Statistical Analysis of Diffusion Tensor Imaging

Diwei Zhou, MSc.

Division of Statistics

School of Mathematical Sciences

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

January 2010

Abstract

This thesis considers the statistical analysis of diffusion tensor imaging (DTI). DTI is an advanced magnetic resonance imaging (MRI) method that provides a unique insight into biological microstructure *in vivo* by directionally describing the water molecular diffusion. We firstly develop a Bayesian multi-tensor model with reparameterisation for capturing water diffusion at voxels with one or more distinct fibre orientations. Our model substantially alleviates the non-identifiability issue present in the standard multi-tensor model. A Markov chain Monte Carlo (MCMC) algorithm is then developed to study the uncertainty of the model parameters based on the posterior distribution. We apply the Bayesian method to Monte Carlo (MC) simulated datasets as well as a healthy human brain dataset. A region containing crossing fibre bundles is investigated using our multi-tensor model with automatic model selection.

A diffusion tensor, a covariance matrix related to the molecular displacement at a particular voxel in the brain, is in the non-Euclidean space of 3×3 positive semidefinite symmetric matrices. We define the sample mean of tensor data to be the Fréchet mean. We carry out the non-Euclidean statistical analysis of diffusion tensor data. The primary focus is on the use of Procrustes size-and-shape space. Comparisons are made with other non-Euclidean techniques, including the log-Euclidean, Riemannian, Cholesky, root Euclidean and power Euclidean methods. The weighted generalised Procrustes analysis has been developed to efficiently interpolate and smooth an arbitrary number of tensors with the flexibility of controlling individual contributions. A new anisotropy measure, Procrustes Anisotropy is defined and compared with other widely

used anisotropy measures. All methods are illustrated through synthetic examples as well as white matter tractography of a healthy human brain.

Finally, we use Giné's statistic to design uniformly distributed diffusion gradient direction schemes with different numbers of directions. MC simulation studies are carried out to compare effects of Giné's and widely used Jones' schemes on tensor estimation. We conclude by discussing potential areas for further research.

Acknowledgements

First, I am most grateful to my supervisors Professor Ian Dryden, Dr. Alexey Koloydenko and Dr. Li Bai for their guidance, insight and encouragement during the course of my PhD studies. Without their insightful commentary and impressive depth of knowledge, this thesis would not have been possible. I am also grateful to Dr. Paul Morgan (Medical University of South Carolina) for providing the brain data, and to Professor Dorothee Auer, Dr. Christopher Tench and Stamatis Sotiropoulos from the EU funded CMIAG Centre at the University of Nottingham for discussions related to my research.

I would not have come this far without the instruction and inspiration gained during my MSc studies in Nottingham. I thank Professor Andy Wood and Dr. Neil Butler (1965-2009) for their advice during my MSc course and PhD application.

I would eternally like to thank my parents for their neverending support and encouragement, not only over the past few years of my PhD, but also throughout the whole of my education and life.

Lastly, many thanks go to all my friends, colleagues and acquaintances in the UK and in China.

The work is supported by the European Commission FP6 Marie Curie programme through the CMIAG Research Training Network.

Declaration

Except where specific reference is made to other sources, the work presented in this thesis is the original work of the author. It has not been submitted, in whole or in part, for any other degree. Certain of the results have already been published.

Signature:

Date:

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The research presented in this thesis is concerned with the statistical analysis of diffusion tensor imaging (DTI). DTI is a relatively new magnetic resonance imaging (MRI) method that captures the directionality and magnitude of water diffusion in biological tissue using the diffusion tensor [Basser *et al*., 1994; Basser and Pierpaoli, 1996; Le Bihan *et al*., 2001]. DTI has been applied into the study of diseases such as multiple sclerosis [Hesseltine *et al*., 2006], schizophrenia [Buchsbaum *et al*., 2006], and stroke [Le Bihan *et al*., 2001]. White matter tractography [Basser *et al*., 2000] is another promising application of DTI for investigating brain study.

1.1 Background to Diffusion MRI

1.1.1 Molecular diffusion

Diffusion (or Brownian motion) is the random motion of molecules due to thermal energy. Hänggi and Marchesoni [2005] gave a good introduction to the history of Brownian motion. Brownian motion was first observed by Robert Brown in 1828 when he monitored the pollen grains moving randomly suspended in water. The theoretical framework that could explain the experimentally observed phenomenon of diffusion was set up by Einstein in 1905. The theory was tested experimentally by Perrin in 1909, who in so doing proved the kinetic molecular theory of gases. Perrin was awarded the 1926 Nobel Prize

Figure 1.1: Diffusion in two different types of medium. (a) isotopic diffusion has similar displacements in all directions (b) anisotropic diffusion in structured medium has greater molecular displacement along one direction.

in physics for this work.

In this thesis, we focus on the water molecular diffusion in the tissue of the human brain. The tissue of the human brain can be divided into three classes, *white matter*, *grey matter* and *cerebrospinal fluid*. In grey matter and cerebrospinal fluid water molecules can diffuse more or less freely in all directions (isotropic diffusion) (see Figure 1.1 (a)). However, in white matter, molecular diffusion is more restricted (see Figure 1.1 (b)). The longitudinally oriented tissue structures, the densely packed axons and in particular their membranes, which are widely assumed to be the main barrier for water diffusion, hinder diffusion perpendicular to the fibres [Kang *et al.*, 2005]. This means that the molecular diffusion is anisotropic, with a preferential direction. The diffusion carries information about the underlying anatomical architecture of living tissues. This is exactly what we try to study using diffusion MRI.

The diffusion of water molecule can be modelled by a probability density function (p.d.f.) $P(\mathbf{r}, t)$ which describes the probability of finding a molecule in a certain position $\mathbf{r}$ at a particular time t, given the initial density $P(\mathbf{r}, 0)$. The diffusion equation is given by [e.g., Crank, 1975]

$$\frac{\partial P(\mathbf{r}, t)}{\partial t} = \nabla \cdot (\mathbf{D} \nabla P(\mathbf{r}, t)), \tag{1.1.1}$$

where $\mathbf{D}$ is the collective *diffusion coefficient*, a 3×3 symmetric positive semi-definite matrix. Some authors assume $\mathbf{D}$ to be positive definite [e.g., Kingsley, 2006a], i.e., eigenvalues of $\mathbf{D}$ are all strictly positive. In our study, eigenvalues of $\mathbf{D}$ can be zeros. In the isotropic medium, $\mathbf{D} = 4d\pi t^{3/2}\mathbf{I}_{3\times 3}$ where $\mathbf{I}_{k\times k}$ is the $k \times k$ identity matrix, and d is a diffusion constant (positive scalar) determined by the temperature, the mass of the molecules, and the nature (viscosity) of the medium. The solution to the diffusion equation (1.1.1) in the isotropic case is given by the Gaussian distribution

$$P(\mathbf{r}, t) = \frac{1}{\sqrt{4\pi t d}} \exp \frac{- \parallel \mathbf{r} \parallel^2}{4td}. \tag{1.1.2}$$

In the presence of anisotropy, diffusion can no longer be characterised by a single scalar, but requires a more advanced mathematical quantity which can fully describe the directional displacement of molecules. For the anisotropic diffusion, the solution to Equation (1.1.1) is given by

$$P(\mathbf{r}, t) = \frac{1}{(4\pi t)^{\frac{3}{2}} |\mathbf{D}|^{\frac{1}{2}}} \exp(-\frac{\mathbf{r}^T \mathbf{D}^{-1} \mathbf{r}}{4t}), \tag{1.1.3}$$

where $\mathbf{D}$ is the *diffusion tensor*. The covariance matrix of this trivariate Gaussian distribution is $2\mathbf{D}t$. The diffusion tensor $\mathbf{D}$ plays an important role in our research. More details about $\mathbf{D}$ will be discussed in Section 1.2.1.

1.1.2 Overview of magnetic resonance imaging

It is helpful to understand the background of MRI before gaining a deeper insight into Diffusion MRI.

Magnetic resonance imaging (MRI) is a widely used medical imaging technique to visualise the structure and function of the body. MRI has particular

advantages. Namely, it is non-invasive, uses non-ionising radiation, and has a high soft-tissue resolution and discrimination in any imaging plane [Hagmann *et al.*, 2006]. As with all medical imaging techniques, MRI is a relatively new technology with its foundations beginning during the year of 1946. The first successful nuclear magnetic resonance experiment was made in 1946 independently by Felix Bloch and Edward Purcell in the United States [Hornak, 2008]. For this discovery Bloch and Purcell were awarded the Nobel Prize for Physics in 1952 [Vlaardingerbroek and den Boer, 1999, p.1]. Paul Lauterbur in 1973 demonstrated magnetic resonance imaging on small test tube samples [Woodward, 2001, p.8]. In 1977, Peter Mansfield at the University of Nottingham developed the echo-planar imaging (EPI) technique. For their discoveries concerning magnetic resonance imaging, Paul Lauterbur and Peter Mansfield were awarded the Nobel Prize in Medicine in 2003 [e.g., see Waxman, 2005, p.16]. The first commercial MR scanner in Europe (from Picker Ltd.) was installed in 1983 in the Department of Diagnostic Radiology at the University of Manchester Medical School [Yudofsky and Hales, 2004, p.91]. MRI has been in widespread use for less than 20 years compared with over 100 years for X-rays. MRI is clearly a young, but growing science.

What is *magnetic resonance imaging*? Let us explain it by explaining its following main ingredients: 1) the *proton spins* without and within a magnetic field, 2) *energetic interaction* by applying an additional electromagnetic pulse, 3) the *MRI signal*, and 4) how the MRI image is obtained.

As elementary particles, neutrons and protons have the intrinsic quantum mechanical property of spin. The proton of the hydrogen (^{1}H) nucleus is of the most interest in clinical MRI. This is because each water molecule (H_2O) has two hydrogen atoms and water is abundant in most parts of the body, and also because the hydrogen (^{1}H) nucleus is one of the most MR-sensitive isotopes.

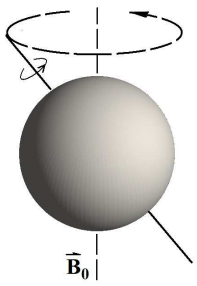

Figure 1.2: Proton precession. The "spinning" proton "rotates about" the axis of the external magnetic field.

According to the Pauli exclusion principle [Casten, 1990, p.16], the net spin of the nucleus is non-zero if the number of neutrons or protons is odd. Therefore hydrogen nuclei have non-zero net spin, and they behave like rotating magnets, represented by vectors. In the absence of an external magnetic field, the protons present no net (that is to say macroscopic) magnetisation because the spins have independent and randomly distributed directions.

However, in the presence of an external static magnetic field $\overrightarrow{B_0}$, the spins align along $\overrightarrow{B_0}$, with the same or opposite orientation, which are called parallel or anti-parallel spins, respectively. According to the Boltzmann statistic, within the $\overrightarrow{B_0}$ magnetic field, there are more parallel spins (low energy state) than anti-parallel spins (high energy state). Due to this slight excess of parallel spins, the net magnetisation is non-zero, and the protons bias toward the lower energy state. Meanwhile, the protons also rotate about the axis of $\overrightarrow{B_0}$ which is called *precession* (see Figure 1.2). The frequency of this rotation is termed *Larmor frequency* (precessional frequency) ω_0 which is proportional to $|\overrightarrow{B_0}|$, i.e. $\omega_0 = \gamma B_0$, where γ is a known constant, the gyroscopic ratio.

Exchange of energy between two systems at a specific frequency is called *resonance*. Magnetic resonance corresponds to the *energetic interaction* between spins and electromagnetic radio frequency (RF) pulse. In response to the RF pulse, the protons absorb the electromagnetic energy of the pulse, which is known as *excitation*. When the system returns from this state of imbalance to

equilibrium (relaxation), two independent forms of energy loss happen : T_1 relaxation and T_2 relaxation. Here T_1 and T_2 represent two time constants in these two relaxations. During T_1 relaxation, or spin-lattice relaxation, the return of the excited nuclei from the high energy to the low energy state is associated with the loss of energy to the surrounding nuclei. T_2 relaxation, or spin-spin relaxation, occurs when the spins in the high and low energy state exchange energy but do not lose energy to the surrounding lattice.

During the relaxation, there is an emission of electromagnetic energy which is the MRI signal detected by the scanner. The MRI signal provides the location of protons and their concentration (proton density). In order to obtain a human readable MRI image, Mansfield and Grannell [1973] proposed to describe the MRI signal in a reciprocal space (the **k** space) and to transform the MRI signal to the spatial domain by the inverse discrete Fourier transform. The resulting **k**-space MRI images have signal intensity proportional to the proton density [Schempp, 1998].

We now review some general concepts of measurement in medical imaging. As mentioned before, an MR *image* is produced by the MRI scanner, and has an intensity that depends on the characteristic of the scanner and the tissue. The MR image comes from a slice of tissue, and this slice has a specified thickness (usually 1-10 *mm*). Each *voxel* (volume element) in the image in fact derives from a cuboidal box-shaped piece of tissue (see Figure 1.3 (a)). If an image consists of a two-dimensional matrix, then each entry of the matrix is called a *pixel* (for picture element), and represents a measurement over a 1-4 mm^2 region. The *image intensity* is the value of each pixel which is characterised by the concentration of protons in the tissue (see Figure 1.3 (b)). A *map* can be formed by computing a scalar quantity for each voxel of the same piece of tissue from one or several MR images, which has the appearance of an image.

