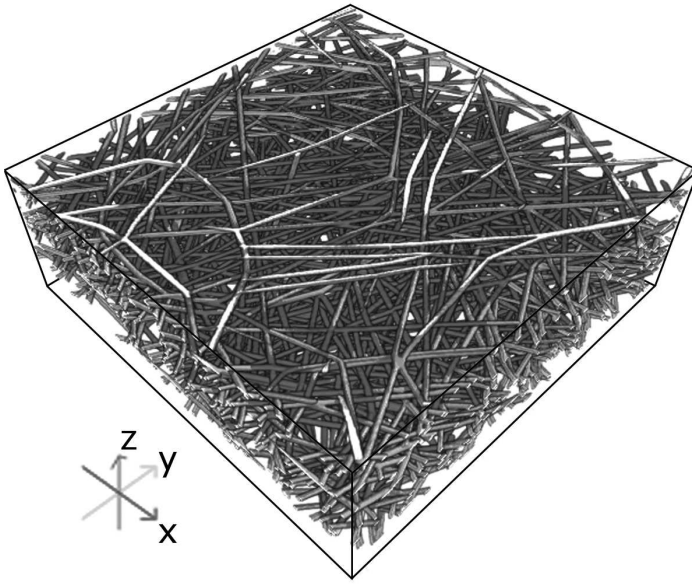
TECHNISCHE UNIVERSITÄT
KAISERSLAUTERN

Fibre Processes and their Applications



Na Zhang

Faserprozesse und ihre Anwendungen

Na Zhang

Vom Fachbereich Mathematik
der Technischen Universität Kaiserslautern
zur Verleihung des akademischen Grades
Doktor der Naturwissenschaften
(Doctor rerum naturalium, Dr. rer. nat.)
genehmigte Dissertation

1. Gutachter: Prof. Dr. Jürgen Franke
2. Gutachter: Prof. Dr. Aila Särkkä

Datum der Disputation: 28. Mai 2013

D 386

Abstract

The main purpose of the study was to improve the physical properties of the modelling of compressed materials, especially fibrous materials. Fibrous materials are finding increasing application in the industries. And most of the materials are compressed for different applications. For such situation, we are interested in how the fibre arranged, e.g. with which distribution. For given materials it is possible to obtain a three-dimensional image via micro computed tomography (μCT). Since some physical parameters, e.g. the fibre lengths or the directions for points in the fibre, can be checked under some other methods from image, it is beneficial to improve the physical properties by changing the parameters in the image.

In this thesis, we present a new maximum-likelihood approach for the estimation of parameters of a parametric distribution on the unit sphere, which is various as some well known distributions, e.g. the von-Mises Fisher distribution or the Watson distribution, and for some models better fit. The consistency and asymptotic normality of the maximum-likelihood estimator are proven. As the second main part of this thesis, a general model of mixtures of these distributions on a hypersphere is discussed. We derive numerical approximations of the parameters in an Expectation Maximization setting. Furthermore we introduce a non-parametric estimation of the EM algorithm for the mixture model. Finally, we present some applications to the statistical analysis of fibre composites.

Chapter 1 establish some fundaments of the study, i.e. the point processes and the fibre processes in 3D, and a briefly introduction of the software MAVI and the models of fibre systems.

The well known von Mises-Fisher distribution and Watson distribution are introduced in Chapter 2. Furthermore we focus on the Schladitz distribution, which we mainly worked with. The Schladitz distribution is a special case of the angular central Gaussian distribution (ACG distribution). Finally in this Chapter a comparison of the Watson distribution and the Schladitz distribution is made.

A maximum-likelihood approach for the estimation of the parameter of Schladitz distribution is introduced in Chapter 3 and consistency and asymptotic normality of the maximum-likelihood estimator are proven. A numerical approximation of the maximum-likelihood estimator is performed for some fibre composites.

Chapter 4 presents a general model of mixtures of Schladitz distributions. We derive numerical approximations of the parameters in an expectation maximization setting. The consistency of the EM estimator is proven. A non-parametric estimation of the EM algorithm for the mixture model is also discussed here.

Acknowledgements

First and foremost I would like to thank my supervisor Prof. Jürgen Franke. It has been an honor to be his PhD student. I have very much appreciated all his contributions of ideas, time and helping me with the funding during my PhD experience. He has taught me how excellent and useful statistic is.

I would also like to thank Dr. Claudia Redenbach. She has supported me in my daily research, supplied me with useful advice and many tips. Additionally, I am also thankful for the excellent example she has provided as a successful woman researcher.

The members of the statistic group have contributed immensely to my PhD time in Kaiserslautern. The group has been a source of friendship as well as good advice and cooperation. I owe thanks to all members of my group.

The members of the image processing group at Fraunhofe ITWM have contributed to the software MAVI and testmaterials for my PhD. I would like to acknowledge all the group members, especially Katja Schladitz and Oliver Wirjadi. They provided useful advice and examples.

My PhD was supported by the $(CM)^2$ (Center for Mathematical and Computational Modelling). I would like to thank all the members of $(CM)^2$, they have helped me to understand the various aspects of mathematics and the possibility of co-work with other subjects.

And, not the least, I thank my parents. They not only brought me into the world, but also constantly showed a great interest in my work and research. I thank my husband Pu Huang for providing unconditional love and care. I also thank my friends (too many to list here but you know who you are!) the providing support and the friendship that I needed.

I would also like to thank all people who supported me in writing this thesis.

Nomenclature

$M(a,b,z)$	Kummer's function of the first kind
$(a)_n$	Rising factorial
$*$	Convolution
λ	Intensity measure
ν_d	Lebesgue measure
$\mathbb{1}_B$	Indicator function of B
$\oplus$	Kronecker product
Φ	Point process
ϕ_{-x}	$\phi_{-x} = \{y + x : y \in \phi\}$
Ψ	Random measure
$\mathbb{R}^d$	$d-$dimensional Euclidean space
$\mathbb{S}$	Convex ring
sec	$\sec(\beta) = \frac{1}{\cos(\beta)}$
$\stackrel{d}{=}$	With the same distribution
$\stackrel{L}{\rightarrow}$	Convergence in law
Var	Variance
$\stackrel{P}{\longrightarrow}$	Convergence in probability
${}_pF_q$	Hypergeometric function
g_σ	Gaussian convolution filter
H	Hessian matrix
h_1	1-dimensional Hausdorff measure
I_ν	Modified Bessel function of the first kind and order ν

$r^{-1}\phi$	$r^{-1}\phi = \{r^{-1}y : y \in \phi\}$
SO_d	Group of all $d \times d$ special orthogonal matrix
tr	Trace of matrix
vec	Vector formed by stacking the columns of matrix
cond	Condition number of normal matrix
det	Determinant

Contents

Chapter 1

Fundaments

1.1 Spatial point processes

A point process is a random collection of points falling in some space. In one dimension, each point represents the time when a particular event occurs. A spatial point process is a random collection of points falling in d-dimensional space ($d \geq 2$). In most applications, each point represents the location of an event. Examples of events include receiving emergency calls in a hospital, sightings of birds in a view point or occurrences of earthquakes or floods. These point processes play a very important role in stochastic geometry, especially as the fundamental components of the fibre processes.

We use the following notations and concepts:

(i) $\mathbb{E}$ is a locally compact space with countable base. The space $\mathbb{E}$ is equipped with the Borel $\sigma-$ field $\mathcal{B}(\mathbb{E})$.

(ii) The space $M = M(\mathbb{E})$ is the set of all locally finite measures on $\mathbb{E}$.

(iii) The space $N = N(\mathbb{E})$ is the set of all locally finite counting measures on $\mathbb{E}$.

(iv) $\mathcal{M}$ is the smallest $\sigma-$algebra on M such that the mapping $\mu \mapsto \mu(B)$ is measurable for all measurable sets $B \in \mathcal{B}(\mathbb{E})$.

(v) $\mathcal{N}$ is the trace of $\mathcal{M}$ on N.

There are several ways of characterizing a point process. Here we use the definition of Stoyan et al. (2008).

Definition 1.1.1 (Random measure, point process)
Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. A random measure on $\mathbb{E}$ is a measurable mapping Ψ from $(\Omega, \mathcal{A}, \mathbb{P})$ into $(M, \mathcal{M})$. A point process on $\mathbb{E}$ is a measurable mapping Φ from a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ into $(N, \mathcal{N})$.

Definition 1.1.2 (Stationary, isotropic random measure)
Let $\mathbb{E} = \mathbb{R}^d$. A random measure Ψ on $\mathbb{R}^d$ is called

- stationary, if $\Psi \stackrel{d}{=} \Psi + x$ for all $x \in \mathbb{R}^d$.
- isotropic, if $\Psi \stackrel{d}{=} \nu\Psi$ for all $\nu \in SO_d$.

where $A \stackrel{d}{=} B$ denotes that A and B have the same distribution. and SO_d denotes the group of all $d \times d$ special orthogonal matrix.

Definition 1.1.3 (Intensity measure of point process)
The intensity measure Λ of a point process Φ is the measure Λ on $(\mathbb{E}, \mathcal{B}(\mathbb{E}))$ defined as

$$\Lambda(B) := \mathbb{E}[\Phi(B)], \qquad B \in \mathcal{B}(\mathbb{E})$$

Definition 1.1.4 (Poisson point process)
Let $\mathbb{E}$ be defined as in Definition 1.1.1. A Poisson process on $\mathbb{E}$ is a simple point process Φ with the following properties:

(i) For pairwise disjoint Borel sets $A_1, \cdots, A_k$ the random variables $\Phi(A_1), \cdots, \Phi(A_k)$ are independent.

(ii) For each $A \in \mathcal{B}(\mathbb{E})$ with $\Lambda(A) < \infty$, the random variable $\Phi(A)$ has a Poisson distribution. Since $\mathbb{E}\Phi(A) = \Lambda(A)$, the parameter of the Poisson distribution is $\Lambda(A)$, hence

$$\mathbb{P}(\Phi(A) = k) = e^{-\Lambda(A)} \frac{\Lambda(A)^k}{k!}.$$

1.2 Spatial fibre processes

The theory of fibre processes is similar to the theory of point processes. In the case of stationarity (the most cases of application models) one basic characteristic is the intensity, i.e. the mean fibre length per unit area. Fibre processes also associate directions to their constituent points and this leads to a further important characteristic: the rose of directions. Statistical investigation begins with estimation of intensity and rose of directions and furthermore consider the suitability of various stochastic models for the sample. Such stochastic models of these systems are called fibre processes.

A fibre is a sufficiently smooth simple curve in the space, of finite length.

Definition 1.2.1 (Fibre)
A fibre γ is a subset of $\mathbb{R}^3$ which is the image of a curve $\gamma(t) = (\gamma_1(t), \gamma_2(t), \gamma_3(t))$ such that:

(i) $\gamma : [0,1] \to \mathbb{R}^3$ is continuously differentiable,

(ii) $|\gamma'(t)|^2 = |\gamma_1'(t)|^2 + |\gamma_2'(t)|^2 + |\gamma_3'(t)|^2 > 0$ for all t,

(iii) the mapping γ is injective, so that a fibre does not intersect itself.

Remark 1.2.2
γ is also used for denoting the measure

$$\gamma(B) = h_1(\gamma \cap B) = \int_0^1 \mathbb{1}_B(\gamma(t)) \sqrt{(\gamma_1'(t))^2 + \gamma_2'(t))^2 + \gamma_3'(t))^2} dt$$

for $B \in \mathcal{B}(\mathbb{R}^3)$, where h_1 is the 1-dimensional Hausdorff measure and $\mathbb{1}_B$ is the indicator function of B. It represents the length of the part of the fibre γ, which lies in B.

Definition 1.2.3 (Fibre system)
A fibre system ϕ is a closed subset of $\mathbb{R}^3$ which

(i) can be represented as

$$\phi = \bigcup_{i=1}^{\infty} \gamma^{(i)}$$

for fibres $\gamma^{(i)}$,

(ii) for any compact set $K \subset \mathbb{R}^3$, $K \cap \gamma^{(i)} \neq \emptyset$ only for finitely many i.

(iii)

$$\gamma^{(i)}((0,1)) \cap \gamma^{(j)}((0,1)) = \emptyset \qquad \text{if } i \neq j.$$

It excludes fibres which are crossing, but it is possible that an endpoint of one fibre touches another fibre (like a T-intersection).

Definition 1.2.4 (Length of the fibre system)
The corresponding length measure of the fibre system ϕ is defined as

$$\phi(B) = \sum_{\gamma^{(i)} \in \phi} \gamma^{(i)}(B) = \sum_{\gamma^{(i)} \in \phi} h_1(\gamma^{(i)} \cap B)$$

for $B \in \mathcal{B}(\mathbb{R}^3)$, i.e. the total length of all fibres of ϕ, which lie in B. Here again h_1 is the 1-dimensional Hausdorff measure, which is equal to the length of the curve.

Note that, the measure ϕ depends only on the closed set that is the union of all the fibres.

The family of all spatial fibre systems $\mathbb{D}$ is endowed with a σ-algebra $\mathcal{D}$ generated by sets of the form

$$\{\phi \in \mathbb{D} : \phi(B) < x\}$$

for $x \in \mathbb{R}$ and $B \in \mathcal{B}(\mathbb{R}^3)$ compact.

Definition 1.2.5 (Fibre process)
A fibre process Φ is a random variable with values in $[\mathbb{D}, \mathcal{D}]$, i.e. a measurable mapping from $[\Omega, \mathcal{A}, P]$ to $[\mathbb{D}, \mathcal{D}]$.

Note that Φ is also used to denote the corresponding random length measure. The distribution of the fibre process P is defined as

$$P(D) = P(\{\omega : \phi(\omega) \in D\})$$

for $D \in \mathcal{D}$.

Definition 1.2.6 (Stationary, isotropic fibre process)
The fibre process Φ is stationary if it has the same distribution as the translated fibre process Φ_x. Thus

$$P(Y) = P(Y_x) \qquad \text{for all } Y \in \mathcal{D} \text{ and all } x \in \mathbb{R}^3$$

where $Y_x = \{\phi \in \mathbb{D} : \phi_{-x} \in Y\}$. It is isotropic if the distribution remains invariant under rotations about the origin. Thus

$$P(Y) = P(rY) \qquad \text{for all } Y \in \mathcal{D} \text{ and } r \text{ rotation about the origin}$$

where $rY = \{\phi \in \mathbb{D} : r^{-1}\phi \in Y\}$. For the notations ϕ_{-x} and $r^{-1}\phi$, we have

$$\begin{aligned} \phi_{-x} &= \{y + x : y \in \phi\} \\ r^{-1}\phi &= \left\{r^{-1}y : y \in \phi\right\} \end{aligned}$$

Definition 1.2.7 (Intensity of fibre process)
The intensity measure Λ of a fibre process is given by

$$\Lambda(B) = \mathbb{E}(\Phi(B)) = \mathbb{E}\left(\sum_{\gamma\in\phi} h_1(\gamma \cap B)\right)$$

for $B \in \mathcal{B}(\mathbb{R}^3)$. If Φ is stationary, the intensity L_V of Φ is the mean fibre length per unit volume, i.e.

$$\Lambda(B) = L_V \nu_3(B) = \mathbb{E}(\Phi(B))$$

where ν_d denotes the Lebesgue measure, in the case $d = 3$ the Lebesgue measure is equal to the volume measure.

Definition 1.2.8 (Rose of directions)
The rose of directions $\mathcal{R}$ is defined as a measure on the projective plane of the spatial system of all straight lines through the origin. The corresponding $\sigma-$algebra $\mathcal{L}$ is the system of all spatial Borel sets which are unions of lines through the origin. The quantity $\mathcal{R}$ of a stationary spatial fibre process Φ is defined by

$$\mathcal{R}(\mathcal{A}) = \frac{1}{L_V \nu_3(B)} \mathbb{E}\left(\int_B \mathbb{1}(l(x) \in A)(x)\Phi(dx)\right) \qquad \text{for } A \in \mathcal{L}$$

If γ is a fibre, $x = \gamma(t)$ for some $0 < t < 1$, then $l(x)$ is the fibre tangent line through x.

1.3 Fibre orientations with MAVI

For the estimation of the parameters of fibre distributions in 3D we need to get the rose of directions, or to put it simply, a sample of fibre directions, which are chosen from the points in the fibre system. We use MAVI to obtain the fibre directions during our works. MAVI is a software tool for 3D image processing, developed by Fraunhofer ITWM (2012).

Firstly it yields a 3D image of fibre system obtained by micro computer tomography (μCT). Then, MAVI measures the local fibre direction in each fibre voxel with use of a method based on partial second derivatives, i.e. the Hessian matrix Wirjadi (2009).

Let σ be the scale parameter of an isotropic Gaussian convolution filter g_σ. Then the Hessian matrix at scale σ for a $d-$dimensional image is defined as

$$H = \begin{pmatrix} f_{x_1x_1} & \cdots & f_{x_1x_d} \\ \vdots & \ddots & \vdots \\ f_{x_dx_1} & \cdots & f_{x_dx_d} \end{pmatrix} \tag{1.1}$$

where

$$f_{x_i x_j} = \frac{\partial^2}{\partial x_i \partial x_j} f * g_\sigma$$

are the partial second derivatives, i.e., more detailed,

$$f_{x_i x_j}(t) = \int_{\mathbb{R}^d} \frac{\partial^2}{\partial x_i \partial x_j} f(t-u) g_\sigma(u) du$$

where g_σ is the density of the d-variate Gaussian distribution with mean 0 and covariance matrix σI_d, I_d denoting the unit matrix. Therefore, H is a matrix of locally smoothed second derivation which can be estimated easier from randomly distributed data. As the order of differentiation is unimportant, the matrix H is symmetric. The second order gradient in direction v can be computed as

$$\nabla_v^2 = v^t H v$$

Low curvatures are expected along a fibre, which implies that ∇_v^2 is minimized if v is the tangent vector to the fibre direction in a fibre point x. Therefore, the local direction $v(x)$ in a pixel x is given by the eigenvector corresponding to the smallest eigenvalue of H in x. In practice, the scale parameter σ should be chosen as the mean radius of the fibres in the data.

Additionally MAVI have also other functions, which are used during our works, such as the function "crop" can crop a large image in some small images. This function is mainly used for measuring the local directions for some materials consistency of several layers. It is possible to measure the directions of corresponding fibres in each layer. An example of such material can be found in 3.3.

Further informations on MAVI can be found in the handbook of ITWM (2012).

1.4 Geometric models for fibre systems

In this section we discuss the models used to simulate the structure of fibre systems in which the fibre positions are represented by sets of random cylinders in 2D or 3D. There are two common geometric models for fibre systems, i.e. Poisson cylinder model and Random Sequential Adsorption (RSA) model. Here we make a brief introduction of RSA model, which we use mainly in our work. The main difference between both models is that Poisson cylinder model produces overlapping fibres and RSA model produces non-overlapping fibres.

Both models start with an empty observation window W. Size, shape and direction of a particle are drawn from given distributions. Then, a position for this object is suggested from a uniform distribution on W. By the Poisson cylinder model, this procedure is repeated until a given number of fibres is obtained. On the other hand it requires a reject-accept process by the RSA model. If the newly introduced particle overlaps with any of the previously placed objects, the proposal is rejected. Otherwise, the proposal is accepted and the particle is placed inside W. This procedure is repeated until either a given volume fraction is obtained or the jamming limit has been reached, which means that it is not possible to place any further particles into W.

The pseudo code of RSA model for simulation fibres reads in Algorithm 1

Algorithm 1 Simulate Fibres of the RSA model

Require: a window W, a natural number $n \geq 0$ as the capacity of fibres in window W (not necessary), a maximal permitted run time $t > 0$ and some fibre parameters for description of the fibre F, e.g. the fibre length l, the fibre radius r.

Ensure: maximal n non-overlapping fibres in the Window W.

```
Simulation
place a fibre in window W and mark F as F_1
while run time <t do
  for i = 1 → n − 1 do
    place a new fibre F in the window W
    if F ∩ F_1 ∩ ··· ∩ F_i = ∅ then
      accept F and mark F as F_{i+1}
    else
      reject F
    end if
  end for
end while
return the fibre system F_1, ··· , F_k (k ≤ n)
```

Chapter 2

Distributions in 3D

We make in this chapter a brief introduction of some 3D distributions, which are used by our research.

Several models for the distribution of spherical data in 3D can be found in the literature. These include the von Mises-Fisher distribution and the Watson distribution. Common to these distribution families is that their probability density elements can be written in the form Ce^T, where C is a normalization constant which varies for every model and T is a function with several parameters for the location and spherical shape of the distribution (see Fisher et al. (1993) for an overview). Furthermore a new distribution called the Schladitz distribution Schladitz et al. (2006) will also be introduced. The Schladitz distribution is a special case of the angular central Gaussian distribution, introduced in Schladitz et al. (2006) and Ohser and Schladitz (2009). The density function of the Schladitz distribution is no more of the form Ce^T. In the end of this subsection we will discuss the differences between the density function of the Watson distribution and of the Schladitz distribution.

In the following we use the polar coordinates (θ, ϕ) (θ the colatitude, ϕ the longitude) to describe a point on the unit sphere in 3D.

2.1 The von Mises-Fisher distribution

Definition 2.1.1 (von Mises-Fisher distribution)
A d-dimensional unit random vector x (i.e. $x \in \mathbb{R}^d$ and $\|x\| = 1$) is said to have d-variate von Mises-Fisher distribution if its density function is given by

$$f_F(x) = C_{F_d}(\kappa) \exp(\kappa \mu^T x) \tag{2.1}$$

where $\kappa \geq 0$, $\|\mu\| = 1$ and the normalization constant $C_{F_d}(\kappa)$ is equal to

$$C_{F_d}(\kappa) = \frac{\kappa^{d/2-1}}{(2\pi)^{d/2} I_{d/2-1}(\kappa)} \tag{2.2}$$

where $I_\nu(\cdot)$ represents the modified Bessel function of the first kind and order ν, given in series form by

$$I_\nu(x) = \sum_{m=0}^{\infty} \frac{1}{m!\Gamma(m+\nu+1)} \left(\frac{x}{2}\right)^{2m+\nu}$$

Especially if $d = 3$, the normalization constant reduces to

$$C_{F_3} = \kappa/(4\pi \sinh \kappa) = \kappa/(2\pi(\exp(\kappa) - \exp(-\kappa)))$$

and in this case, using polar coordinates, the density f_F is given by

$$f_F(\theta, \phi) = C_{F_3}(\kappa) \exp(\kappa \mu^T x(\theta, \phi))$$

with

$$x(\theta, \phi) = \begin{pmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{pmatrix}.$$

The parameter μ is the location parameter called the mean direction, and the distribution shows the rotational symmetry about this direction. κ is a shape parameter called the concentration parameter, which characterizes how strongly the unit vectors drawn according to the distribution $f_F(x)$ are concentrated about the mean direction μ. The greater the value of κ, the stronger the concentration of the distribution about the mean direction. In extreme cases the distribution is uniform on the sphere $\mathbb{S}^{d-1}$ for $\kappa = 0$, and the distribution tends to total concentration at μ, i.e. the point density as $\kappa \to \infty$.

Fisher et al. (1981) describes a method to simulate unit vectors from the von Mises-Fisher distribution. The pseudo code reads as in Algorithm 2.

Figure 2.1 shows points sampled from von Mises-Fisher distributions on the sphere with different parameter κ. 1000 points were simulated using Fisher's method. The top subfigures show the 1000 points distributed on the unit sphere and in the bottom subfigures, a density estimation is constructed and displayed. The areas with the colour more approach to yellow describes more intensive the points in these areas distributed and with the colour approach to red otherwise. Note that for large κ, the von Mises-Fisher distribution is stark polar distribution, there exist as many points by the south pole as by the north pole, but unfortunately it shows only the points by the north pole in our figure.

An estimation method based on a series of independent measurements drawn from a von Mises-Fisher distribution is given by Banerjee et al. (2006) by maximizing the log-likelihood.

Let $\mathbb{X}$ be a finite set of sample unit vectors drawn independently from a von Mises-Fisher distribution, i.e.

$$\mathbb{X} = \left\{ x_i^T \in \mathbb{S}^{d-1} | x_i^T \text{ drawn from } f(\theta, \phi) \text{ for } 1 \leq i \leq n \right\}$$

Then the maximum likelihood estimates for the parameters μ, κ are given by

$$\begin{aligned} \hat{\mu} &= \frac{r}{\|r\|} = \frac{\sum_{i=1}^n x_i}{\|\sum_{i=1}^n x_i\|} \\ \hat{\kappa} &= A_d^{-1}(\bar{r}) = A_d^{-1}\left(\frac{\|\sum_{i=1}^n x_i\|}{n}\right) \end{aligned}$$

where

$$r = \sum_{i=1}^n x_i \text{ and } A_d(\kappa) = \frac{I_{d/2}(\kappa)}{I_{d/2-1}(\kappa)}$$

Algorithm 2 Simulate data from the von Mises-Fisher distribution (Fisher's Method)

Require: a natural number $n > 0$, a non-negative real number κ and a direction with the polar coordinates (α, β)

Ensure: n unit vectors drawn from the von Mises-Fisher distribution with the parameter κ, rotational symmetry about the axis (α, β)

Main-step: Simulation

$\lambda = \exp(-2\kappa)$;

for $i = 1 \to n$ **do**

$\theta_i = 2\arcsin(\sqrt{-\log(R_1(1-\lambda)+\lambda)/(2\kappa)})$;

$\phi_i = 2\pi R_2$;

end for

Post-step: rotate the data towards its mean direction

calculate the rotation matrix

$$A = \begin{pmatrix} \cos\alpha\cos\beta & \cos\alpha\sin\beta & -\sin\alpha \\ -\sin\beta & \cos\beta & 0 \\ \sin\alpha\cos\beta & \sin\alpha\sin\beta & \cos\alpha \end{pmatrix}$$

calculate (θ_i', ϕ_i') from

$$\begin{pmatrix} \sin\theta_i'\cos\phi_i' \\ \sin\theta_i'\sin\phi_i' \\ cos\theta_i' \end{pmatrix} = A \cdot \begin{pmatrix} \sin\theta_i\cos\phi_i \\ \sin\theta_i\sin\phi_i \\ cos\theta_i \end{pmatrix}$$

return (θ_i', ϕ_i')

where R_1, R_2 are independent pseudo-random numbers, which are uniformly distributed in the interval $[0, 1]$

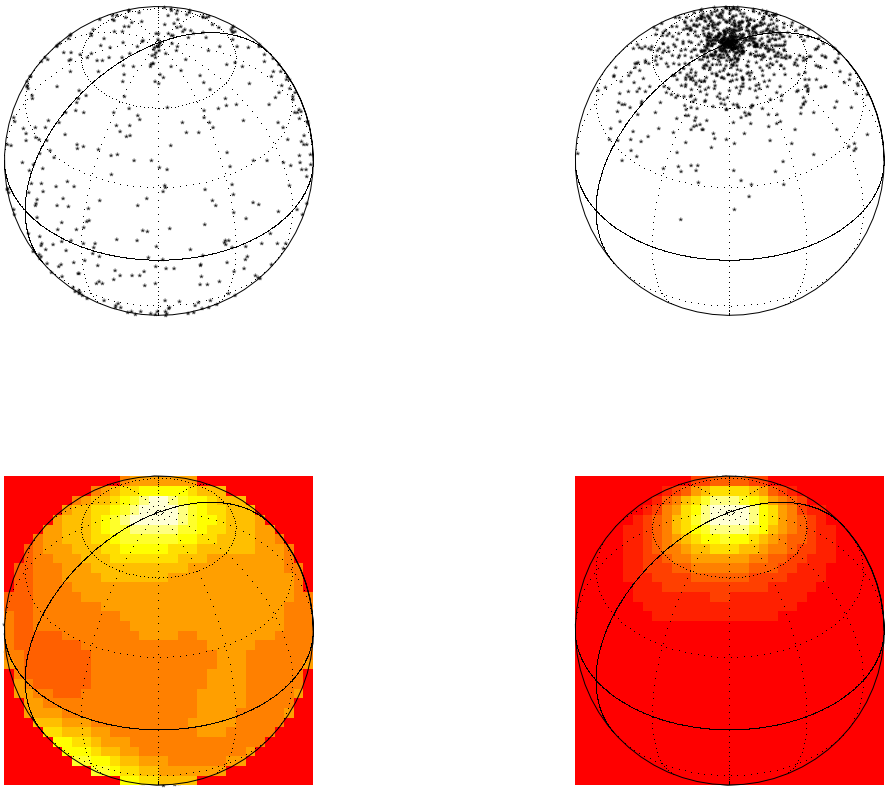


Figure 2.1: Simulation data from von Mises-Fisher distributions using the Fisher's Method for 1000 points with $\kappa = 2$ (left) and $\kappa = 50$ (right).

