are concentrated on the unit sphere. One could either use a rotational Maxwell distribution [13] as a local model or the approximation in [4]. The data for the baboon and the chimpanzee show similar traits.

Progress can be made by applying a gauge transform in (18), aimed at scaling the input for H-BFGS. Let λ_0 be the largest eigenvalue of

$$A = \begin{pmatrix} 0 & x_1 & x_2 & x_3 \\ x_1 & 0 & x_3 & x_2 \\ x_2 & x_3 & 0 & x_1 \\ x_3 & x_2 & x_1 & 0 \end{pmatrix}.$$

We can derive an ODE for the function $E = C \cdot \exp(-\lambda_0 t)$ from (18). The function E is guaranteed to have smaller values than C. Furthermore, the ratio $(\partial_i \bullet \tilde{c})/\tilde{c} = C_i/C_0 = D_i/D_0$ is invariant. Despite being able to compute $\log(\tilde{c})$ using the gauge transformation, MLE becomes very unstable due to the numerical accuracy required. Finding the MLE from a random starting point using H-BFGS proved intractable. However, to provide a suitable starting point for H-BFGS, we use the asymptotic formula of [14], which gives $\hat{x}_1 = 5543.102$, $\hat{x}_2 = 3753.025$, $\hat{x}_3 = -3685.298$. We refined this result with H-BFGS and found the MLE $\hat{x}_1 = 5543.106$, $\hat{x}_2 = 3753.078$, $\hat{x}_3 = -3685.242$ corresponding to a log-likelihood of $\hat{\ell} = 10.59342$. Calculating the log-likelihood using HGM and the asymptotic values yields $\hat{\ell} = 10.52366$. Hence, H-BFGS finds a slightly better MLE than the asymptotic formula.

5.3. ***Astronomy, geology, and materials science.*** Astronomical applications of the matrix Fisher model on SO(3) are often concerned with the orbits of near-earth objects [20; 29]. Such objects are comets or asteroids in an elliptic orbit around the sun with the sun in their focus. The data comes as sets of vectors in $\mathbb{R}^3$ taking the sun as the origin. The first vector, X_1, is the perihelion direction, which points to the location on the orbit closest to the sun. The second vector, X_2, is the unit normal to the orbit. Together

with their cross product these vectors form a right handed set. Therefore, they define a rotation matrix. Questions of astronomical interest are whether the perihelion direction is uniformly distributed on the sphere and whether the orbit orientations are uniform on $\mathrm{SO}(3)$. To answer the latter question the Raleigh statistic can be used [20; 29].

Sei et al. [29] studied a dataset of rotations representing 151 comets and 6496 asteroids. They computed maximum likelihood estimates using the holonomic gradient method and also series expansions. The Raleigh statistic for the dataset was calculated and the null hypothesis of a uniform distribution was strongly rejected. Further, the hypothesis of the data originating from a Fisher distribution on a Stiefel manifold was tested against the hypothesis of $\mathrm{SO}(3)$, and the evidence strongly suggested to reject the Stiefel manifold.

Rotations arise in geology and earth sciences in the study of earthquake epicenters [13] and the analysis of plate tectonics [6]. Davis and Titus [6] studied a dataset of the deformation of a shear zone in northern Idaho. However, this was done in the context of invalidating a geology inspired model that had been used previously to explain the shear deformations.

Kagan [13] studied rotational data describing the earthquake focal mechanism orientation. Various models, including the Fisher model, were discussed in this article. However, the Fisher model was dismissed due to the difficulty of normalization for small spread data as discussed in Remark 2.2. The alternative model used in [13] was a rotational Maxwell distribution as a local approximation. Our results offer a chance to revisit the Fisher model.

One important source of rotational data is materials science, where patterns from electron backscatter diffraction (EBSD) are analyzed (see, e.g., [2]). This type of data provides information about the orientation of grains within a material. Crystal orientation has important implications on the properties of polycrystalline materials. One issue with EBSD data is the fact that orientations of the crystals can only be determined within a coset of the crystallographic group the grain belongs to. This is due to the fact that a crystal is a lattice and every lattice comes with certain translational and rotational symmetries. Orientations can only be determined up to the rotational invariance of the lattice. Hence, the data, although giving information about rotations, is strictly speaking not on $\mathrm{SO}(3)$, but on its quotient by a discrete symmetry subgroup. To adapt our analysis, an appropriate parametrization or embedding for such a quotient needs to be found. This, however, is beyond the scope of this paper and is left for future work. Before going to such manifolds, we start with Lie groups.