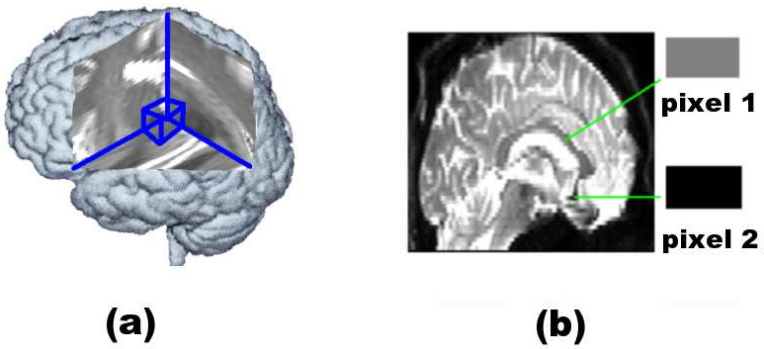

Figure 1.3: Elements in a medical image. (a) Voxel (blue) in the brain cutout adapted from Mäkelä *et al*. [2001]. ' 'Voxel' stands for a volume element which corresponds to a cuboidal piece of tissue. In Diffusion MRI, the actual voxel size is commonly between 1 and 10 mm^3 (the voxel shown in this Figure has been magnified for demonstration). (b) 'Pixel' refers to a picture element. The image intensity of pixel 1 is different from that of pixel 2, hence the difference in their brightness levels.

A device capable of producing MR images is called an 'MRI scanner'. An MRI scanner has its own reference frame, see Figure 1.4 (a). Most human brain MR images are acquired and displayed in three main orientations: coronal (frontal-anterior), sagittal (left-right), axial (inferior-superior) orientations (see Figure 1.4 (b)).

In conventional MRI, T_1 and T_2-weighted images are two widely used image contrasts due to their sensitivities to specific tissue properties. Low T_1 values are displayed bright whereas low T_2 values are displayed dark. Abnormal tissues tend to have higher signal intensities in T_1 and T_2-weighted images due to the increased water content.

1.1.3 Diffusion-weighted imaging

How does water molecular diffusion affect the MRI signal? Diffusion-weighted imaging (DWI) introduces *diffusion weighting* along a specific direction and is a modification of the conventional MRI. The pulsed-gradient spin-echo (PGSE) sequence [Stejskal and Tanner, 1965; Callaghan, 1995], a phase step approach,

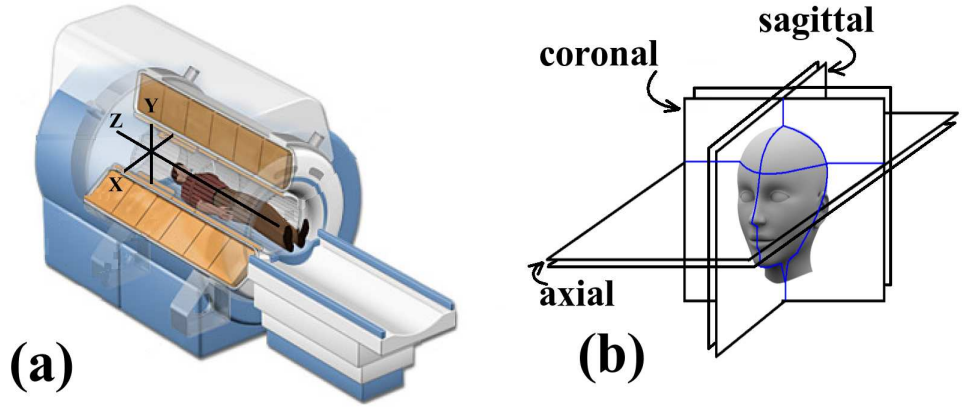

Figure 1.4: Scanner reference frame and viewing orientations. (a) Scanner reference frame $[x, y, z]$ (adapted from Coyne [1999]) and (b) three viewing orientations of the human head (adapted from Hiroshi [2009]).

has been used commonly for diffusion-weighted MRI. We have discussed that in an external static magnetic field $\overrightarrow{B_0}$ maintained by the MRI scanner the net magnetisation of the spins is non-zero (see Section 1.1.2). After applying a 90° RF pulse, the spins rotate perpendicularly to $\overrightarrow{B_0}$ in a transverse plane and then precess along $\overrightarrow{B_0}$ at the Larmor frequency. Gradually the spins precessions to dephase and the net magnetisation decays. At time t^*, the 180° RF pulse is applied which attenuates the phase of each spin. The spins return to phase at time TE (echo time) and the 'spin echo' occurs. A DWI experiment is carried out by applying diffusion gradients along specific directions (*diffusion gradient directions*) during two periods, i.e. 0 to TE/2 and TE/2 to TE respectively. Let $\mathbf{g}_i$ be the ith diffusion gradient direction, where $\mathbf{g}_i \in \mathbb{R}P^2$ ($\mathbf{g}_i \equiv -\mathbf{g}_i$, and $\| \mathbf{g}_i \| = 1$), $i = 1, ..., N$, and $\mathbb{R}P^k$ is the real-projective space of unsigned directions in $\mathbb{R}^{k+1}$. A known parameter b (called the b-value) characterises the strength of the diffusion gradient. The diffusion-weighted MRI signal can be described in a $\mathbf{q}$-space. A reciprocal space vector $\mathbf{q}$ is relative to the diffusion gradient direction and a delta function of molecular displacement from the Fourier transform. More details of $\mathbf{q}$-space imaging have been discussed in Callaghan [1995].

DWI gives a reduced MRI signal (signal attenuation) in areas where the dif-

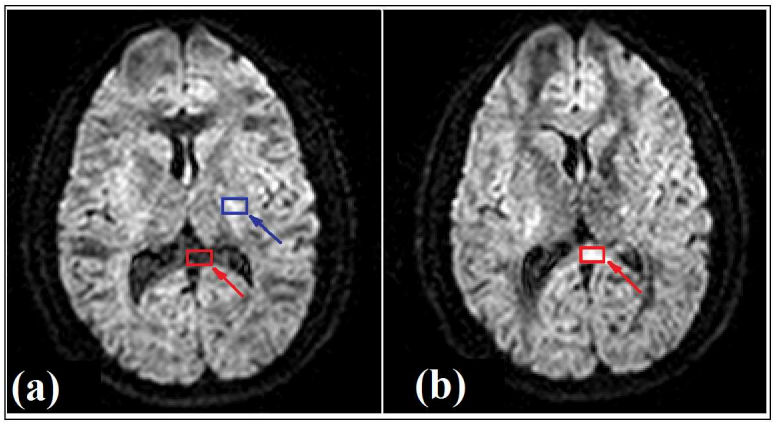

Figure 1.5: Two raw DWI images (axial view) from FSL package along two particular diffusion gradient directions. In (a) red region (darker) has higher diffusion than that in the blue region (brighter); red region in (a) has higher diffusion than that in (b).

fusion in that direction is higher [e.g., Kingsley, 2006c]. Figure 1.5 shows two diffusion weighted images of a healthy human brain (inferior-superior or axial view), obtained by applying two distinct diffusion gradients. These two images are viewed using the packages FSL (Analysis Group, FMRIB, Oxford, UK) [Smith *et al.*, 2004]. In FSL, the image is effectively threshold at a minimum value and these voxels with lower values will be rendered transparent. For this dataset, FSL automatically thresholds all values at 0. In Figure 1.5 (a), the red region is darker than blue region, because the diffusion in red region is higher than that in blue region. Comparing the same region (red) in Figure 1.5 (a) and (b), the one in (a) is darker (higher diffusion) than that in (b) (lower diffusion).

DWI introduces different contrasts in MRI which provides a particular type of visualisation of clinical images. However, DWI is very sensitive to the choice of acquisition parameters in the scanner. Diffusion tensor imaging which is the next generation of diffusion weighted imaging (DWI) will be discussed mainly in this thesis.

1.2 Principles of Diffusion Tensor Imaging

We have seen (Section1.1.1) that water molecular diffusion is anisotropic due to the biological barriers (such as cell walls and nerve fibres). Water can diffuse easily along one direction rather than others. This anisotropic diffusion is of great interest because it carries much information about the underlying anatomical architecture of living tissues. Diffusion tensor imaging (DTI) is such a powerful diffusion MRI modality that directionally describes the water molecular diffusion using the diffusion tensor $\mathbf{D}$.

1.2.1 Diffusion tensor models

The diffusion tensor is a 3×3 covariance matrix which is estimated at each voxel in the brain, and is obtained by fitting a physically-motivated model on measurements from the Fourier transform of the molecule displacements [Basser *et al.*, 1994; Alexander, 2005].

In DTI, the water molecules at a voxel diffuse according to a multivariate Gaussian model centred on the voxel and with covariance matrix $2\mathbf{D}$. The displacement of a water molecule $\mathbf{x} \in \mathbb{R}^3$ has probability density function [Callaghan, 1995]

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{3}{2}}|2\mathbf{D}|^{\frac{1}{2}}} \exp(-\frac{1}{2}\mathbf{x}^T (2\mathbf{D})^{-1} \mathbf{x}) \tag{1.2.1}$$

where

$$\mathbf{D} = \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{pmatrix}. \tag{1.2.2}$$

is the diffusion tensor $\mathbf{D}$ which is a 3×3 symmetric positive semi-definite real

matrix. An introduction to tensors is in Simmonds [1994]. Thus, a diffusion tensor $\mathbf{D}$ is uniquely determined by the six parameters $D_{ij} \in \mathbb{R}$, $i, j = x, y, z$ which depend on the scanner reference frame, and are usually unknown and need to be estimated from experimental data. The MR scanner has a set of magnetic field gradients applied along N diffusion gradient directions $\mathbf{g}_1, ..., \mathbf{g}_N$ with scanner gradient parameter b mentioned in Section 1.1.3. The data at a voxel consist of signals $\mathbf{S} = (S_0, S_1, S_2, ..., S_N)$ which are related to the Fourier transform of the displacements in axial directions $\mathbf{g}_j \in \mathbb{R}P^2, j = 1, ..., N$ and the reading S_0 is obtained with no gradient ($b = 0$). The Fourier transform in axial direction $\mathbf{g} \in \mathbb{R}P^2$ of the multivariate Gaussian displacement is given by

$$\begin{aligned} \mathcal{F}(\mathbf{g}) &= \int \exp(i\sqrt{b}\mathbf{g}\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \exp(-b\mathbf{g}^{\mathrm{T}}\mathbf{D}\mathbf{g}) \end{aligned} \tag{1.2.3}$$

and the theoretical model for the mean μ_j of the resulting diffusion attenuated MR signal S_i is

$$\begin{aligned} \mu_j &= S_0 \mathcal{F}(\mathbf{g}) \\ &= S_0 \exp(-b\mathbf{g}_j^T \mathbf{D}\mathbf{g}_j), j = 1, ..., N, \end{aligned} \tag{1.2.4}$$

where $S_0 > 0$ is the unattenuated MR signal (i.e. $b = 0$) and is assumed to be error-free.

Asymmetry of $f(\mathbf{x})$ results in complex-valued measurements $\mathcal{F}(\mathbf{g})$ in Equation (1.2.3) [e.g. Alexander, 2005], often only the magnitude of $\mathcal{F}(\mathbf{g})$ is retained. When the distribution of measurement noise in complex MR data are assumed to be bivariate Gaussian with independent components and equal variance σ^2, the magnitude follows the Rician distribution [Gudbjartsson and Patz, 1995; Basu *et al.*, 2006], which converges to a Gaussian distribution as the signal to

noise ratio (SNR) $S_0/\sigma \rightarrow \infty$ [e.g. Alexander, 2005]. A theoretical framework has been presented for evaluating the effects of noise on estimation of diffusion tensor at moderate to high SNR (20 to 100) [Anderson, 2001]. The non-Gaussian diffusion behaviour that appears in the complex tissue microstructure which can only be studied by applying a higher b-value [Basser, 2002]. In our study, we consider Gaussian diffusion behaviour from the DWI dataset with lower b-value (see Section 2.4.2.1).

For each voxel, the noise of the measured signal attenuation is denoted as ε_i. It is commonly assumed that ε_i's are independent and identically distributed (i.i.d.) Gaussian variables, $\varepsilon_i \sim \mathbf{N}(0, \sigma^2)$. Thus, S_i along the ith diffusion gradient direction is given by

$$S_i = S_0 \exp(-b\mathbf{g}_i^{\mathrm{T}}\mathbf{D}\mathbf{g}_i) + \varepsilon_i, i = 1, ..., N. \tag{1.2.5}$$

Consequently, $S_i, i = 1, ..., N$ are independent Gaussian variables, i.e., $S_i \sim \mathbf{N}(\mu_i, \sigma^2)$. Here, the six coefficients of $\mathbf{D}$ and the variance σ^2 of the noise are unknown parameters.