A simple approximation of κ is

$$\hat{\kappa} = \frac{\bar{r}d - \bar{r}^3}{1 - \bar{r}^2} \tag{2.3}$$

Sra (2012) improved this approximation by performing a few iterations of Newton's method, where the initial value κ_0 is calculated from (2.3), and the estimation is obtained from computing the following two Newton steps

$$\begin{aligned} \kappa_1 &= \kappa_0 - \frac{A_d(\kappa_0) - \bar{r}}{1 - A_d(\kappa_0)^2 - \frac{p-1}{\kappa_0} A_d(\kappa_0)} \\ \kappa_2 &= \kappa_1 - \frac{A_d(\kappa_1) - \bar{r}}{1 - A_d(\kappa_1)^2 - \frac{p-1}{\kappa_1} A_d(\kappa_1)} \end{aligned}$$

2.2 The Watson distribution

Definition 2.2.1 (the Watson distribution)
A $d-$dimensional unit random vector x (i.e. $x \in \mathbb{R}^d and \|x\| = 1$) is said to have $d-$variate

Watson distribution if its density function is given by

$$f_W(x) = C_{W_d}(\kappa)\exp(\kappa(\mu^T x)^2) \tag{2.4}$$

where $\|\mu\| = 1$ and the normalization constant $C_{W_d}(\kappa)$ is equal to $M\left(\frac{1}{2}, \frac{d}{2}, \kappa\right)^{-1}$. $M\left(\frac{1}{2}, \frac{d}{2}, \kappa\right)$ is the confluent hypergeometric function known as Kummer's function of the first kind, which is defined as

$$M(a, b, z) = \sum_{n=0}^{\infty} \frac{(a)_n z^n}{(b)_n n!}$$

where $(a)_n = a(a+1)(a+2)\cdots(a+n-1)$ is the rising factorial.

Especially if $d = 3$, the normalization constant reduces to

$$C_{W_3} = 1 \Bigg/ \left(4\pi \int_0^1 \exp(-\kappa u^2) du\right)$$

Again, for $d = 3$, we write $f_W(\theta, \phi) = f_W(x(\theta, \phi))$ if we use polar coordinates.

The Watson distribution is the basic model for undirected lines distributed with rotational symmetry in either bipolar or girdle form, which is determined by the shape parameter κ. For positive values of κ the distribution is bipolar. The larger the value of κ the more the distribution is concentrated round the axis μ. For negative values of κ the distribution is girdle. The larger the value of $|\kappa|$ the more the distribution is concentrated round the great circle in the plane orthogonal to the axis μ. The parameter μ is a location parameter, the distribution has rotational symmetry about this axis and it is called principal axis in the bipolar case and polar axis in the girdle case.

In extreme cases the distribution is uniform on the sphere $\mathbb{S}^{d-1}$ for $\kappa = 0$, the Watson distribution tends to total concentration at the two ends (the two intersection points of the principal axis and the unit sphere) for $\kappa \to \infty$, and the distribution tends to a uniform distribution on the great circle in the plane normal to the polar axis for $\kappa \to -\infty$.

Best and Fisher (1986) was giving a method to simulate unit vectors from the Watson distribution, which is still in use. The pseudo code reads in Algorithm 3.

Figure 2.2 shows points sampled from the Watson distribution on the unit sphere with different parameter κ. 1000 points were simulated under the method of Algorithm 3. The left subfigures show the 1000 points distributed on the unit sphere and the right subfigures display the density estimation. Note that the Watson distribution describes the axial directions, which are identical for the upper half-sphere and lower half-sphere, therefore we consider here only the upper half-sphere.

An estimation based on a series of independent measurements drawn from a Watson distribution is calculated by Fisher et al. (1993) by maximizing the log-likelihood.

Let $\mathbb{X}$ be a finite set of sample unit vectors drawn independently from a Watson distribution, i.e.

$$\mathbb{X} = \left\{x_i^T \in \mathbb{S}^{d-1} | x_i^T \text{ drawn from } f_W(\theta, \phi; \mu, \kappa) \text{ for } 1 \le i \le n\right\}$$

The corresponding log-likelihood is

$$\begin{aligned} l(\mu, \kappa) &= \sum_{i=1}^{n} \log\left(M\left(\frac{1}{2}, \frac{d}{2}, \kappa\right)^{-1} \cdot e^{\kappa(\mu^T x)^2}\right) \\ &= n\left(\kappa \mu^T S \mu - \log M\left(\frac{1}{2}, \frac{d}{2}, \kappa\right)\right) \end{aligned}$$

Algorithm 3 Simulate data from the Watson distribution

Require: a natural number $n > 0$, a real number κ and a direction with the polar coordinates (α, β)

Ensure: n unit vectors drawn from the Watson distribution with the parameter κ, rotational symmetry about the axis (α, β)

Main-step: Simulation

for $i = 1 \to n$ **do**

if $\kappa > 0$ **then**

$C = 1/(\exp(\kappa) - 1)$;

$U = R_1, V = R_2$;

$S = 1/\kappa \log(U/C + 1)$;

while $V > \exp(\kappa S^2 - \kappa S)$ **do**

$U = R_3, V = R_4$;

$S = 1/\kappa \log(U/C + 1)$;

end while

$\theta_i = \arccos S$;

$\phi_i = 2\pi R_0$;

else

$C_1 = \sqrt{-\kappa}$;

$C_2 = \arctan C_1$;

$U = R_1, V = R_2$;

$S = (1/C_1) \cdot \tan(C_2 \cdot U)$;

while $V > (1 - \kappa S^2) \cdot \exp(\kappa S^2)$ **do**

$U = R_3, V = R_4$;

$S = (1/C_1) \cdot \tan(C_2 \cdot U)$;

end while

$\theta_i = \arccos S$;

$\phi_i = 2\pi R_0$;

end if

end for

Post-step: rotate the data towards its mean direction

calculate the rotation matrix

$$A = \begin{pmatrix} \cos\alpha\cos\beta & \cos\alpha\sin\beta & -\sin\alpha \\ -\sin\beta & \cos\beta & 0 \\ \sin\alpha\cos\beta & \sin\alpha\sin\beta & \cos\alpha \end{pmatrix}$$

calculate (θ_i', ϕ_i') from

$$\begin{pmatrix} \sin\theta_i'\cos\phi_i' \\ \sin\theta_i'\sin\phi_i' \\ cos\theta_i' \end{pmatrix} = A \cdot \begin{pmatrix} \sin\theta_i\cos\phi_i \\ \sin\theta_i\sin\phi_i \\ cos\theta_i \end{pmatrix}$$

return (θ_i', ϕ_i')

where R_0, R_1, R_2, R_3, R_4 are independent pseudo-random numbers, which are uniformly distributed in the interval $[0, 1]$.

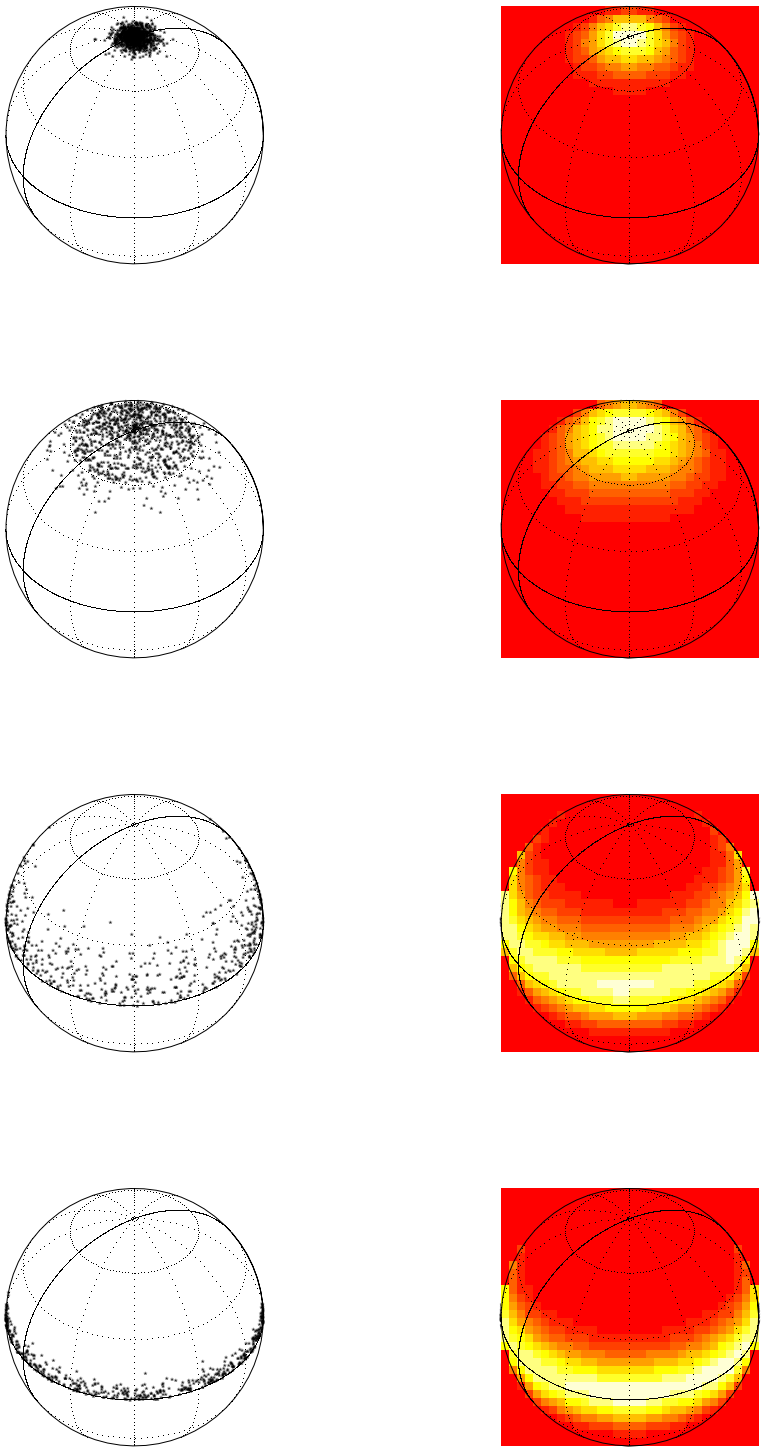


Figure 2.2: Simulation data from the Watson distributions for 1000 points with $\kappa = 20$, $\kappa = 2$, $\kappa = -2$ and $\kappa = -20$ (from top to bottom).

where $S = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T$. After some simple steps by the maximizing $l(\mu, \kappa)$, it leads to the following parameter estimates for the parameter μ (see Fisher et al. (1993))

$$\hat{\mu} = \begin{cases} s_d & \text{if } \hat{\kappa} > 0 \\ s_1 & \text{if } \hat{\kappa} < 0 \end{cases}$$

where $\pm s_1, \cdots, \pm s_d$ are normalised eigenvectors of the matrix S corresponding to the eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_d$. The concentration estimate $\hat{\kappa}$ is obtained by solving

$$\frac{M'(\frac{1}{2}, \frac{d}{2}, \hat{\kappa})}{M(\frac{1}{2}, \frac{d}{2}, \hat{\kappa})} = \hat{\mu}^T S \hat{\mu} =: r \tag{2.5}$$

where

$$M'\left(\frac{1}{2}, \frac{d}{2}, \hat{\kappa}\right) = \frac{\partial M(\frac{1}{2}, \frac{d}{2}, \hat{\kappa})}{\partial \hat{\kappa}}$$

A reasonable approximation to the solution of (2.5) for $d = 3$, which was given by Fisher et al. (1993), is

$$\hat{\kappa} = \begin{cases} 3.75 \times (3s_3 - 1) & 0.333 < s_3 \leq 0.36 \\ 3.34 \times (3s_3 - 1) & 0.36 < s_3 \leq 0.65 \\ 0.7 + 1/(1 - s_3) & 0.65 < s_3 \leq 0.99 \\ 1/(1 - s_3) & s_3 > 0.99 \end{cases}$$

for bipolar cases and

$$\hat{\kappa} = \begin{cases} -1/(2s_1) & s_1 \leq 0.06 \\ -(0.961 - 7.08s_1 + 0.466/s_1) & 0.06 < s_1 \leq 0.333 \end{cases}$$

for girdle cases.

Sra and Karp (2011) solved (2.5) using a root-finding method (e.g. Newton-Raphson). A slightly more general equation is to be considered

$$g(a, c; \kappa) := \frac{M'(a, c, \kappa)}{M(a, c, \kappa)} = r$$

Starting at κ_0, Newton-Raphson solves the equation $g(a, c; \kappa) - r = 0$ by iterating

$$\kappa_{n+1} = \kappa_n - \frac{g(a, c; \kappa_n) - r}{g'(a, c; \kappa_n)} \qquad n = 0, 1, \cdots$$

After several steps of simplify by rewriting $g'(a, c; \kappa)$ we can get

$$g'(a, c; \kappa) = (1 - c/\kappa) g(a, c; \kappa) + (a/\kappa) - (g(a, c; \kappa))^2.$$

Furthermore, Sra and Karp (2011) provided also an approximation of this solution, i.e.

$$\kappa = \frac{cr - a}{r(1 - r)} + \frac{r^2}{2cr(1 - r)}$$

2.3 The angular central Gaussian distribution and the Schladitz distribution

Definition 2.3.1 (angular central Gaussian (ACG) distribution)
A d-dimensional unit random vector x (i.e. $x \in \mathbb{R}^d$ and $\|x\| = 1$) is said to have d-variate angular central Gaussian distribution if its density function is given by

$$f(x;\Sigma) = \alpha_d^{-1}[det(\Sigma)]^{-1/2}(x^T\Sigma^{-1}x)^{-d/2} \tag{2.6}$$

where $\alpha_d = 2\pi^{d/2}/\Gamma(d/2)$ is the surface area of the d-dimensional unit sphere S_d, and especially if $d = 3$, $\alpha_3 = 4\pi$, and Σ is a $d \times d$ symmetric positive-definite matrix.

Remark 2.3.2
The ACG density has a number of attractive properties, which are discussed by Tyler (1987)

(i) The ACG density is unique only up to factor of Σ, and there are various methods of normalization, e.g. $tr(\Sigma^{-1}\Sigma) = d$ described by Tyler (1987) or $\det\Sigma = 1$.

(ii) It is simple to transform between the Gaussian distributions and the ACG distributions. If W is a sample from a $d-$dimensional Gaussian $\mathcal{N}_d(0,\Sigma)$ distributed, then

$$X = \left\{ X_i : X_i = \frac{w_i}{\|w_i\|}, w_i \in W \right\}$$

is a sample from the density (2.6). If X is a sample from the d-dimensional $ACG(\Sigma)$ distributed, then

$$W = \left\{ w_i : w_i = \Sigma^{-1/2}x/(x^{-1}\Sigma^{-1}x)^{1/2}, x_i \in X \right\}$$

is a sample from a Gaussian distribution $\mathcal{N}_d(0,\Sigma)$.

(iii) The ACG distribution has the geometric interpretation of a uniform distribution on the ellipse $\left\{X : x^T\Sigma^{-1}x = 1\right\}$.

Definition 2.3.3 (the Schladitz distribution)
A three-dimensional unit random vector x (i.e. $x \in \mathbb{R}^d and \|x\| = 1$) with altitude $\theta \in [0,\pi)$ and longitude $\phi \in [0, 2\pi)$ in polar coordinates is said to have a Schladitz distribution if its probability density function is given by

$$f(\theta,\phi) = \frac{1}{4\pi}\frac{\beta\sin\theta}{(1+(\beta^2-1)\cos^2\theta)^{3/2}}, \tag{2.7}$$

Remark 2.3.4
The density can be interpreted as the compression of a uniform distribution on the unit sphere by a factor β which serves as an anisotropy parameter for the model. Depending on the value of β both axial and girdle distributions with rotational symmetry around the z-axis $(0,0,1)^T$ can be realised: $\beta = 1$ describes the isotropic case, i.e., a uniform distribution on the sphere, $\beta < 1$ describes a bipolar axial model with the principal axis $(0,0,1)^T$, and $\beta > 1$ describes a girdle model with the principal plane in the x-y plane.

Theorem 2.3.5
The Schladitz distribution is a special case of the ACG distribution with $d = 3$ *and* 3×3 *diagonal matrix* $\Sigma = diag(1, 1, \beta^{-2})$.

Proof It follows directly from replacing $d = 3$, $\Sigma = diag(1, 1, \beta^{-2})$ in (2.6) and the translation of the vector x from the direction cosines to the polar coordinates (θ, ϕ). □

In the following we write A for the Schladitz distribution instead of Σ.

The density described in (2.7) applies only for the case that the sample is rotationally symmetric about the z-axis. We discuss here the density function of the Schladitz distribution with variant principal axis.

Theorem 2.3.6
Let $\mu = (\theta_0, \phi_0)$ *be a three-dimensional unit vector as the principal axis, then the density function of the corresponding Schladitz distribution is given by*

$$p(\theta, \phi) = \frac{1}{4\pi} \frac{\beta \sin\theta}{(1 + (\beta^2 - 1)(\sin\theta_0 \sin\theta \cos(\phi_0 - \phi) + \cos\theta_0 \cos\theta)^2)^{3/2}}$$

Proof We include the rotation as part of the model parameters, the density function with rotation can be shown that

$$f(x; \beta, \theta_0, \phi_0) = f(x; A)$$

taking $A = Q^T_{\theta_0,\phi_0} \mathrm{diag}(1, 1, \beta^{-2}) Q_{\theta_0,\phi_0}$, where the rotation matrix Q_{θ_0,ϕ_0} is given by

$$Q_{\theta_0,\phi_0} = \begin{pmatrix} \cos\theta_0 \cos\phi_0 & \cos\theta_0 \sin\phi_0 & -\sin\theta_0 \\ -\sin\phi_0 & \cos\phi_0 & 0 \\ \sin\theta_0 \cos\phi_0 & \sin\theta_0 \sin\phi_0 & \cos\theta_0 \end{pmatrix}$$

and

$$det(A) = det(Q^T_{\theta_0,\phi_0}) det(\mathrm{diag}(1, 1, \beta^{-2})) det(Q_{\theta_0,\phi_0}) = \beta^{-2}$$

Therefore

$$\begin{aligned} p(\theta, \phi) &= \frac{1}{4\pi} [det(A)]^{-1/2} (x^T (Q^T_{\theta_0,\phi_0} \mathrm{diag}(1, 1, \beta^{-2}) Q_{\theta_0,\phi_0})^{-1} x)^{-3/2} \\ &= \frac{\beta}{4\pi} \left(\left(\left(Q^{-1}_{\theta_0,\phi_0}\right)^T x \right)^T \mathrm{diag}(1, 1, \beta^2) \left(\left(Q^{-1}_{\theta_0,\phi_0}\right)^T x \right) \right)^{-3/2} \\ &= \frac{1}{4\pi} \frac{\beta \sin\theta}{(1 + (\beta^2 - 1)(\sin\theta_0 \sin\theta \cos(\phi_0 - \phi) + \cos\theta_0 \cos\theta)^2)^{3/2}} \end{aligned}$$

□

This theorem can also be proved under the use of the theorem of transformation for densities.

Proof Since $\det Q_{\theta_0,\phi_0} = 1$, Q_{θ_0,ϕ_0} is invertible and it yields

$$Q^{-1}_{\theta_0,\phi_0} = \begin{pmatrix} \cos\theta_0 \cos\phi_0 & -\sin\theta_0 & \sin\theta_0 \cos\phi_0 \\ \cos\theta_0 \sin\phi_0 & \cos\phi_0 & \sin\theta_0 \sin\phi_0 \\ -\sin\phi_0 & 0 & \cos\theta_0 \end{pmatrix} = Q^T_{\theta_0,\phi_0}$$

Let (θ', ϕ') be a three-dimensional Schladitz distribution with principal axis z-axis, and (θ, ϕ) comes from (θ', ϕ') with the rotation to the principal axis μ, then we have

$$\begin{aligned}
\begin{pmatrix} \sin\theta'\cos\phi' \\ \sin\theta'\sin\phi' \\ \cos\theta' \end{pmatrix} &= \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} \\
&= Q_{\theta_0,\phi_0} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} \\
&= \begin{pmatrix} \cos\theta_0\cos\phi_0 x + \cos\theta_0\sin\phi_0 y - \sin\theta_0 z \\ -\sin\theta_0 x + \cos\theta_0 y \\ \sin\theta_0\cos\phi_0 x + \sin\theta_0\sin\phi_0 y + \cos\theta_0 z \end{pmatrix}
\end{aligned}$$

Therefore

$$\begin{aligned}
\theta' &= \arccos(z') \\
&= \arccos(\sin\theta_0\cos\phi_0\sin\theta\cos\phi + \sin\theta_0\sin\phi_0\sin\theta\sin\phi + \cos\theta_0\cos\theta) \\
&= g_1(\theta, \phi) \\
\phi' &= \arctan(y'/x') \\
&= \arctan\left(\frac{-\sin\theta_0\sin\theta\cos\phi + \cos\theta_0\sin\theta\sin\phi}{\cos\theta_0\cos\phi_0\sin\theta\cos\phi + \cos\theta_0\sin\phi_0\sin\theta\sin\phi - \sin\theta_0\cos\theta}\right) \\
&= g_2(\theta, \phi)
\end{aligned}$$

Then

$$\begin{aligned}
& p(\theta, \phi) \\
= \; & \frac{1}{2}\frac{\beta\sin\theta}{(1 + (\beta^2 - 1)(\sin\theta_0\cos\phi_0\sin\theta\cos\phi + \sin\theta_0\sin\phi_0\sin\theta\sin\phi + \cos\theta_0\cos\theta)^2)^{3/2}} \\
& \cdot \left|\det\left(\frac{\partial(g_1(\theta,\phi), g_2(\theta,\phi))}{\partial(\theta,\phi)}\right)\right| \\
= \; & \frac{1}{2}\frac{\beta\sin\theta}{(1 + (\beta^2 - 1)(\sin\theta_\mu\sin\theta\cos(\phi_0 - \phi) + \cos\theta_0\cos\theta)^2)^{3/2}} \\
& \cdot \left|\det\left(\frac{\partial(g_1(\theta,\phi), g_2(\theta,\phi))}{\partial(\theta,\phi)}\right)\right|
\end{aligned}$$

Hence, the matrix

$$\frac{\partial(g_1(\theta,\phi), g_2(\theta,\phi))}{\partial(\theta,\phi)}$$

and the matrix $Q^{-1}_{\theta_0,\phi_0}$ describe both the transformation of data from $\mu-$axis to $z-$axis. The single difference is the first matrix with the Cartesian coordinates and the second matrix with the polar coordinates. Therefore

$$\left|\det\left(\frac{\partial(g_1(\theta,\phi), g_2(\theta,\phi))}{\partial(\theta,\phi)}\right)\right| = \left|\det Q^{-1}_{\theta_0,\phi_0}\right| = 1$$

□

For the estimation of the $d-$dimensional matrix Σ we use here the maximum likelihood estimation. Given a random sample $\{x_j, 1 \leq j \leq n\}$ from a distribution having density of (2.6) with unknown Σ, the likelihood function for Σ is proportional to

$$L(\Sigma) = |\Sigma|^{-\frac{n}{2}} \prod_{i=1}^{n} (x_j^T \Sigma^{-1} x_j)^{-\frac{d}{2}}$$

Tyler (1987) proved that a maximum likelihood estimate for Σ exists if $n > d(d-1)$ and that it is unique up to multiplication by a positive scalar. It corresponds to a solution to the equation

$$\hat{\Sigma} = dn^{-1} \sum_{i=1}^{n} \frac{x_j x_j^T}{x_j^T \hat{\Sigma}^{-1} x_j} \tag{2.8}$$

Furthermore it can also be proven that the maximum likelihood estimate for Σ is asymptotically normal. Let $\hat{\Sigma}_0$ represent the solution to (2.8) which is normalized so that $tr(\Sigma^{-1}\hat{\Sigma}_0) = d$, then we have

$$\sqrt{n}(\hat{\Sigma}_0 - \Sigma) \xrightarrow{L} \mathcal{N}(0, C(\Sigma))$$

where

$$C(\Sigma) = (1 + 2d^{-1}) \left\{ (I + K_{d,d})(\Sigma \oplus \Sigma) - 2d^{-1} vec(\Sigma) vec(\Sigma)^T \right\}$$

and $\xrightarrow{L}$ denotes convergence in law. As the notation, for matrix Σ, $tr(\Sigma)$ denotes the trace of Σ, $vec(\Sigma)$ represents the vector formed by stacking the columns of Σ, the product $\oplus$ refers to the Kronecker product, and $K_{d,d}$ is the commutation matrix, which satisfies

$$K_{d,d} vec(\Sigma) = vec(\Sigma^T)$$

Note that the Σ are not unique. We use some standardization, e.g. $det(\Sigma) = 1$ or $tr(\Sigma^{-1}\hat{\Sigma}_0) = d$ as mentioned above to make it unique.

The pseudo code of simulation reads in Algorithm 4 (see Altendorf (2012))

Figure 2.3 shows points sampled from the Schladitz distribution on the unit sphere with different parameter β. 1000 points were simulated under the method of Algorithm 4. The top subfigures show the 1000 points distributed on the unit sphere and the bottom subfigures display the density estimation. The same as by the Watson distribution, we consider here also only the upper half-sphere.

We believe that the Schladitz model is useful for applications since, for instance, many materials are compressed during their production process. Examples of application of the distribution given by (2.7) can be found in Louis et al. (2011); Redenbach and Vecchio (2011); Schladitz et al. (2006).

A maximum likelihood approach for the estimation of parameter β in the Schladitz distribution will be discussed in Chapter 3.

2.4 Differences between the Watson distribution and the Schladitz distribution

With some slight abuse of notation, we write θ for the colatitude of a random direction as well as for the corresponding argument of its probability density.

Algorithm 4 Simulate data from the Schladitz distribution

Require: a natural number $n > 0$, a non-negative real number β and a direction with the polar coordinates (θ_0, ϕ_0)

Ensure: n unit vectors drawn from the Schladitz distribution with the parameter β, rotational symmetry about the axis (θ_0, ϕ_0)

Main-step: Simulation

$\phi = 2 * \pi * R_1$

$\xi = 2 * R_2 - 1$

$\eta = \sqrt{1 - \xi^2}$

if $|\beta - 1| > eps$ **then**

$norm = \sqrt{\xi^2 - \xi^2 * \beta^2 + \beta^2}$

$\xi = \xi / norm$

$\eta = \eta * \beta / norm$

end if

$\theta = \arccos(\xi)$

Post-step: rotate the data with its mean direction

calculate the rotation matrix

$$A = \begin{pmatrix} \cos\theta_0 \cos\phi_0 & \cos\theta_0 \sin\phi_0 & -\sin\theta_0 \\ -\sin\phi_0 & \cos\phi_0 & 0 \\ \sin\theta_0 \cos\phi_0 & \sin\theta_0 \sin\phi_0 & \cos\theta_0 \end{pmatrix}$$

calculate (θ_i', ϕ_i') from

$$\begin{pmatrix} \sin\theta_i' \cos\phi_i' \\ \sin\theta_i' \sin\phi_i' \\ cos\theta_i' \end{pmatrix} = A \cdot \begin{pmatrix} \sin\theta_i \cos\phi_i \\ \sin\theta_i \sin\phi_i \\ cos\theta_i \end{pmatrix}$$

return (θ_i', ϕ_i')

where R_1, R_2 are independent pseudo-random numbers, which are uniformly distributed in the interval $[0, 1]$.