6. Compact Lie groups

The Fisher model on $\mathrm{SO}(n)$ generalizes naturally to other compact Lie groups. We define the Fisher distribution and the normalizing constant as in (1) and (2), but with integration over the Haar measure on the Lie group. In this section, we introduce these objects and their holonomic representation. In particular, we establish the analogue of Theorem 3.3 for compact Lie groups. This opens up the possibility of applying algebraic analysis to data sampled from manifolds other than $\mathrm{SO}(n)$ provided these have the structure of a group.

Let G be a compact connected Lie group and fix a real representation $\pi : G \to \mathrm{GL}_n(\mathbb{R})$. We can assume that π is injective, i.e., the representation is faithful. We note that any compact Lie group admits a faithful representation [25, Section 8.3.4]. The matrix group $\pi(G) \subset \mathbb{R}^{n\times n}$ is a closed algebraic subvariety (see [25, Section 8.7]). If one starts with a complex representation instead, the situation can be studied in the polynomial ring over $\mathbb{C}$.

For our algebraic approach, the ambient setting is the complex affine space $X := \mathbb{C}^{n\times n}$. The complexification $G_{\mathbb{C}}$ of our group G is a complex connected reductive algebraic group [25, Section 8.7.2]. The extension $\pi : G_{\mathbb{C}} \to X$ is a closed embedding. Its image, the matrix group $\pi(G_{\mathbb{C}})$, is the complex affine variety in X, cut out by the same polynomials as the ones defining $\pi(G)$. We denote by I_G the ideal generated by these polynomials in $\mathbb{C}[X]$. The quotient ring $\mathbb{C}[G] := \mathbb{C}[X]/I_G$ is the ring of polynomial functions on the group $\pi(G_{\mathbb{C}})$.

Let $\mathfrak{g}$ denote the complex Lie algebra of $G_{\mathbb{C}}$. This is the complexification of the real Lie algebra of the given Lie group G. We write $U(\mathfrak{g})$ for the universal enveloping algebra of $\mathfrak{g}$. For any affine variety, one can define the ring of algebraic differential operators on that variety. This is generally a complicated object, but things are quite nice in our case.

Let D_G denote the ring of differential operators on $G_{\mathbb{C}}$. We have natural inclusions

$$\mathfrak{g} \subset U(\mathfrak{g}) \subset D_G \qquad \text{and} \qquad \mathbb{C}[G] \subset D_G.$$

These inclusions exhibit desirable properties. Namely, we have canonical isomorphisms

$$D_G \;\cong\; \mathbb{C}[G] \otimes U(\mathfrak{g}) \;\cong\; U(\mathfrak{g}) \otimes \mathbb{C}[G]. \tag{24}$$

This holds because left (or right) invariant vector fields of $G_{\mathbb{C}}$ trivialize the tangent bundle. Recall that $G_{\mathbb{C}}$ acts on $X = \mathbb{C}^{n\times n}$ by left matrix multiplication via π. Through this action, elements in the Lie algebra $\mathfrak{g}$ induce vector fields on X. This gives an injective map

$$\phi \,:\, U(\mathfrak{g}) \hookrightarrow D_{n^2}. \tag{25}$$

We now proceed to describing the algebra map ϕ explicitly. Fix an arbitrary element $\xi \in \mathfrak{g}$. Let $-M_\xi$ be the $n \times n$ matrix corresponding to ξ via the inclusion $\mathfrak{g} \hookrightarrow \mathfrak{gl}(n)$. The following is the vector field encoding the Lie algebra action of M_ξ on the space $\mathfrak{gl}(n) \simeq \mathbb{C}^{n\times n}$:

$$\phi(\xi) \;=\; \sum_{i,\,j\,=1}^{n} (M_\xi)_{ij} \cdot \sum_{k=1}^{n} t_{jk}\partial_{ik} \quad \in D_{n^2}. \tag{26}$$

Example 6.1. Fix $G = \mathrm{SO}(n)$ and let $\pi : G \to \mathrm{GL}_n(\mathbb{R})$ be the standard representation on $\mathbb{R}^n$. The associated Lie algebra $\mathfrak{g}$ is the space of skew-symmetric $n \times n$ matrices over $\mathbb{C}$. A canonical basis of $\mathfrak{g}$ consists of the rank 2 matrices $e_{ij} - e_{ji}$ for $1 \le i < j \le n$. The operator $P_{ij} \in D_{n^2}$ in Theorem 3.3 is Fourier dual to the vector field (26) if we take $\xi = e_{ji} - e_{ij}$.