Alternatively, we can assume that $\log(S_i/S_0)$ follows a Gaussian distribution, i.e.,

$$\log(S_i/S_0) \sim N(-b\mathbf{g}_i^{\mathrm{T}}\mathbf{D}\mathbf{g}_i, \sigma_{\mathrm{G}}^2), i = 1, ..., N. \tag{1.2.6}$$

Then S_i follows a Log-Gaussian distribution, i.e., $S_i \sim$ Log-Gaussian$(\mu_i^*, \sigma_{L,i}^2)$, where

$$\mu_i^* = E(S_i) = S_0 \exp(-b\mathbf{g}_i^{\mathrm{T}}\mathbf{D}\mathbf{g}_i + \frac{1}{2}\sigma_{\mathrm{G}}^2), \tag{1.2.7}$$

and

$$\sigma_{L,i}^2 = \mathrm{Var}(S_i) = S_0^2 \exp(-2b\mathbf{g}_i^{\mathrm{T}}\mathbf{D}\mathbf{g}_i + \sigma_{\mathrm{G}}^2)[\exp(\sigma_{\mathrm{G}}^2) - 1]. \tag{1.2.8}$$

1.2.2 Tensor estimations

For a measured set of signal attenuations $\mathbf{S} = (S_1, S_2, ..., S_N)$ along N diffusion gradient directions $\mathbf{g}_1, ..., \mathbf{g}_N$, there are a variety of approaches for estimating the diffusion tensor under the assumption of Gaussian distribution of the noise, such as the maximum likelihood estimation (MLE) and least squares methods.

Under the assumption of a Gaussian distribution for the noise ε_i and the DTI model in Equation (1.2.5), the p.d.f. of S_i is given by

$$f(S_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\left(S_i - \mu_i\right)^2\right\}, i = 1, ..., N, \tag{1.2.9}$$

and $S_1, ..., S_N$ are independent. The *likelihood* function of $\mathbf{D}$ (with 6 unknown parameters) and σ^2 given $\mathbf{S}$ is

$$\begin{aligned} L(\mathbf{D}, \sigma^2 \mid \mathbf{S}) &= \prod_{i=1}^{N} f(S_i) \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}[S_i - S_0 \exp\left(-b\mathbf{g}_i^{\mathrm{T}}\mathbf{D}\mathbf{g}_i\right)]^2\right\}. \end{aligned} \tag{1.2.10}$$

Furthermore, the *log-likelihood* function is

$$\begin{aligned} l(\mathbf{D}, \sigma^2 \mid \mathbf{S}) &= \log L(\mathbf{D}, \sigma^2 \mid \mathbf{S}) \\ &= -\frac{N}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}[S_i - S_0 \exp\left(-b\mathbf{g}_i^{\mathrm{T}}\mathbf{D}\mathbf{g}_i\right)]^2. \end{aligned} \tag{1.2.11}$$

Thus, the maximum likelihood estimators (m.l.e.) of six coefficients of $\mathbf{D}$ and σ^2 can be calculated by maximising $L(\mathbf{D}, \sigma^2 \mid \mathbf{S})$ in Equation (1.2.10) or $l(\mathbf{D}, \sigma^2 \mid \mathbf{S})$ in Equation (1.2.11).

Least squares methods can also be used for estimating the diffusion tensor. The

objective of least squares methods is to minimise the *sum of squared residuals* (RSS) from fitting the estimated model to the data. RSS for the single tensor model in Equation (1.2.5) is given by[Koay *et al.*, 2006]

$$RSS = \sum_{i=1}^{N} [S_i - S_0 \exp(\mathbf{X}_i \boldsymbol{\beta})]^2 \tag{1.2.12}$$

where $\mathbf{X}_i$ is the ith row of

$$\mathbf{X} = -b \begin{pmatrix} g_{1x}^2 & g_{1y}^2 & g_{1z}^2 & 2g_{1x}g_{1y} & 2g_{1y}g_{1z} & 2g_{1x}g_{1z} \\ : & : & : & : & : & : \\ g_{Nx}^2 & g_{Ny}^2 & g_{Nz}^2 & 2g_{Nx}g_{Ny} & 2g_{Ny}g_{Nz} & 2g_{Nx}g_{Nz} \end{pmatrix}, \tag{1.2.13}$$

and $\boldsymbol{\beta} = (D_{xx}, D_{yy}, D_{zz}, D_{xy}, D_{yz}, D_{xz})^{\mathrm{T}}$ is the vector representation of $\mathbf{D}$.

Nonlinear least squares (NLLS) estimator of $\mathbf{D}$ can be obtained by minimising RSS in Equation (1.2.12). To solve the NLLS problem, most routines need a good starting value. However, the single tensor model is transformably linear, and its linearised version does not require initialisation.

In the following, we apply *linear least squares* (LLS) method by linearising Equation (1.2.12) as follows:

$$RSS_{LLS} = \sum_{i=1}^{N} (y_i - \mathbf{X}_i \boldsymbol{\beta})^2 \tag{1.2.14}$$

where y_i is the ith element of $\mathbf{y} = [\log(S_1/S_0), ..., \log(S_N/S_0)]^{\mathrm{T}}$. Hence, the LLS estimator of $\mathbf{D}$ is

$$\boldsymbol{\beta}^{LLS} = [\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}, \tag{1.2.15}$$

which is unique if $\mathbf{X}$ is of full rank and $N \geq 6$.

The LLS estimates of $\mathbf{D}$ are equivalent to the ML estimates of $\mathbf{D}$ under the log-linear model (1.2.6) [Lindgren, 2003], although this does not take into ac-

count the positive semi-definiteness.

In DTI studies, the diffusion tensor is assumed to be symmetric positive semi-definite. However, the positive semi-definiteness of **D** may not be satisfied with LLS, NLLS and MLE methods. Constrained methods, such as constrained nonlinear least squares method (CNLLS), have been adapted to ensure the positive semi-definiteness of **D** [Koay *et al*., 2006; Lenglet *et al*., 2009].

1.2.3 Eigensystem of the diffusion tensor

The eigensystem of the diffusion tensor plays an important role in DTI study. The spectral decomposition of **D** is given by

$$\begin{aligned} \mathbf{D} &= \mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^T \\ &= \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \begin{pmatrix} \mathbf{v}_1^{\mathrm{T}} \\ \mathbf{v}_2^{\mathrm{T}} \\ \mathbf{v}_3^{\mathrm{T}} \end{pmatrix}, \end{aligned} \tag{1.2.16}$$

where $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$ are the three eigenvalues of **D**. Three column vectors $\mathbf{v}_i = (v_{ix}, v_{iy}, v_{iz})^{\mathrm{T}}$, $i = 1, 2, 3$ are the corresponding eigenvectors which are of unit length and orthogonal. The eigensystem of **D** provides the available information about the local diffusion in a clearly interpretable manner. Namely, the eigenvectors $\mathbf{v}_i$, $i = 1, 2, 3$ and eigenvalues λ_i, $i = 1, 2, 3$ correspond to main diffusion directions and associated diffusivities (strengths of diffusion) respectively. Thus, in particular λ_1 is the maximum diffusivity over all directions. Moreover, in voxels with a single dominant fibre orientation, $\mathbf{v}_1$ (the *principal eigenvector*) is believed to be aligned with this orientation.

In a particular voxel, the eigensystem of the diffusion tensor may be interpreted graphically as an ellipsoidal surface (*diffusion ellipsoid*) with semi-major

axis oriented along the $\mathbf{v}_1$ direction and semi-minor axes oriented along $\mathbf{v}_2$ and $\mathbf{v}_3$. The lengths of the axes in this diffusion ellipsoid are given by the square roots of the corresponding eigenvalues of each eigenvector, with semi-major axis length $\sqrt{\lambda_1}$ and semi-minor axis lengths $\sqrt{\lambda_2}$ and $\sqrt{\lambda_3}$.

In cases of isotropic diffusion, the diffusion ellipsoid takes on a spherical shape, i.e. $\lambda_1 \approx \lambda_2 \approx \lambda_3$ (see Figure 1.6 (a)). There are two extreme cases of anisotropic diffusion. For linear anisotropic diffusion, $\lambda_1 \gg \lambda_2 \approx \lambda_3$ (see Figure 1.6 (b)), the diffusion ellipsoid degenerates into a line pointing in the $\mathbf{v}_1$ direction. In the case of planar anisotropic diffusion, the diffusion ellipsoid becomes oblate, meaning that $\lambda_1 \approx \lambda_2 \gg \lambda_3$ (see Figure 1.6 (c)).

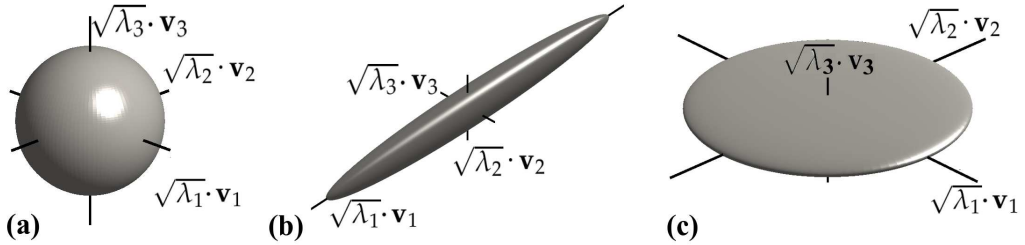

Figure 1.6: Diffusion ellipsoid is defined by eigensystem of **D**. (a) Spherical ellipsoid: isotropic diffusion $\lambda_1 \approx \lambda_2 \approx \lambda_3$. There are two extreme cases of anisotropic diffusion: (b) linear diffusion ($\lambda_1 \gg \lambda_2 \approx \lambda_3$) and (c) planar diffusion ($\lambda_1 \approx \lambda_2 \gg \lambda_3$).

1.2.4 Tensor reparameterisations

The positive semi-definiteness of **D** ensures the eigenvalues, which are proportional to the diffusivities along the main diffusion directions, are non-negative. However the MLE, LLS and NLLS estimators of **D** may have negative eigenvalues. To ensure the positive semi-definiteness, some reparameterisations of **D** have been proposed.

In DTI, symmetry of **D** ensures that its eigenvalues are real, whereas positive semi-definiteness further ensures that they are non-negative. Hence, the

square roots of the eigenvalues are real and proportional to the diffusivities along the main diffusion directions. Since the parameterisation 1.2.2 does not ensure (semi-) positive-definiteness, estimators of $\mathbf{D}$ (unless appropriately constrained) may have negative eigenvalues. This has been acknowledged in particular for the Maximum Likelihood (ML) and Least Square (LS) estimators [e.g., Koay *et al.*, 2006]. We will introduce a new reparameterisation in Section 2.2.2 to ensure the positive semi-definite symmetric structure of our estimator. Now we review some commonly known parameterisations that fulfil these conditions.

The *Cholesky decomposition* [e.g., Koay *et al.*, 2006] provides one such a reparameterisation by decomposing $\mathbf{D}$ into a lower triangular matrix $\mathbf{L}$ and its transpose, $\mathbf{L}^T$, i.e.,

$$\mathbf{D} = \mathbf{L}\mathbf{L}^T = \begin{pmatrix} L_{xx} & 0 & 0 \\ L_{xy} & L_{yy} & 0 \\ L_{xz} & L_{yz} & L_{zz} \end{pmatrix} \begin{pmatrix} L_{xx} & L_{xy} & L_{xz} \\ 0 & L_{yy} & L_{yz} \\ 0 & 0 & L_{zz} \end{pmatrix}, \tag{1.2.17}$$

where $L_{ii} \in \mathbb{R}^+, i = x, y, z$, and $\mathbf{D}$ is guaranteed to be positive semi-definite. Note that Cholesky decomposition is unique if and only if $\mathbf{D}$ is strictly positive definite.

Another reparameterisation of $\mathbf{D}$ is based on the spectral decomposition in Equation (1.2.16). Let $\lambda_i = \exp \gamma_i, i = 1, 2, 3$ where $\gamma_i \in \mathbb{R}$. The spectral decomposition of $\mathbf{D}$ is then given by

$$\begin{aligned} \mathbf{D} &= \mathbf{E}\mathbf{\Lambda}\mathbf{E}^T \\ &= \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{pmatrix} \begin{pmatrix} \exp\gamma_1 & 0 & 0 \\ 0 & \exp\gamma_2 & 0 \\ 0 & 0 & \exp\gamma_3 \end{pmatrix} \begin{pmatrix} \mathbf{v}_1^{\mathrm{T}} \\ \mathbf{v}_2^{\mathrm{T}} \\ \mathbf{v}_3^{\mathrm{T}} \end{pmatrix}, \end{aligned}$$

and $\exp \gamma_i > 0$ ensures the positive definiteness of $\mathbf{D}$. Note that the spectral

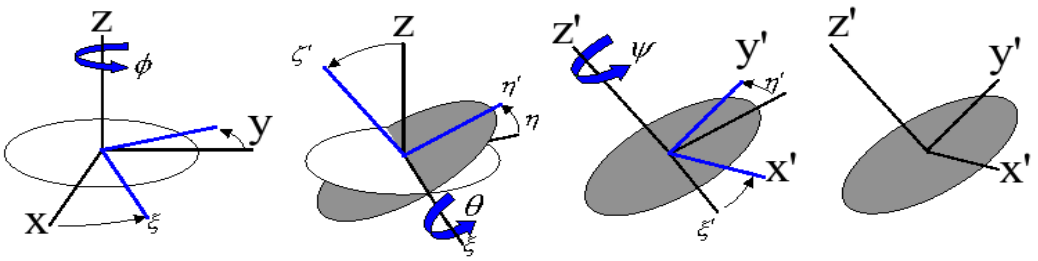

Figure 1.7: Rotation from the scanner reference frame to the diffusion reference frame. The rotation matrix **E** is defined in Equation (1.2.4). The scanner reference frame: $[x, y, z]$. The diffusion reference frame: $[x', y', z']$.