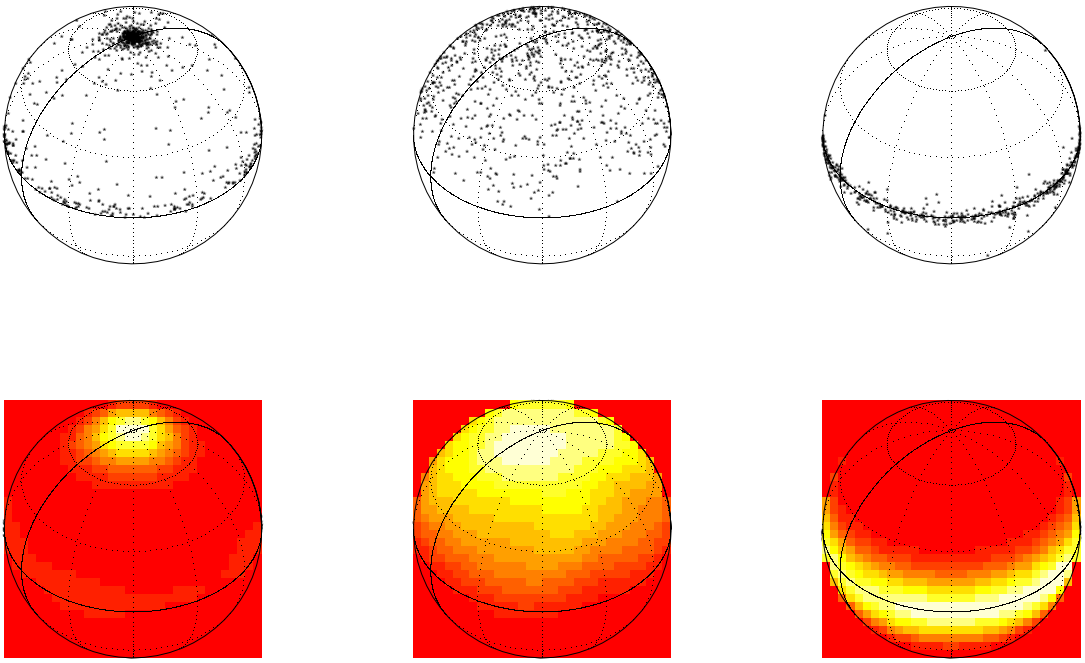


Figure 2.3: Simulation data from the Schladitz distributions for 1000 points with $\beta = 0.1$ (left), $\beta = 0.9$ (middle) and $\beta = 10$ (right).

We discuss here the differences between the Watson distribution and the Schladitz distribution. The idea is to find the parameter κ of the Watson distribution and the parameter β of the Schladitz distribution with the same expected value of θ, optimally to obtain some parameter-relationships between the two distributions, and then to compare the variances of θ of the two distributions with the same expected value of θ. Mark that both the Watson and the Schladitz distribution with principal axis $(0, 0, 1)^T$ are rotationally symmetric, i.e. the density in polar coordinates depends only on θ and not on ϕ.

For the Watson distribution with principal/polar axis $(0, 0, 1)^T$, the marginal distribution of ϕ is uniform and the marginal distribution of θ is given by

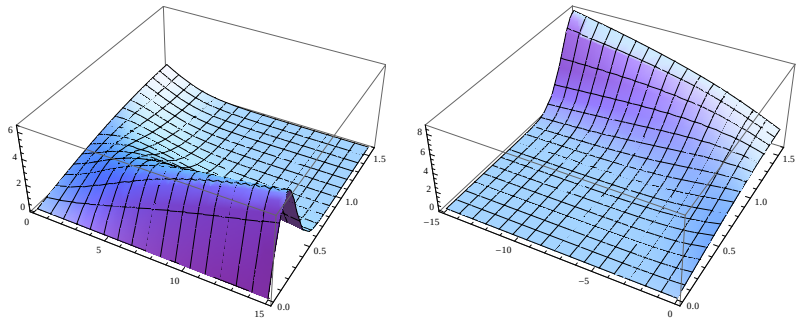
$$f_W(\theta) = \frac{1}{M\left(\frac{1}{2}, \frac{3}{2}, \kappa\right)} \exp(\kappa \cos^2 \theta) \sin \theta \tag{2.9}$$

for $0 \leq \theta < \pi/2$. These results follow at once from (2.4). We rewrite $M(a, b, z)$ as ${}_1F_1(a, b, z)$, where ${}_pF_q$ describes the whole family of hypergeometric functions with

$${}_pF_q(a_1, \cdots, a_p; b_1, \cdots, b_q; k) = \sum_{z=0}^{\infty} \prod_{i=1}^{p} \frac{\Gamma(z + a_i)}{\Gamma(a_i)} \prod_{j=1}^{q} \frac{\Gamma(b_j)}{\Gamma(z + b_j)} \frac{k^z}{z!}$$

for $p, q \in \mathbb{N}_0$.

Figure 2.4 represents the marginal density function of θ in 3D for bipolar distributions and girdle distributions, respectively.



(a) bipolar cases (b) girdle cases

Figure 2.4: The marginal density function of θ of Watson distribution in 3D

Table 2.1: Expected value $E_{f_W(\theta)}$ of Watson distribution for some given κ from -10000 to 1000

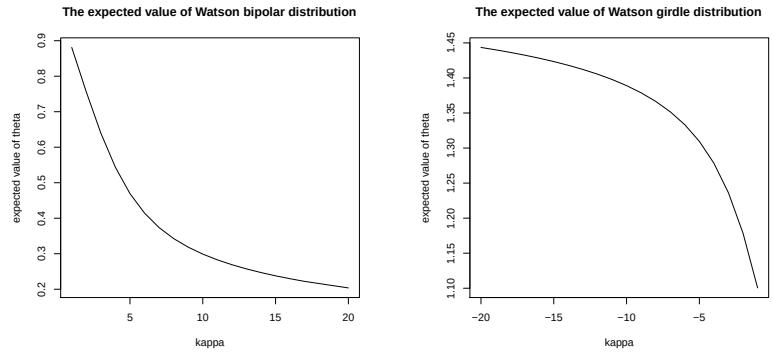
κ	-10000	-1000	-100	-50	-20
$E_{f_W(\theta)}$	1.56515	1.55295	1.51428	1.49074	1.44354
κ	-15	-14	-13	-12	-11
$E_{f_W(\theta)}$	1.42339	1.41808	1.41215	1.40546	1.39785
κ	-10	-9	-8	-7	-6
$E_{f_W(\theta)}$	1.38908	1.37882	1.36661	1.35176	1.33326
κ	-5	-4	-3	-2	-1
$E_{f_W(\theta)}$	1.30955	1.2783	1.23608	1.17839	1.10047
κ	1	2	3	4	5
$E_{f_W(\theta)}$	0.881031	0.755741	0.639628	0.543377	0.469393
κ	6	7	8	9	10
$E_{f_W(\theta)}$	0.414441	0.373631	0.342678	0.318482	0.298975
κ	11	12	13	14	15
$E_{f_W(\theta)}$	0.282815	0.269121	0.257305	0.246961	0.237797
κ	20	50	100	1000	10000
$E_{f_W(\theta)}$	0.203721	0.126638	0.089075	0.028039	0.00886271

The expected value of θ is given by

$$\begin{aligned} E_{f_W(\theta)} &= \int_0^{\pi/2} \frac{\theta}{M\left(\frac{1}{2}, \frac{3}{2}, \kappa\right)} \exp(\kappa \cos^2 \theta) \sin \theta d\theta \\ &= \frac{{}_2F_2\left(\frac{1}{2}, 1; \frac{3}{2}, \frac{3}{2}; \kappa\right)}{{}_1F_1(0.5, 1.5, \kappa)} \end{aligned} \tag{2.10}$$

For each given κ, (2.10) can be simplified and numerically calculated. Table 2.1 shows the expected value for some given κ from -10000 to 10000. Furthermore Figure 2.5 describes the expected value function graphically for the Watson distribution.

Turning now to the Schladitz distribution with principal axis $(0, 0, 1)^T$, the marginal distribu-



(a) bipolar cases (b) girdle cases

Figure 2.5: Expected value of θ of Watson distribution for parameter $0 < \kappa \leq 20$ of bipolar cases and parameter $-20 \leq \kappa < 0$ of girdle cases, respectively

tion of θ is given by

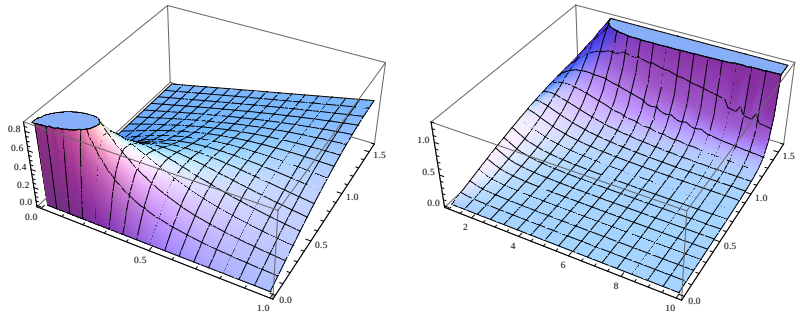
$$f(\theta) = \frac{1}{2} \frac{\beta \sin \theta}{(1 + (\beta^2 - 1) \cos^2 \theta)^{3/2}}$$

for $0 \leq \theta < \pi$. This result follows at once from (2.7). Since the standard Schladitz distribution is symmetric, we can rewrite the marginal distribution of θ, for consideration only by the upper half sphere, as

$$f_S(\theta) = \frac{\beta \sin \theta}{(1 + (\beta^2 - 1) \cos^2 \theta)^{3/2}}$$

for $0 \leq \theta < \pi/2$.

Figure 2.6 represents the density function of the Schladitz distribution in 3D.



(a) bipolar cases (b) girdle cases

Figure 2.6: Density function of Schladitz distribution in 3D

The corresponding expected value for all θ from 0 to $\pi/2$ is then given by

$$\begin{aligned} E_{f_S(\theta)} &= \int_0^{\pi/2} \theta \cdot \frac{\beta \sin \theta}{(1 + (\beta^2 - 1) \cos^2 \theta)^{3/2}} d\theta \\ &= \frac{\beta \sec^{-1}(\beta)}{\sqrt{|1 - \beta^2|}} \end{aligned}$$

Table 2.2: Expected value $E_{f_S(\theta)}$ of the Schladitz distribution for some given β from 0.01 to 100

β	0.01	0.02	0.03	0.04	0.05
$E_{f_S(\theta)}$	0.0529856	0.0921198	0.126041	0.15659	0.184644
β	0.06	0.07	0.08	0.09	0.1
$E_{f_S(\theta)}$	0.210719	0.235159	0.258209	0.280052	0.30083
β	0.2	0.3	0.4	0.5	0.6
$E_{f_S(\theta)}$	0.467941	0.589289	0.683807	0.760346	0.823959
β	0.7	0.8	0.9	1	2
$E_{f_S(\theta)}$	0.877852	0.924196	0.964534	1	1.2092
β	3	4	5	6	7
$E_{f_S(\theta)}$	1.30563	1.36134	1.39768	1.42325	1.44224
β	8	9	10	11	12
$E_{f_S(\theta)}$	1.4569	1.46855	1.47804	1.48591	1.49256
β	13	14	15	50	100
$E_{f_S(\theta)}$	1.49824	1.50315	1.50743	1.55111	1.56087

where $\sec(\beta) = \frac{1}{\cos(\beta)}$.

Table 2.2 shows the expected value of the Schladitz distribution for some given β from 0.01 to 100. Furthermore Figure 2.7 describes it graphically.

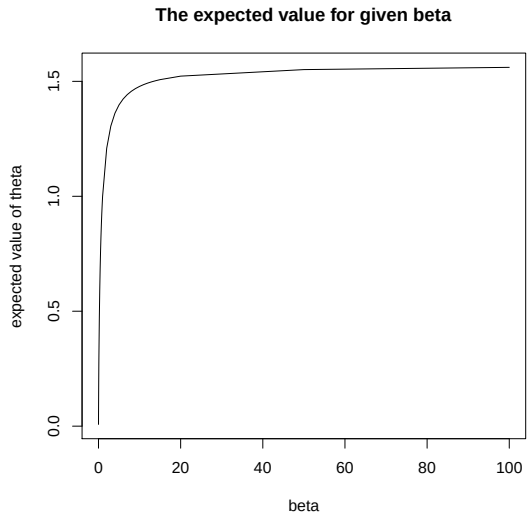


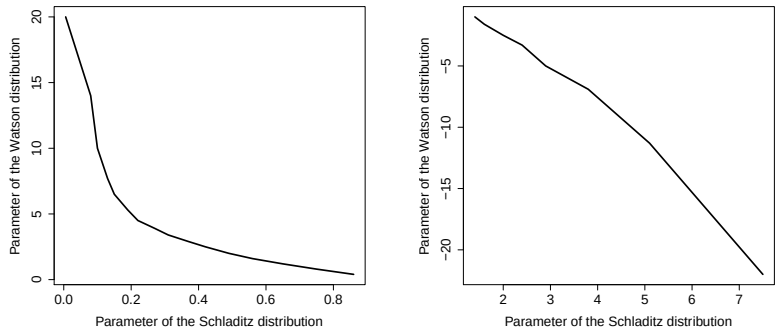
Figure 2.7: Expected value of θ of the Schladitz distribution for parameter $0 < \beta < 100$

To compare the Watson distribution and the Schladitz distribution, we need to find the parameter κ of the Watson distribution and the parameter β of the Schladitz distribution for each given expected value of θ, i.e to solve κ and β with

$$E_{f_W(\theta)} = E_{f_S(\theta)} \tag{2.11}$$

Table 2.3 shows the solutions of (2.11) as comparison of the parameter of the Watson distribution and the Schladitz distribution for some given expected values.

Furthermore Figure 2.8 shows the relationship between the parameter of the Watson distribution and the parameter of the Schladitz distribution with the same expected values of θ.



(a) bipolar cases (b) girdle cases

Figure 2.8: Plots of the parameter of the Watson distribution and the parameter of the Schladitz bipolar distribution with the same expected values of θ

We consider first the bipolar cases. It can be observed from Figure 2.8 that for not so small expected values of θ ($E_{f_W(\theta)} = E_{f_S(\theta)} \geq 0.5$), namely $\beta \geq 0.22$ and $0 < \kappa \leq 4.5$, which is mainly used by our applications, it is possible to fit with a polynomial curve the relationship between the parameter of the Watson distribution and the parameter of the Schladitz distribution.

We assume that there exist a $n-$degree polynomial $p(\beta)$ approximated $\kappa(\beta)$ that

$$\kappa(\beta) \approx p(\beta) = p_1\beta^n + p_2\beta^{n-1} + \cdots p_n\beta + p_{n+1}$$

The question is how to determine the degree n and the corresponding coefficients $p_1, p_2, \cdots, p_{n+1}$. Obviously we obtain better results for larger n, but for applications we prefer smaller n. Therefore the smallest n to make sure that the polynomial fit shows as acceptable gap between the original curve and the approximated curve will be chosen.

We use here the method of linear least squares to find the best-fitting curve $p(\beta)$. As the input data, a set of points (show in Table 2.3 from $E_{f_W(\theta)}/E_{f_S(\theta)} = 0.5$ to $E_{f_W(\theta)}/E_{f_S(\theta)} = 1.5$ except $E_{f_W(\theta)}/E_{f_S(\theta)} = 1$) is given and as (β_i, κ_i) $(1 \leq i \leq 18)$ marked. Linear least squares fitting proceeds by finding the curve to minimize the sum of the squares err^2 of a set of n data points (here: $n = 18$), which err^2 defined as

$$err^2 = \sum_{i=1}^{n} (\kappa_i - p(\beta_i))^2$$

Table 2.4 shows the parameters for some given n and the value of corresponding err. Furthermore Figure 2.9 represents the original $\kappa(\beta)$ (black curve) and the approximated polynomial $p(\beta)$ (grey curve) graphically for degree n between 1 and 4. It can be observed not only from err of Table 2.4, but also from Figure 2.9 that there exist no obviously differences between the original function $\kappa(\beta)$ and the approximating polynomial $p(\beta)$ for $n = 3$, therefore we get

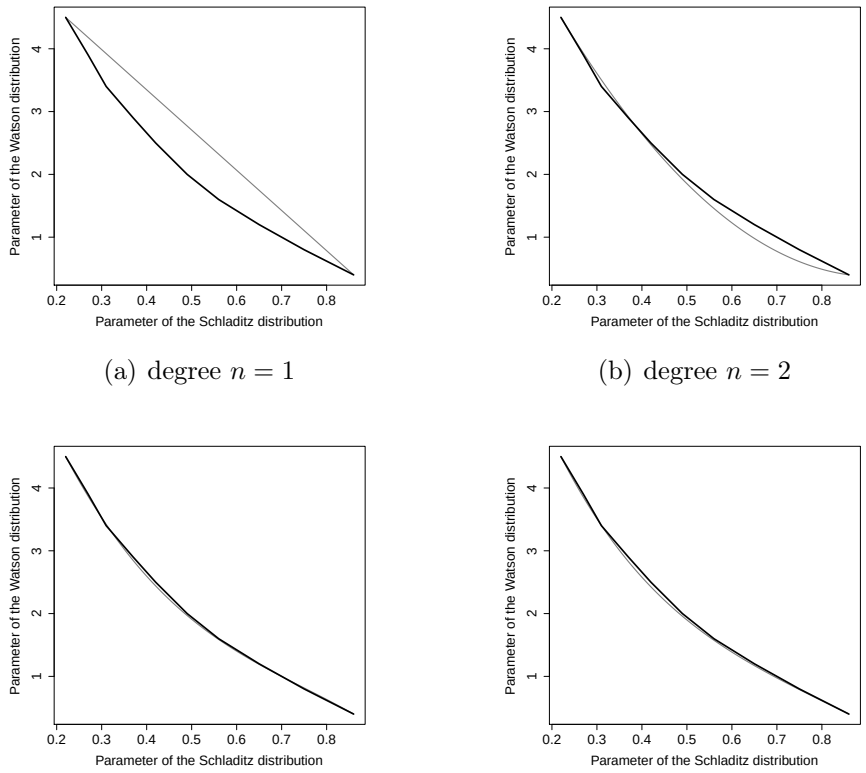
$$p_b(\beta) = -9.9272 \times \beta^3 + 23.2292 \times \beta^2 - 21.7868 \times \beta + 8.2679$$

Table 2.3: The comparison of the parameter of the Watson distribution (κ) and the Schladitz distribution (β) for some given expected values

$E_{f_W(\theta)}/E_{f_S(\theta)}$	κ	β
0.1	>1000	0.02
0.2	20	0.06
0.25	14	0.08
0.3	10	0.1
0.35	7.7	0.13
0.4	6.5	0.15
0.45	5.3	0.19
0.5	4.5	0.22
0.55	3.9	0.27
0.6	3.4	0.31
0.65	2.9	0.37
0.7	2.5	0.42
0.75	2	0.49
0.8	1.6	0.56
0.85	1.2	0.65
0.9	0.8	0.75
0.95	0.4	0.86
1	-	1
1.1	-1	1.4
1.15	-1.6	1.6
1.2	-2.5	2
1.25	-3.3	2.4
1.3	-5	2.9
1.35	-6.9	3.8
1.4	-11.3	5.1
1.45	-22	7.5
1.5	<-50	13.5

Table 2.4: Parameters of polynomial approximation the relationship of κ and β of bipolar distribution

n	p_1	p_2	p_3	p_4	p_5	p_6	err
1	-6.2451	5.3801					0.8623
2	7.2414	-13.9529	7.1228				0.2051
3	-9.9272	23.2292	-21.7868	8.2679			0.0608
4	4.7670	-20.1569	30.9631	-24.2033	8.5290		0.0589
5	-34.4720	96.5614	-113.4219	75.8966	-34.4133	9.4010	0.0566



(a) degree $n = 1$ (b) degree $n = 2$

(c) degree $n = 3$ (d) degree $n = 4$

Figure 2.9: Plots of the original function $\kappa(\beta)$ and the approximated polynomial $p(\beta)$ for the bipolar cases

for the bipolar cases.

Turning now to the girdle cases, it is quite similar to the bipolar distributions. Table 2.5 shows the parameters of approximated polynomial of given degree n and Figure 2.10 represents the $\kappa(\beta)$ and $p(\beta)$ graphically. It can also observed that the minimal degree of the polynomial $p(\beta)$ approximating well is equal to 2, therefore

$$p_g(\beta) \approx -0.2572 \times \beta^2 - 1.1042 \times \beta + 0.8241$$

for the girdle cases.

Up to here we obtain already the parameter of the Waston distribution and the Schladitz distribution with the same expected value of θ, and this relationship is also polynomially approximated. Then we want to investigate the differences between the two distributions by comparison the variances of θ with the same expected value.

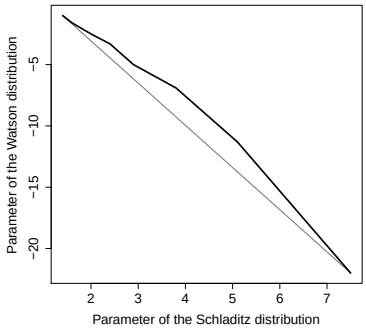
Table 2.6 shows the variances of θ for given κ from Table 2.3 of Watson distribution. For each set of parameter we simulated 1000 directions with given parameter κ and the principal/polar axis $(0, 0, 1)^T$, calculated the variance of θ, this procedure was repeated 100 times to obtain the mean values $\mathrm{Var}_{f_W(\theta)}$. $\overline{\mathrm{Var}}_{f_W(\theta)}$ is the theoretical variance directly from the density function of the Watson distribution

$$\overline{\mathrm{Var}}_{f_W(\theta)} = E\theta^2 - (E_{f_W(\theta)})^2$$

For comparison with the results of the Watson distribution, Table 2.7 shows the corresponding

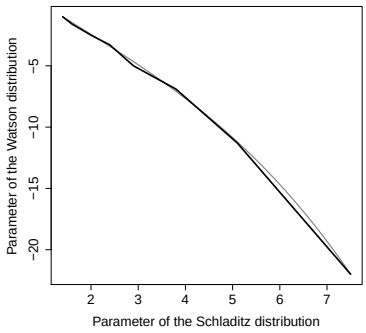
Table 2.5: Parameters of polynomial approximation the relationship of κ and β of girdle distribution

n	p_1	p_2	p_3	p_4	p_5	p_6	err
1	-3.3469	4.4703					2.5894
2	-0.2572	-1.1042	0.8241				0.5901
3	-0.0314	0.1534	-2.6443	2.4346			0.4157
4	0.0094	-0.1891	1.0466	-4.6546	3.9418		0.3969
5	0.0213	-0.4148	2.9426	-9.6490	12.1948	-5.8383	0.3411



(a) degree $n = 1$

(b) degree $n = 2$



(c) degree $n = 3$

(d) degree $n = 4$

Figure 2.10: Plots of the original function $\kappa(\beta)$ and the approximated polynomial $p(\beta)$ for the girdle cases

parameters of the Schladitz distribution, where $\mathrm{Var}_{f_S(\theta)}$ and $\overline{\mathrm{Var}}_{f_S(\theta)}$ denote the experimental and theoretical value of the variance, respectively. $\overline{\mathrm{Var}}_{f_S(\theta)}$ is given by

$$\begin{aligned}\overline{\mathrm{Var}}_{f_S(\theta)} &= E\theta^2 - (E_{f_S(\theta)})^2 \\ &= \int_0^{\pi/2} \frac{\theta^2 \beta \sin\theta}{(1+(\beta^2-1)\cos^2\theta)^{3/2}} d\theta - \left(\frac{\beta \sec^{-1}(\beta)}{\sqrt{|\beta^2-1|}}\right)^2\end{aligned}$$

Table 2.6: The experimental and theoretical variances of θ for given κ from Table 2.3 of the Watson distribution

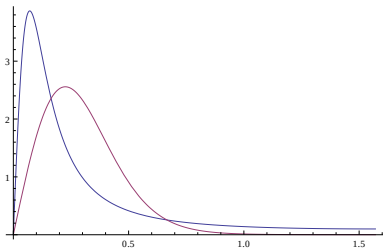
$E_{f_W(\theta)}$	κ	$\mathrm{Var}(\theta)$	$\overline{\mathrm{Var}}_{f_W(\theta)}$
0.1	10000	3.07×10^{-4}	2.1464×10^{-5}
0.2	20	0.0119	0.0118813
0.3	10	0.0277	0.0278665
0.4	6.5	0.0540	0.0546655
0.5	4.5	0.0935	0.0933197
0.6	3.5	0.1210	0.119921
0.7	2.5	0.1440	0.144415
0.8	1.5	0.1578	0.156778
0.9	1	0.1563	0.156069
1.1	-1	0.1169	0.116892
1.2	-2.5	0.0804	0.0800775
1.3	-5	0.0433	0.0433026
1.4	-12	0.0163	0.0163113
1.5	-50	0.0037	0.00369329

$$= \beta \int_0^{\pi/2} \frac{\theta^2 \sin\theta}{(1+(\beta^2-1)\cos^2\theta)^{3/2}} d\theta - \frac{\beta^2(\sec^{-1}(\theta))^2}{|\beta^2-1|}$$

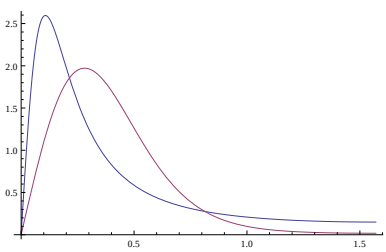
Figure 2.11 and 2.13 show the plots of the density functions of the Watson distribution (blue curves) and the Schladitz distribution (red curves) with the same expected values of bipolar distributions and girdle distributions, respectively. For the girdle distributions and bipolar distributions with not so concentrated cases ($E_{f_W(\theta)} = E_{f_S(\theta)} \geq 0.7$), there is just slight differences between the two distributions. Otherwise for the cases that $E_{f_W(\theta)} = E_{f_S(\theta)} < 0.7$, the density function of the Watson distribution is more concentrated as to comparisons with the density function of the Schladitz distribution. This is also consistent with the results of Table 2.6 and Table 2.7, i.e. the density function of the Watson distributions has smaller variance than the Schladitz distributions for the given same expected value of θ. Variance of the Schladitz distribution always large than that of the Watson distribution. Difference, however, pretty small for medium-sized values of the expected value of θ, but more prominent for small and large values. Figure 2.4 shows the plot of relative scale $\sqrt{\frac{\overline{\mathrm{Var}}_{f_S(\theta)}}{\overline{\mathrm{Var}}_{f_W(\theta)}}}$ against $Ef_S(\theta) = Ef_W(\theta)$.