As seen in [12, Section 1.3], the morphism of varieties $\pi : G_{\mathbb{C}} \to X$ induces a pushforward functor of D-modules $\pi_+ : \mathrm{Mod}(D_G) \to \mathrm{Mod}(D_{n^2})$ satisfying the following key property.

Theorem 6.2. *If we regard* $\mathbb{C}[G]$ *as a left* D_G*-module, then we have the isomorphism*

$$\pi_+(\mathbb{C}[G]) \;\cong\; D_{n^2}/\langle\, I_G,\, \phi(\mathfrak{g})\,\rangle.$$

In particular, this quotient is a regular holonomic simple D_{n^2}*-module.*

Proof. By (24), we have the following isomorphism of right D_G-modules:

$$\mathbb{C}[G] \;\cong\; \mathbb{C} \otimes_{U(\mathfrak{g})} D_G. \tag{27}$$

On the right, $\mathbb{C}$ denotes the trivial representation of the universal enveloping algebra $U(\mathfrak{g})$.

Let $D_{G\to X} := \mathbb{C}[G] \otimes_{\mathbb{C}[X]} D_X$ denote the transfer bimodule. This is a left D_G-module and a right D_X-module. Since the action of $\mathfrak{g}$ extends to the whole space X, we have $\mathbb{C}[G] \cong \mathbb{C}[X]/I_G$ as $\mathfrak{g}$-modules, and the left $U(\mathfrak{g})$-structure of $D_{G\to X}$ is induced by the Leibniz rule via the map (25) on the second factor. We obtain the isomorphism of bimodules

$$D_{G\to X} \;\cong\; D_X/(I_G \cdot D_X). \tag{28}$$

By (27) and (28), we have the following isomorphisms of right D_X-modules:

$$\begin{aligned}\pi_+(\mathbb{C}[G]) := \mathbb{C}[G] \otimes_{D_G} D_{G\to X} &\;\cong\; (\mathbb{C} \otimes_{U(\mathfrak{g})} D_G) \otimes_{D_G} D_{G\to X} \\ &\;\cong\; \mathbb{C} \otimes_{U(\mathfrak{g})} D_X/(I_G \cdot D_X) \;\cong\; D_X/((I_G + \phi(\mathfrak{g})) \cdot D_X).\end{aligned}$$

The fist claim now follows by switching to left D_X-modules. By Kashiwara's Equivalence Theorem [12, Section 1.6], the module $D_X/\langle I_G, \phi(\mathfrak{g})\rangle$ is regular holonomic and simple. □

Remark 6.3. The assumption that G is compact is not needed in Theorem 6.2. The proof works for any representation $\pi : H \to \mathrm{GL}_n(\mathbb{C})$ of a complex connected algebraic group such that $\pi(H)$ is closed in $\mathbb{C}^{n\times n}$. Such a representation exists for all semisimple groups H. Another natural setting is that of orbits of a compact group G acting linearly on a real vector space, with left-invariant measures used in Corollary 6.5. In our view, the theory of orbitopes [27] should be of interest for statistical inference with data sampled from orbits.

Remark 6.4. Here is a more conceptual argument for Theorem 6.2. The D-module $M = D_{n^2}/\langle I_G,\, \phi(\mathfrak{g})\rangle$ is equivariant and supported on $\pi(G_{\mathbb{C}})$ (see [12, Section 11.5]). By Kashiwara's Equivalence Theorem, it is the pushforward of a coherent equivariant D-module on $G_{\mathbb{C}}$. This is a direct sum of copies of the module $\mathbb{C}[G]$, by the Riemann–Hilbert Correspondence. Hence, M is a direct sum of copies of $\pi_+(\mathbb{C}[G])$. The existence of a unique left-invariant measure on G implies that there is only one such summand in M.

Let μ_π be the distribution on $\mathbb{R}^{n\times n}$ given by integration against the Haar measure on G. The following corollary generalizes [16, Theorem 1] from $\mathrm{SO}(n)$ to other Lie groups G.