decomposition rotates the diagonalised tensor $\mathbf{\Lambda}$ from the scanner reference frame $[x, y, z]$ to the diffusion frame $[x', y', z']$ at each voxel by multiplying $\mathbf{\Lambda}$ on the left by a rotation matrix **E** and on the right by $\mathbf{E}^T$. This rotation is described in Figure 1.7. Then the rotation matrix **E** can be given by

$$\mathbf{E} = \begin{pmatrix} \cos\psi\cos\phi - \cos\theta\sin\phi\sin\psi & \cos\psi\sin\phi + \cos\theta\cos\phi\sin\psi & \sin\psi\sin\theta \\ -\sin\psi\cos\phi - \cos\theta\sin\phi\cos\psi & -\sin\psi\sin\phi + \cos\theta\cos\phi\cos\psi & \cos\psi\sin\theta \\ \sin\theta\sin\phi & -\sin\theta\cos\phi & \cos\theta \end{pmatrix},$$

where $\psi, \phi \in [0, 2\pi)$ and $\theta \in [0, \pi)$. Therefore the reparameterisation of **D** is given by a matrix of the six unknown parameters ψ, ϕ, θ, γ_1, γ_2 and γ_3 which guarantees the positive semi-definiteness of **D**. Clearly, this representation is not unique even if the eigenvalues are distinct and ordered. Indeed, composition $\mathbf{E} \circ \mathbf{A}$ where **E** is any rotation and **A** is the rotation by π of any of the principal planes of **D** yields the same tensor $\mathbf{D} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^{\mathrm{T}}$. Note also that this representation has singularities. E.g., if $\theta = 0$, then

$$\mathbf{E} = \begin{pmatrix} \cos\psi\cos\phi - \sin\phi\sin\psi & \cos\psi\sin\phi + \cos\phi\sin\psi & 0 \\ -\sin\psi\cos\phi - \sin\phi\cos\psi & -\sin\psi\sin\phi + \cos\phi\cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

and ϕ and ψ are not identifiable. But $\phi + \psi$ is identifiable, i.e.,

$$\mathbf{E} = \begin{pmatrix} \cos(\psi+\phi) & sin(\psi+\phi) & 0 \\ -\sin(\psi+\phi) & \cos(\psi+\phi) & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

This is one of many possible parameterisations of the rotation matrices. Also $\mathbf{E}$ can contain reflection. Then $\mathbf{E}$ describes an orthogonal transformation (i.e. rotation and reflection) and $\mid \mathbf{E} \mid = \pm 1$.

1.3 Visualisation Methods

1.3.1 Tensor-derived image contrasts

A number of tensor-derived scalar indices have been devised to combine the diffusion information contained in the eigensystem of $\mathbf{D}$. Image contrasts based on these scalar indices can be used for displaying the diffusion properties in the tissue and the neuroanatomy. In this section, we will first introduce four commonly used *diffusion anisotropy indices* (DAI), namely mean diffusivity (MD), fractional anisotropy (FA), relative anisotropy (RA) and volume ratio (VR). Then, geometrical measures of diffusion and the fibre orientation maps will be introduced.

Mean Diffusivity (MD)

An estimate that summarises the average diffusion properties of a voxel is given by the average of the eigenvalues of $\mathbf{D}$. It is normally referred to as the *mean*

diffusivity (MD) (see Figure 1.8(a)) of the voxel:

$$MD = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}. \tag{1.3.1}$$

MD is greater in cerebrospinal fluid (CSF) and smaller in organised brain tissue and has been suggested for CSF-related disease studies [Narr *et al.*, 2009].

To characterise tissue anisotropy in *vivo*, a wide range of diffusion anisotropy indices [Basser *et al.*, 2000] have been investigated in terms of the eigenvalues of **D**. All anisotropy indices which depend on **D** only through its eigenvalues are rotationally invariant. Here we introduce three commonly used anisotropy indices.

Fractional Anisotropy (FA)

Fractional anisotropy (FA) is one of the most popular DAI's in the MRI community. FA gives an estimate of the proportion of the 'magnitude' of **D** that can be ascribed to anisotropic diffusion [Kingsley, 2006b]. The definition of FA is given by

$$FA = \frac{\sqrt{3[(\lambda_1 - MD)^2 + (\lambda_2 - MD)^2 + (\lambda_3 - MD)^2]}}{\sqrt{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}}. \tag{1.3.2}$$

FA ranges from 0 for complete isotropy to 1 for a linear anisotropy. For example, when $\lambda_1 \gg \lambda_2 = \lambda_3$, FA$\approx$ 1. When $\lambda_1 = \lambda_2 \gg \lambda_3$, FA$\approx 1/\sqrt{2}$. FA maps are therefore intuitive to interpret when the white matter is rendered white and grey matter - dark (see Figure 1.8(b)).

Relative Anisotropy (RA)

Relative anisotropy (RA) is a normalised standard deviation of the eigenvalues,

and is given by equation below [Le Bihan *et al.*, 2001]

$$RA = \frac{\sqrt{(\lambda_1 - MD)^2 + (\lambda_2 - MD)^2 + (\lambda_3 - MD)^2}}{\sqrt{3MD}}. \tag{1.3.3}$$

RA varies between 0 (isotropic diffusion) and $\sqrt{2}$ (anisotropic diffusion). RA represents the ratio of the anisotropic part of **D** to its isotropic part (see Figure 1.8(c)).

Volume Ratio (VR)

Volume Ratio (VR) defines the ratio between the volume of the diffusion ellipsoid and the volume of a sphere of diameter MD [Wheeler-Kingshott *et al.*, 2003]. VR can be represented as

$$VR = \frac{\lambda_1\lambda_2\lambda_3}{MD^3}. \tag{1.3.4}$$

As the range of VR is from 1 (isotropic diffusion) to 0 (anisotropic diffusion), many authors prefer to use $(1 - VR)$ (see Figure 1.8(d)).

Thus, in FA, RA and 1-VR maps, the brighter the voxel, the higher its degree of anisotropy. The main differences between FA, RA and VR lie in their sensitivity to anisotropy: FA is more sensitive to low and VR to high values of diffusion anisotropy, and RA scales linearly for different levels of anisotropy [Sundgren *et al.*, 2004].

Geometrical measures of diffusion

The relationship between the eigenvalues of **D** can be used for classification of the diffusion tensor according to its geometry. How close the tensor is to the generic cases of line, plane and sphere is of interest. It is possible to classify the diffusion tensor according to the generic shape of the diffusion ellipsoid. By us-

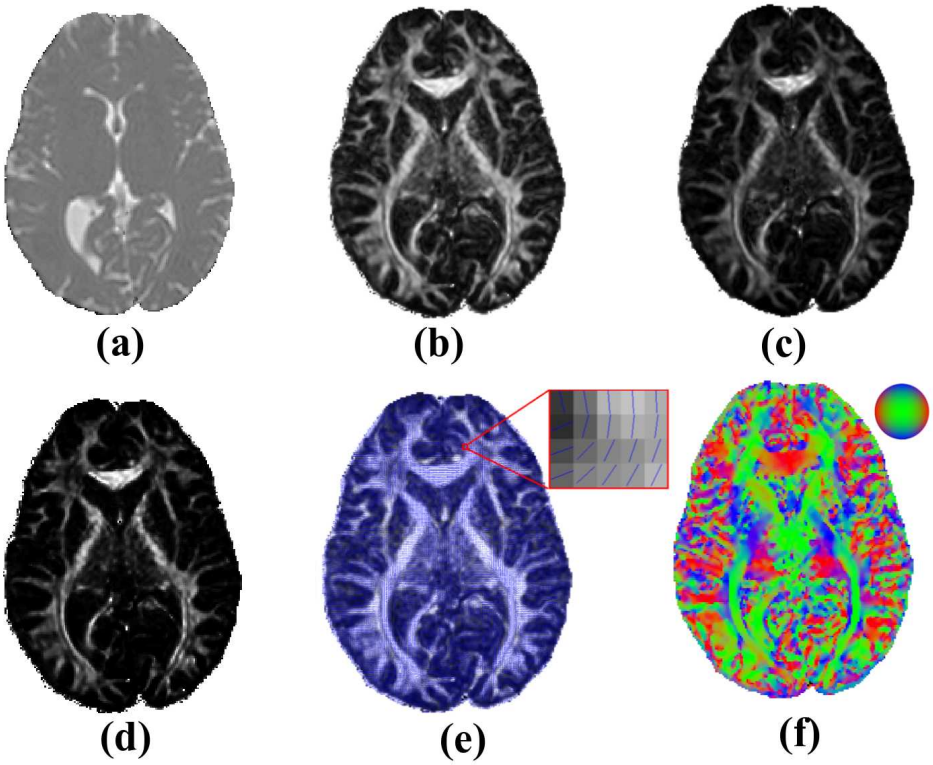

Figure 1.8: Image contrasts from diffusion tensor imaging. (a) MD map, (b) FA map, (c) RA map, (d) 1-VR map, (e) $\mathbf{v}_1$ orientation line map with FA background, and (f) $\mathbf{v}_1$ orientation colour coded map (where the colours red, green and blue represent diffusion in the x,y,z axes respectively).

ing the largest eigenvalue λ_1 for normalisation, the following shape measures are obtained for detection of the linear, planar and spherical profiles, respectively [Westin *et al.*, 2002]

$$c_l = \frac{\lambda_1 - \lambda_2}{\lambda_1}, \quad c_p = \frac{\lambda_2 - \lambda_3}{\lambda_1}, \quad c_s = \frac{\lambda_3}{\lambda_1}.$$

All these three measures evidently range from 0 to 1, and

$$c_l + c_p + c_s = 1.$$

Alternatively, measures can be normalised with the sum of eigenvalues $\lambda_1 + \lambda_2 + \lambda_3$ [Westin *et al.*, 1997]

$$c_l^* = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2 + \lambda_3}, \quad c_p^* = \frac{2(\lambda_2 - \lambda_3)}{\lambda_1 + \lambda_2 + \lambda_3}, \quad c_s^* = \frac{3\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3}$$

To ensure the measures remain in the [0, 1] range and sum up to one, the scaling factors of 2 and 3 have been introduced in the planar and spherical cases, respectively. The typical ranges of tensor-derived quantities of the human brain *in vivo* have been reported in [Pierpaoli *et al.*, 1996].

Fibre orientation maps

To demonstrate spatial, directional diffusion anisotropy, more sophisticated methods, such as *line* and *colour coded orientation maps*, have been produced based on $\mathbf{v}_1$. Recall that the principal eigenvector $\mathbf{v}_1$ of $\mathbf{D}$ represents the main diffusion direction and dominant fibre orientation at each voxel with a single dominant fibre orientation. It is a common practice to display these vectors at each voxel as lines superimposed on FA-coloured images (see Figure 1.8(e)).

Another advanced method that is commonly used in a medical context is

the colour coded of the main fibre orientation. In this method, the information contained in the tensor is reduced to $\mathbf{v}_1$, and a colour is assigned to each voxel using $\mathbf{v}_1 = (v_{1x}, v_{1y}, v_{1z})^T$. Namely, the absolute value of the v_{1x}, v_{1y}, v_{1z} components are used as intensities of the red, green, and blue channels. Therefore a red pixel in such a map corresponds to a vector oriented in the left-right direction, green - in the anterior-posterior direction and blue - in the feet-head direction (see Figure 1.8(f)).

1.3.2 Fibre tractography

Assuming there is a single dominant orientation of fibre bundles at any given voxel, the principal eigenvector of the diffusion tensor describing the orientation of the water diffusion is aligned with the dominant orientation of fibre bundles at the voxel. This local fibre orientation information provided directly by DTI can be used to reconstruct the pathways of major white matter structures through the brain. This three-dimensional reconstruction process is known as *fibre tractography*. A number of methods [Nucifora *et al.*, 2007; McGraw *et al.*, 2004; Basser *et al.*, 2000] have been developed to infer fibre structure and connection between tissues or between brain regions. These methods fall into two main groups: deterministic approaches (or streamline approaches) that define a single route of connection for each start voxel, and probabilistic approaches that attempt to establish the spatially distributed degree of connection in the region of interest from a given start voxel.

Deterministic Fibre Tractography

In deterministic fibre tractography, the implicit underlying assumption is that the principal eigenvector $\mathbf{v}_1$ forms a tangent to the space curve traced out by the white matter tract [Basser *et al.*, 2000]. There are three important steps in

this approach.

First, we have to extract the local fibre orientation at each voxel. While using $\mathbf{v}_1$ as an estimate of fibre orientation is straightforward, this estimate becomes unreliable for isotropic or planar diffusion profiles. In these extreme cases, the entire diffusion tensor will be needed.

The second step is to propagate a single pathway bi-directionally from a seeding point (usually a voxel in the centre of the image), by moving in a direction that is parallel with $\mathbf{v}_1$. However, this discrete path is not good because fibre tracts are believed to be sufficiently smooth. The simplest way to solve this issue is to convert the discrete voxel information into a continuous tracking trajectory (see Figure 1.9(a)). Therefore, sub-voxel estimates of the tensor are required for this approach. These are obtained either by interpolation of the raw DWI images before re-estimation of the diffusion tensor [Conturo *et al.*, 1999] or by interpolation of the diffusion tensors directly at appropriately chosen sub-voxel locations within the imaged region [Basser *et al.*, 2000].