Table 2.7: The experimental and theoretical variances for given β from Table 2.3 of the Schladitz distribution

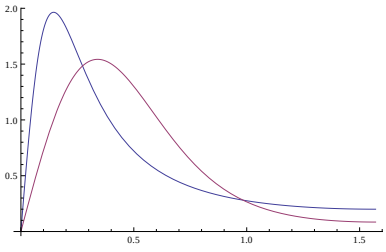
$E_{fS(\theta)}$	β	$\mathrm{Var}(\theta)$	$\overline{\mathrm{Var}}_{fS(\theta)}$
0.1	0.02	0.0347	0.0342804
0.2	0.06	0.0784	0.0794357
0.3	0.1	0.1106	0.108972
0.4	0.15	0.1337	0.133414
0.5	0.2	0.1495	0.149154
0.6	0.3	0.1648	0.165424
0.7	0.4	0.1699	0.170456
0.8	0.55	0.1690	0.168103
0.9	0.75	0.1580	0.15758
1.0	1	0.1411	0.141593
1.1	1.5	0.1125	0.112716
1.2	2	0.0919	0.090867
1.3	3	0.0628	0.0626816
1.4	5	0.0361	0.0354929
1.5	13.5	0.0090	0.00923502



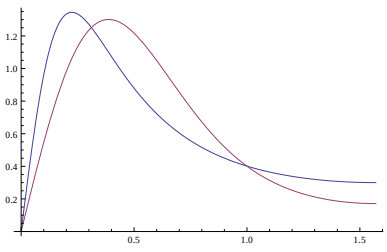
(a) $E_{f_W(\theta)} = E_{f_S(\theta)} = 0.3$



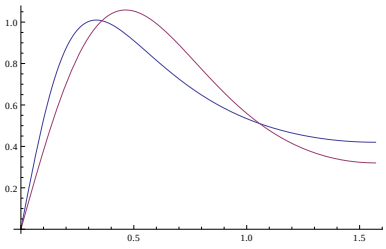
(b) $E_{f_W(\theta)} = E_{f_S(\theta)} = 0.4$



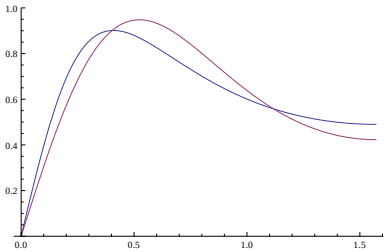
(c) $E_{f_W(\theta)} = E_{f_S(\theta)} = 0.5$



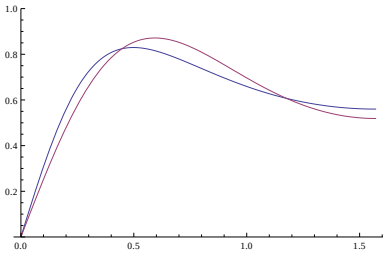
(d) $E_{f_W(\theta)} = E_{f_S(\theta)} = 0.6$



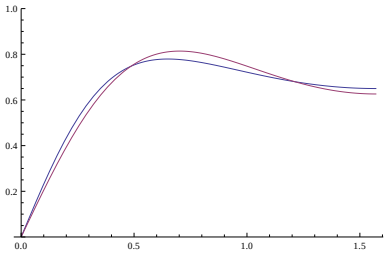
(e) $E_{f_W(\theta)} = E_{f_S(\theta)} = 0.7$



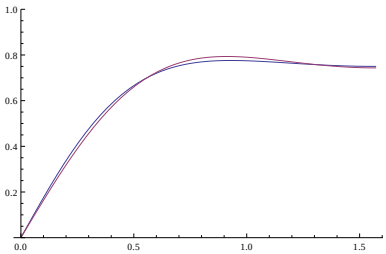
(f) $E_{f_W(\theta)} = E_{f_S(\theta)} = 0.75$



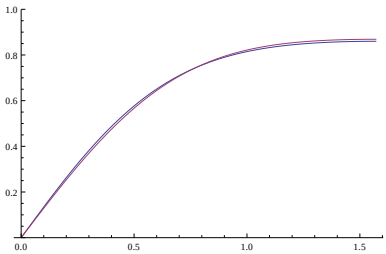
(g) $E_{f_W(\theta)} = E_{f_S(\theta)} = 0.8$



(h) $E_{f_W(\theta)} = E_{f_S(\theta)} = 0.85$



(i) $E_{f_W(\theta)} = E_{f_S(\theta)} = 0.9$



(j) $E_{f_W(\theta)} = E_{f_S(\theta)} = 0.95$

Figure 2.11: Plots of the density functions of the Watson bipolar distribution (blue curves) and the Schladitz bipolar distribution (red curves) with the same given expected values for bipolar cases

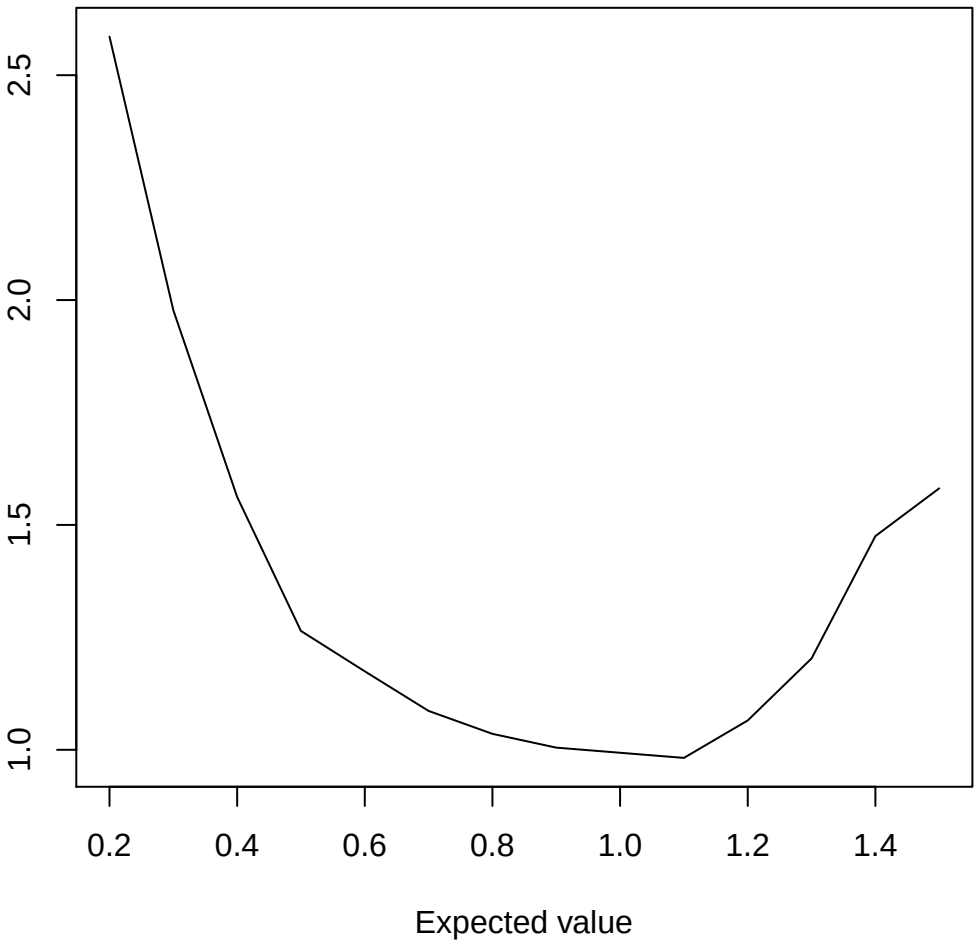
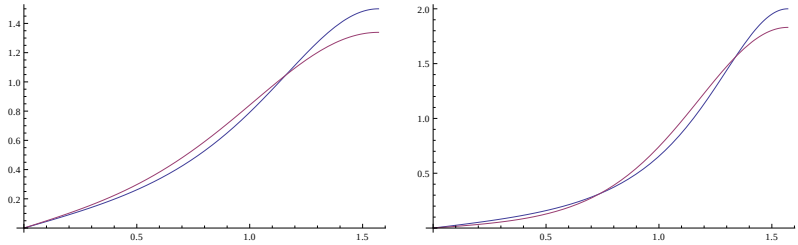
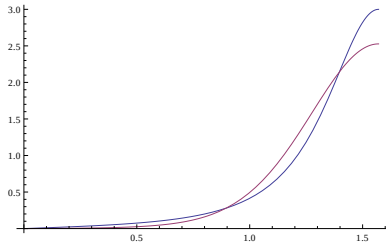


Figure 2.12: Plot of relative scale $\sqrt{\frac{\overline{\mathrm{Var}}_{f_S(\theta)}}{\overline{\mathrm{Var}}_{f_W(\theta)}}}$ against $Ef_S(\theta) = Ef_W(\theta)$

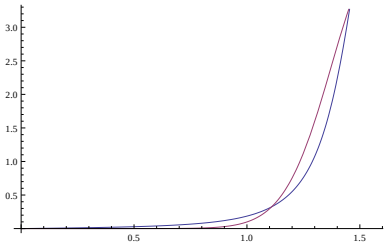


(a) $E_{f_W(\theta)} = E_{f_S(\theta)} = 1.1$

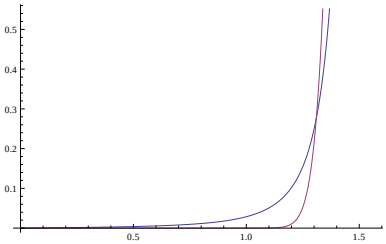
(b) $E_{f_W(\theta)} = E_{f_S(\theta)} = 1.2$



(c) $E_{f_W(\theta)} = E_{f_S(\theta)} = 1.3$



(d) $E_{f_W(\theta)} = E_{f_S(\theta)} = 1.4$



(e) $E_{f_W(\theta)} = E_{f_S(\theta)} = 1.5$

Figure 2.13: Plots of the density functions of the Watson girdle distribution (blue curves) and the Schladitz girdle distribution (red curves) with the same given expected values for girdle cases

Chapter 3

Maximum-likelihood Estimation of the Parameter of the Schladitz Distribution

In this chapter, we discuss maximum-likelihood estimation of the parameter β and derive properties of the resulting estimator. The methods are evaluated using simulated data and are applied to fibre direction distributions estimated from three-dimensional image data. Finally, we discuss statistical tools to evaluate the goodness-of-fit of the parametric model.

3.1 Maximum-likelihood estimation of β

Let $(\theta_1, \phi_1), (\theta_2, \phi_2), \cdots, (\theta_n, \phi_n)$ be a sample of independent and identically distributed unit vectors with density (2.7) given in polar coordinates. Owing to the rotational symmetry of f around the z-axis, the estimation of β requires only the sample $\theta_1, \ldots, \theta_n$. In practice, the assumption of rotational symmetry has to be justified, e.g., by testing the hypothesis that the sample $\phi_1, \ldots, \phi_n$ is drawn from a uniform distribution on $[0, 2\pi)$.

Now we restrict attention to the sample $\theta_1, \theta_2, \cdots, \theta_n$. We will write $p(\theta \,|\, \beta)$ for the density of θ w.r.t. a given β, i.e.,

$$p(\theta \,|\, \beta) := \frac{1}{2} \frac{\beta \sin\theta}{(1 + (\beta^2 - 1)\cos^2\theta)^{3/2}}.$$

With $\Theta = (\theta_1, \ldots, \theta_n)$, the log-likelihood function is given by

$$\begin{aligned} L(\beta \,|\, \Theta) &= \ln(\prod_{i=1}^{n} p(\theta_i \,|\, \beta)) \\ &= n \ln\left(\frac{1}{2}\right) + n \ln(\beta) + \sum_{i=1}^{n} \left(\ln(\sin\theta_i) - \frac{3}{2} \ln(1 + (\beta^2 - 1)\cos^2\theta_i) \right) \end{aligned}$$

For $\beta \to 0$ we have $L(\beta \,|\, \Theta) \to -\infty$, so this case does not cause any problems in the maximum likelihood estimation. In practice, we may assume that β is contained in a compact interval $[a, b]$ such that the log-likelihood function for a given sample is bounded.

To compute the extreme values of $L(\beta \,|\, \Theta)$, we compute the derivative w.r.t. β.

$$\frac{\partial L(\beta \,|\, \Theta)}{\partial \beta} = \frac{n}{\beta} - 3\beta \sum_{i=1}^{n} \frac{\cos^2\theta_i}{1 + (\beta^2 - 1)\cos^2\theta_i} \tag{3.1}$$

Since we were not able to compute the roots of (3.1) analytically, we will use a numerical procedure based on Newton's method in the following.

For $\beta > 0$ we can define $g(\beta) := -\beta\, \partial L/\partial\beta$ which makes the first term independent of β. Finding the roots of (3.1) is equivalent to finding the roots of g. It turns out that g has a unique root in a suitable bounded interval. This result is summarized in the following Lemma 3.1.1. As a consequence, we may assume that the maximum-likelihood estimator $\hat{\beta}_n$ of β_0 exists.

Lemma 3.1.1
With probability 1, g is increasing and continuous in $[0, \infty)$,$g(0) = \lim_{\beta\to 0} g(\beta) = -n < 0$ and $g(\beta) > 0$ for any $\beta > \beta^* = (T_{(n)}/2)^{1/2}$, where $T_{(n)} = \max\{T_1, \cdots, T_n\}$ with $T_i = \tan^2\theta_i$, $i = 1, \cdots, n$. In particular, with probability 1, $g(\beta_0) = 0$ for some unique $\beta_0 \in (0, \beta^*]$.

Proof With probability 1 we have $\cos^2\theta_i \notin \{0, 1\}$, $i = 1, \cdots, n$. Hence, g is a continuous function with $g(0) = -n$ and $g'(\beta) > 0$ for any $\beta > 0$, i.e. g is strictly increasing. Define

$$X_i(\beta) = \beta^2 \frac{\cos^2\theta_i}{\sin^2\theta_i + \beta^2\cos^2\theta_i} = \frac{1}{1 + \frac{1}{\beta^2}T_i}$$

with $T_i = \tan^2\theta_i$. With $T_{(n)} = \max\{T_1, \cdots, T_n\}$ we obtain

$$X_i(\beta) \geq \frac{1}{1 + \frac{1}{\beta^2}T_{(n)}}, \quad i = 1, \cdots, n$$

Hence

$$g(\beta) = 3\sum_{i=1}^{n} X_i(\beta) - n \geq 3n\left(\frac{1}{1 + \frac{1}{\beta^2}T_{(n)}} - \frac{1}{3}\right) > 0$$

for $\beta > (T_{(n)}/2)^{1/2}$. Consequently, there exists a unique root of g which is contained in $(0, \beta^*]$. □

Using Newton's method we can approximate the roots of (3.1) iteratively by

$$\beta_{t+1} = \beta_t - \frac{g(\beta_t)}{g'(\beta_t)} \tag{3.2}$$

where the derivative $g'(\beta)$ of $g(\beta)$ is

$$g'(\beta) = \sum_{i=1}^{n} \frac{3/2\beta\sin^2(2\theta_i)}{(1 + (\beta^2 - 1)\cos^2\theta_i)^2}$$

Given an initial value β_{init} for β this methods allows to find a root of (3.1) in finite time. Since g has only a single root, the solution obtained by Newton's method should be insensitive to the choice of β_{init}.

The maximum-likelihood estimator $\hat{\beta}_n$ is consistent and asymptotically normal. These properties are summarized in the following theorems.

Theorem 3.1.2 (Consistency of the maximum-likelihood estimator)
Let $(\theta_1, \phi_1), \cdots, (\theta_n, \phi_n)$ be independent and identically distributed with density (2.7) with parameter β_0, where $0 < a \leq \beta_0 \leq b < \infty$. Then $\hat{\beta}_n$ is consistent, i.e. $\hat{\beta}_n$ converges to β_0 in probability.

Theorem 3.1.3 (Asymptotic normality of the maximum-likelihood estimator)
Under the assumptions of Theorem 3.1.2 we have

$$n^{1/2}(\hat{\beta}_n - \beta_0) \xrightarrow{L} \mathcal{N}(0, \frac{4}{5\beta_0^2}), \quad \text{for } n \to \infty,$$

where $\xrightarrow{L}$ denotes convergence in law.

The consistency and asymptotic normality of the maximum likelihood estimator (Theorems 3.1.2 and 3.1.3) follow from (Serfling, 2001, Chapter 4.2.2) if the following conditions hold.

(i) For each $0 < a \leq \beta \leq b < \infty$, the derivatives

$$\frac{\partial \ln p(\theta \mid \beta)}{\partial \beta}, \frac{\partial^2 \ln p(\theta \mid \beta)}{\partial \beta^2}, \frac{\partial^3 \ln p(\theta \mid \beta)}{\partial \beta^3}$$

exist for all $\theta \in [0, \pi)$.

(ii) For each $0 < a \leq \beta_0 \leq b < \infty$, there exist functions $f(\theta)$, $h(\theta)$ and $H(\theta)$ (possibly depending on β_0) such that for β in a neighborhood $N(\beta_0)$ the relations

$$\left|\frac{\partial p(\theta \mid \beta)}{\partial \beta}\right| \leq f(\theta), \left|\frac{\partial^2 p(\theta \mid \beta)}{\partial \beta^2}\right| \leq h(\theta), \left|\frac{\partial^3 p(\theta \mid \beta)}{\partial \beta^3}\right| \leq H(\theta)$$

hold for all $\theta \in [0, \pi)$, and

$$\int_0^\pi f(\theta) d\theta < \infty, \quad \int_0^\pi h(\theta) d\theta < \infty, \quad E_\beta H(\theta) < \infty \quad \text{for } \beta \in N(\beta_0). \tag{3.3}$$

(iii) For each $0 < a \leq \beta \leq b < \infty$, we have $0 < I(\mathcal{P}_\beta) < \infty$, where

$$I(\mathcal{P}_\beta) = E_\beta \left(\left(\frac{\partial \ln p(\theta \mid \beta)}{\partial \beta} \right)^2 \right)$$

is the Fisher information.

In this case, the variance of the limiting normal distribution is $I(\mathcal{P}_{\beta_0})$.

Proof (i) The partial derivatives of the log-density function are given by

$$\begin{aligned}
\frac{\partial \ln p(\theta \mid \beta)}{\partial \beta} &= \frac{1}{\beta} - \frac{3\beta \cos^2 \theta}{1 + (\beta^2 - 1)\cos^2 \theta} \\
\frac{\partial^2 \ln p(\theta \mid \beta)}{\partial \beta^2} &= -\frac{1}{\beta^2} - \frac{3\cos^2 \theta(\sin^2 \theta - \beta^2 \cos^2 \theta)}{(1 + (\beta^2 - 1)\cos^2 \theta)^2} \\
\frac{\partial^3 \ln p(\theta \mid \beta)}{\partial \beta^3} &= \frac{2}{\beta^3} - \frac{6\beta \cos^4 \theta(\beta^2 \cos^2 \theta - 3\sin^2 \theta)}{(1 + (\beta^2 - 1)\cos^2 \theta)^3}
\end{aligned}$$

Since $1 + (\beta^2 - 1)\cos^2\theta = \sin^2\theta + \beta^2\cos^2\theta$ is positive for all $\theta \in [0, \pi)$, the derivatives exist for all θ. In the following we write $A(\theta, \beta) = 1 + (\beta^2 - 1)\cos^2\theta$ for simplicity.

(ii) We have

$$\begin{aligned}
\left|\frac{\partial p(\theta \mid \beta)}{\partial \beta}\right| &= \frac{1}{2}\left|\frac{\sin\theta}{A(\theta,\beta)^{3/2}} - \frac{3\beta^2 \sin\theta\cos^2\theta}{A(\theta,\beta)^{5/2}}\right| \\
&\leq \frac{1}{2}\frac{\sin\theta}{A(\theta,\beta)^{3/2}}\left(1 + 3\beta^2\frac{\cos^2\theta}{A(\theta,\beta)}\right) \\
&\leq \frac{1}{2}\frac{\sin\theta}{A(\theta,\beta)^{3/2}}\left(1 + 3\beta^2\frac{\cos^2\theta}{\beta^2\cos^2\theta}\right) \\
&= \frac{2\sin\theta}{A(\theta,\beta)^{3/2}} = \frac{2\sin\theta}{A(\theta,\beta)^{3/2}} \cdot (\sin^2\theta + \cos^2\theta) \\
&= \left(\frac{2\sin^3\theta}{A(\theta,\beta)^{3/2}} + 2\sin\theta\cos^2\theta\frac{1}{A(\theta,\beta)^{1/2}}\frac{1}{A(\theta,\beta)}\right) \\
&\leq \left(\frac{2\sin^3\theta}{(\sin^2\theta)^{3/2}} + 2\sin\theta\cos^2\theta\frac{1}{(\sin^2\theta)^{1/2}}\frac{1}{(\beta^2\cos^2\theta)}\right) \\
&= 2 + \frac{2}{\beta^2} \leq 2 + \frac{2}{a^2} = f(\theta)
\end{aligned}$$

By similar computations we get

$$\left|\frac{\partial^2 p(\theta \mid \beta)}{\partial \beta^2}\right| \leq \frac{12}{a} + \frac{12}{a^3} = h(\theta)$$

and

$$\left|\frac{\partial^3 p(\theta \mid \beta)}{\partial \beta^3}\right| \leq \frac{102}{a^2} + \frac{102}{a^4} = H(\theta)$$

Obviously, (3.3) is fulfilled with this choice of f, h and H.

(iii) It can be shown that

$$\begin{aligned}
I(\mathcal{P}_{\beta_0}) &= E_{\beta_0}\left(\left(\frac{\partial \ln p(\theta \mid \beta_0)}{\partial \beta}\right)^2\right) \\
&= \int_0^\pi \left(\frac{1 - (1 + 2\beta_0^2)\cos^2\theta}{\beta_0 + \beta_0(\beta_0^2 - 1)\cos^2\theta}\right)^2 p(\theta \mid \beta_0) d\theta \\
&= \int_0^\pi \frac{\left(-1 + (1 + 2\beta_0^2)\cos^2\theta\right)^2 \sin\theta}{2\beta_0\left(1 + (\beta_0^2 - 1)\cos^2\theta\right)^{7/2}} d\theta.
\end{aligned}$$

Evaluation of this integral yields $I(\mathcal{P}_{\beta_0}) = 4/5\beta_0^2$. Hence, $0 < I(\mathcal{P}_{\beta_0}) < \infty$ and the variance of the normal distribution is obtained.

□

Table 3.1: Means and standard deviations of $\hat{\beta}_n$ for various β_0.

	β_0	0.1	0.5	0.9	1	2	5
n=1000	mean	0.1001	0.4992	0.9011	1.0012	2.0054	5.0031
	sd	0.0034	0.0183	0.0311	0.0361	0.0713	0.1699
n=100	mean	0.1018	0.5073	0.8865	1.0188	1.9757	5.0624
	sd	0.0121	0.0541	0.0932	0.1084	0.4566	0.5861
n=10	mean	0.1100	0.5615	0.9900	1.0808	2.1115	5.2449
	sd	0.0386	0.2072	0.3498	0.3803	0.7665	1.9413

3.2 Evaluation of the numerical estimator

For small β_0, a good initial value β_{init} is obtained by setting $\beta = 0$ in the denominator of the sum in (3.1) which yields

$$\beta_{\text{init}} = \left(\frac{n}{3 \sum_{i=1}^{n} \cot^2 \theta_i} \right)^{\frac{1}{2}} \tag{3.4}$$

This β_{init} is suitable for the case of a bipolar distribution, but it turned out that it also yields good results for the girdle case, which means in the case $\beta_0 > 1$.

In order to evaluate our estimation procedure we chose several values of β_0. For each β_0 we simulated a series of unit vectors (θ_1, ϕ_1), (θ_2, ϕ_2), $\ldots$, (θ_n, ϕ_n), and computed β_{init} using (3.4). The application of Newton's method with this initial value then yields an estimate for β_0. The procedure is stopped as soon as $|\beta_{t+1} - \beta_t| < \epsilon$ for a given $\epsilon > 0$. Here, we have used the machine epsilon with the value of $\epsilon = 1.11 \times 10^{-16}$ on 64-bit machines. Table 3.1 shows the comparison of the results for various given β_0 and different sample sizes. For each set of parameters, the procedure was repeated 100 times to get the mean value and standard deviation of $\hat{\beta}_n$. The results indicate that our method works well for large sample sizes (such as $n = 1000$ and $n = 100$). However, even for the small sample size $n = 10$ the estimated means do not differ too much from the true parameter. As expected, the root of (3.1) is actually not dependent on the choice of the initial value of (3.2).

The methods described above apply only for the case that the sample is rotationally symmetric about the z-axis. In applications, different symmetry axes may occur. The axis of symmetry can be estimated using the methods described in Fisher et al. (1993):

Let $(\theta_1, \phi_1), (\theta_2, \phi_2), \ldots, (\theta_n, \phi_n)$ denote the given directions with $\theta_i \in [0, \pi), \phi_i \in [0, 2\pi)$. Using the direction cosines (x_i, y_i, z_i) given by

$$x_i = \sin(\theta_i)\cos(\phi_i), \quad y_i = \sin(\theta_i)\sin(\phi_i), \quad z_i = \cos(\theta_i)$$

we define the orientation matrix

$$T = \begin{pmatrix} \sum x_i^2 & \sum x_i y_i & \sum x_i z_i \\ \sum x_i y_i & \sum y_i^2 & \sum y_i z_i \\ \sum x_i z_i & \sum y_i z_i & \sum z_i^2 \end{pmatrix} \tag{3.5}$$

In the bipolar case, an estimate of the principal axis is the eigenvector $\hat{u} = (x_0, y_0, z_0)^T$ corresponding to the largest eigenvalue of the matrix T. In the girdle case, the polar axis,

the direction perpendicular to the principal plane, can be estimated using the eigenvector $\hat{u} = (x_0, y_0, z_0)^T$ corresponding to the smallest eigenvalue of T.

In both cases, a sample $(\theta'_1, \phi'_1), (\theta'_2, \phi'_2), \ldots, (\theta'_n, \phi'_n)$ with rotational symmetry about the z-axis is obtained by a suitable rotation yielding the transformed direction cosines

$$\begin{pmatrix} x'_i \\ y'_i \\ z'_i \end{pmatrix} = \begin{pmatrix} \cos\theta_0 \cos\phi_0 & \cos\theta_0 \sin\phi_0 & -\sin\theta_0 \\ -\sin\phi_0 & \cos\phi_0 & 0 \\ \sin\theta_0 \cos\phi_0 & \sin\theta_0 \sin\phi_0 & \cos\theta_0 \end{pmatrix} \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} \tag{3.6}$$

where $\theta_0 = \arccos(z_0)$ and $\phi_0 = \arctan(y_0/x_0)$.

The corresponding pseudo code can be found as Algorithm 5.

From this sample the value of β can be estimated using the method introduced above.

3.3 Applications

In the following, we will present an application of our method to the statistical analysis of the fibre direction distribution in fibre composites. The direction distribution of the fibres is estimated from a three-dimensional image of the material obtained by micro computer tomography (μCT). A desirable approach would be to determine the direction of each single fibre contained in the image. Unfortunately, in most cases the resolution which can be obtained by μCT imaging is not sufficient to allow for a separation of single fibres. As an alternative, the local fibre direction in each fibre voxel can be estimated using the methods introduced in Altendorf and Jeulin (2009); Eberly et al. (1994); Wirjadi (2009). Since fibre directions in neighbouring voxels are obviously not independent, the assumptions for the derivation of the maximum likelihood estimator $\hat{\beta}_n$ are not satisfied.

Therefore, we start by investigating the error introduced by the dependence within a sample on the estimation of β_0. For this purpose, random systems consisting of n non-overlapping cylinders were simulated using a random sequential adsorption procedure Redenbach and Vecchio (2011). Fibre directions are drawn according to the density (2.7). The cylinders were discretised into three-dimensional images of size 256^3 voxels. For the length and radius of the fibres the values 50 and 2 voxels were chosen. From the images, the fibre direction in each voxel was estimated using the method based on the Hessian matrix which is presented in Wirjadi (2009) and which is implemented in the MAVI software package ITWM (2012).

Finally, β was estimated on the resulting (dependent) sample of directions. The results for several choices of the parameter β are summarised in Table 3.2. Except for $\beta_0 = 5$ a slight overestimation of β_0 is observed. Again, the results for $n = 100$ and $n = 1000$, which are comparable in most cases, are better than for $n = 10$. For the extreme cases $\beta_0 = 0.1$ and $\beta_0 = 5$ the estimated parameters are biased towards the uniform distribution ($\beta_0 = 1$). Experience shows that this is a typical effect of the image analysis method applied.