Corollary 6.5. *The annihilator in* D_{n^2} *of this distribution equals*

$$\mathrm{ann}_{D_{n^2}}(\mu_\pi) \;=\; \langle\, I_G,\, \phi(\mathfrak{g})\,\rangle.$$

Proof. Since $\mathrm{supp}(\mu_\pi) = \pi(G)$, we have $I_G \subset \mathrm{ann}_{D_{n^2}}(\mu_\pi)$. Since μ_π is a left-invariant distribution, we have also $\phi(\mathfrak{g}) \subset \mathrm{ann}_{D_{n^2}}(\mu_\pi)$. By Theorem 6.2, the D-ideal $\langle I_G, \phi(\mathfrak{g}) \rangle$ is a maximal left ideal in D_{n^2}, since its quotient is simple. It is therefore equal to $\mathrm{ann}_{D_{n^2}}(\mu_\pi)$. □

The following observation establishes the connection to statistics, as in [16, Section 4].

Remark 6.6. The Fourier–Laplace transform of μ_π has a complex analytic continuation to a holomorphic function on $\mathbb{C}^{n\times n}$ by the Paley–Wiener–Schwartz Theorem, namely

$$c(\Theta) \;=\; \int_G \exp(\mathrm{tr}(\Theta^{\mathrm{t}}\pi(Y)))\mu(dY). \tag{29}$$

This is the normalizing constant of the Fisher distribution on the group $\pi(G) \subset \mathrm{GL}_n(\mathbb{R})$. Note that this can be defined for a complex representation $\pi(G) \subset \mathrm{GL}_n(\mathbb{C})$ as well.

The Fourier transform, denoted by $(\bullet)^{\mathcal{F}}$, switches the operators t_{ij} and ∂_{ij} in the Weyl algebra D_{n^2}, with a minus sign involved. We consider the image of the D-ideal in Corollary 6.5 under this automorphism of D_{n^2}. This image is a D-ideal J_π that is defined over $\mathbb{R}$:

$$J_\pi \;=\; \langle I_G, \phi(\mathfrak{g}) \rangle^{\mathcal{F}}. \tag{30}$$

The following result generalizes Theorem 3.3 to compact Lie groups other than $\mathrm{SO}(n)$.

Corollary 6.7. *The D-module D_{n^2}/J_π is simple holonomic and* $\mathrm{ann}_{D_{n^2}}(c(\Theta)) = J_\pi$.

Proof. By Corollary 6.5, Remark 6.6, and the defining property of the Fourier transform, we see that J_π annihilates the integral in (29). The proof concludes by recalling that the Fourier transform induces an auto-equivalence on the category of (holonomic) D_{n^2}-modules. □

We saw in Section 5 that sampling from $\mathrm{SO}(3)$ is ubiquitous in the applied sciences. It would be worthwhile to explore such scenarios also for other matrix groups $\pi(G)$, and to apply holonomic methods to maximum likelihood estimation in their Fisher model. An example of such a model is the complex matrix Fisher distribution for unitary groups [20, Section 13.2.4].

One promising context for data applications is the unitary group in quantum physics. The following example is as an invitation to mathematical physicists to develop this further.

Example 6.8. The compact group $G = \mathrm{SU}(2)$ consists of complex 2×2 matrices of the form

$$\begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix}, \qquad \text{with } |\alpha|^2 + |\beta|^2 \;=\; 1. \tag{31}$$

Note that G is a double cover of $\mathrm{SO}(3)$. While the odd-dimensional (complex) representations of G descend to real-valued representations of $\mathrm{SO}(3)$, this is not true for the even-dimensional (spin) representations. Consider the standard representation $G \subset \mathbb{C}^{2\times 2}$.

The complexification of the matrix group in (31) is simply the group $\mathrm{SL}_2(\mathbb{C}) \subset \mathbb{C}^{2\times 2}$. The associated

(maximal, holonomic) ideal J_π is generated by four operators:

$$d = \det(\partial) - 1, \qquad h = t_{11}\partial_{11} + t_{12}\partial_{12} - t_{21}\partial_{21} - t_{22}\partial_{22},$$
$$e = t_{21}\partial_{11} + t_{22}\partial_{12}, \qquad f = t_{11}\partial_{21} + t_{12}\partial_{22}.$$