The third step is termination of the propagation. There are two commonly used criteria. The first is that tracking is terminated if the tract enters into a low-anisotropy region (the threshold is chosen to differentiate white matter from grey matter). The second is to stop tracking if a sharp bend occurs. That means we fix a threshold of the angle that the path can turn through between one step to the next. Figure 1.9(b) shows an example of the deterministic fibre tractography in the *corpus callosum* region of interest starting from a voxel displayed in green. We set 0.2 and 64° for FA and the angular thresholds respectively. This tractography is obtained with the software package *Camino* [Cook *et al.*, 2006].

Probabilistic Fibre Tractography

Uncertainty of the estimates of the main diffusion direction and fibre orientation exists due to the measurement noise, patient movement and imaging arti-

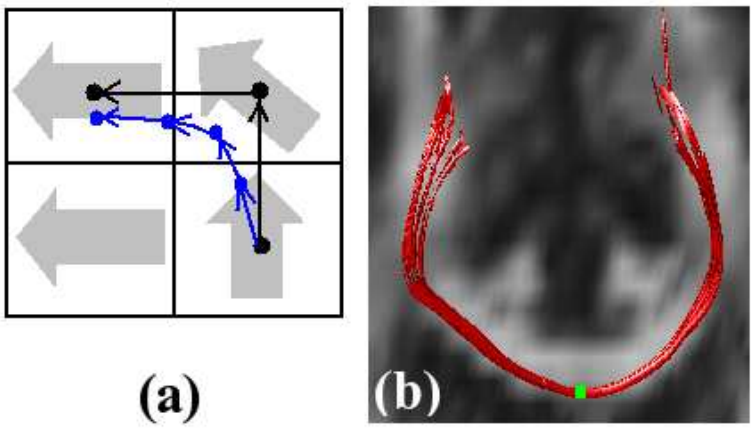

Figure 1.9: Deterministic fibre tractography. (a) Primitive discrete tracking (black) based on $\mathbf{v}_1$ (grey) is converted into a smooth trajectory (blue). (b) An example of deterministic tracking (red) in the *corpus callosum* region starting from voxel shown in green, obtained with the software package Camino.

facts. This is the main criticism levied at the deterministic fibre tractography. Probabilistic fibre tracking incorporates a distribution of possible orientations into the tracking methods which allows one to assign a confidence measure to each fibre trajectory. Instead of reconstructing a single trajectory from a given seeding voxel, probabilistic tractography propagates a large number of pathways from the seeding voxel. The probabilistic approaches can generate probabilistic maps of fibre connectivity between brain regions using a voxel-based connectivity index which is sensitive to structural changes [Behrens *et al.*, 2003]. This voxel-based connectivity may reflect not only the integrity and coherence of white matter tracts, but also tract geometry and length [Ciccarelli *et al.*, 2006].

Criteria for the termination of probabilistic tracking is slightly different from the deterministic approach. The latter uses both thresholds on anisotropy and angular deviation. Probabilistic tractography method uses the angular deviation between successive steps. This allows the reconstruction of trajectory from and into low-anisotropy regions such as grey matter.

1.4 Applications

Compared with conventional MRI, DTI can provide four types of new information: new contrasts, white matter morphology, refined information about anatomical locations, and connectivity [Mori, 2007]. We have mentioned some image contrasts derived from the eigensystem of **D** in Section 1.3.1, such as widely used FA map and fibre orientation maps. White matter morphology is concerned with shape and size of white matter tracts, and DTI can be used to assess these characteristics [Mori, 2007, p.132]. Deformation of tracts due to tumour growth can also be captured. By providing an anatomical template of the white matter, DTI improves our ability to understand the location of abnormality [Pomara *et al.*, 2001] in terms of white matter anatomy and functions. Recall that DTI can measure white matter tract direction and thereby create the possibility of non-invasive digital reconstruction of neuronal connectivity [Tuch *et al.*, 2003]. Since knowledge of neuronal connectivity is tremendously important for interpreting functional MRI (fMRI) data, DTI can also serve to complement fMRI [Olesen *et al.*, 2003]. In this section, we review clinical applications of DTI for human brain studies which use this new modality.

1.4.1 Normal brain development

The human brain needs 10 or 12 years after birth to complete its general development. The most dramatic changes during the maturation is in myelination in the first few years of life [Paus *et al.*, 1999]. Large groups of myelinated axons, which connect various regions in the brain, appear visibly as white matter. Conventional MRI signal changes in a T_2-weighted image may reflect a decrease in brain water content and an increase in white matter myelination during brain maturation in children [Dong *et al.*, 2004] (see an example

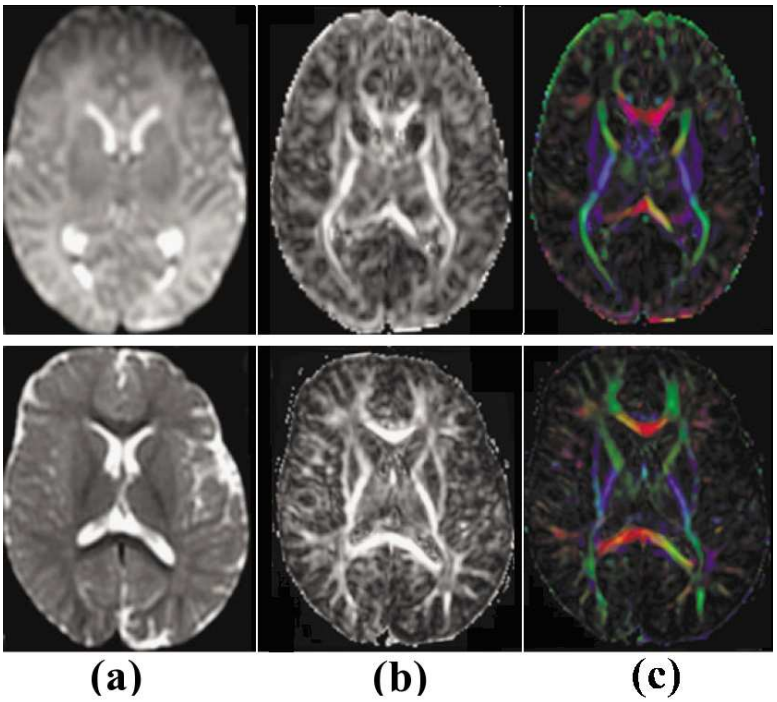

Figure 1.10: Comparison of T_2-weighted images and tensor-derived images from diffusion MRI data. The T_2-weighted images (a), FA (b) and fibre orientation colour coded (scaling with FA) (c) maps of a three-day-old boy (top row) and a one-year-old boy (bottom row) without brain abnormalities on clinical MRI are from [Dong *et al.*, 2004]. During the first year of brain maturation, the white matter signal decreases on T_2-weighted images; development of central and peripheral white matter structures are identified with increasing anisotropy on FA and fibre orientation maps.

in Figure 1.10 (a)). DTI with anisotropy analysis and fibre orientation detection provides more information in assessing brain development than conventional MRI image contrast [Le Bihan *et al.*, 2001]. The diffusion anisotropy increases with increasing gestational age in some white matter regions in previous studies [e.g., Hüppi *et al.*, 1998]. Increasing FA continues into childhood and adolescence (see Figure 1.10(b)). This finding has been shown to correlate with progressing myelination and increasing organisation of white matter fibre tracts (see Figure 1.10(c)). Different observation from these normal development changes in water diffusion may be an early marker of brain injury and abnormality in newborns [Dong *et al.*, 2004].

1.4.2 White matter diseases

DTI has already shown its potential in white matter diseases, such as multiple sclerosis [Tievsky *et al.*, 1999], leukoencephalopathy, Wallerian degeneration [Dong *et al.*, 2004], and Alzheimer's disease [Medina and Gaviria, 2008].

There has also been an increasing interest in the use of DTI to investigate various tumour components, and to assess tumoural invasion from normal tissue [Mori *et al.*, 2002]. In general, a brain tumour is an abnormal growth of cells within the brain or inside the skull. The goal of brain tumour surgery is to maximise tumour resection while avoiding important adjacent brain structures since their inadvertent resection can lead to devastating neurological consequences. Despite the information provided by conventional MRI in detecting the location and in characterising the extent of tumours, radiological specification and grading of a brain tumour is still limited [Sundgren *et al.*, 2004]. Fibre tractography can delineate deformation of white matter tracts in brain tumour patients which is helpful for neurosurgical planning and postoperative assessment [Mori, 2007].

Figure 1.11 [Lazar *et al.*, 2006] shows image contrasts and fibre tractographies of a patient before and after surgical resection of a brain tumour. Figure 1.11 (a) is the preoperative fibre orientation colour coded (scaling with FA) map which allows differentiation of the white matter tracts situated in the tumour vicinity: the corpus callosum, cingulum bundles, and the corona radiata. The tumour is visible as a dark region in the colour maps (white arrow). Tractographies of the cortico-spinal tracts before and after surgery (8 months after surgical resection) for the patient are shown in Figure 1.11 (b) and (c). Preoperatively, in Figure 1.11 (b) at least a portion of the callosal body appears interrupted by the tumour mass. The colour of the fibre trajectories relates to local anisotropy, with yellow indicating high anisotropy and dark red, low anisotropy. Deviation

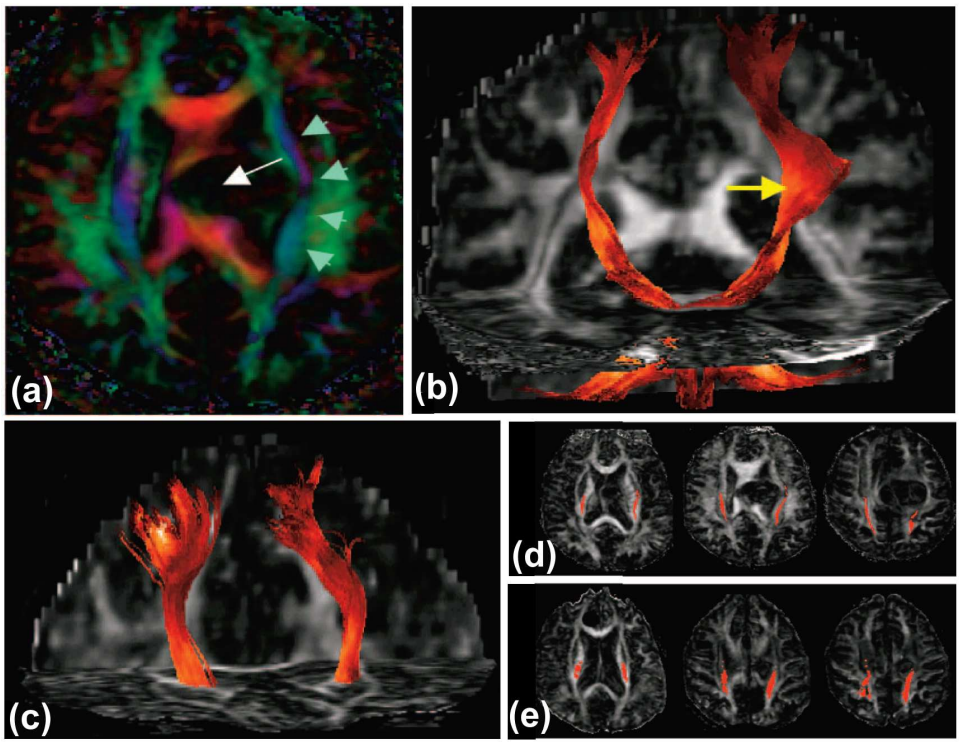

Figure 1.11: Image contrasts and fibre tractographies of a patient before and after surgical resection of brain tumour from [Lazar *et al.*, 2006]. (a): the pre-operative fibre orientation colour coded (scaling with FA) map. (b) and (c) are tractographies of the cortico-spinal tracts before and after surgery. (d) and (e) are the corresponding tract positions superimposed onto axial FA maps.

and deformation of the ipsilateral tract in the tumour proximity are apparent in the preoperative tractogram (yellow arrow). The appearance of the corpus callosum is improved, as shown by higher interhemispheric tract symmetry in Figure 1.11 (c). The corresponding tract positions are shown superimposed onto axial FA maps in (d) and (e), respectively.

1.4.3 Brain connectivity

DTI provides directional information of fibre orientation at the voxel which can be used for brain connectivity studies. DTI is able to describe anatomical connectivity [Lori *et al.*, 2002] by grouping together neighbouring voxels based on a similarity measure characterising voxels' relative principal diffusion directions.

This anatomical connectivity study groups voxels corresponding to anatomical tracts. DTI can also be used for providing the route of connection between regions by tracing axonal trajectories [Hagmann *et al.*, 2003]. Hagmann *et al.* [2008] applied network analysis into the large-scale structural brain connectivity of human cerebral cortex. In Hagmann's paper a comparison of diffusion imaging and functional MRI (fMRI) data was made for revealing a close relationship between structural and functional connections in the regions forming the structure cores.

To visualise anatomic connections between different parts of the brain on an individual basis is an important potential application of DTI. Brain tissue is composed of white matter and grey matter. Most normal brain functions require that specific grey matter regions communicate with each other via white matter fibre pathways. Therefore, studies of the white matter connections between different parts of the brain is critical to understanding human brain function and the establishment of networks underlying cognitive processes [Lori *et al.*, 2002]. However, this issue is difficult, as one has to infer continuity of fibre orientation from voxel to voxel. In out study, we will focus on fibre tractography. One may deal with fibres kissing, crossing and branching or merging (see Figure 1.12).