As an example of application, we examine the method for three different samples of fibre materials. The first sample is a silica gel provided by the Fraunhofer ISC. It consists of slightly bent fibres which are almost parallel to the x-y plane (see Figure 3.1). The analysis of this material is based on a cylindrical sample imaged by μCT in a $1023 \times 1023 \times 228$ voxel image with a voxel edge length of 10.31 μm. The second and third sample are glass fibre composites provided by the Institut für Verbundwerkstoffe GmbH in Kaiserslautern.

Algorithm 5 Estimation of the Parameter of Beta-distribution

Require: A set of n unit vectors with the polar coordinates (θ_i, ϕ_i), as well as the direction cosines (x_i, y_i, z_i) for $1 \leq i \leq n$, the type of the distribution

Ensure: Principal axis of bipolar distributions or polar axis of girdle distributions, as well as the estimated β

Pre-step: rotate the data with its principle/polar axis

calculate the orientation matrix

$$T = \begin{pmatrix} \sum x_i^2 & \sum x_i y_i & \sum x_i z_i \\ \sum x_i y_i & \sum y_i^2 & \sum y_i z_i \\ \sum x_i z_i & \sum y_i z_i & \sum z_i^2 \end{pmatrix}$$

if bipolar distribution **then**

principal axis = the eigenvector corresponding to the largest eigenvalue of T

else if girdle distribution **then**

polar axis = the eigenvector corresponding to the smallest eigenvalue of T

end if

rotate the original data with the principle axis for the bipolar distribution or with the polar axis for the girdle distribution and get (x_i', y_i', z_i') as well as (θ_i', ϕ_i') for $1 \leq i \leq n$

Main-step: Newton's Method

initial β_0

repeat

$$\begin{aligned} \beta_1 &= \beta_0 \\ g(\beta_1) &= 3\beta_1^2 \sum_{i=1}^{n} \frac{\cos^2 \theta_i'}{1 + (\beta_1^2 - 1)\cos^2 \theta_i'} - n \\ g'(\beta_1) &= \sum_{i=1}^{n} \frac{3/2\beta_1 \sin^2(2\theta_i')}{(1 + (\beta_1^2 - 1)\cos^2 \theta_i')^2} \\ \beta_0 &= \beta_1 - \frac{g(\beta_1)}{g'(\beta_1)} \end{aligned}$$

until $|\beta_1 - \beta_0| < \epsilon$

return β_0, principle/polar axis

Table 3.2: Estimation using image data. Means and standard deviations of $\hat{\beta}_n$ for various β_0.

β_0		0.1	0.5	0.9	1	2	5
n=1000	mean	0.1506	0.5781	0.8810	1.0323	2.0381	2.7879
	sd	0.0123	0.0531	0.0286	0.0490	0.0501	0.0777
n=100	mean	0.1447	0.5619	0.9215	1.0378	2.1820	2.9026
	sd	0.0143	0.0187	0.0404	0.1033	0.1291	0.2883
n=10	mean	0.2312	0.5576	0.9834	1.4728	2.8853	2.4812
	sd	0.0495	0.1068	0.3763	0.4925	0.7418	0.7429

Table 3.3: p-values and test statistic for varied materials of Kolmogorov-Smirnov Test to test the uniform distribution of ϕ_i

material	silica gel	GF30	GF60 bottom	GF60 middle	GF60 top
p-value	0.5428	0.7095	0.9300	0.7108	0.7520
test statistic	0.8009	0.7011	0.5427	0.7003	0.6752

Sample GF 30 consists of fibres which are almost parallel to the y-axis (see Figure 3.2). It was analyzed using a μCT image consisting of 710 $\times$ 1000 $\times$ 1000 voxels with a voxel edge length of 1 μm. Sample GF60 consists of three layers of fibres with different preferred directions (see Figure 3.2). For this sample, an image with 760 $\times$ 1000 $\times$ 1000 voxels and a voxel edge length of 3 μm was used.

For the analysis, ten cubes of size 128^3 voxels were selected from each image. For GF60, ten cubes for each layer were considered. Due to the small thickness of the middle layer of GF60, the size of the cubes was reduced to 60^3 voxels in this case. The local fibre directions for each subvolume were computed using the MAVI software package. From the resulting sample of unit vectors, the principal axis (or polar axis) was estimated and the samples were rotated according to (3.6). Then the parameter β was estimated from the transformed direction data using the maximum-likelihood estimator presented above. The results are summarised in Table 3.4. The estimated values are both reasonable and relatively stable on the different samples for each material. For the middle layer of GF60, the standard deviation of the estimate is comparably high. This is due to the small sample sizes which were analysed in this case.

The uniform distribution of ϕ_i' (i.e. the spherical coordinates after transformation 3.6 w.r.t. the estimated symmetry axis) can be verified using the Kolmogorov-Smirnov test. The p-values and the test statistics for the different materials can be found in Table 3.3. In all cases, the null hypothesis cannot be rejected i.e. the ϕ_i' samples do not show a significant deviation from the uniform distribution on $[0, 2\pi)$.

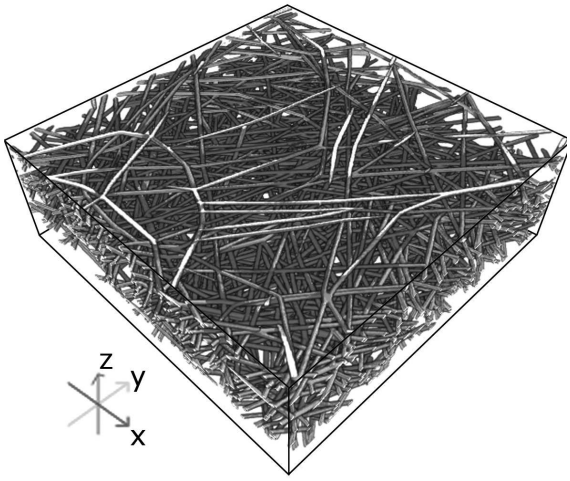


Figure 3.1: Volume rendering of the silica gel fibres. Visualised are 600 × 600 × 200 voxels. Image: Fraunhofer ITWM

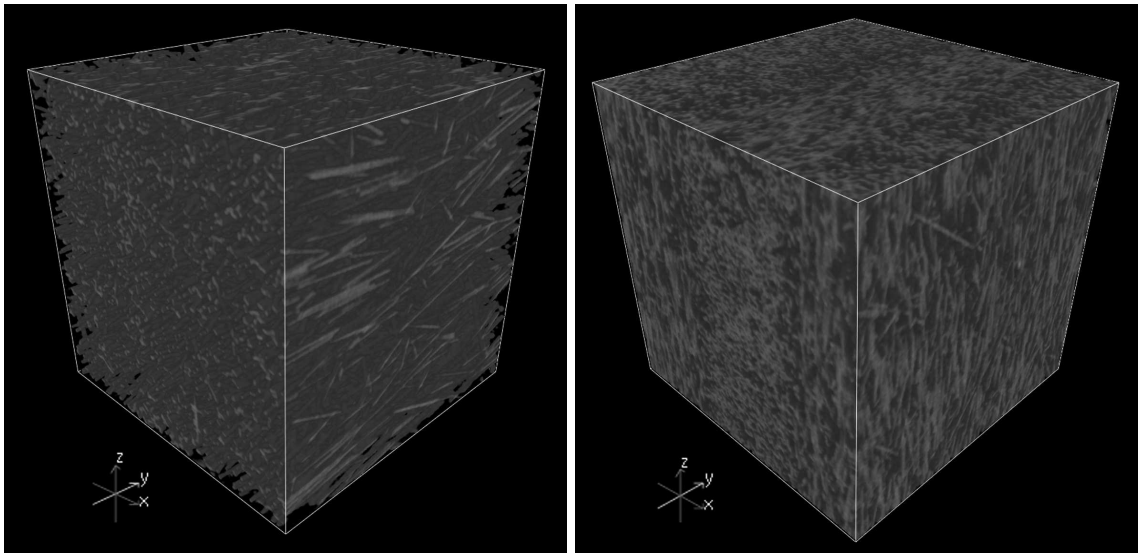


Figure 3.2: Volume rendering of the glass fibre composites: GF30 (400^3 voxels, left) and GF60 (400^3 voxels, right). Image: Institut für Verbundwerkstoffe GmbH, Kaiserslautern

3.4 Goodness-of-fit

For testing the goodness-of-fit of a spherical distribution model, Fisher et al. (1993) propose the use of QQ-plots which consist of a colatitude plot and a longitude plot. The longitude plot is based on the rotated sample $\phi'_1, \ldots, \phi'_n$ according to (3.6). The plot is obtained by plotting $x_i = \phi'_i/2\pi$ against $y_i = (i - 0.5)/n$ and shows if the data is rotationally symmetric about the main axis.

The colatitude plot is obtained by plotting the sample $\theta'_1, \ldots, \theta'_n$ obtained also via (3.6) against $y_i = F^{-1}((i - 0.5)/n)$, where F^{-1} is the inverse function of the distribution function

Table 3.4: Estimation of β for the fibre materials.

$\hat{\beta}$	silica gel	GF30	GF60 bottom	GF60 middle	GF60 top
mean	2.5698	0.3762	0.4121	0.6455	0.4083
min	2.4950	0.3422	0.3970	0.4739	0.3921
max	2.6940	0.4156	0.4372	0.7894	0.4392
sd	0.0670	0.0241	0.0133	0.1288	0.0152

of θ. If θ is distributed with density (2.7) then

$$F^{-1}(x) = \arccos\left(\left(1 - \frac{\beta^2 x(4x-4)}{-1+4x-4\beta^2 x-4\beta^2+4\beta^2 x^2}\right)^{\frac{1}{2}}\right)$$

Figure 3.3 and Figure 3.4 show the longitude and the colatitude plots of the silica gel and glass fibres, respectively.

For comparison, we also fitted a Watson distribution to the observed direction distributions. The estimated values of κ, obtained as the means of the estimates for the ten samples, for the materials are $\hat{\kappa} = -3.92$ for the silica gel, $\hat{\kappa} = 2.60$ for GF30, and $\hat{\kappa} = 2.35$, 1.83 and 2.32 for the bottom, middle and top layer of GF60, respectively.

Since the Watson distribution is rotationally symmetric around μ, the longitude plots coincide with those for the Schladitz distribution. As indicated by the Komogorov-Simirnov test, no significant deviation from the uniform distribution can be detected. The colatitude plots obtained for the Watson distribution are also show in Figure 3.4. The fibre direction distributions in both GF30 and the top and bottom layers of GF60 are more closely fitted by the Schladitz distribution than by the Watson distribution. Only in the middle layer of GF60 the Watson distribution yields the better fit. For the silica gel fibres none of the two distribution families in clearly superior to the other.

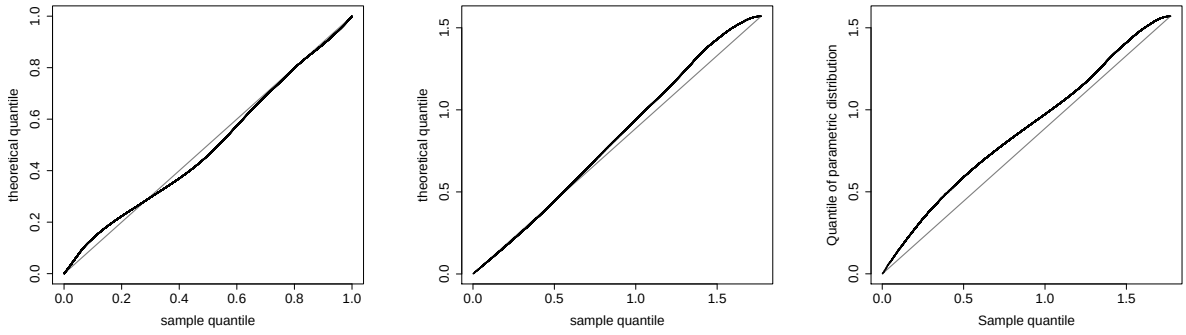


Figure 3.3: QQ-Plots of the silica gel fibres. Left: longitude plot, Middle: colatitude plot for the Schladitz distribution, Right: colatitude plot for the Watson distribution

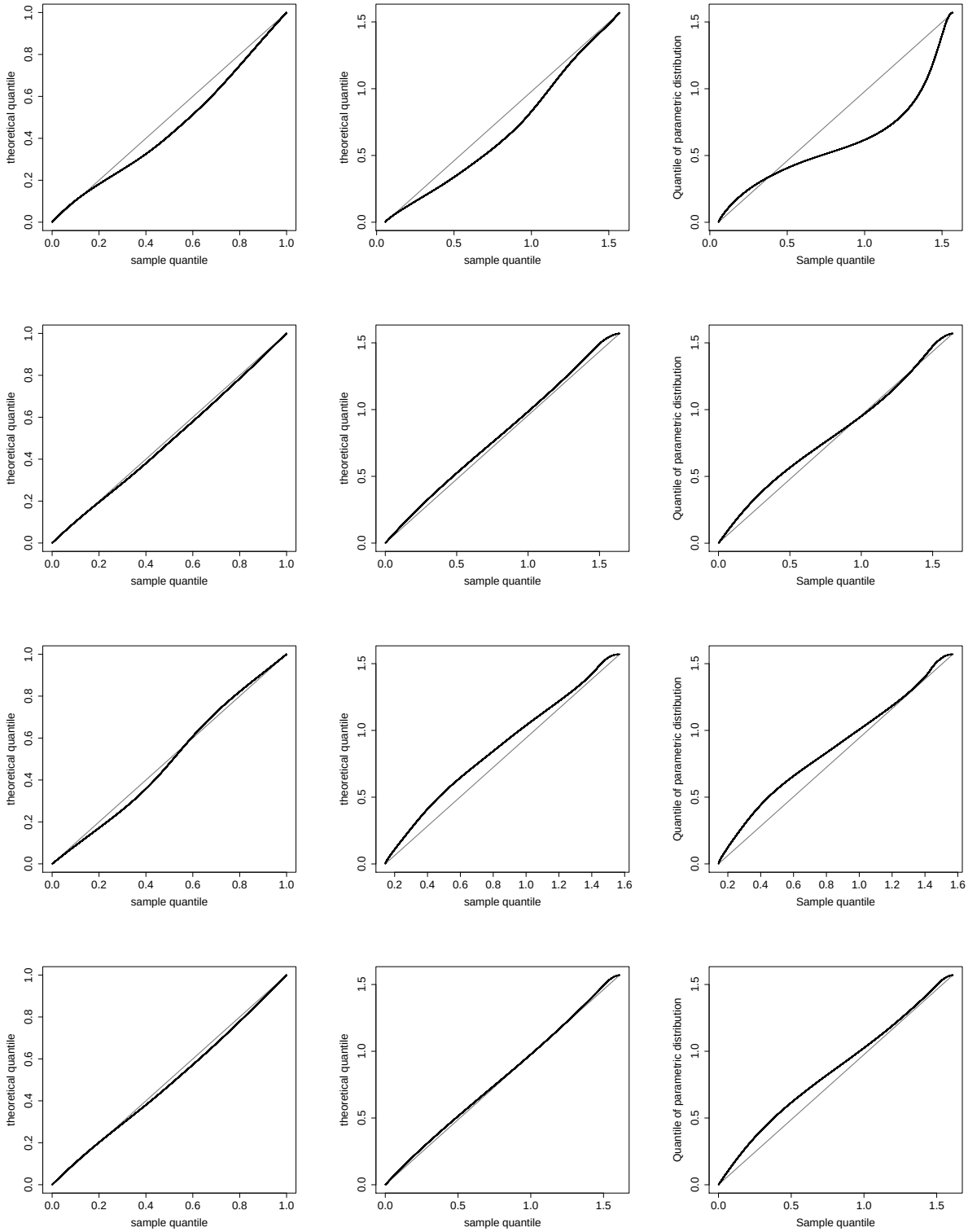


Figure 3.4: QQ-Plots of the glass fibres: GF30, GF60 bottom, GF60 middle, GF60 top (from top to bottom). Left: longitude plot, Middle: colatitude plot for the Schladitz distribution, Right: colatitude plot for the Watson distribution

3.5 Discussion

Our numerical procedure which is based on Newton's method was evaluated on simulated data as well as on fibre systems discretized in 3D images. The application to the modelling of the fibre direction distribution in composite materials shows that this distribution is well suited for applications. In particular, the direction distribution in most samples is fitted more closely by our distribution than by the Watson distribution. The question whether the use of methods from robust statistics can provide improved results when estimating β from image data, is subject to future research.

Chapter 4

Mixtures of Schladitz Distributions

In this chapter, we present a general model given by mixtures of Schladitz distributions. A maximum-likelihood method including the numerical approach of an expectation maximization setting is used. The consistency of the maximum likelihood estimate, as well as the consistency of the expectation maximization estimate are proven. A non-parametric estimation of the mixture model is also discussed.

4.1 Introduction

In nature, some materials of fibres do not show one dominant direction, but the distributions seem to be multimodal. In such cases, distributions which are mixtures of simple unimodal distributions, provide better fits. We use a mixture of several distributions, where each distribution has its own principal/polar axis and parameters. Some mixture models for the distributions on the unit sphere have already been discussed, e.g. Banerjee et al. (2006) proposed a generative mixture-model approach to clustering directional data based on the von Mises-Fisher distribution, Figueiredo and Gomes (2006) provided a good identification of a mixture of bipolar Watson distribution and Peel et al. (2001) described a model and a identification w.r.t. a mixture of Kent distribution. For the compressed materials, the Schladitz distribution shows better fitting as other distributions. Hence the clustering or identification of a mixture of the Schladitz distribution is of interest.

4.2 Mixtures of ACG Distributions

Since the Schladitz distribution is a special case of the ACG distribution, we consider here a general model for directional data as a mixture of m ACG distributions. Let x_j $(1 \leq j \leq n)$ be the angular central Gaussian distribution component with the parameter Σ_k if $Q_j = k$, where $Q_1, \cdots, Q_n$ independent and identically distributed with values in $\{1, \cdots, m\}$. We define the mixture proportions π_k $(1 \leq k \leq m)$ as $\pi_k = \mathbb{P}(Q_j = k)$. Therefore $\Phi = (\pi_1, \cdots, \pi_m, \Sigma_1, \cdots, \Sigma_m)$ is the set of unknown parameters we need to estimate. Mark that π_m is only used as an abbreviation for $1 - \pi_1 - \pi_2 - \cdots - \pi_{m-1}$. The corresponding density for x_j of this mixture model is then given by

$$g(x; \Phi) = \sum_{k=1}^{m} \pi_k f(x; \Sigma_k)$$

where $f(x, \Sigma)$ denotes the ACG density given by (2.6).

We define indicator variables $S_{jk} \in \{0, 1\}$ such that

$$S_{jk} = \begin{cases} 1 & \text{if and only if } Q_j = k \\ 0 & \text{otherwise} \end{cases}$$

Then

$$\begin{aligned} E[S_{jk} \,|\, x_j] &= P[S_{jk} = 1 \,|\, x_j] \\ &= P[Q_j = k \,|\, x_j] \\ &= \frac{\pi_k f(x_j; \Sigma_k)}{g(x_j; \Phi)} \\ &= \frac{\pi_k f(x_j; \Sigma_k)}{\sum_{l=1}^{m} \pi_l f(x_j; \Sigma_l)} \end{aligned}$$

We assume that $\Phi = (\pi_k, \Sigma_k) \in \Theta$, where the parameter set Θ satisfies for $k = 1, \cdots, m$

$$0 < \delta \leq \pi_k \leq 1 - \delta < 1 \tag{4.1}$$

for some $\delta > 0$, and

$$0 < \underline{\lambda} \leq \lambda_{\min}(\Sigma_k) \leq \lambda_{\max}(\Sigma_k) \leq \overline{\lambda} < \infty \tag{4.2}$$

where $\lambda_{\min}(\Sigma_k)$ and $\lambda_{\max}(\Sigma_k)$ is the minimal and maximal eigenvalue of Σ_k, respectively.

The log-likelihood function of the complete data of the mixture model is given by

$$\begin{aligned} L_c(\Phi \,|\, Q, X) &= \sum_{j=1}^{n} \ln(\prod_{j=1}^{n} \pi_{Q_j} f(x_j; \Sigma_{Q_j})) \\ &= \sum_{j=1}^{n} \ln \pi_{Q_j} + \sum_{j=1}^{n} \ln f(x_j; \Sigma_{Q_j}) \\ &= \sum_{j=1}^{n} \ln \pi_{Q_j} + \sum_{j=1}^{n} \sum_{k=1}^{m} S_{jk} \ln f(x_j; \Sigma_k) \end{aligned} \tag{4.3}$$

Since $x_1, \cdots, x_n$ are observed, but $Q_1, \cdots, Q_n$ respective S_{jk}, $k = 1, \cdots, m, j = 1, \cdots, n$ are hidden, we replace S_{jk} by

$$S^*_{jk} = E^*[S_{jk} \,|\, x_j] = \frac{\pi^*_k f(x_j; \Sigma^*_k)}{\sum_{l=1}^{m} \pi^*_l f(x_j; \Sigma^*_l)} \tag{4.4}$$

where E^* denotes that we calculate the conditional expectation pretending that

$$\Phi^* = (\pi^*_1, \cdots, \pi^*_m, \Sigma^*_1, \cdots, \Sigma^*_m)$$

is the true parameter of the data. Then, we get the following approximation of (4.3):

$$Q[\Phi \,|\, \Phi^*] = \sum_{j=1}^{n} \sum_{k=1}^{m} S^*_{jk} \ln(\pi_k f(x_j; \Sigma_k)) \tag{4.5}$$

Now we start with some arbitrary initial guess $\Phi^{(0)} = \Phi^*$ and get an update $\Phi^{(1)}$ by maximizing Q w.r.t. Φ. Then, we replace $\Phi^{(0)}$ by $\Phi^{(1)}$ and iterate. The maximization step is made easier by the observation that

$$\begin{aligned} Q[\Phi \mid \Phi^{(i)}] &= \sum_{j=1}^{n}\sum_{k=1}^{m} S_{jk}^{(i)} \ln(\pi_k f(x_j, \Sigma_k)) \\ &= \sum_{j=1}^{n}\sum_{k=1}^{m} S_{jk}^{(i)} \ln \pi_k + \sum_{j=1}^{n}\sum_{k=1}^{m} S_{jk}^{(i)} \ln(f(x_j, \Sigma_k)) \end{aligned}$$

which consists of two unrelated summands which can be separately maximized w.r.t. π_k respective Σ_k. To take into account the constraint $\pi_1 + \cdots + \pi_k = 1$, we introduce the Lagrange multiplier λ and get for the first term

$$L_1 = \sum_{j=1}^{n}\sum_{k=1}^{m} S_{jk}^{(i)} \ln(\pi_k) + \lambda \sum_{k=1}^{m} (\pi_k - 1)$$

Differentiating with reference to each π_k we obtain

$$\sum_{j=1}^{n} \frac{1}{\pi_k} S_{jk}^{(i)} + \lambda = 0$$

or

$$\sum_{j=1}^{n} S_{jk}^{(i)} = -\lambda \pi_k$$

Summing both sides over k, we get that

$$\sum_{k=1}^{m}\sum_{j=1}^{n} S_{jk}^{(i)} = -\lambda \sum_{k=1}^{m} \pi_k$$

and, as $S_{j1}^{(i)} + \cdots S_{jm}^{(i)} = 1, \pi_1 + \cdots + \pi_m = 1$, it results $\lambda = -n$, therefore

$$\pi_k = \frac{1}{n} \sum_{k=1}^{n} S_{jk}^{(i)}$$

The second term decomposes into m sums involving only one ACG parameter matrix Σ_k each, such that we have to maximize

$$\sum_{j=1}^{n} S_{jk}^{(i)} \ln f(x_j; \Sigma_k) = \max_{\Sigma_k}!, \quad k = 1, \cdots, m$$

These expression are similar to the log-likelihood from an i.i.d. ACG sample, but the single data terms $\ln f(x_j; \Sigma_k)$ are weighted by the conditional probability $S_{jk}^{(i)}$ that x_j is generated by the $k-$th regime.

Iterating these considerations, we get the EM algorithm for a mixture of ACG distributions.

E-step: Compute $Q[\Phi \mid \Phi^{(i)}]$ for given $\Phi^{(i)}$ and mixture components $1 \leq k \leq m$ as in (4.5):

$$Q[\Phi \mid \Phi^{(i)}] = \sum_{j=1}^{n} \sum_{k=1}^{m} S_{jk}^{(i)} \ln(\pi_k f(x_j, \Sigma_k))$$

with

$$S_{jk}^{(i)} = \frac{\pi_k^{(i)} f(x_j; \Sigma_k^{(i)})}{\sum_{l=1}^{m} \pi_l^{(i)} f(x_j; \Sigma_l^{(i)})}$$

M-step: Update the parameter estimated $\Phi^{(i+1)}$ with

$$\Phi^{(i+1)} = \underset{\Phi \in \Theta}{\operatorname{argmax}}\, Q[\Phi \mid \Phi^{(i)}]$$

i.e.

(i) calculate $\pi_k^{(i+1)}$ by

$$\pi_k^{(i+1)} = \frac{1}{n} \sum_{j=1}^{n} S_{jk}^{(i)}$$

(ii) calculate $\Sigma^{(i+1)}$ by maximizing

$$\sum_{j=1}^{n} S_{jk}^{(i)} \ln f(x_j; \Sigma_k^{(i)}), \quad k = 1, \cdots, m$$

In the following, we focus on the properties of this procedure, in particular the consistency of the maximum likelihood estimate and the numerical convergence of the EM algorithm towards this estimate. These properties are summarized in Theorem 4.2.1 and Theorem 4.2.2.

Let Φ^* denote the true value of the parameter, and let $\hat{\Phi}$ denote the maximum likelihood estimate of Φ^*. Further let $\Phi^{(i)}$ denote the numerical approximation of $\hat{\Phi}$ after i iterations of the EM algorithm, and let $\hat{\Phi}_{\mathsf{EM}} = \Phi^{(i_f)}$ denote the parameter approximation which we get after running the EM algorithm i_f times until some stopping criterion is satisfied. For the purpose of verifying that $\hat{\Phi}_{\mathsf{EM}}$ converges to the true value Φ^* we need to prove that $\Phi^{(i)}$ converges to $\hat{\Phi}$ for $i \to \infty$ such that $\hat{\Phi}_{\mathsf{EM}} \approx \hat{\Phi}$ for large enough i_f, and that $\hat{\Phi}$ converges to Φ^*.

Theorem 4.2.1 (Consistency of the maximum likelihood estimate of the mixture model of ACG distributions)
Let x_j be independent and identically distributed from a mixture of m ACG distributions with the parameter $\Phi^* = ((\pi_k^*)_{1 \leq k \leq m}, (\Sigma_k^*)_{1 \leq k \leq m})$, where Σ_k^* a symmetric positive-definite $d \times d$ matrix parameter and $\sum_{k=1}^{m} \pi_k^* = 1$. Let Θ satisfy 4.1 and 4.2. Then $\hat{\Phi}$ is consistent, i.e.,

$$\hat{\Sigma}_k \xrightarrow[n \to \infty]{P} \Sigma_k^*$$
$$\hat{\pi}_k \xrightarrow[n \to \infty]{P} \pi_k^*$$

where $\xrightarrow{P}$ denotes convergence in probability.

Theorem 4.2.2 (Convergence of the expectation maximization estimate of the mixture model of ACG distributions)
Under the assumption of Theorem 4.2.1, if $\hat{\Phi}$ *is the unique maximum of the log-likelihood in* Θ *and if it lies in the interior of* Θ*, then* $\Phi^{(i)}$ *converges to* $\hat{\Phi}$ *for* $i \to \infty$ *and, in particular,*

$$\Sigma_k^{(i)} \to \hat{\Sigma}_k, \quad \pi_k^{(i)} \to \hat{\pi}_k, \quad k = 1, \cdots, m, \quad i \to \infty$$

Mark that the uniqueness assumption on $\hat{\Phi}$, Θ is standard in the formulation of consistency results for maximum likelihood estimates. If Θ is so large that it contains several local maxima of the likelihood, then estimates can always target a local maximum instead of the global maximum. To counter this effect, in practice one starts the estimation algorithm with several initial values for the parameter.