A computation shows that $\operatorname{rank} J_\pi = 2$ and $\operatorname{Sing}(J_\pi) = \{\Theta \in \mathbb{C}^{2\times 2} \mid \det(\Theta) = 0\}$. The Lie algebra operators e, f, h ensure that every holomorphic solution to J_π is SL_2-invariant. By [19], every solution has the form $\Theta \mapsto \phi(\det(\Theta))$, for some analytic function ϕ in a domain of $\mathbb{C}^*$. This is annihilated by d (hence, by J_π) if and only if $\phi(x)$ is annihilated by

$$x\partial^2 + 2\partial - 1 \in D_1.$$

This has only one (up to scaling) entire solution ϕ, with series expansion at $x = 0$ given by

$$\phi(x) = \sum_{n=0}^{\infty} \frac{1}{n! \cdot (n+1)!}\, x^n.$$

By comparing constant terms, we conclude that $c(\Theta) = \phi\big(\det(\Theta)\big)$. It is straightforward to generalize the above considerations to the fundamental representation of the special unitary group $\mathrm{SU}(m)$ for any $m \geq 1$. In that setting, we find that $\operatorname{rank}(J_\pi) = m$.

In conclusion, the D-ideal J_π is an interesting object that deserves further study, not just for the rotation group $\mathrm{SO}(n)$, but for arbitrary Lie groups G. Sections 3 and 6 offer numerous suggestions for future research. For instance, what is the holonomic rank of J_π? Furthermore, it would be desirable to experiment with data sampled from groups G other than $\mathrm{SO}(3)$, so as to broaden the applicability of algebraic analysis in statistical inference.

Acknowledgments

We thank Mathias Drton for helpful discussions on statistics, Ralf Hielscher for discussions on materials science, and Max Pfeffer for improvements in our numerical methods. We are grateful to Nobuki Takayama and his collaborators for many insightful discussions, and to Charles Wang for getting us started on the material for $\mathrm{SO}(n)$.

References

[1] D. Andres, "dmodloc_lib: a Singular:Plural library for localization of algebraic D-modules and applications", available at https://www.singular.uni-kl.de/Manual/4-1-3/sing_661.htm#SEC715.

[2] F. Bachmann, R. Hielscher, P. E. Jupp, W. Pantleon, H. Schaeben, and E. Wegert, "Inferential statistics of electron backscatter diffraction data from within individual crystalline grains", *J. Appl. Crystallogr.* **43** (2010), 1338–1355.

[3] M. A. Bingham, D. J. Nordman, and S. B. Vardeman, "Finite-sample investigation of likelihood and Bayes inference for the symmetric von Mises–Fisher distribution", *Comput. Statist. Data Anal.* **54**:5 (2010), 1317–1327.

[4] M. A. Bingham, D. J. Nordman, and S. B. Vardeman, "Bayes inference for a tractable new class of non-symmetric distributions for 3-dimensional rotations", *J. Agric. Biol. Environ. Stat.* **17**:4 (2012), 527–543.

[5] M. Brandt, J. Bruce, T. Brysiewicz, R. Krone, and E. Robeva, "The degree of SO(n, $\mathbb{C}$)", pp. 229–246 in *Combinatorial algebraic geometry*, edited by G. G. Smith and B. Sturmfels, Fields Inst. Commun. **80**, Fields Inst. Res. Math. Sci., Toronto, ON, 2017.

[6] J. R. Davis and S. J. Titus, "Modern methods of analysis for three-dimensional orientational data", *J. Struct. Geol.* **96** (2017), 65–89.

[7] T. Downs, J. Liebman, and W. Mackay, "Statistical methods for vectorcardiogram orientations", pp. 216–222 in *Proceedings of the XIth International Symposium on Vectorcardiography*, North-Holland, Amsterdam, 1974.

[8] P. Görlach, C. Lehn, and A.-L. Sattelberger, "Algebraic analysis of the hypergeometric function ${}_1F_1$ of a matrix argument", *Beitrğe Algebra Geom.* (published online November 2020).

[9] D. R. Grayson and M. E. Stillman, "Macaulay2, a software system for research in algebraic geometry", available at http://www.math.uiuc.edu/Macaulay2/.

[10] H. Hashiguchi, Y. Numata, N. Takayama, and A. Takemura, "The holonomic gradient method for the distribution function of the largest root of a Wishart matrix", *J. Multivariate Anal.* **117** (2013), 296–312.