1.5 Thesis Aims and Organisation

The three main aims of this thesis are (1) to model water molecular diffusion in the regions containing one or more dominant fibre orientations and carry out statistical inference of parameters of interest, (2) to define statistics of diffusion tensor data using different metrics and apply these metrics to analyse the tensor field, and (3) to design diffusion gradient directions with statistics of

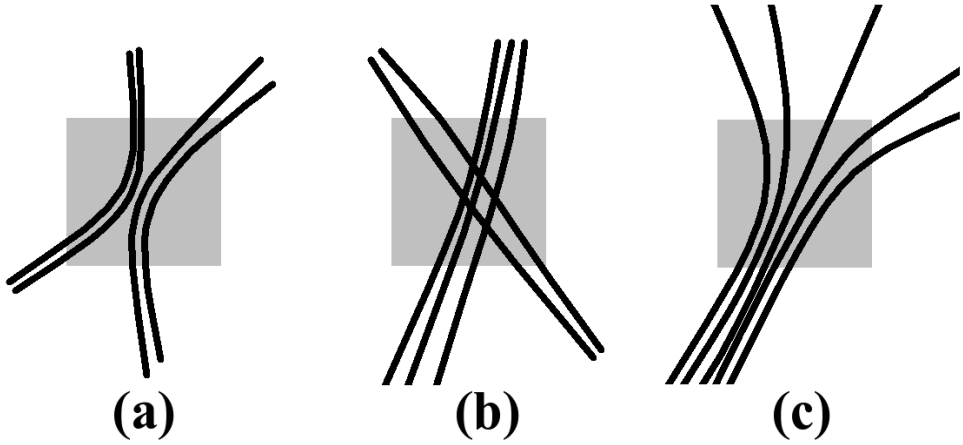

Figure 1.12: Three patterns of fibre bundles. (a) fibres are kissing (b) crossing and (c) branching or merging at the voxel.

uniformity. We also aim to apply all our methodologies to simulated diffusion image data and real data from the human brain.

This thesis is divided into five chapters. In Chapter 1, we briefly review the background to MRI and diffusion-weighted imaging (DWI), have provided the requisite knowledge of DTI, such as the diffusion tensor model and diffusion tensor eigensystem, and introduce the main applications of DTI.

In Chapter 2, a multi-tensor model is developed for diffusion tensor image data. A new parameterisation method guarantees the symmetric positive semi-definiteness of the diffusion tensor. We set up the Bayesian framework for parameter estimation from the multi-tensor model. To investigate the uncertainty of fibre orientations, Markov chain Monte Carlo (MCMC) methods are used to obtain the credible cone of fibre orientations. A model selection criterion has also been applied to a healthy human brain dataset.

Since diffusion MRI is a relatively low resolution modality, advanced tensor processing methods such as non-Euclidean interpolation, have been considered. Yet, reliable and accurate estimation of the highly complex white matter architecture of the brain remains a challenge despite the many advances in modelling, processing, and analysis of diffusion MRI data. Once a sample of

diffusion tensors is available (e.g. generated with MCMC method mentioned in Chapter 2) we wish to estimate the population mean and then carry out statistical inference. In Chapter 3, the statistical analysis of diffusion tensor data is considered which takes into account the non-Euclidean nature of the space of tensors (positive semi-definite symmetric matrices). We define what is meant by a mean covariance matrix in a non-Euclidean space, using the Fréchet mean. We also review some recently proposed techniques based on matrix logarithms and also consider estimators based on matrix decompositions, such as the Cholesky decomposition and the matrix square root. We propose an alternative approach, Procrustes analysis, to process the tensor field of the human brain.

The accuracy of DTI measurements depends on the diffusion gradient direction scheme applied. It is normally assumed that the optimum set will have uniformly distributed directions over a sphere due to equal noise level along each gradient direction. However, there appear to be several distinct notions of uniformity in this context. For example, Jones *et al*. [1999] proposed to arrange the directions uniformly in 3-dimensional space using the analogy of repulsive forces. In Chapter 4, we design a series of uniform schemes with various numbers of directions by optimising the spatial uniformity of the directions. By spatial uniformity we mean that the points have equal probability density over the sphere with respect to the surface area measure [Mardia and Jupp, 2000, p167].

Finally, in Chapter 5 we conclude by discussing the main results drawn from this research. Also possible areas for further work are presented.

In this thesis, the results and graphics have been obtained using the MATLAB (R2009a, The MathWorks, Inc., Natick, Massachusetts, USA). The packages FSL (Analysis Group, FMRIB, Oxford, UK) [Smith *et al*., 2004] and Camino

(Centre for Medical Image Computing, University of London, UK) [Cook *et al.*, 2006] are used for viewing DWI image data and fibre tracking respectively.

Chapter 2

Bayesian Multi-tensor Model with Reparameterisation

2.1 Introduction

In DTI study if fibre bundles at a voxel orientate along only one dominant direction, then water molecules at that voxel will diffuse along that dominant fibre orientation. Recall that the diffusion behaviour at that voxel can be characterised with the single tensor model (see Section 1.2.1).

$$\begin{aligned} S_i &= \mu_i + \varepsilon_i \\ &= S_0 \exp(-b\mathbf{g}_i^{\mathrm{T}}\mathbf{D}\mathbf{g}_i) + \varepsilon_i, i = 1, ..., N \end{aligned} \tag{2.1.1}$$

where $S_0 > 0$ is the signal attenuation without diffusion gradient applied (i.e. $b = 0$). Notation used in the single tensor model is summarised in Table 2.1.

A variety of approaches for estimating the diffusion tensor, such as maximum likelihood estimation (MLE) and least squares methods have been discussed in Section 1.2.1.

Table 2.1: Glossary of terms in single tensor model

Term	Meaning	Properties
$\mathbf{g}_i$	the ith diffusion gradient direction	$\mathbf{g}_i \in \mathbb{R}P^2, \mathbf{g}_i \equiv -\mathbf{g}_i$, and $\parallel \mathbf{g}_i \parallel = 1$
b	an known parameter characterising the strength of diffusion gradient	$b \geq 0$
ε_i	the noise of the measured signal	*i.i.d.* Gaussian variables, $\varepsilon_i \sim \mathbf{N}(0, \sigma^2)$
μ_i	the mean of the measured diffusion weighted signal	$\mu_i > 0$
S_i	the measured diffusion weighted signal corresponding to $\mathbf{g}_i$	independent Gaussian variables, $S_i \sim \mathbf{N}(\mu_i, \sigma^2)$
S_0	the signal without diffusion gradient	$S_0 > 0$
N	the total number of diffusion gradient directions	$N = 1, 2, 3, ...$
$\mathbf{D}$	the diffusion tensor	a 3×3 symmetric and positive semi-definite matrix

2.1.1 Multiple compartment model

The single tensor model defined by Equation (2.1.1) is commonly applied under the assumption that the principal eigenvector $\mathbf{v}_1$ is aligned along the dominant fibre orientation in a voxel. However, there are regions in the brain where more than one distinct fibre orientation is captured in a single voxel.

Molecular diffusion which takes place in $m \geq 1$ distinct compartments enclosed in a single voxel, is commonly assumed to have no molecular exchange between compartments. Therefore, such diffusion is commonly described by a mixture of (trivariate centred) Gaussian distributions, in which diffusion in compartment j would be represented by a Gaussian with covariance matrix $2t\mathbf{D}_j$, $j = 1, ..., m$. Using linearity of the Fourier transform in this case, the mean μ_i of the ith measured diffusion-weighted signal can be modelled as

[Alexander, 2005]:

$$\mu_i = \sum_{j=1}^{m} a_j S_0 \exp(-b\mathbf{g}_i^{\mathrm{T}}\mathbf{D}_j\mathbf{g}_i),\ i = 1, ..., N. \tag{2.1.2}$$

where the weights $a_j \in (0, 1]$, $\sum_{j=1}^{m} a_j = 1$, of the individual compartments are also known as 'volume fractions' which generally also need to be estimated. We can assume that the noise of the measured signal attenuation ε_1, ..., ε_n are i.i.d. Gaussian variables, i.e. $\varepsilon_i \sim \mathbf{N}(0, \sigma^2)$. Thus, S_i along the ith diffusion gradient direction is given by

$$S_i = \sum_{j=1}^{m} a_j S_0 \exp(-b\mathbf{g}_i^{\mathrm{T}}\mathbf{D}_j\mathbf{g}_i) + \varepsilon_i,\ i = 1, ..., N. \tag{2.1.3}$$

Appropriately constrained nonlinear optimisation has been used to fit the model to the data [Alexander, 2005; Lenglet *et al.*, 2009].

General problems accompany the use of multiple compartment models [Alexander, 2006]. First, the choice of the number of compartments m presents a model selection problem. Second, the constrained nonlinear fitting procedure is unstable and starting-point dependent due to the local minima in the objective function. Thirdly there is the problem of non-identifiability of the parameters which does not appear to have been fully appreciated previously in the literature. In Section 2.1.2 below we address the third problem with these models.

2.1.2 Non-identifiability of multiple compartment model

The term identifiability has been used in economics [e.g. Hsiao, 1983], engineering, statistics [e.g. Paulino and de Bragança Pereira, 1994] and mathematical biology [e.g. Cobelli and DiStefano, 1980; van den Hof, 1998]. In this section, we first introduce the identifiability of parameters and a model. Then we

will prove that the multiple compartment model based on Equation (2.1.2) is nonidentifiable.

Definition 2.1. Consider a model $\{P_{\boldsymbol{\theta}}\}_{\boldsymbol{\theta}\in\Theta}$ where parameters $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$, and $\{P_{\boldsymbol{\theta}}\}_{\boldsymbol{\theta}\in\Theta}$ is a family of probability distributions describing observed data. A point $\boldsymbol{\theta}$ in the parameter space Θ of the model is called identifiable if for any other point $\boldsymbol{\theta}' \in \Theta$, the equality of the distributions $P_{\boldsymbol{\theta}}$ and $P_{\boldsymbol{\theta}'}$ implies $\boldsymbol{\theta} = \boldsymbol{\theta}'$.

Definition 2.2. Model $\{P_{\boldsymbol{\theta}}\}_{\boldsymbol{\theta}\in\Theta}$ is said to be identifiable if every point $\boldsymbol{\theta} \in \Theta$ is identifiable relative to this model.

In light of Definition 2.1 and Definition 2.2, the multiple compartment model (with the assumption that S_0 is observed noise-free, and the noise is homoscedastic Gaussian) is written as follows: $\{P_{\boldsymbol{\theta}}\}$ where $\boldsymbol{\theta} = \{a_1, a_2, ..., a_m, \mathbf{D}_1, \mathbf{D}_2, ..., \mathbf{D}_m, \sigma^2\}$ and $P_{\boldsymbol{\theta}}$ is the multivariate Gaussian distribution with mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2, ...\mu_N)^{\mathrm{T}}$ and the covariance matrix $\sigma^2\mathbf{I}_{N\times N}$.

Let $\theta_1 = \{a_1, a_2, ..., a_m, \mathbf{D}_1, \mathbf{D}_2, ..., \mathbf{D}_m\}$ be a setting of parameters in the multiple compartment model (2.1.2). That is,

$$\mu_i = \sum_{j=1}^{m} a_j S_0 \exp(-b\mathbf{g}_i^{\mathrm{T}}\mathbf{D}_j\mathbf{g}_i),\ i = 1, ..., N, \tag{2.1.4}$$

where $a_j \in (0, 1]$, $\sum_{j=1}^{m} a_j = 1$, and $\mathbf{D}_j$ is the diffusion tensor of the jth compartment, $j = 1, ..., m$. Let $\lambda_{1,j} \geq \lambda_{2,j} \geq \lambda_{3,j} \geq 0$ be the three eigenvalues of $\mathbf{D}_j$. Note that identifiability will be considered up to permutations on the indexes of the components.

Lemma 2.1. If $b > 0$, and c_j, $j = 1, ..., m$ satisfy

$$\frac{a_j}{\exp(b\lambda_{3,j})} \leq c_j \leq 1, j = 1, ..., m, \tag{2.1.5}$$

and

$$\sum_{j=1}^{m} c_j = 1, \tag{2.1.6}$$

then the matrix $\mathbf{D}_j - \frac{\log(a_j/c_j)}{b}\mathbf{I}_{3\times 3}, j = 1, ..., m$ is symmetric and positive semi-definite (diffusion tensors).

Proof. It is obvious that $\mathbf{D}_j^* = \mathbf{D}_j - \frac{\log(a_j/c_j)}{b}\mathbf{I}_{3\times 3}$ is symmetric.

Since $\mathbf{D}_j, j = 1, ..., m$ is symmetric, the eigendecomposition of $\mathbf{D}_j$ is given by

$$\begin{aligned} \mathbf{D}_j &= \mathbf{E}_j \boldsymbol{\Lambda}_j \mathbf{E}_j^{\mathrm{T}} \\ &= (\mathbf{v}_{1,j}, \mathbf{v}_{2,j}, \mathbf{v}_{3,j}) \begin{pmatrix} \lambda_{1,j} & 0 & 0 \\ 0 & \lambda_{2,j} & 0 \\ 0 & 0 & \lambda_{3,j} \end{pmatrix} \begin{pmatrix} \mathbf{v}_{1,j}^{\mathrm{T}} \\ \mathbf{v}_{2,j}^{\mathrm{T}} \\ \mathbf{v}_{3,j}^{\mathrm{T}} \end{pmatrix}, \end{aligned} \tag{2.1.7}$$

where $\lambda_{1,j} \geq \lambda_{2,j} \geq \lambda_{3,j} \geq 0$ are three eigenvalues of $\mathbf{D}_j$, and $\mathbf{v}_{i,j}, i = 1, 2, 3$ is the eigenvector corresponding to $\lambda_{i,j}$.