We begin with the proof of Theorem 4.2.2. It follows from Wu (1983) if the following conditions hold.

(C1) $Q[\Phi \,|\, \Phi^*] \geq Q[\Phi^* \,|\, \Phi^*]$ for all $\Phi, \Phi^* \in \Theta$

(C2) $H[\Phi^* \,|\, \Phi^*] \geq H[\Phi \,|\, \Phi^*]$ for all $\Phi, \Phi^* \in \Theta$

(C3) $Q(\Phi \,|\, \Phi^*)$ is continuous in both Φ and Φ^*

where

$$H[\Phi \,|\, \Phi^*] = Q[\Phi \,|\, \Phi^*] - \sum_{j=1}^{n} \ln \left(\sum_{l=1}^{M} \pi_l f(x_j; \Sigma_l) \right)$$

It is simple to see that the difference term between $H[\Phi \,|\, \Phi^*]$ and $Q[\Phi \,|\, \Phi^*]$ is just the log-likelihood function of the incomplete density of the mixture model. Note that, an EM algorithm with assumptions (C1) and (C2) is a generalized EM algorithm (GEM algorithm).

Proof We begin with (C2). We use the notation introduced in the context of (4.4) and (4.5). Since $\sum_{k=1}^{m} S_{jk}^* = 1$, we have

$$\begin{aligned} H[\Phi \,|\, \Phi^*] &= \sum_{j=1}^{n} \sum_{k=1}^{m} S_{jk}^* \ln(\pi_k f(x_j, \Sigma_k)) \\ &\quad - \sum_{j=1}^{n} \sum_{k=1}^{m} S_{jk}^* \ln \left(\sum_{l=1}^{m} \pi_l f(x_j; \Sigma_l) \right) \\ &= \sum_{j=1}^{n} \sum_{k=1}^{m} S_{jk}^* \cdot \ln \left(\frac{\pi_k f(x_j, \Sigma_k)}{\sum_{l=1}^{m} \pi_l f(x_j; \Sigma_l)} \right) \end{aligned}$$

From (4.1) it yields

$$\frac{\pi_k f(x_j, \Sigma_k)}{\sum_{l=1}^{m} \pi_l f(x_j; \Sigma_l)} = E[S_{jk} \,|\, x_j; \Phi] = \hat{S}_{jk}$$

Therefore

$$H[\Phi \,|\, \Phi^*] = \sum_{j=1}^{n} \sum_{k=1}^{m} S_{jk}^* \ln \hat{S}_{jk}$$

It follows as in Dempster et al. (1977) and Franke et al. (2011) that

$$\sum_{k=1}^{m} S_{jk}^{*} \ln\left(\frac{S_{jk}^{*}}{\hat{S}_{jk}}\right) \geq 0$$

with equality if and only if $\hat{S}_{jk} = S_{jk}^{*}$. Hence,

$$H[\Phi^{*} \,|\, \Phi^{*}] \geq H[\Phi \,|\, \Phi^{*}]$$

with equality if and only if $\hat{S}_{jk} = S_{jk}^{*}$.

(C1) We define a mapping describing the iteration

$$\begin{aligned} M &: \Phi^{(i)} \to \Phi^{(i+1)} \\ M(\Phi^{*}) &= \operatorname*{argmax}_{\Phi \in \Theta} Q(\Phi \,|\, \Phi^{*}) \end{aligned}$$

which obviously satisfies

$$Q(M(\Phi^{*}) \,|\, \Phi^{*}) \geq Q(\Phi^{*} \,|\, \Phi^{*})$$

for all $\Phi^{*} \in \Theta$

(C3) Since $0 < \delta \leq \pi_k \leq 1 - \delta$ for all k, it is sufficient to proof that $f(x; \Sigma_k)$ is positive and bounded.

We have, as f is a density on the unit sphere,

$$x^T \Sigma_k^{-1} x \leq \frac{1}{\lambda_{\min}(\Sigma_k)} \cdot \|x\|^2 = \frac{1}{\lambda_{\min}(\Sigma_k)} < \infty$$

and

$$x^T \Sigma_k^{-1} x \geq \frac{1}{\lambda_{\max}(\Sigma_k)} \cdot \|x\|^2 = \frac{1}{\lambda_{\max}(\Sigma_k)} > 0$$

Since Σ_k is symmetric matrix

$$\det(\Sigma_k) = \lambda_1(\Sigma_k) \cdot \lambda_2(\Sigma_k) \cdot \ldots \cdot \lambda_d(\Sigma_k)$$

where $\lambda_1, \ldots, \lambda_d$ are the eigenvalues of Σ_k. Therefore

$$\lambda_{\min}^{d}(\Sigma_k) \leq \det(\Sigma_k) \leq \lambda_{\max}^{d}(\Sigma_k)$$

It follows

$$\begin{aligned} f(x; \Sigma_k) &= \frac{1}{\alpha_d} \cdot \frac{1}{(\det(\Sigma_k))^{1/2}} \cdot \frac{1}{(x^T \Sigma_k^{-1} x)^{d/2}} \\ &\geq \frac{1}{\alpha_d} \cdot \frac{1}{(\lambda_{\max}(\Sigma_k))^{d/2}} \cdot (\lambda_{\min}(\Sigma_k))^{d/2} \\ &= \frac{1}{\alpha_d} \left(\frac{\lambda_{\min}(\Sigma_k)}{\lambda_{\max}(\Sigma_k)}\right)^{d/2} \\ &= \frac{1}{\alpha_d} (\operatorname{cond}(\Sigma_k))^{-d/2} \\ &\geq \frac{1}{\alpha_d} \left(\frac{\underline{\lambda}}{\overline{\lambda}}\right)^{d/2} > 0 \end{aligned}$$

where $\text{cond}(\Sigma_k)$ it the condition number of the normal matrix Σ_k, and $\underline{\lambda} \leq \lambda_{\min}(\Sigma_k)$, $\overline{\lambda} \geq \lambda_{\max}(\Sigma_k)$ by assumption on Θ.

Analogously

$$f(x;\Sigma_k) \leq \frac{1}{\alpha_d}\left(\frac{\lambda_{\max}(\Sigma_k)}{\lambda_{\min}(\Sigma_k)}\right)^{d/2} \leq \frac{1}{\alpha_d}\left(\frac{\overline{\lambda}}{\underline{\lambda}}\right)^{d/2} < \infty$$

for all Σ_k. □

We recall the log-likelihood function based on the incomplete observation from the mixture model as

$$\begin{aligned} l_{\mathsf{inc}}(\Phi \,|\, x) &= Q(\Phi^* \,|\, \Phi) - H(\Phi^* \,|\, \Phi) \\ &= \sum_{j=1}^{n} \rho(x_j;\Phi) \\ &= \sum_{j=1}^{n} \ln\left(\sum_{k=1}^{m} \pi_k f(x_j;\Sigma_k)\right) \end{aligned}$$

where $\rho(x;\Phi) = \ln\left(\sum_{k=1}^{m} \pi_k f(x;\Sigma_k)\right)$

Theorem 4.2.1 follows from Franke et al. (2011) if the following conditions hold.

(C4) Θ is compact

(C5) $\rho(x;\Phi)$ is continuous in Φ and $r(\Phi) = \mathbb{E}(\rho(x_1;\Phi)) < \infty$.

(C6) $r(\Phi)$ is continuous.

(C7) $\rho_0(x;\Phi) = \rho(x;\Phi) - r(\Phi)$ satisfies a uniform Lipschitz condition

$$|\rho_0(x;\Phi) - \rho_0(x;\Phi^*)| \leq L(x)\,\|\Phi - \Phi^*\|$$

for all $\Phi, \Phi^* \in \Theta$, and the non-negative function L satisfying $\mathbb{E}L(x_1) < \infty$.

(C8) For $n \to \infty$

$$\frac{1}{n}\sum_{j=1}^{n} \rho(x_j;\Phi) \xrightarrow{P} r(\Phi) \qquad \text{for all } \Phi \in \Theta$$

Proof of Theorem 4.2.1 We have to check (C4) - (C8).

(C4) follows immediately from (4.4) and (4.5).

(C5) It follows from the proof of (C3) that $\rho(x;\Phi)$ is continuous and $f(x;\Sigma_k)$ is uniformly bounded in $\|x\| = 1$, $\Sigma_k \in \Theta$, then $\rho(x;\Phi)$ is bounded. Therefore $r(\Phi) = E(\rho(x_1;\Phi)) < \infty$.

(C6) $r(\Phi)$ is continuous as $\rho(x;\Phi)$ is uniformly continuous in x.

(C7) It is sufficient to proof that $\rho_0(x;\Phi)$ and $\rho(x,\Phi)$ are differentiable w.r.t. Φ and the derivative is bounded uniformly in $\|x\| = 1$.

Since π_k and $f(x;\Sigma_k)$ is positive and bounded, the ln-function is differentiable. And we have

$$\begin{aligned}\frac{\partial\rho}{\partial\pi_k} &= \frac{1}{\sum_{l=1}^m \pi_l f(x;\Sigma_l)}\cdot f(x,\Sigma_k) < \infty \\ \frac{\partial\rho}{\partial\Sigma_k} &= \frac{\pi_k}{\sum_{l=1}^m \pi_l f(x;\Sigma_l)}\cdot \frac{\partial f(x,\Sigma_k)}{\partial\Sigma_k}\end{aligned}$$

$f(x,\Sigma_k)$ continuous differentiable w.r.t. Σ_k as long as we stay in Θ if we look at the definition of f and take into account $\|x\| = 1$.

(C8) follows directly from the law of large numbers as $x_1,\cdots,x_n$ are independent and identically distributed.

□

Up to here, the consistency of parameter estimates for a mixture of general ACG distributions is proven. Then the mixture model of Schladitz distributions has also consistent parameter estimates as the Schladitz distribution is a special case of the ACG distribution with $d = 3$ and $\Sigma = \mathrm{diag}(1,1,\beta^{-2})$.

4.3 Mixtures of the Schladitz Distributions

Here we come back to the Schladitz distribution. We consider a general model of a mixture of m Schladitz distributions. For $1 \le k \le m$, let $p(\theta,\phi\,|\,\beta_k,\mu_k)$ be the Schladitz distribution component with the parameter β_k and principal/polar axis μ_k, which described in Section 2.3. The probability density function of this mixture model is

$$f(\theta,\phi\,|\,\Phi) = \sum_{k=1}^m \pi_k p(\theta,\phi\,|\,\beta_k,\mu_k) \tag{4.6}$$

where $\Phi = (\pi_1,\cdots,\pi_m,\beta_1,\cdots,\beta_m,\mu_1,\cdots,\mu_m)$ and $\sum_{k=1}^m \pi_k = 1$.

Let $(\theta_1,\phi_1),(\theta_2,\phi_2),\cdots,(\theta_n,\phi_n)$ be a sample of independent and identically distributed unit vectors from the mixture of Schladitz distributions. Furthermore, let $S_{jk}, j = 1,\cdots,n, k = 1,\cdots,m$ be the indicator variables as in the previous section, i.e. in particular, $S_{jk} = 1$, $S_{jl} = 0, l \neq k$, if (θ_j,ϕ_j) is generated by the $k-$th regime.

Again, we consider the maximum likelihood estimate $\hat{\Phi}$ which we get from maximizing the incomplete log-likelihood

$$\sum_{j=1}^n \ln f(\theta_j,\phi_j\,|\,\Phi) = \sum_{j=1}^n \ln\left[\sum_{k=1}^m \pi_k p(\theta_j,\phi_j\,|\,\beta_k,\mu_k)\right]$$

The corresponding complete log-likelihood assuming knowledge of the S_{jk} or, equivalently, of the state variables $Q_j = k$ if and only if $S_{jk} = 1$, is given by

$$\begin{aligned}L_c(\Phi\,|\,(\theta_1,\phi_1),\cdots,(\theta_n,\phi_n),Q) &= \sum_{j=1}^n \ln\pi_{Q_j} + \sum_{j=1}^n\sum_{k=1}^m S_{jk}\ln p(\theta_j,\phi_j\,|\,\beta_k,\mu_k) \\ &= \sum_{j=1}^n\sum_{k=1}^m S_{jk}\ln(\pi_k p(\theta_j,\phi_j\,|\,\beta_k,\mu_k)) \end{aligned}\tag{4.7}$$

(compare (4.3)). As in the previous section for mixtures of general ACG distributions, we replace S_{jk} in (4.7) by the posterior probability of $S_{jk} = 1$ given the data, we apply the EM algorithm, and we get the iterative scheme.

E-step: Compute $Q[\Phi \mid \Phi^{(i)}]$:

$$Q[\Phi \mid \Phi^{(i)}] = \sum_{j=1}^{n}\sum_{k=1}^{m} S_{jk}^{(i)} \ln\left(\pi_k p(\theta_j, \phi_j \mid \beta_k, \mu_k)\right)$$

with

$$S_{jk}^{(i)} = \frac{\pi_k^{(i)} p\left(\theta_j, \phi_j \mid \beta_k^{(i)}, \mu_k^{(i)}\right)}{\sum_{l=1}^{m} \pi_l^{(i)} p\left(\theta_j, \phi_j \mid \beta_l^{(i)}, \mu_l^{(i)}\right)}$$

M-step: Update from i to $i+1$

(i) calculate $\pi_k^{(i+1)}$ by

$$\pi_k^{(i+1)} = \frac{1}{n}\sum_{j=1}^{n} S_{jk}^{(i)} \tag{4.8}$$

(ii) calculate $\beta_k^{(i+1)}, \mu_k^{(i+1)}$ by

$$\left(\beta_k^{(i+1)}, \mu_k^{(i+1)}\right) = \underset{\beta_k, \mu_k}{\operatorname{argmax}} \sum_{j=1}^{n} S_{jk}^{(i)} \ln p(\theta_j, \phi_j \mid \beta_k, \mu_k), \quad k = 1, \cdots, m$$

Now, we discuss two methods for solving the maximization task of (ii).

Since $p(\theta_j, \phi_j \mid \beta_k, \mu_k)$ is calculated pretending that (θ_j, ϕ_j) belongs to the $k-$th group, we can rotate the data about the principal/polar axis μ_k under the method described in (3.6) and obtain a new sample $(\theta_{j\mid k}, \phi_{j\mid k})$, $j = 1, \cdots, n$. Under this transformation, μ_k is changed to $(0,0,1)^T$ such that the density $p\left(\theta_{j\mid k}, \phi_{j\mid k} \mid \beta_k\right)$ does no longer depend on μ_k and has rotational symmetry around the $z-$axis. Let T denote the term to be maximized in (ii) of the M-step, we have

$$\begin{aligned}
T &= \sum_{k=1}^{m}\sum_{j=1}^{n} S_{jk}^{(i)} \ln p(\theta_j, \phi_j \mid \beta_k, \mu_k) \\
&= \sum_{k=1}^{m}\sum_{j=1}^{n} S_{jk}^{(i)} \ln p\left(\theta_{j\mid k}, \phi_{j\mid k} \mid \beta_k\right) \\
&= \sum_{k=1}^{m}\sum_{j=1}^{n} S_{jk}^{(i)} \ln\left(\frac{1}{4\pi}\frac{\beta_k \sin\theta_{j\mid k}}{\left(1+(\beta_k^2-1)\cos^2\theta_{j\mid k}\right)^{3/2}}\right) \\
&= \sum_{k=1}^{m}\sum_{j=1}^{n} S_{jk}^{(i)} \left(\ln\left(\frac{\sin\theta_{j\mid k}}{4\pi}\right) + \ln(\beta_k) - \frac{3}{2}\ln\left(1+(\beta_k^2-1)\cos^2\theta_{j\mid k}\right)\right)
\end{aligned}$$

Mark that T depends not only on β_k but via the transformation $\theta_j \to \theta_{j\,|\,k}$ also on μ_k.

We introduce here two different methods to get $\left(\beta_k^{(i+1)}, \mu_k^{(i+1)}\right)$. One of them takes into account the constants $\mu_k^T \mu_k = 1$, as μ_k is a unit vector, by means of a Lagrangian term. Adding it to T, we get

$$\begin{aligned} T^* &= \sum_{k=1}^{m}\sum_{j=1}^{n} S_{jk}^{(i)} \ln p_k(\theta_{j\,|\,k}, \phi_{j\,|\,k} \,|\, \beta_k) + \sum_{k=1}^{m} \lambda_k (1 - \mu_k^T \mu_k) \\ &= \sum_{k=1}^{m}\sum_{j=1}^{n} S_{jk}^{(i)} \left(\ln\left(\frac{\sin\theta_{j\,|\,k}}{4\pi}\right) + \ln(\beta_k) - \frac{3}{2}\ln\left(1 + \left(\beta_k^2 - 1\right)\cos^2\theta_{j\,|\,k}\right)\right) \\ &\quad + \sum_{k=1}^{m} \lambda_k (1 - \mu_k^T \mu_k) \end{aligned}$$

To obtain the maximum of T^* we set each partial derivative of T^* to zero. By differentiating with respect to each β_k we have

$$\sum_{j=1}^{n} S_{jk}^{(i)} = 3\beta_k^2 \sum_{j=1}^{n} \frac{S_{jk}^{(i)} \cos^2\theta_{j|k}}{1 + (\beta_k^2 - 1)\cos^2\theta_{j|k}} \tag{4.9}$$

This is similar to the problem of the estimation of parameter β as described in Algorithm 5. By differentiating with respect to each λ_k we have

$$\mu_k^T \mu_k = 1 \tag{4.10}$$

By differentiation with respect to each μ_k we have

$$\frac{1}{2}\sum_{j=1}^{n} T_\mu = \lambda_k \mu_k \tag{4.11}$$

where T_μ comes from differentiating of T w.r.t. μ_k, i.e.

$$T_\mu = \sum_{j=1}^{n} \frac{\partial \ln p(\theta_j, \phi_j \,|\, \beta_k, \mu_k)}{\partial \mu_k} \cdot S_{jk}^{(i)} \tag{4.12}$$

As a result of Section 2.4, we believe that if we replace the density function $p(\theta_j, \phi_j \,|\, \beta_k, \mu_k)$ of (4.12) in the density function of Watson distribution, the estimate is still acceptable. Therefore

$$\begin{aligned} T_\mu &\approx \sum_{ij=1}^{n} \frac{\partial \ln \exp(\beta_k (v_j^T \mu_k)^2)}{\partial \mu_k} \cdot S_{jk}^{(i)} \\ &= 2\sum_{j=1}^{n} \beta_k (v_j^T \mu_k) v_j S_{jk}^{(i)} \end{aligned} \tag{4.13}$$

where $v_j = (x_j, y_j, z_j)^T$ as Cartesian representation of the point on the unit sphere with polar coordinates (θ_j, ϕ_j).

Table 4.1: Means and standard deviations of $\hat{\pi}, \hat{\beta}$ and $\hat{\mu}$ for mixture of two bipolar Schladitz distributions with various mixing proportions (Algorithm 6). We simulate 2000 Schladitz distributed unit vectors with the parameter $\beta = 0.1$, some of them are distributed with the principal axis $(0, 0, 1)^T$, the others are distributed with the principal axis $(1, 0, 0)^T$, and the mixing ratios are written as $\pi_1 : \pi_2$ for $0.5 : 0.5$, $0.25 : 0.75$ and $0.2 : 0.8$.

	$\pi_1 : \pi_2$	0.5:0.5		0.25:0.75		0.2:0.8	
$\hat{\pi}$	mean	0.5004	0.4996	0.2206	0.7794	0.3980	0.6020
	sd	0.0083	0.0083	0.0073	0.0073	0.0835	0.0835
$\hat{\beta}$	mean	0.1045	0.1036	0.0993	0.1070	0.6882	0.0830
	sd	0.0053	0.0071	0.0096	0.0038	0.2137	0.0086
$\hat{\mu}$	mean	0.0041	1.0000	0.0065	1.0000	0.6815	0.9999
		0.0053	0.0043	0.0083	0.0049	0.0106	0.0039
		1.0000	0.0044	1.0000	0.0029	0.6836	0.0088
	sd	0.0036	e-5	0.0090	e-5	0.2441	e-5
		0.0036	0.0026	0.0056	0.0038	0.0139	0.0035
		e-5	0.0034	e-5	0.0031	0.1264	0.0041

From (4.11) and (4.13) we have

$$\beta_k \left\| \sum_{j=1}^{n} (v_j^T \mu_k) v_j S_{jk}^{(i)} \right\| = \lambda_k$$

Therefore

$$\mu_k = \frac{\sum_{j=1}^{n} (v_j^T \mu_k) v_j S_{jk}^{(i)}}{\left\| \sum_{j=1}^{n} (v_j^T \mu_k) v_j S_{jk}^{(i)} \right\|} \tag{4.14}$$

From equations (4.8), (4.9)and (4.14) we have the first EM algorithm for a mixture of Beta-distributions with Lagrangian, which is shown in Algorithm 6.

Table 4.1 shows the adequacy of this method for various mixing proportions. We simulate 2000 β-distributed unit vectors with the parameter $\beta = 0.1$, part of them are distributed with the principle axis $(0, 0, 1)^T$, the other part of them are distributed with the principle axis $(1, 0, 0)^T$, and the mixing ratios are tested of $0.5 : 0.5$, $0.25 : 0.75$ and $0.2 : 0.8$. We estimate the parameters of the mixture distributions with the method described in Algorithm 6. For each set of parameters, the procedure was repeated 10 times to get the mean value and standard deviation of π_k, β_k and μ_k for $k = 1, 2$.

The results indicate that this method works well for not so large ratio of two distributions, smaller or equal than $1 : 3$. Note that for the case of $0.2 : 0.8$ it yields sometimes good estimators, but most of the time not, therefore the standard deviation is very large, and mostly the estimated principal/polar axes are also not expected.

To show the adequacy of this method for various differences of the original parameter of the two distributions we simulate 2000 unit vectors of β-distributions, a half of them are distributed with parameter $\beta = 0.1$ and the principal axis $(0, 0, 1)^T$, the other 1000 vectors are distributed with the principal axis $(1, 0, 0)^T$, but with various parameter β between 0.1 and 0.4. We estimate

Algorithm 6 Estimation of the Parameter of Schladitz Distribution (with Lagrangian)

Require: A set of n unit vectors with the polar coordinates (θ_j, ϕ_j), as well as the direction cosines (x_j, y_j, z_j) for $1 \leq j \leq n$, the number of different distributions m

Ensure: The mixture proportions π_k, the principal axis of bipolar distributions or polar axis of girdle distributions μ_k , as well as the estimated parameter β_k for each $1 \leq k \leq m$

initial π_k, β_k, μ_k with $\sum_{k=1}^{m} \pi_k = 1$ and $\|\mu_j\| = 1$ for each j

repeat

E-step

for $j = 1$ **to** n **do**

rotate the original data (θ_j, ϕ_j) to $(\theta_{j\,|\,k}, \phi_{j\,|\,k})$ with μ_k

for $k = 1$ **to** m **do**

$$p(\theta_{j\,|\,k}|\beta_k) = \frac{1}{4\pi} \frac{\beta_k \sin\theta_{j\,|\,k}}{\left(1 + (\beta_k^2 - 1)\cos^2\theta_{j\,|\,k}\right)^{3/2}}$$

end for

for $k = 1$ **to** m **do**

$$S_{jk} = \frac{\pi_k p(\theta_{j\,|\,k}|\beta_k)}{\sum_{l=1}^{m} \pi_l p(\theta_{j\,|\,k}\,|\,\beta_l)}$$

end for

end for

M-step

for $k = 1$ **to** m **do**

$$\begin{aligned} \pi_k &= \frac{1}{n}\sum_{j=1}^{n} S_{jk} \\ \mu_k &= \frac{\sum_{j=1}^{n}((x_j, y_j, z_j)^T \mu_k)(x_j, y_j, z_j) S_{jk}}{\left\|\sum_{j=1}^{n}((x_j, y_j, z_j)^T \mu_k)(x_j, y_j, z_j) S_{jk}\right\|} \end{aligned}$$

rotate the original data (θ_j, ϕ_j) to $(\theta_{j\,|\,k}, \phi_{j\,|\,k})$ with μ_k

estimate β_k with the Main-step of Algorithm5 (replace θ_j in $\theta_{j\,|\,k}$)

end for

until there is no more significant improvement

return π_k, β_k, μ_k

Table 4.2: Means and standard deviations of $\hat{\pi}, \hat{\beta}$ and $\hat{\mu}$ Algorithm 6 for mixture of two bipolar Schladitz distributions with various original parameter (Algorithm 6). We simulate 2000 Schladitz distributed unit vectors, a half of them are distributed with the principal axis $(0,0,1)^T$ with the parameter $\beta = 0.1$, the other half of them are distributed with the principal axis $(1,0,0)^T$ with various parameter of $\beta = 0.1$, $\beta = 0.2$, $\beta = 0.3$ and $\beta = 0.4$.

β		0.1		0.2		0.3		0.4	
$\hat{\pi}$	mean	0.5004	0.4996	0.4306	0.5694	0.3350	0.6650	0.2372	0.7628
	sd	0.0083	0.0083	0.0177	0.0177	0.0233	0.0233	0.0217	0.0217
$\hat{\beta}$	mean	0.1045	0.1036	0.0949	0.2529	0.0842	0.4566	0.0749	0.7806
	sd	0.0053	0.0071	0.0036	0.0144	0.0039	0.0284	0.0749	0.0155
$\hat{\mu}$	mean	0.0041	1.0000	0.0052	0.9998	0.0036	0.9991	0.0215	0.5655
		0.0053	0.0043	0.0037	0.0106	0.0040	0.0095	0.0041	0.0182
		1.0000	0.0044	1.0000	0.0126	1.0000	0.0283	0.9997	0.8188
	sd	0.0036	e-5	0.0020	e-4	0.0018	0.0016	0.0067	0.0786
		0.0036	0.0026	0.0035	0.0071	0.0026	0.0097	0.0043	0.0119
		e-5	0.0034	e-5	0.0070	e-5	0.0307	e-4	0.0644

the parameters of the mixture distributions with the method described in Algorithm 6. To each set of parameters, the procedure was repeated 10 times to get the mean value and standard deviation for each parameter. The result is shown in Table 4.2.

It indicates that this method works well for mixtures of concentrated bipolar distributions, i.e. for small β. For a mixture of a concentrated bipolar distribution and a bipolar distribution with $\beta \geq 0.4$, Algorithm 6 will recognize it as a concentrated bipolar distribution and an isotropic distribution, this is not the result we expected.

Since there are many coincide points in the case of the mixture of two or more girdle distributions, the method of 6 yields very unexpected results for such cases, therefore we improve it through calculate the "weighted" principal/polar axis in each M-step. The idea is that we change the original directions with multiplicate with the posterior probability $S_{jk}^{(i)}$ in each M-step.

The pseudo code shows in Algorithm 7.

For this method we make the same tests as for Algorithm 6.

Table 4.3 indicates that Algorithm 7 works also well for the case of $0.25 : 0.75$, especially for the estimation of principal/polar axis, which is impossible for the method of Algorithm 6. But this method is still not suitable for very large difference between the numbers of two distributions, for the case of $0.1 : 0.9$, which is not described in Table 4.3, it recognize just one distribution with the principle axis of the group of 1800 unit vectors, i.e. ignore the unit vectors of distribution of very small proportion.

Table 4.4 indicate that Algorithm 7 works well for almost all the cases of the mixture of bipolar distributions of the same number of vectors, especially for the case of $\beta_2 = 0.4$, $\beta_2 = 0.5$ or even larger β_2 it makes acceptable results. But it still have the tendency that one of the estimated distribution is more concentrated, the other is more isotropic for large β_2. And there are also some bias between the original simulated principle axis and the estimated principle axis.