[11] P. Heinzner, "Geometric invariant theory on Stein spaces", *Math. Ann.* **289**:4 (1991), 631–662.

[12] R. Hotta, K. Takeuchi, and T. Tanisaki, *D-modules, perverse sheaves, and representation theory*, Progress in Mathematics **236**, Birkhäuser, Boston, 2008.

[13] Y. Y. Kagan, "Double-couple earthquake source: symmetry and rotation", *Geophys. J. Int.* **194**:2 (2013), 1167–1179.

[14] C. G. Khatri and K. V. Mardia, "The von Mises–Fisher matrix distribution in orientation statistics", *J. Roy. Statist. Soc. Ser. B* **39**:1 (1977), 95–106.

[15] C. Koutschan, "HolonomicFunctions: A Mathematica package for dealing with multivariate holonomic functions, including closure properties, summation, and integration", available at https://www3.risc.jku.at/research/combinat/software/ergosum/RISC/HolonomicFunctions.html.

[16] T. Koyama, "The annihilating ideal of the Fisher integral", preprint, 2015. arXiv 1503.05261

[17] T. Koyama, H. Nakayama, K. Nishiyama, and N. Takayama, "The holonomic rank of the Fisher–Bingham system of differential equations", *J. Pure Appl. Algebra* **218**:11 (2014), 2060–2071.

[18] V. Levandovskyy and J. Martin Morales, "dmod_lib: A Singular:Plural library for algorithms for algebraic D-modules", available at https://www.singular.uni-kl.de/Manual/4-1-3/sing_537.htm#SEC591.

[19] D. Luna, "Fonctions différentiables invariantes sous l'opération d'un groupe réductif", *Ann. Inst. Fourier* (*Grenoble*) **26**:1 (1976), 33–49.

[20] K. V. Mardia and P. E. Jupp, *Directional statistics*, Wiley Series in Probability and Statistics **494**, Wiley, Chichester, 2009.

[21] R. J. Muirhead, *Aspects of multivariate statistical theory*, Wiley, New York, 1982.

[22] 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", *Adv. in Appl. Math.* **47**:3 (2011), 639–658.

[23] J. Nocedal and S. J. Wright, *Numerical optimization*, 2nd ed., Springer, 2006.

[24] M. J. Prentice, "Orientation statistics without parametric assumptions", *J. Roy. Statist. Soc. Ser. B* **48**:2 (1986), 214–222.

[25] C. Procesi, *Lie groups: an approach through invariants and representations*, Springer, 2007.

[26] D. Rancourt, L.-P. Rivest, and J. Asselin, "Using orientation statistics to investigate variations in human kinematics", *J. Roy. Statist. Soc. Ser. C* **49**:1 (2000), 81–94.

[27] R. Sanyal, F. Sottile, and B. Sturmfels, "Orbitopes", *Mathematika* **57**:2 (2011), 275–314.

[28] A.-L. Sattelberger and B. Sturmfels, "D-Modules and Holonomic Functions", preprint, 2019. arXiv 1910.01395

[29] T. Sei, H. Shibata, A. Takemura, K. Ohara, and N. Takayama, "Properties and applications of Fisher distribution on the rotation group", *J. Multivariate Anal.* **116** (2013), 440–455.

[30] N. Takayama, "Gröbner basis for rings of differential operators and applications", pp. 279–344 in *Gröbner bases*, edited by T. Hibi, Springer, 2013.

[31] N. Takayama, T. Koyama, T. Sei, H. Nakayama, and K. Nishiyama, "hgm: an R package for the holonomic gradient method", available at https://cran.r-project.org/web/packages/hgm/hgm.pdf.

Received 2019-12-03. Revised 2020-04-09. Accepted 2020-04-30.

MICHAEL F. ADAMER: michael.adamer@bsse.ethz.ch
D-BSSE, ETH Zürich, 4058 Basel, Switzerland

ANDRÁS C. LŐRINCZ: lorincza@hu-berlin.de
Humboldt-Universität zu Berlin, Institut für Mathematik, 10099 Berlin, Germany

ANNA-LAURA SATTELBERGER: anna-laura.sattelberger@mis.mpg.de
Max Planck Institute for Mathematics in the Sciences, 04103 Leipzig, Germany

BERND STURMFELS: bernd@math.berkeley.edu
University of California, Berkeley, CA 94720, United States

and

Max Planck Institute for Mathematics in the Sciences, 04103 Leipzig, Germany