Consequently, the eigendecomposition of $\mathbf{D}_j^*$ is given by

$$\begin{aligned} \mathbf{D}_j^* &= \mathbf{D}_j - \frac{\log(a_j/c_j)}{b}\mathbf{I}_{3\times 3} \\ &= \mathbf{E}_j \boldsymbol{\Lambda}_j \mathbf{E}_j^{\mathrm{T}} - \mathbf{E}_j \frac{\log(a_j/c_j)}{b}\mathbf{I}_{3\times 3}\mathbf{E}_j^{\mathrm{T}} \\ &= \mathbf{E}_j \begin{pmatrix} \lambda_{1,j} - \frac{\log(a_j/c_j)}{b} & 0 & 0 \\ 0 & \lambda_{2,j} - \frac{\log(a_j/c_j)}{b} & 0 \\ 0 & 0 & \lambda_{3,j} - \frac{\log(a_j/c_j)}{b} \end{pmatrix} \mathbf{E}_j^{\mathrm{T}}. \end{aligned} \tag{2.1.8}$$

Hence, $\lambda_{i,j}^* = \lambda_{i,j} - \frac{\log(a_j/c_j)}{b}$ is the ith eigenvalue of $\mathbf{D}_j^*$. It is simple to see that $\lambda_{1,j}^* \geq \lambda_{2,j}^* \geq \lambda_{3,j}^*$.

Since $c_1, ..., c_m$ satisfy inequality (2.1.5), we know

$$\begin{aligned} & \exp(b\lambda_{3,j}) \geq a_j/c_j \\ \iff \quad & b\lambda_{3,j} \geq \log\left(a_j/c_j\right) \\ \iff \quad & \lambda_{3,j} - \tfrac{\log\left(a_j/c_j\right)}{b} \geq 0 \\ \iff \quad & \lambda^*_{3,j} \geq 0. \end{aligned} \tag{2.1.9}$$

Hence, it is true that $\lambda^*_{1,j} \geq \lambda^*_{2,j} \geq \lambda^*_{3,j} \geq 0$, i.e., three eigenvalues of $\mathbf{D}^*_j$ are nonnegative. This completes the proof. □

For the multiple compartment model (2.1.2), we can always find a distinct set of parameter values $\theta_2 = \{c_1, c_2, ..., c_m, \mathbf{D}_1 - h_1\mathbf{I}_{3\times 3}, \mathbf{D}_2 - h_2\mathbf{I}_{3\times 3}, ..., \mathbf{D}_m - h_m\mathbf{I}_{3\times 3}\}$ where $h_j = \frac{\log(a_j/c_j)}{b}$, $\frac{a_j}{\exp(b\lambda_{j,3})} \leq c_j \leq 1$ and $\sum_{j=1}^m c_j = 1$ which give the same model. Hence, $\theta_2 \neq \theta_1$ but $P_{\theta_1} = P_{\theta_2}$ and therefore we have the result: The multiple compartment model (2.1.2) is non-identifiable.

Example 2.1. Let us fix $b = S_0 = 1$, and $m = 2$. Let $a_1 = 0.2$, then $a_2 = 0.8$. $\mathbf{D}_1$ and $\mathbf{D}_2$ are set as follows:

$$\mu_i = 0.2 \times \exp\left\{-\mathbf{g}_i^{\mathrm{T}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \mathbf{g}_i\right\} + 0.8 \times \exp\left\{-\mathbf{g}_i^{\mathrm{T}} \begin{pmatrix} 4 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{g}_i\right\}. \tag{2.1.10}$$

Since $\mathbf{g}_i^{\mathrm{T}}\mathbf{g}_i = 1$, μ_i in Equation (2.1.10) can also be written as follows:

$$\begin{aligned}\mu_i &= 0.1 \times \exp\left\{-\mathbf{g}_i^{\mathrm{T}}\begin{pmatrix} 1-\log 2 & 0 & 0 \\ 0 & 2-\log 2 & 0 \\ 0 & 0 & 3-\log 2 \end{pmatrix}\mathbf{g}_i\right\} + \\ & \quad 0.9 \times \exp\left\{-\mathbf{g}_i^{\mathrm{T}}\begin{pmatrix} 4+\log\frac{9}{8} & 0 & 0 \\ 0 & 5+\log\frac{9}{8} & 0 \\ 0 & 0 & 1+\log\frac{9}{8} \end{pmatrix}\mathbf{g}_i\right\}.\end{aligned}$$

Evidently, we have two distinct settings of parameters which result in the identical model. Thus, the multiple compartment model is non-identifiable in this example, i.e., the solution to the parameter in multiple compartment model is not unique.

2.2 Bayesian Multi-tensor Model with Reparameterisation

2.2.1 A multi-tensor model

In this subsection, we will develop a new model for capturing the diffusion behaviours at a voxel based on the multiple compartment model which is essentially identifiable.

If there is no diffusion gradient applied for the experiment i.e. $b = 0$, then it is clear that

$$\begin{aligned}\mu_i &= \sum_{j=1}^{m} a_j S_0 \\ &= S_0. \end{aligned} \tag{2.2.1}$$

Consider a general DWI experiment with diffusion gradients applied i.e. $b > 0$. From the multiple compartment model, we can derive

$$\begin{aligned}
\mu_i &= \sum_{j=1}^{m} a_j S_0 \exp(-b\mathbf{g}_i^{\mathrm{T}}\mathbf{D}_j\mathbf{g}_i) \\
&= \sum_{j=1}^{m} S_0 \exp\left\{-b\mathbf{g}_i^{\mathrm{T}}\mathbf{D}_j\mathbf{g}_i + \log a_j\right\} \\
&= \sum_{j=1}^{m} S_0 \exp\left\{-b\mathbf{g}_i^{\mathrm{T}}\mathbf{D}_j\mathbf{g}_i - b\mathbf{g}_i^{\mathrm{T}}(-\frac{\log a_j}{b})\mathbf{I}_{3\times 3}\mathbf{g}_i\right\} \\
&= \sum_{j=1}^{m} S_0 \exp\left\{-b\mathbf{g}_i^{\mathrm{T}}(\mathbf{D}_j - \frac{\log a_j}{b}\mathbf{I}_{3\times 3})\mathbf{g}_i\right\}, i = 1, ..., N. \qquad (2.2.2)
\end{aligned}$$

Let $q_j = -\frac{\log a_j}{b}, j = 1, ..., m$. Since $0 < a_j \leq 1$ and $b > 0$, $q_j \geq 0$. Now we define

$$\mathbf{D}_j^* = \mathbf{D}_j + q_j\mathbf{I}_{3\times 3}, j = 1, ..., m. \qquad (2.2.3)$$

Applying Lemma 2.1 (with $c_j = 1$) gives that $\mathbf{D}_j^*$ is symmetric and positive semi-definite, if $\mathbf{D}_j$ are positive semi-definite.

Hence, $\mathbf{D}_j^*$ can be a diffusion tensor which is used to capture the diffusion behaviour in the jth compartment.

Now we can define the **multi-tensor model** given by

$$\mu_i = \begin{cases} \sum_{j=1}^{m} S_0 \exp(-b\mathbf{g}_i^{\mathrm{T}}\mathbf{D}_j^*\mathbf{g}_i) & \text{if } b > 0 \\ S_0 & \text{if } b = 0. \end{cases} \qquad (2.2.4)$$

In this multi-tensor model (2.2.4), the diffusion tensor $\mathbf{D}_j^*$ not only captures the diffusion property in the jth compartment, but also contains the fraction of contribution to the overall signal at the voxel.

Since $\mathbf{D}_j^*$ is a general diffusion tensor, we can replace it with $\mathbf{D}_j$, and rewrite

the multi-tensor model to be

$$\mu_i = \begin{cases} \sum_{j=1}^{m} S_0 \exp(-b\mathbf{g}_i^{\mathrm{T}}\mathbf{D}_j\mathbf{g}_i) & \text{if } b > 0 \\ S_0 & \text{if } b = 0. \end{cases} \tag{2.2.5}$$

So, we largely circumvent the problem of non-identifiability by not including the volume fraction explicitly.

2.2.2 A new reparameterisation

In order to guarantee the symmetric positive semi-definiteness of $\mathbf{D}_j$, we reparameterize $\mathbf{D}_j$ by decomposing it to a general 3×3 matrix $\mathbf{Q}_j$ and $\mathbf{Q}_j^{\mathrm{T}}$, i.e., $\mathbf{D}_j = \mathbf{Q}_j\mathbf{Q}_j^{\mathrm{T}}$ where

$$\mathbf{Q}_j = \begin{pmatrix} Q_{11,j} & Q_{12,j} & Q_{13,j} \\ Q_{21,j} & Q_{22,j} & Q_{23,j} \\ Q_{31,j} & Q_{32,j} & Q_{33,j} \end{pmatrix}. \tag{2.2.6}$$

Note that this is over-parameterised now.

Then the multi-tensor model is given by

$$\mu_i = \begin{cases} \sum_{j=1}^{m} S_0 \exp(-b\mathbf{g}_i^{\mathrm{T}}\mathbf{Q}_j\mathbf{Q}_j^{\mathrm{T}}\mathbf{g}_i) & \text{if } b > 0 \\ S_0 & \text{if } b = 0. \end{cases} \tag{2.2.7}$$

Note that $\mathbf{Q}_j$ and $\mathbf{Q}_j\mathbf{R}_j$ where $\mathbf{R}_j \in O(3)$ result in the same model, and $O(3)$ is the space of 3×3 orthogonal matrices. That is,

$$\begin{aligned} \mathbf{D}_j &= \mathbf{Q}_j\mathbf{Q}_j^{\mathrm{T}} \\ &= (\mathbf{Q}_j\mathbf{R}_j)(\mathbf{Q}_j\mathbf{R}_j)^{\mathrm{T}}, j = 1, ..., m. \end{aligned} \tag{2.2.8}$$

So the terms in $\mathbf{Q}_j$ are essentially identifiable only up to $\mathbf{R}_j$. Note that the

Cholesky decomposition is then a special case of our parameterisation, i.e., $\mathbf{Q}_j = \mathbf{L}_j$ where $\mathbf{L}_j$ is lower triangular with positive diagonal entries.

The advantage of this reparameterisation is that we can carry out unconstrained optimisation methods to solve the multi-tensor model. The penalty is that the dimension of the parameter space increases from 6 for $\mathbf{D}_j$ to 9 for $\mathbf{Q}_j$. However, since most voxels in our studies have at most two compartments, these few extra parameters will not influence much our DTI studies. In fact, we will control these extra degrees of freedom by specifying appropriate prior distributions when performing Bayesian estimation (see next section 2.2.3).

2.2.3 Bayesian framework

Bayesian methods provide a complete paradigm for both statistical inference and decision making under uncertainty, solving many of the difficulties faced by conventional statistical methods, and extending the applicability of statistical methods [Bernardo and Smith, 2009]. In particular, Bayesian methods make it possible to incorporate scientific prior information in the analysis (by means of the prior distribution) and may be applied to problems whose structure is too complex for conventional methods to be able to handle. Bayesian estimation methods can be used for the DTI study, in particular for estimating multiple diffusion tensors [Behrens *et al*., 2007]. In Behrens *et al*. [2007], a partial volume model was used, where the signal is split into infinitely anisotropic components for fibre orientations, and single isotropic components. The parameters of interest are fibre orientation and volume fractions from each diffusion population which are different from the parameters (the decomposition of diffusion tensors) used in our multi-tensor model (see Section 2.2.1).

Bayes' theorem, the basis of Bayesian inference, tells us that the posterior probability density function for parameters posterior to the data collection is

proportional to the product of the density for parameters prior to the data collection and the likelihood for parameters given the data. That is,

$$\textit{posterior density} \propto \textit{likelihood function} \times \textit{prior density}. \tag{2.2.9}$$

Bayes' theorem precisely specifies how the modification of the uncertainty about the parameters in the light of evidence should be made.

In our DTI study, the essence of the Bayesian approach is as follows:

Treat the collection of the unknown parameters $\Theta = \{\mathbf{Q}_1, ..., \mathbf{Q}_m, \sigma^2\}$ as a random vector and encode all prior information (i.e. before measuring the signal intensities $\mathbf{S} = (S_1, S_2, ..., S_N)^{\mathrm{T}}$) about Θ into the prior distribution of Θ. Then, update this prior information and thus produce the posterior probability distribution of Θ and to draw appropriate inference based on this distribution. Thus, there are four key steps to the Bayesian approach:

1. Specification of the likelihood function $L(\Theta \mid \mathbf{S})$;
2. determination of prior distributions of $\mathbf{Q}_1$,...,$\mathbf{Q}_m$ and σ^2 respectively, and specification of the full joint distribution of Θ;
3. calculation of posterior distribution, $p(\Theta \mid \mathbf{S})$ from Bayes' Theorem;
4. drawing inferences from $p(\Theta \mid \mathbf{S})$.