An obvious advantage of Algorithm 7 with compare to Algorithm 6 is that it also works well for the mixture of two girdle distributions. We simulate 2000 β-distributed unit vectors with the parameter $\beta = 2$, some of them are distributed with the principal axis $(0,0,1)^T$, the others

Algorithm 7 Estimation of the Parameter of Schladitz Distribution (with "weighted" principal/polar axis)

Require: A set of n unit vectors with the polar coordinates (θ_j, ϕ_j), as well as the direction cosines (x_j, y_j, z_j) for $1 \leq j \leq n$, the number of different distributions m and the types of the distributions.

Ensure: The mixture proportions π_k, the principle axis of bipolar distributions or polar axis of girdle distributions μ_k , as well as the estimated parameter β_k for each $1 \leq k \leq m$

initial π_k, β_k, μ_k with $\sum_{k=1}^{m} \pi_k = 1$ and $\|\mu_k\| = 1$ for each k

repeat

E-step

for $j = 1$ **to** n **do**

rotate the original data (θ_j, ϕ_j) to $(\theta_{j|k}, \phi_{j|k})$ with μ_k

for $k = 1$ **to** m **do**

$$p(\theta_{j\,|\,k} \,|\, \beta_k) = \frac{1}{4\pi} \frac{\beta_k \sin \theta_{j\,|\,k}}{\left(1 + (\beta_k^2 - 1)\cos^2 \theta_{j\,|\,k}\right)^{3/2}}$$

end for

for $k = 1$ **to** m **do**

$$S_{jk} = \frac{\pi_k p(\theta_{j\,|\,k} \,|\, \beta_k)}{\sum_{l=1}^{m} \pi_l p(\theta_{j\,|\,k} \,|\, \beta_l)}$$

end for

end for

M-step

for $k = 1$ **to** m **do**

$$\pi_j = \frac{1}{n} \sum_{j=1}^{n} S_{jk}$$

for $j = 1$ **to** n **do**

$$(x_{jk}^*, y_{jk}^*, z_{jk}^*)^T = (x_j, y_j, z_j)^T S_{jk}$$

end for

calculate the principal/polar axis μ_k with the method of Pre-step of Algorithm 5 ($(x_{jk}^*, y_{jk}^*, z_{jk}^*)$ instead of (x_j, y_j, z_j))

rotate the original data (θ_j, ϕ_j) to $(\theta_{j\,|\,k}, \phi_{j\,|\,k})$ with μ_k

estimate β_k with the main step of Algorithm 5 ($\theta_{j\,|\,k}$ instead of θ_j)

end for

until there is no more significant improvement

return π_k, β_k, μ_k

Table 4.3: Means and standard deviations of $\hat{\pi}, \hat{\beta}$ and $\hat{\mu}$ for mixture of two bipolar Schladitz distributions with various mixing proportions (Algorithm 7). We simulate 2000 Schladitz distributed unit vectors with the parameter $\beta = 0.1$, some of them are distributed with the principal axis $(0, 0, 1)^T$, the others are distributed with the principal axis $(1, 0, 0)^T$, and the mixing ratios are written as $\pi_1 : \pi_2$ for $0.5 : 0.5$, $0.25 : 0.75$ and $0.2 : 0.8$.

	$\pi_1 : \pi_2$	0.5:0.5		0.25:0.75		0.2:0.8	
$\hat{\pi}$	mean	0.4990	0.5010	0.2236	0.7764	0.1720	0.8280
	sd	0.0119	0.0119	0.0044	0.0044	0.0091	0.0091
$\hat{\beta}$	mean	0.1072	0.1046	0.1008	0.1057	0.0991	0.1053
	sd	0.0044	0.0068	0.0096	0.0041	0.0081	0.0031
$\hat{\mu}$	mean	0.0035	1.0000	0.0047	1.0000	0.0036	1.0000
		0.0048	0.0044	0.0067	0.0051	0.0045	0.0034
		1.0000	0.0043	1.0000	0.0044	1.0000	0.0037
	sd	0.0028	e-5	0.0038	e-5	0.0030	e-5
		0.0035	0.0024	0.0050	0.0020	0.0038	0.0028
		e-5	0.0026	e-5	0.0032	e-5	0.0027

Table 4.4: Means and standard deviations of $\hat{\pi}, \hat{\beta}$ and $\hat{\mu}$ for mixture of two bipolar Schladitz distributions with various original parameter (Algorithm 7). We simulate 2000 Schladitz distributed unit vectors, a half of them are distributed with the principal axis $(0, 0, 1)^T$ with the parameter $\beta = 0.1$, the other half of them are distributed with the principal axis $(1, 0, 0)^T$ with various parameter of $\beta = 0.1$, $\beta = 0.2$, $\beta = 0.3$ and $\beta = 0.4$.

	β_2	0.1		0.2		0.3		0.4	
$\hat{\pi}$	mean	0.4990	0.5010	0.4367	0.5633	0.3515	0.6485	0.2364	0.7636
	sd	0.0119	0.0119	0.0099	0.0099	0.0192	0.0192	0.0326	0.0326
$\hat{\beta}$	mean	0.1072	0.1046	0.0949	0.2424	0.0865	0.4514	0.0732	0.7155
	sd	0.0044	0.0068	0.0051	0.0108	0.0040	0.0332	0.0039	0.0655
$\hat{\mu}$	mean	0.0035	1.0000	0.0030	0.9998	0.0018	0.9996	0.0030	0.9959
		0.0048	0.0044	0.0026	0.0106	0.0014	0.0129	0.0044	0.0196
		1.0000	0.0043	1.0000	0.0100	1.0000	0.0160	1.0000	0.0637
	sd	0.0028	e-5	0.0019	0.0002	0.0013	0.0003	0.0019	0.0066
		0.0035	0.0024	0.027	0.0101	0.0010	0.0134	0.0033	0.0157
		e-5	0.0026	e-5	0.0094	e-5	0.0100	e-5	0.0631

Table 4.5: Means and standard deviations of $\hat{\pi}, \hat{\beta}$ and $\hat{\mu}$ for mixture of two girdle Schladitz distributions with various mixing proportions (Algorithm 7). We simulate 2000 Schladitz distributed unit vectors with the parameter $\beta = 2$, some of them are distributed with the principal axis $(0, 0, 1)^T$, the others are distributed with the principal axis $(1, 0, 0)^T$, and the mixing ratios are written as $\pi_1 : \pi_2$ for $0.5 : 0.5$, $0.25 : 0.75$ and $0.1 : 0.9$.

	$\pi_1 : \pi_2$	0.5:0.5		0.25:0.75		0.1:0.9	
$\hat{\pi}$	mean	0.4993	0.5007	0.3362	0.6638	0.2143	0.7857
	sd	0.0235	0.0235	0.0229	0.0229	0.0150	0.0150
$\hat{\beta}$	mean	2.3308	2.3508	2.2601	2.3904	2.0115	2.4132
	sd	0.1076	0.0833	0.02170	0.0675	0.03663	0.1047
$\hat{\mu}$	mean	0.0086	0.9944	0.0246	0.9986	0.0239	0.9994
		0.0392	0.0826	0.0472	0.0419	0.2723	0.0187
		0.9973	0.0169	0.9934	0.0206	0.8985	0.0249
	sd	0.0061	0.0092	0.0268	0.0013	0.0269	e-4
		0.0634	0.0659	0.1021	0.0252	0.3132	0.0104
		0.0058	0.0140	0.0189	0.0085	0.1796	0.0157

are distributed with the principal axis $(1, 0, 0)^T$, and the mixing ratios are written as $\pi_1 : \pi_2$ for $0.5 : 0.5$, $0.25 : 0.75$ and $0.1 : 0.9$. Table 4.5 shows that it yield acceptable results even for the case of $0.1 : 0.9$, but the estimated principle axis is with large bias. During the test of so large mixing proportion, it can be found that it sometimes yields very good results (not only for the parameter β, but also for the principal axis), but sometimes yields a mixture of a girdle distribution and an isotropic distribution. Therefore the standard deviation for such case is large.

4.4 Applications

In the following, we present an application for both methods described in Section 4.3. Firstly, we use the same samples of materials as in Section 3.3 and check which model fits to which sample better, a single distribution or a mixture model. Then we consider some material, which the mixture model fits better and compare the results of the two methods.

We use again glass fibre composites GF30 and GF60 described in Section 3.3. For the analysis, we selected ten cubes of size 128^3 voxels (for GF60 middle layer this is reduced to 60^3 voxels) and estimated the parameters of the mixture model of Schladitz distributions using Algorithm 6, as well as Algorithm 7. The results are summarised in Appendix Table A.1- Table A.8. The estimated values are a mixture of two bipolar distributions with quite similar parameters (like tube 7 in Table A.1), a mixture of two bipolar distributions with two different $\hat{\beta}$ values and quasi the same principal axis (like tube 1 in Table A.1),or a mixture of a bipolar distribution and a quasi isotropic distribution (like tube 4 in Table A.5). It can be considered in all three cases as a single bipolar distribution with the parameter β, which in the first case is equal to the estimated $\hat{\beta}$, in the second case is somehow between the two estimated $\hat{\beta}$ and in the third case is somehow larger than the estimated $\hat{\beta}$ of the bipolar distribution. Here "somehow" depends on the estimated $\hat{\pi}$, which means the portion of the directions which lie in each group. Compare to the result of the estimated value of a single Schladitz distribution in Table 3.4, there is no large difference in the

Table 4.6: Estimated parameters of the mixture model of Schladitz distributions using Algorithm 6 for foam

foam	$\hat{\pi}$	$\hat{\beta}$	$\hat{\mu}$
A	0.5383	0.7755	(0.1661,0.7646,0.6227)
	0.4617	0.5672	(0.2005,0.9694,0.1413)
B	0.5795	0.4420	(0.2251,0.9736,0.0381)
	0.4205	0.8594	(0.1585 0.7322,0.6624)
C	0.7968	0.7999	(0.1193,0.6563,0.7450)
	0.2032	0.1663	(0.1391,0.6567,0.7412)
D	0.8862	0.6776	(0.1629,0.6739,0.7206)
	0.1138	0.2253	(0.1272,0.6060,0.7852)
E	0.5443	0.7799	(0.1009,0.6636,0.7412)
	0.4557	0.3447	(0.1268,0.6426,0.7556)
F	0.8204	0.6947	(0.1366,0.6402,0.7559)
	0.1796	0.2911	(0.1320,0.5376,0.8328)
G	0.2554	0.8163	(0.1967,0.7527,0.6283)
	0.7446	0.5068	(0.2608,0.9617,0.0844)
H	0.7895	0.7163	(0.1307,0.6172,0.7759)
	0.2105	0.1925	(0.1033,0.4381,0.8930)

estimated values. Therefore we believe that a single Schladitz distribution fits better to each material, i.e. GF30, GF60 bottom, GF60 middle and GF60 top.

Algorithm 6 and Algorithm 7 yield not always the same result, e.g. the estimated values of GF30 tube 8 using Algorithm 6 is of the first case and using Algorithm 7 is of the second case. This situation is not rare in our tests, which means both methods are not stable.

The Schladitz distribution shows prominent advantages for the compressed fibre system, especially for the compressed glass fibre, as compared to e.g. the Watson distribution or other axial distributions. Usually a single distribution or a layer distribution (like GF60) fits better for the compressed glass fibre, and it is difficult to find application examples for the mixture of Schladitz distributions. It showed in Schlachter (2012) and the production processes that a mixture model fits the normal directions of walls in ceramic foams. Therefore we examine the method for the sample of ceramic foams instead of the fibres.

Here we consider the directions of the walls in eight $2.2 \times 5 \times 5$cm ceramic foam. Since the real direction for the walls in the edge of the foam is difficult to obtain , we consider here only the walls inside the foam. For this sample, an image with $650 \times 650 \times 650$ voxels was used.

The parameters π_k, β_k and μ_k $(1 \leq k \leq m)$ were estimated from the direction data using the two methods presented above. The results are summarised in Table 4.6 (using Algorithm 6) and Table 4.7 (using Algorithm 7).

Furthermore, Figure 4.1 and Figure 4.2 show the density plots of the original data and the data with the estimated parameters of the ceramic foam. The results of our methods showed a mixture of a bipolar distribution with the principal axis somewhere in the x-y plane and a quasi isotropic distribution.

The fitting of mixture of Watson distributions to these data has already been discussed by Schlachter (2012) and the results are summarised in Appendix Table A.9. With compare to the

Table 4.7: Estimated parameter of the mixture model of Schladitz distributions using Algorithm 7

foam	$\hat{\pi}$	$\hat{\beta}$	$\hat{\mu}$
A	0.6321	1.0076	(0.1703,0.7162,0.6768)
	0.3679	1.1218	(0.1533,0.7293,0.6668)
B	0.4339	0.8129	(0.1349,0.6701,0.7300)
	0.5661	0.4774	(0.2251,0.9376,0.2650)
C	0.7656	0.8326	(0.1177,0.6764,0.7271)
	0.2344	0.1815	(0.1404,0.6750,0.7244)
D	0.8609	0.6939	(0.1642,0.6844,0.7103)
	0.1391	0.2450	(0.1531,0.6679,0.7283)
E	0.7253	0.7186	(0.1503,0.6910,0.7070)
	0.2747	0.2535	(0.1621,0.6849,0.7103)
F	0.6990	0.7693	(0.1427,0.6751,0.7238)
	0.3010	0.3381	(0.1646,0.6585,0.7343)
G	0.7347	0.5266	(0.2551,0.9283,0.2704)
	0.2653	0.7786	(0.1839,0.7398,0.6472)
H	0.9122	0.6360	(0.1217,0.5803,0.8053)
	0.0878	0.0920	(0.1034,0.5589,0.8228)

results of the mixture of Watson distributions, it can be found that our methods recognize the principal/polar axis not so well, if the points of the individual distributions intersect each other, i.e the groups are not clearly separated on the sphere in such cases. The improvement of this point is still a topic for the further research.

4.5 A smooth non-parametric estimate for spherical densities using Gaussian mixtures

Based on Magder and Zeger (1996) let $f(x;b)$ be a spherical density of the form

$$f(x;b) = c(\|b\|)\Phi(b^T x)$$

where

$$b = \beta\mu \in \mathbb{R}^3$$

$\beta = \|b\| \geq 0$ is the concentration parameter and $\mu = \frac{b}{\|b\|} \in \mathbb{S}^2$ is the direction parameter. Then we have

$$f(x;b) = c(\beta)\Phi(\beta\mu^T x)$$

As examples for this form we have

von Mises-Fisher distribution $\beta = \kappa$, $\Phi(u) = \exp(u)$.

Watson distribution $\beta = \sqrt{|\kappa|}$, $\Phi(u) = \exp(\mathrm{sgn}(u)u^2)$

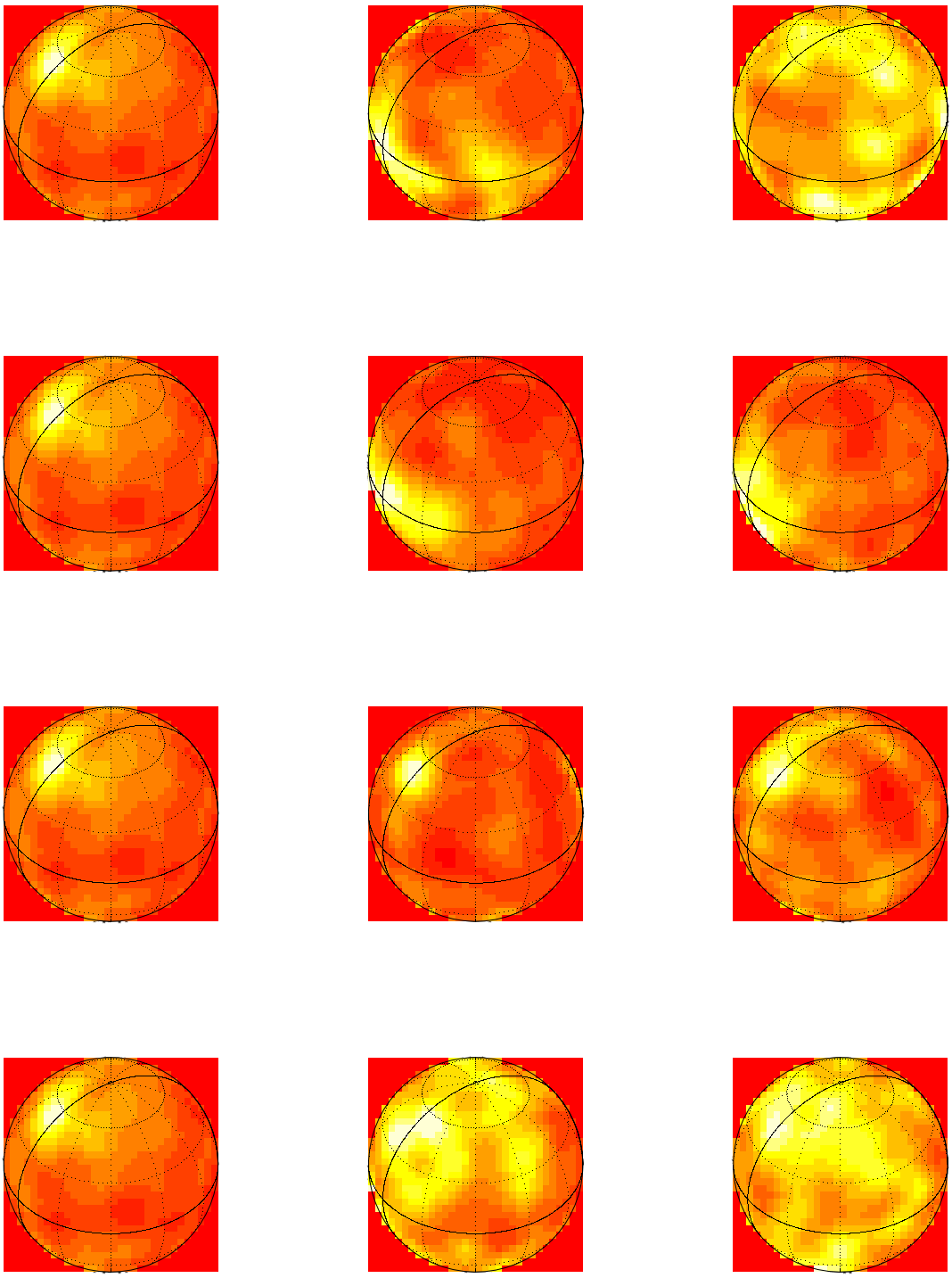


Figure 4.1: The density plots of foam A-D. Left: original data, middle: data with the estimated parameters of Algorithm 6, right:data with the estimated parameters of Algorithm 7

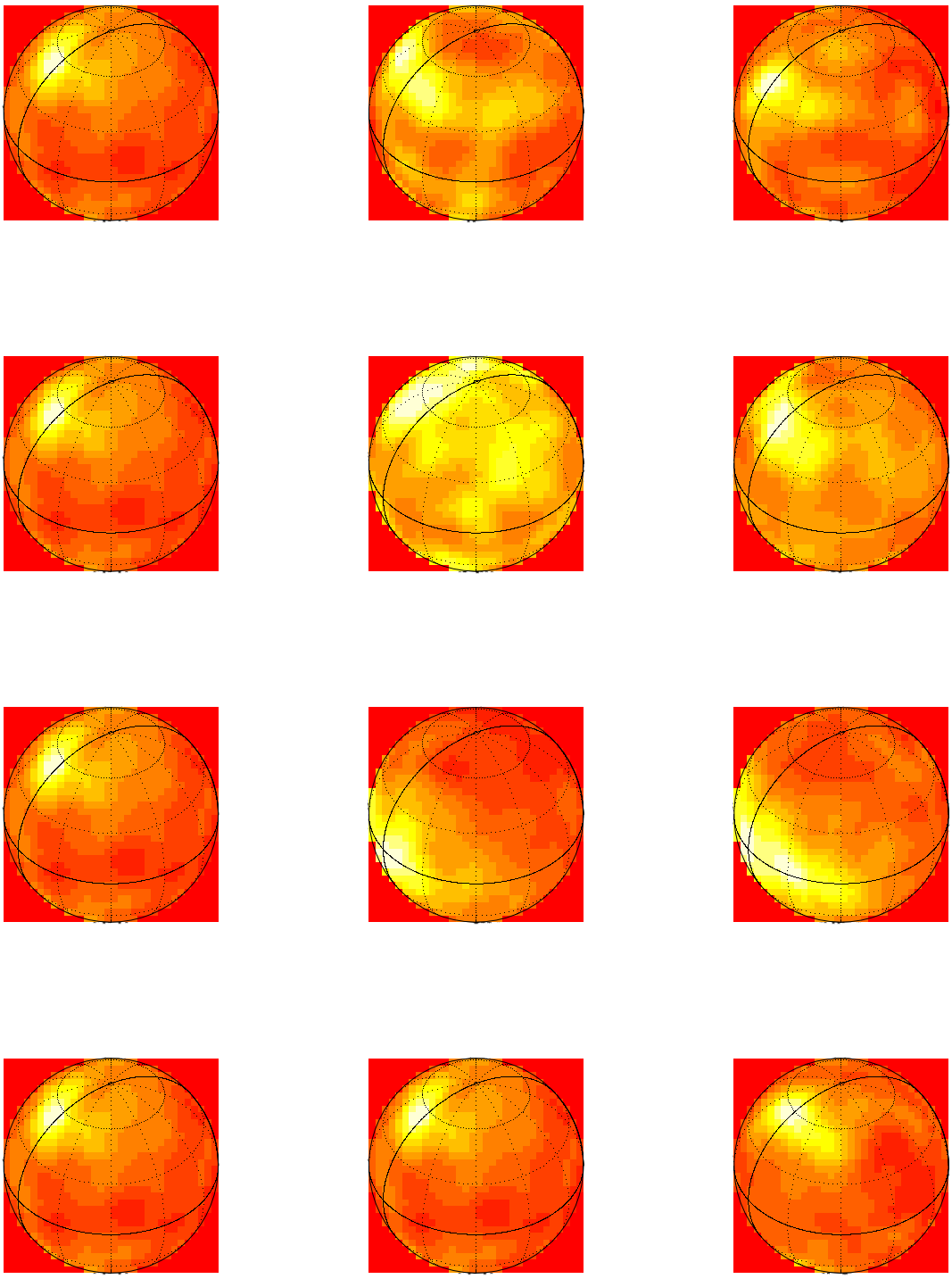


Figure 4.2: The density plots of foam E-H. Left: original data, middle: data with the estimated parameters of Algorithm 6, right:data with the estimated parameters of Algorithm 7

Mark that the Watson parametrization is slightly different than usual for the sake of identifiability of the parameters $\beta \geq 0, \mu \in \mathbb{S}^2$.

For $\mu^T x > 0$, we have

$$\Phi(\beta \mu^T x) = e^{\beta^2 (\mu^T x)^2} = e^{\kappa (\mu^T x)^2}$$

i.e. $\kappa = \beta^2 > 0$.

For $\mu^T x < 0$, we have

$$\Phi(\beta \mu^T x) = e^{-\beta^2 (\mu^T x)^2} = e^{\kappa (\mu^T x)^2}$$

i.e. $\kappa = -\beta^2 < 0$.

Choosing f is a matter of convenience, depending only on the question "if we have axial data or not". In the first case, the Watson density would be appropriate, in the second case, the von Mises-Fisher density. However, we could use other axial or non-axial densities instead.

We do not use f as a model for our data. It only serves as an auxiliary quantity for construction of a nonparametric density estimate, similar to using the Gaussian density as kernel in a Rosenblatt-Parzen density estimate on the plane.

We now randomize b by letting $\mathcal{L}(b) = \mathcal{N}_3(0, \Sigma)$ as in Magder and Zeger (1996). The only difference is that we restrict our conditions to $\mathbb{E}b = 0$, which is more in line with the spherical setup.

Then we consider the Gaussian mixture G with some mixture distribution H on the covariance matrix Σ, such that we get as density of b

$$g(b) = \int \varphi_{(0,\Sigma)}(b) dH(\Sigma)$$

where $\varphi_{(0,\Sigma)}(b)$ is the density of $\mathcal{N}_3(0, \Sigma)$.

Given data $x_1, \cdots, x_n$ i.i.d. from some spherical distribution, which is a mixture of simple distribution with various patameters.

We approximate it by a two-step procedure with hidden data $B_1, \cdots, B_n$ from g. Let $(x_j; B_j)$ $(j = 1, \cdots, n)$ be i.i.d. with $\mathcal{L}(B_j)$ has density $g(b)$ and $\mathcal{L}(x_j | B_j)$ has density $f(x; B_j)$.

The parameter of the mixture density $g(b)$ is the mixture distribution H. The density of the observed data x_j is, then, given by

$$\int_{\mathbb{R}^3} f(x_j; b) g(b) db$$

i.e. a continuous mixture of the densities $f(x; b)$ and the likelihood is

$$\begin{aligned} L(H|X) &= \prod_{j=1}^{n} \int_{\mathbb{R}^3} f(x_j; b) g(b) db \\ &= \prod_{j=1}^{n} \iint_{\mathbb{R}^3} f(x_j; b) \varphi_{0,\Sigma}(b) dH(\Sigma) db \end{aligned} \tag{4.15}$$

It turns out that the maximum is assumed at a discrete distribution H with finite support if we constrain Σ to be bounded away from 0, e.g. require $\det \Sigma \geq h$ for some $h > 0$.

h has the usual interpretation as a parameter controlling the smoothness of the final estimate. The smaller h, the rougher the spherical density estimate will be. In the limit $h \to 0$, we would

get the empirical distribution of the data. Mark that (4.15) is only a pseudo likelihood as we do not assume that $\mathcal{L}(x_j|B_j = b)$ really has density $f(x;b)$.

Let Γ_h denote the set of distributions H over the covariance matrices Σ which have all mass concentrated on $\{\Sigma, \det\Sigma \geq h\}$, then we have

Theorem 4.5.1
Assume that for almost all x, f(x;b) is bounded and continuous in b and $f(x;b) \to 0$ *for* $\|b\| \to \infty$. *For any* $h > 0$, *L(H/x) is maximized over* Γ_h *by a discrete measure with finite support* $\{\Sigma_1, \cdots, \Sigma_M\}$ *for some* $M \leq n$. *Hence, the maximum likelihood estimate of g is almost surely of the form*

$$\hat{g}(b) = \sum_{k=1}^{M} \hat{\pi}_k \varphi_{(0,\hat{\Sigma}_k)}(b)$$

where all $\hat{\Sigma}_k$ satisfy $\det\hat{\Sigma}_k = h$.

Proof The result follows from Theorem 2 of Magder and Zeger (1996) as, checking the proof of said Theorem, it is easily seen that fixing the means of the Gaussian to 0 does not change the arguments. The main property of the Gaussian family that it is invariant under convolution may still be used to apply a general result on mixtures by Lindsay (1983). □

Mark that Magder and Zeger (1996) use in their proof that $f(x_j;b) \to 0$ for $\|b\| \to \infty$ for $j = 1, \cdots n$, which obviously holds almost surely under our condition on f.

Many well-know spherical distributions have a density proportional to $\Phi(b^T x)$ for some given function Φ. Furthermore, the conditions on $f(x;b)$ follows immediately from conditions on Φ and the scaling factor:

Corollary 4.5.2
If $f(x;b) = c(\|b\|)\Phi(b^T x)$, *then the Theorem 4.5.1 holds if* $c(\beta)\Phi(\beta u)$ *is continuous and bounded in u and* β *and* $c(\beta)\Phi(\beta u) \to 0$ *for* $\beta \to \infty$ *for almost all* $u \in [-1, 1]$.

These conditions are satisfied for our two main examples, the Watson and the von Mises-Fisher distribution.

Lemma 4.5.3
The condition of the Corollary 4.5.2 is satisfied for

$$\Phi(u) = e^{\mathrm{sgn}(u)u^2}$$

i.e. for the Watson density.

Proof For the sphere in $\mathbb{R}^3$, the Watson density has the form

$$f(x;b) = \frac{1}{M_\kappa} e^{\kappa(\mu^T x)^2} = \frac{1}{M_\kappa} e^{\beta^2 \mathrm{sgn}(\mu^T x)(\mu^T x)^2}$$

where

$$M_\kappa = c\int_{-1}^{1} e^{\kappa t^2} dt = c\int_{-1}^{1} e^{\mathrm{sgn}(\kappa)\beta^2 t^2} dt$$

with some universal constant c depending only on the dimension, here $c = 3$.

(i) We first show that $f(x;b) \to 0$ for $\|b\|^2 = \beta^2 = |\kappa| \to \infty$ for all x except for $x = \pm\mu$ in case of $\kappa > 0$ and except for $x \perp \mu$ in case of $\kappa < 0$, which both are sets of measure 0 on the sphere.

(a) We first consider the case $\kappa > 0$, $x \neq \pm\mu$. Then

$$|\mu^T x| = \tau < 1$$

$$\begin{aligned}
\frac{1}{f(x;b)} &= \frac{M_\kappa}{e^{\kappa\tau^2}} \\
&= c \int_{-1}^{1} e^{\kappa(t^2-\tau^2)} dt \\
&= c \int_{-\tau}^{\tau} e^{\kappa(t^2-\tau^2)} dt + 2c \int_{\tau}^{1} e^{\kappa(t^2-\tau^2)} dt \\
&\geq 2c \int_{\tau}^{1} e^{\kappa(t^2-\tau^2)} dt \\
&\geq 2c \int_{\tau}^{1} e^{\kappa(t-\tau)^2} dt \\
&= 2c \int_{0}^{1-\tau} e^{\kappa s^2} ds \\
&\geq 2c \int_{0}^{\delta} e^{\kappa s^2} ds + 2ce^{\kappa\delta^2} \int_{\delta}^{1-\tau} ds \\
&\geq 2c(1-\tau-\delta)e^{\kappa\delta^2} \to \infty
\end{aligned}$$

for $\kappa \to \infty$, for all suitably small $\delta > 0$. Therefore $f(x;b) \to 0$ for $\kappa \to \infty$.

(b) Now, for $\kappa < 0$, $\mu^T x \neq 0$, We have

$$|\mu^T x| = \tau > 0$$

and

$$\begin{aligned}
\frac{1}{f(x;b)} &= \frac{M_\kappa}{e^{-|\kappa|\tau^2}} \\
&= c \int_{-1}^{1} e^{|\kappa|(\tau^2-t^2)} dt \\
&\geq 2c \int_{0}^{\tau} e^{|\kappa|(\tau^2-t^2)} dt
\end{aligned}$$

$$\begin{aligned} &\geq 2c \int_0^{\tau} e^{|\kappa|(\tau-t)^2} dt \\ &= 2c \int_0^{\tau} e^{|\kappa|s^2} ds \to \infty \end{aligned}$$

for $|\kappa| \to \infty$ as in (i), and again $f(x;b) \to 0$ for $\kappa \to \infty$.

(ii) The continuity of $f(x;b)$ is immediate. The boundedness follows from the continuity, the fact that $\mu \in \mathbb{S}^2$ which is compact and that $\beta > 0$ and $f(x;b) \to 0$ for $\beta \to \infty$.

□

Lemma 4.5.4
The condition of the Corollary 4.5.2 is satisfied for

$$\Phi(u) = e^u$$

i.e. for the von Mises-Fisher distribution.

Proof The density of the von Mises-Fisher distribution is

$$f(x;b) = c\kappa^{1/2} \frac{1}{I_{1/2}(\kappa)} e^{\kappa \mu^T x} = c \frac{1}{I_\kappa} e^{\kappa \mu^T x}$$

for $\kappa \geq 0, \mu = 1$, where

$$I_\kappa = \frac{1}{\sqrt{\kappa}} I_{1/2}(\kappa) = \frac{1}{2\pi\sqrt{\kappa}} \int_0^{2\pi} \cos\left(\frac{\theta}{2}\right) e^{\kappa \cos\theta} d\theta$$

or, alternatively,

$$I_\kappa = \frac{1}{\pi\sqrt{\kappa}} \int_0^{\pi} \cos\theta e^{\kappa \cos(2\theta)} d\theta.$$

Using

$$\begin{aligned} \int_{\pi/2}^{\pi} \cos\theta e^{\kappa \cos(2\theta)} d\theta &= \int_0^{\pi/2} \cos(\theta + \pi/2) e^{\kappa \cos(2\theta + \pi)} d\theta \\ &= \int_0^{\pi/2} -\sin\theta e^{-\kappa \cos(2\theta)} d\theta \end{aligned}$$

we also have

$$I_\kappa = \frac{1}{\pi\sqrt{\kappa}} \int_0^{\pi/2} (\cos\theta e^{\kappa \cos(2\theta)} - \sin\theta e^{-\kappa \cos(2\theta)}) d\theta \tag{4.16}$$

We define $\mu^T x = \tau$, $|\tau| \leq 1$, therefore

$$\tau = \cos(2\theta^*)$$

for some unique $\theta^* \in [0, \pi/2]$. Assume $|\mu^T x| \neq 1$, i.e. $\theta^* \in (0, \pi/2)$. We have

$$\frac{1}{f(x;b)} = \text{const} \cdot \frac{I_\kappa}{e^{\kappa \mu^T x}} = \text{const} \cdot \frac{I_\kappa}{e^{\kappa \cos(2\theta^*)}}$$

We use (4.16) for I_κ, and we consider only the first term , as the second one can be dealt with analogously.

$$\begin{aligned}
& \frac{1}{\sqrt{\kappa}} \int_0^{\pi/2} \cos\theta e^{\kappa\cos(2\theta)} e^{-\kappa\cos(2\theta^*)} d\theta \\
= & \frac{1}{\sqrt{\kappa}} \int_0^{\pi/2} \cos\theta e^{\kappa(\cos(2\theta)-\cos(2\theta^*))} d\theta \\
\geq & \frac{1}{\sqrt{\kappa}} \int_0^{\theta^*} \cos\theta e^{\kappa(\cos(2\theta)-\cos(2\theta^*))} d\theta \\
\geq & \frac{1}{\sqrt{\kappa}} \int_0^{\theta^*} e^{\kappa(\cos(2\theta)-\cos(2\theta^*))} d\theta \cdot \cos\theta^*
\end{aligned}$$

As $\cos(2\theta)$ decreases on $[0, \pi/2]$, $\cos(2\theta) - \cos(2\theta^*) \geq 0$ in the exponent for $\theta \leq \theta^*$, and we have for small enough $\delta > 0$

$$\begin{aligned}
& \frac{1}{\sqrt{\kappa}} \int_0^{\pi/2} \cos\theta e^{\kappa\cos(2\theta)} e^{-\kappa\cos(2\theta^*)} d\theta \\
\geq & \frac{1}{\sqrt{\kappa}} \int_0^{\theta^*-\delta} e^{\kappa(\cos(2\theta)-\cos(2\theta^*))} d\theta \cdot \cos\theta^* \\
\geq & \frac{1}{\sqrt{\kappa}} (\theta^* - \delta)(\theta^* - \delta) \cdot e^{\kappa(\cos(2\theta^*-2\delta)-\cos(2\theta^*))} \cdot \cos\theta^* \to \infty \quad \text{for } \kappa \to \infty
\end{aligned}$$

as $\frac{1}{\sqrt{\kappa}} e^{\kappa z} \to \infty$ for $\kappa \to \infty$ if $z > 0$. □

From Theorem 4.5.1 together with the last two lemmas, we get a simple nonparametric estimate of the density $f(x)$ of the x_j of the form

$$\hat{f}_h(x) = \sum_{k=1}^{M} \hat{\pi}_k \int f(x;b) \phi_{(0,\hat{\Sigma}_k)}(b) db$$

where $\hat{\pi}_k$, $\hat{\Sigma}_k$ can be derived by maximizing the likelihood (4.15) under the constraint $\det \hat{\Sigma}_k = h$. Magder and Zeger (1996) showed hat this approach works in a setting where the numerical calculation is rather simple. In our case, the likelihood is a rather involved function of $\hat{\Sigma}_k$, $k = 1, \cdots, M$, such that constrained maximization is rather cumbersome. Therefore, we do not present simulations here, but this has to be postponed to future work.

Appendix A

Application Results

Table A.1: Estimated parameters of the mixture model of Schladitz distributions using Algorithm 6 for GF30

tube	$\hat{\pi}$	$\hat{\beta}$	$\hat{\mu}$
1	0.5341	0.2448	(0.0295,0.9794,0.1999)
	0.4659	0.3781	(0.0160,0.9753,0.2203)
2	0.1820	0.1128	(0.0012,0.7696,0.6386)
	0.8180	0.5038	(0.0500 0.9954,0.0813)
3	0.6472	0.1609	(0.0366,0.8149,0.5784)
	0.3528	0.2246	(0.0229,0.8359,0.5483)
4	0.8770	0.3861	(0.0059,0.9673,0.2534)
	0.1230	0.5298	(0.0037,0.9996,0.0275)
5	0.9853	0.3944	(0.0055,0.9697,0.2443)
	0.0417	0.5272	(0.0036,0.9995,0.0324)
6	0.2736	0.4088	(0.1301,0.9443,0.3024)
	0.7264	0.4400	(0.0399,0.9940,0.1017)
7	0.5963	0.2580	(0.0311,0.9770,0.2109)
	0.4037	0.2580	(0.0311,0.9770,0.2109)
8	0.5064	0.3173	(0.0015,0.9984,0.0561)
	0.4936	0.3177	(0.0013,0.9984,0.0559)
9	0.6234	0.3044	(0.1396,0.9805,0.1382)
	0.3766	0.3499	(0.1488,0.9811,0.1238)
10	0.5081	0.2069	(0.0436,0.9982,0.0404)
	0.4919	0.2076	(0.0437,0.9982,0.0407)

Table A.2: Estimated parameters of the mixture model of Schladitz distributions using Algorithm 7 for GF30

tube	$\hat{\pi}$	$\hat{\beta}$	$\hat{\mu}$
1	0.2843	0.2457	(0.0092,0.8981,0.4398)
	0.7157	0.4345	(0.0154,0.9977,0.0656)
2	0.6289	0.2479	(0.1099,0.9909,0.0780)
	0.3711	0.3481	(0.1397 0.9797,0.1437)
3	0.6789	0.4020	(0.0337,0.9728,0.2291)
	0.3211	0.4211	(0.0485,0.9981,0.0390)
4	0.4360	0.3711	(0.1268,0.9777,0.1677)
	0.5640	0.4718	(0.0317,0.9791,0.0404)
5	0.0747	0.1543	(0.1029,0.9631,0.2486)
	0.9253	0.5970	(0.1948,0.9446,0.2641)
6	0.7217	0.3499	(0.1621,0.9241,0.1198)
	0.2783	0.6426	(0.0461,0.9901,0.0176)
7	0.4337	0.4932	(0.2743,0.9614,0.0216)
	0.5663	0.4936	(0.2739,0.9615,0.0218)
8	0.2134	0.3231	(0.1530,0.9881,0.0182)
	0.7866	0.4955	(0.2448,0.9686,0.0430)
9	0.5317	0.3375	(0.1640,0.9865,0.0003)
	0.4683	0.4763	(0.1764,0.9831,0.0483)
10	0.3582	0.4691	(0.0051,0.9935,0.0130)
	0.6418	0.4692	(0.0051,0.9934,0.0131)

Table A.3: Estimated parameters of the mixture model of Schladitz distributions using Algorithm 6 for GF60 bottom layer

tube	$\hat{\pi}$	$\hat{\beta}$	$\hat{\mu}$
1	0.3418	0.1755	(0.0372,0.0271,0.9989)
	0.6582	0.1759	(0.0372,0.0272,0.9989)
2	0.6481	0.1520	(0.0560,0.0198,0.9982)
	0.3519	0.1522	(0.0560 0.0198,0.9982)
3	0.6384	0.1835	(0.0694,0.0389,0.9968)
	0.3616	0.1841	(0.0695,0.0389,0.9968)
4	0.6357	0.1476	(0.0096,0.0046,0.9999)
	0.3643	0.1477	(0.0093,0.0046,0.9999)
5	0.0223	0.0310	(0.0028,0.0018,1.0000)
	0.9777	0.1251	(0.0107,0.0171,0.9998)
6	0.3327	0.0893	(0.0033,0.0018,1.0000)
	0.6673	0.1596	(0.0031,0.0043,1.0000)
7	0.0396	0.0150	(0.0100,0.0094,0.9999)
	0.9604	0.1622	(0.0013,0.0016,1.0000)
8	0.3688	0.1486	(0.0119,0.0018,0.9999)
	0.6312	0.1524	(0.0136,0.0017,0.9999)
9	0.6344	0.1042	(0.0362,0.0084,0.9993)
	0.3656	0.2230	(0.0328,0.0162,0.9993)
10	0.6251	0.1773	(0.0284,0.0172,0.9994)
	0.3749	0.1774	(0.0284,0.0172,0.9994)

Table A.4: Estimated parameter of the mixture model of Schladitz distributions using Algorithm 7 for GF60 bottom layer

tube	$\hat{\pi}$	$\hat{\beta}$	$\hat{\mu}$
1	0.3889	0.1767	(0.0284,0.0150,0.9995)
	0.6111	0.1777	(0.0284,0.0186,0.9994)
2	0.1741	0.0537	(0.0219,0.0126,0.9997)
	0.8259	0.1717	(0.0210 0.0072,0.9998)
3	0.9196	0.1781	(0.0549,0.0292,0.9981)
	0.0804	0.2091	(0.0082,0.0034,1.0000)
4	0.3696	0.1832	(0.0061,0.0014,1.0000)
	0.6304	0.2063	(0.0930,0.0400,0.9949)
5	0.9611	0.2391	(0.0580,0.0901,0.9942)
	0.0389	0.5958	(0.5902,0.4875,0.6434)
6	0.8743	0.1161	(0.0161,0.0182,0.9997)
	0.1257	0.5008	(0.3795,0.1062,0.9191)
7	0.0775	0.0297	(0.0287,0.0034,0.9996)
	0.9225	0.1217	(0.0283,0.0080,0.9996)
8	0.3287	0.1404	(0.0044,0.0021,1.0000)
	0.6713	0.1616	(0.1004,0.0047,0.9949)
9	0.6030	0.3129	(0.0161,0.1411,0.9899)
	0.3970	0.3168	(0.0160,0.1447,1.9894)
10	0.2162	0.0948	(0.0046,0.0073,1.0000)
	0.7838	0.1897	(0.0113,0.0587,0.9982)

Table A.5: Estimated parameters of the mixture model of Schladitz distributions using Algorithm 6 for GF60 middle layer

tube	$\hat{\pi}$	$\hat{\beta}$	$\hat{\mu}$
1	0.6148	0.4875	(0.0283,0.9811,0.1913)
	0.3852	0.5391	(0.0309,0.9769,0.2113)
2	0.5325	0.4544	(0.0152,0.9470,0.3209)
	0.4675	0.4546	(0.0152 0.9470,0.3209)
3	0.3172	0.3530	(0.0678,0.9977,0.0062)
	0.6828	0.4479	(0.0802,0.9967,0.0094)
4	0.2637	0.3434	(0.1653,0.9847,0.0557)
	0.7363	0.9161	(0.2099,0.9244,0.3184)
5	0.1047	0.1799	(0.1358,0.9294,0.3432)
	0.8953	0.4519	(0.1985,0.9315,0.3049)
6	0.3306	0.2203	(0.0401,0.9777,0.2064)
	0.6694	0.5810	(0.1355,0.9676,0.2129)
7	0.1148	0.4757	(0.2051,0.9785,0.0206)
	0.8852	0.9904	(0.1764,0.9075,0.3812)
8	0.3561	0.2088	(0.0172,0.9997,0.0181)
	0.6439	0.8310	(0.0589,0.9542,0.2932)
9	0.6185	0.5837	(0.0435,0.8474,0.5292)
	0.3815	0.5840	(0.0435,0.8474,0.5292)
10	0.2589	0.4370	(0.1329,0.9745,0.1807)
	0.7411	0.6013	(0.1620,0.9565,0.2424)

Table A.6: Estimated parameter of the mixture model of Schladitz distributions using Algorithm 7 for GF60 middle layer

tube	$\hat{\pi}$	$\hat{\beta}$	$\hat{\mu}$
1	0.6652	0.3724	(0.1794,0.9315,0.3165)
	0.3348	0.4852	(0.1423,0.9398,0.3106)
2	0.1749	0.1915	(0.0864,0.9747,0.2060)
	0.8251	0.4971	(0.1456 0.9657,0.2152)
3	0.2890	0.2354	(0.0419,0.9988,0.0250)
	0.7110	0.8488	(0.0519,0.8990,0.4348)
4	0.3140	0.4544	(0.0151,0.9470,0.3209)
	0.6860	0.4545	(0.0152,0.9470,0.3209)
5	0.2788	0.4923	(0.6357,0.7663,0.0928)
	0.7212	0.7600	(0.5823,0.7687,0.2646)
6	0.6299	0.5895	(0.1903,0.9567,0.2202)
	0.3701	0.5895	(0.1902,0.9567,0.2202)
7	0.2304	0.3709	(0.2272,0.9650,0.1311)
	0.7696	0.9727	(0.2684,0.8664,0.4210)
8	0.3772	0.2016	(0.4488,0.8798,0.1564)
	0.6228	0.4832	(0.0391,0.9945,0.0969)
9	0.5875	0.5837	(0.0435,0.8474,0.5292)
	0.4125	0.5839	(0.0435,0.8474,0.5292)
10	0.1418	0.2153	(0.0365,0.8685,0.4944)
	0.8582	0.5571	(0.0195,0.8861,0.4631)

Table A.7: Estimated parameters of the mixture model of Schladitz distributions using Algorithm 6 for GF60 top layer

tube	$\hat{\pi}$	$\hat{\beta}$	$\hat{\mu}$
1	0.2307	0.0566	(0.0111,0.0083,0.9999)
	0.7693	0.2245	(0.0726,0.0100,0.9973)
2	0.1082	0.0605	(0.0166,0.0029,0.9999)
	0.8918	0.1810	(0.0288 0.0038,0.9996)
3	0.8560	0.1209	(0.0081,0.0315,0.9995)
	0.1440	0.1547	(0.0036,0.0334,0.9994)
4	0.3854	0.1339	(0.0323,0.0789,0.9964)
	0.6146	0.1340	(0.0323,0.0789,0.9964)
5	0.6196	0.1587	(0.0236,0.0177,0.9996)
	0.3804	0.1587	(0.0236,0.0177,0.9996)
6	0.6299	0.1708	(0.0440,0.0214,0.9988)
	0.3701	0.1712	(0.0440,0.0214,0.9988)
7	0.6466	0.1893	(0.0798,0.0202,0.9966)
	0.3534	0.1896	(0.0795,0.0214,0.9966)
8	0.8096	0.1392	(0.0123,0.0130,0.9998)
	0.1904	0.3433	(0.0676,0.0574,0.9961)
9	0.6273	0.2367	(0.0793,0.0215,0.9966)
	0.3727	0.2369	(0.0794,0.0215,0.9966)
10	0.6388	0.1740	(0.0484,0.0124,0.9987)
	0.3612	0.1742	(0.0484,0.0124,0.9987)

Table A.8: Estimated parameter of the mixture model of Schladitz distributions using Algorithm 7 for GF60 top layer

tube	$\hat{\pi}$	$\hat{\beta}$	$\hat{\mu}$
1	0.7772	0.1412	(0.0307,0.0004,0.9995)
	0.2228	0.1923	(0.0058,0.0038,1.0000)
2	0.2026	0.0492	(0.0110,0.0143,0.9998)
	0.7974	0.1520	(0.0117 0.0415,0.9991)
3	0.0642	0.0175	(0.0039,0.0007,1.0000)
	0.9358	0.1515	(0.0026,0.0002,1.0000)
4	0.6078	0.1174	(0.0018,0.0568,0.9984)
	0.3922	0.1185	(0.0005,0.0019,1.0000)
5	0.1220	0.0488	(0.0149,0.0215,0.9997)
	0.8780	0.1949	(0.0028,0.0340,0.9994)
6	0.6527	0.1344	(0.0121,0.0707,0.9974)
	0.3473	0.1444	(0.0010,0.0151,0.9999
7	0.9282	0.1714	(0.0519,0.0134,0.9986)
	0.0718	0.2122	(0.0186,0.0080,0.9998)
8	0.6189	0.2367	(0.0793,0.0215,0.9966)
	0.3811	0.2369	(0.0795,0.0216,0.9966)
9	0.7875	0.1710	(0.0563,0.0281,0.9980)
	0.2125	0.1793	(0.0019,0.0037,1.0000)
10	0.1267	0.0486	(0.0390,0.0075,0.9992)
	0.8733	0.2007	(0.0823,0.0071,0.9966)

Table A.9: Estimated parameters of the mixture model of Watson distributions obtained by Schlachter (2012)

foam	$\hat{\pi}$	$\hat{\kappa}$	$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\mu_3}$
	0.70106	-3.37221	0.99693	0.02202	
A	0.29894	2.00992	0.07826	-0.10440	
	0		-0.00333	-0.99429	
	0.52873	-3.8979	0.99127	0.06363	
B	0.47127	2.22254	0.10226	-0.14811	
	0		-0.08318	-0.98692	
	0.91062	-3.72463	0.99843	0.01589	
C	0.08938	2.47566	0.03009	0.08681	
	0		0.04719	-0.99610	
	0.68042	-4.03783	0.99871	-0.04822	-0.01831
D	0.07318	1.61011	-0.04063	-0.93259	0.29206
	0.24640	2.16700	-0.03037	0.35771	-0.95622
	0.29282	2.25718	0.05423	0.03945	0.02967
E	0.32573	2.25303	-0.96368	-0.25042	-0.60079
	0.38145	2.30360	-0.26150	-0.96733	0.79885
	0.75360	-3.25646	0.99855	-0.02775	-0.04618
F	0.08847	1.30225	-0.05001	-0.63990	-0.30016
	0.15792	1.76363	0.01981	-0.76796	0.95277
	0.44085	-3.84855	0.99090	-0.14631	0.07890
G	0.26493	1.99174	-0.08742	-0.29611	-0.31956
	0.29422	2.03488	-0.10232	-0.94388	0.94420
	0.83012	-4.34502	0.99815	0.07007	0.02233
H	0.11168	2.04129	0.05384	-0.98441	0.03885
	0.05820	1.80007	0.02835	0.16131	-0.99900

Appendix B

Curriculum Vitae

since 06/2009	PhD Student in Mathematics (Mathematics Department of the University of Kaiserslautern, Germany)
05/2009	Diplom Degree in Mathematics (Clausthal University of Technology, Germany)
11/2004 – 05/2009	Study of Business Mathematics (Clausthal University of Technology, Germany)
07/2004	Bachelor Degree in Medicine (Peking University Health Science Center, China)
10/1999 – 07/2004	Study of Clinical Medicine (Peking University Health Science Center, China)
07/1999	High School Diploma in Beijing, China

Lebenslauf

Juni 2009 - heute	PhD im Fach Mathematik (Fachbereich Mathematik, Technische Universität Kaiserslautern)
Mai 2009	Diplom der Wirtschaftsmathematikerin (Technische Universität Clausthal, Deutschland)
Nov. 2004 – Mai 2009	Studium der Wirtschaftsmathematik (Technische Universität Clausthal, Deutschland)
Juli 2004	Bachelor der Humanmedizin (Peking University Health Science Center, China)
Okt. 1999 – Juli 2004	Studium der Medizin (Peking University Health Science Center, China)
Juli 1999	Abitur in Beijing, China

Bibliography

H. Altendorf. 3D Morphological Analysis and Modeling of Random Fiber Networks - applied on Glass Fiber Reinforced Composites. PhD thesis, Technische Universität Kaiserslautern, 2012.

H. Altendorf and D. Jeulin. 3D directional mathematical morphology for analysis of fiber orientations. Image Anal Stereol, 28:143–153, 2009.

A. Banerjee, I. S Dhillon, J. Ghosh, and S. Sra. Clustering on the unit hypersphere using von Mises-Fisher distributions. Journal of Machine Learning Research, 6(2):1345, 2006.

D. J Best and N. I Fisher. Goodness-of-fit and discordancy tests for samples from the watson distribution on the sphere. Australian Journal of Statistics, 28(1):13–31, 1986.

A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the em algorithm. Journal of the Royal Statistical Society. Series B (Methodological), pages 1–38, 1977.

D. Eberly, R. Gardner, B. Morse, S. Pizer, and C. Scharlach. Ridges for image analysis. J. Mathematical Imaging and Vision, 4(4):353–373, 1994.

A. Figueiredo and P. Gomes. Performance of the em algorithm on the identification of a mixture of distributions defined on the hypersphere. REVSTAT-Statistical Journal, 4(2):111–130, 2006.

N. I. Fisher, T. Lewis, and M. E. Willcox. Tests of discordancy for samples from fisher's distribution on the sphere. Journal of the Royal Statistical Society. Series C (Applied Statistics), 30(3):pp. 230–237, 1981. ISSN 00359254. URL `http://www.jstor.org/stable/2346346`.

N. I Fisher, T. Lewis, and B. J. J. Embleton. Statistical analysis of spherical data. Cambridge Univ Pr, 1993. ISBN 0521456991.

J. Franke, J.P. Stockis, J. Tadjuidje-Kamgaing, and WK Li. Mixtures of nonparametric autoregressions. Journal of Nonparametric Statistics, 23(2):287–303, 2011.

Fraunhofer ITWM. Mavi - modular algorithmus for volume images, 2012.

B.G. Lindsay. The geometry of mixture likelihoods, part ii: The exponential family. The Annals of Statistics, 11(3):783–792, 1983.

A. K. Louis, M. Riplinger, M. Spiess, and E. Spodarev. Inversion algorithms for the spherical radon and cosine transform. Inverse Problems, 27:035015, 2011.

L.S. Magder and S.L. Zeger. A smooth nonparametric estimate of a mixing distribution using mixtures of gaussians. Journal of the American Statistical Association, 91(435):1141–1151, 1996.

J. Ohser and K. Schladitz. 3D images of materials structures: processing and analysis. Wiley-VCH, 2009.

D. Peel, W. J Whiten, and G. J McLachlan. Fitting mixtures of kent distributions to aid in joint set identification. Journal of the American Statistical Association, 96(453):56–63, 2001.

C. Redenbach and I. Vecchio. Statistical analysis and stochastic modelling of fibrecomposites. Composites Science and Technology, 71:107–112, 2011.

Anna-Lena Schlachter. Segmentierung und statistische analyse des wandsystems in schäumen. Master's thesis, Technische Universität Kaiserslautern, 2012.

K. Schladitz, S. Peters, D. Reinel-Bitzer, A. Wiegmann, and J. Ohser. Design of acoustic trim based on geometric modeling and flow simulation for non-woven. Computational materials science, 38(1):55–56, 2006. ISSN 0927-0256.

R.J. Serfling. Approximation theorems of mathematical statistics. Wiley Series in Probability and Statistics. Wiley, 2001. ISBN 9780471219279. URL http://books.google.de/books?id=a30-JfZboegC.

S. Sra. A short note on parameter approximation for von mises-fisher distributions: and a fast implementation of i s (x). Computational Statistics, 27(1):177–190, 2012.

Suvrit Sra and Dmitrii Karp. The multivariate watson distribution: Maximum-likelihood estimation and other aspects. eprint arXiv:1104.4422, 96(453):56–63, 2011.

P.D. Stoyan, D.W.S. Kendall, and J. Mecke. Stochastic Geometry and Its Applications. Wiley-interscience paperback series. John Wiley & Sons, 2008. ISBN 9780470743645. URL http://books.google.de/books?id=Le3kQgAACAAJ.

D.E. Tyler. Statistical analysis for the angular central gaussian distribution on the sphere. Biometrika, 74(3):579–589, 1987.

O. Wirjadi. Models and Algorithms for Image-Based Analysis of Microstructures. PhD thesis, Technische Universität Kaiserslautern, 2009.

CF Wu. On the convergence properties of the em algorithm. The Annals of Statistics, 11(1): 95–103, 1983.