2.2.3.1 Single tensor estimation

First let us consider the simple case where there is only one dominant fibre orientation at a voxel. This means water molecules diffuse mainly along one direction (the fibre orientation). Then, the single tensor model (multi-tensor

model with $m = 1$) is given by

$$\begin{aligned} S_i &= \mu_i + \varepsilon_i \\ &= S_0 \exp(-b\mathbf{g}_i^{\mathrm{T}}\mathbf{Q}\mathbf{Q}^{\mathrm{T}}\mathbf{g}_i) + \varepsilon_i,\ i = 1, ..., N. \end{aligned} \tag{2.2.10}$$

where ε_i's are i.i.d. Gaussian noise, $\varepsilon_i \sim \mathbf{N}(0, \sigma_1^2)$. Consequently, S_i's are independent Gaussian variables, i.e. $S_i \sim \mathbf{N}(\mu_i, \sigma_1^2)$. Thus, the parameter space of the single tensor model is $\Theta_1 = (\mathbf{Q}, \sigma_1^2)$ where $\mathbf{Q}$ has nine unknown entries.

The likelihood function of the parameters plays a very important role in Bayesian analysis. In our study, the likelihood function of Θ_1 represents the information about Θ_1 coming from the measured DWI image data.

Since S_i's are independent and follow $N(\mu_i, \sigma_1^2)$, the likelihood function of $\Theta_1 = (\mathbf{Q}, \sigma_1^2)$ given signal intensity data is given by

$$\begin{aligned} L(\mathbf{Q}, \sigma_1^2 \mid \mathbf{S}) &= \prod_{i=1}^{N} f(S_i \mid \mathbf{Q}, \sigma_1^2) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma_1^2}}\right)^N \exp\{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{N}[S_i - S_0 \exp(-b\mathbf{g}_i^{\mathrm{T}}\mathbf{Q}\mathbf{Q}^{\mathrm{T}}\mathbf{g}_i^{\mathrm{T}})]^2\}. \end{aligned} \tag{2.2.11}$$

We assume that a *priori* the individual parameters $Q_{11}, Q_{12}, \ldots, Q_{33}, \sigma_1^2$ are distributed independently, and we will choose their prior distributions according to our initial beliefs. Thus, we assume that the prior distribution of $\mathbf{Q}$ is a multivariate Gaussian distribution, i.e.

$$vec(\mathbf{Q}) \sim \mathbf{N}_9(vec(\mathbf{I}_{3\times 3}), \xi^2\mathbf{I}_{9\times 9}) \tag{2.2.12}$$

where $vec(\mathbf{A})$ vectorises an $n \times m$ matrix $\mathbf{A}$ to be a column vector by stacking the columns of $\mathbf{A}$, and $\mathbf{I}_{3\times 3}$ and $\mathbf{I}_{9\times 9}$ are the 3×3 and 9×9 identity matrices.

Or, specifically, the prior distribution of $\mathbf{Q}$ is

$$p(\mathbf{Q}) = (2\pi\xi^2)^{-\frac{9}{2}} \exp\left\{-\frac{1}{2\xi^2}(vec(\mathbf{Q}) - vec(\mathbf{I}_{3\times 3}))^{\mathrm{T}}(vec(\mathbf{Q}) - vec(\mathbf{I}_{3\times 3}))\right\}. \tag{2.2.13}$$

By choosing ξ to be large, we make the prior virtually noninformative.

For the variance of Gaussian noise σ_1^2, a suitable prior is [Lee, 1991]

$$\sigma_1^2 \sim \textit{Inv-Gamma}(\alpha, \beta), \tag{2.2.14}$$

i.e.,

$$p(\sigma_1^2) = \frac{\beta^\alpha}{\Gamma(\alpha)}(\sigma_1^2)^{-\alpha-1}\exp(-\beta/\sigma_1^2) \tag{2.2.15}$$

where $\Gamma()$ is the Gamma function, and $\alpha, \beta \in \mathbb{R}^+$.

Since the prior distributions of $\mathbf{Q}$ and σ_1^2 are independent, the full joint prior distribution is given by

$$\begin{aligned} p(\mathbf{Q}, \sigma_1^2) &= p(\mathbf{Q})p(\sigma_1^2) \\ &\propto \xi^{-9}(\sigma_1^2)^{-\alpha-1}\frac{\beta^\alpha}{\Gamma(\alpha)}\exp\{-\frac{1}{2\xi^2}(vec(\mathbf{Q}) - vec(\mathbf{I}_{3\times 3}))^{\mathrm{T}}(vec(\mathbf{Q}) - vec(\mathbf{I}_{3\times 3})) - \\ &\quad \beta/\sigma_1^2\}. \end{aligned} \tag{2.2.16}$$

By Bayes' theorem (2.2.9), the posterior distribution of parameters (σ_1 and $\mathbf{Q}$) has density given by

$$\begin{aligned} p(\mathbf{Q}, \sigma_1^2 \mid \mathbf{S}) \propto\ & (\sigma_1^2)^{-\frac{N}{2}-\alpha-1}\exp\{-\frac{1}{2\sigma_1^2}(2\beta + RSS_1) - \\ & \frac{1}{2\xi^2}trace((\mathbf{Q} - \mathbf{I}_{3\times 3})(\mathbf{Q} - \mathbf{I}_{3\times 3})^{\mathrm{T}})\}, \end{aligned} \tag{2.2.17}$$

where $RSS_1 = \sum_{i=1}^{N}[S_i - S_0\exp(-b\mathbf{g}_i^{\mathrm{T}}\mathbf{Q}\mathbf{Q}^{\mathrm{T}}\mathbf{g}_i)]^2$ is the sum of squared residuals.

Consequently, the log-posterior distribution of parameters (σ_1 and $\mathbf{Q}$) has

the following form:

$$\log p(\mathbf{Q}, \sigma_1^2 \mid \mathbf{S}) \propto -(\frac{N}{2} + \alpha + 1)\log\sigma_1^2 + \{-\frac{1}{2\sigma_1^2}(2\beta + RSS_1) - \frac{1}{2\xi^2} trace((\mathbf{Q} - \mathbf{I}_{3\times 3})(\mathbf{Q} - \mathbf{I}_{3\times 3})^{\mathrm{T}})\}. \quad (2.2.18)$$

The estimates of σ_1 and $\mathbf{Q}$ (and, subsequently estimates of $\mathbf{D}$) can be obtained by maximising the log-posterior density (see Equation (2.2.18)). I.e.,

$$(\hat{\mathbf{Q}}, \hat{\sigma_1}^2) = \arg\max_{\mathbf{Q},\sigma_1^2} \left\{ p(\mathbf{Q}, \sigma_1^2 \mid \mathbf{S}) \right\}. \quad (2.2.19)$$

Then, $\hat{\mathbf{D}} = \hat{\mathbf{Q}}\hat{\mathbf{Q}}^{\mathrm{T}}$. In practice, we use a function *fminsearch* in MATLAB (R2009a, The MathWorks, Inc., Natick, Massachusetts, USA) to solve this optimisation problem. The function *fminsearch* uses the Nelder-Mead simplex search method [Lagarias *et al*., 1998]. The function *fminsearch* finds the minimum of a scalar function of several variables, starting at an initial estimate. We use the linear least squares estimation to produce the initial estimates of the parameters. The algorithm converges quickly for this study. One can also use random starting points when run the Nelder-Mead simplex search method to improve the optimisation. The conjugate gradient method is a more advaced numerical algorithm for optimisation in which the objective function is decreasing or increasing most rapidly [Daniel, 1967].

2.2.3.2 Double tensor estimation

In some regions of brain, there may be two distinct fibre orientations. Two diffusion tensors, then, need to be employed to characterise diffusion behaviours at such a voxel containing crossing fibres. Let us set $m = 2$ in the multi-tensor

model, then the double tensor model is given by

$$\begin{aligned} S_i &= \mu_i + \varepsilon_i \\ &= S_0[\exp(-b\mathbf{g}_i^{\mathrm{T}}\mathbf{Q}_1\mathbf{Q}_1^{\mathrm{T}}\mathbf{g}_i) + \exp(-b\mathbf{g}_i^{\mathrm{T}}\mathbf{Q}_2\mathbf{Q}_2^{\mathrm{T}}\mathbf{g}_i)] + \varepsilon_i, \end{aligned} \tag{2.2.20}$$

where $i = 1, 2, ...N$, ε_i's are i.i.d. Gaussian noise, with zero mean and σ_2^2 as the variance, i.e. $S_i \sim N(\mu_i, \sigma_2^2)$, $i = 1, 2, ..., N$.

Then, the likelihood function of $\Theta_2 = (\mathbf{Q}_1, \mathbf{Q}_2, \sigma_2^2)$ given the measured signal intensities $\mathbf{S}$ is

$$L(\mathbf{Q}_1, \mathbf{Q}_2, \sigma_2^2 \mid \mathbf{S}) = (\frac{1}{\sqrt{2\pi\sigma_2^2}})^N \exp\{-\frac{1}{2\sigma_2^2} RSS_2\}, \tag{2.2.21}$$

where $RSS_2 = \sum_{i=1}^{N} [S_i - S_0 \exp(-b\mathbf{g}_i^{\mathrm{T}}\mathbf{Q}_1\mathbf{Q}_1^{\mathrm{T}}\mathbf{g}_i) - S_0 \exp(-b\mathbf{g}_i^{\mathrm{T}}\mathbf{Q}_2\mathbf{Q}_2^{\mathrm{T}}\mathbf{g}_i)]^2$ is the residual sum of squares between the observed and the expected signals.

Now let us state our prior beliefs of the parameters $\mathbf{Q}_1$, $\mathbf{Q}_2$ and σ_2^2. When jointed into a single vector $\mathbf{H}_{\mathbf{Q}_1,\mathbf{Q}_2}^{\mathrm{T}} = (vec(\mathbf{Q}_1)^{\mathrm{T}}, vec(\mathbf{Q}_2)^{\mathrm{T}})$, $\mathbf{Q}_1$, $\mathbf{Q}_2$ become an 18-dimensional random vector. Thus, we need to specify the joint distribution of $(\mathbf{H}_{\mathbf{Q}_1,\mathbf{Q}_2}, \sigma_2^2)$. Since σ_2^2 is assumed to be distributed independently of $\mathbf{H}_{\mathbf{Q}_1,\mathbf{Q}_2}$, we first focus on the specification of the prior distribution for $\mathbf{H}_{\mathbf{Q}_1,\mathbf{Q}_2}$.

$$vec(\mathbf{Q}_1) \sim \mathbf{N}_9(vec(\mathbf{I}_{3\times3}), \xi_1^2\mathbf{I}_{9\times9}) \tag{2.2.22}$$

$$vec(\mathbf{Q}_2) \sim \mathbf{N}_9(vec(\mathbf{I}_{3\times3}), \xi_1^2\mathbf{I}_{9\times9}) \tag{2.2.23}$$

$$vec(\mathbf{Q}_1 - \mathbf{Q}_2) \sim \mathbf{N}_9(\mathbf{0}_{9\times1}, \xi_2^2\mathbf{I}_{9\times9}). \tag{2.2.24}$$

The given formulae (2.2.22), (2.2.23), (2.2.24) do not uniquely specify the joint distribution of $\mathbf{H}_{\mathbf{Q}_1,\mathbf{Q}_2}$ unless we explicitly state that $\mathbf{H}_{\mathbf{Q}_1,\mathbf{Q}_2}$ are jointly (18-variate) normal with the parameters specified by (2.2.22), (2.2.23), (2.2.24) above

[e.g. Melnick and Tenenbein, 1982].

Thus, we must state that

$$\mathbf{H}_{\mathbf{Q}_1,\mathbf{Q}_2} \sim \mathbf{N}_{18}(\mathbb{I}, \mathbf{\Sigma}), \tag{2.2.25}$$

where $\mathbb{I}^{\mathrm{T}} = (vec(\mathbf{I}_{3\times3})^{\mathrm{T}}, vec(\mathbf{I}_{3\times3})^{\mathrm{T}})$ and

$$\mathbf{\Sigma} = \begin{pmatrix} \xi_1^2\mathbf{I}_{9\times9} & (\xi_1^2 - 0.5\xi_2^2)\mathbf{I}_{9\times9} \\ (\xi_1^2 - 0.5\xi_2^2)\mathbf{I}_{9\times9} & \xi_1^2\mathbf{I}_{9\times9} \end{pmatrix}. \tag{2.2.26}$$

Let ρ be the coefficient of correlation between $\mathbf{Q}_{ij,1}$ and $\mathbf{Q}_{ij,2}$, $i, j \in 1, 2, 3$, then $\rho = 1 - \frac{\xi_2^2}{2\xi_1^2}$. Since $-1 \leq \rho = 1 - \frac{\xi_2^2}{2\xi_1^2} \leq 1$, $\xi_2^2 \leq 4\xi_1^2$. Note that $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are independent if and only if $2\xi_1^2 = \xi_2^2$.

For the variance of Gaussian noise σ_2^2, again we use an inverse gamma prior. I.e.

$$\sigma_2^2 \sim \textit{Inv-Gamma}(\alpha, \beta). \tag{2.2.27}$$

The logarithm of the probability density function of the posterior distribution is given by

$$\log p(\mathbf{Q}_1, \mathbf{Q}_2, \sigma_2^2 \mid \mathbf{S}) \propto -(\frac{N}{2} + \alpha + 1)\log\sigma_2^2 - \frac{1}{2\sigma_2^2}(2\beta + RSS_2) - (A + B) \tag{2.2.28}$$

where

$$A + B = \frac{1}{2}(\mathbf{H}_{\mathbf{Q}_1,\mathbf{Q}_2} - \mathbb{I})^{\mathrm{T}}\mathbf{\Sigma}^{-1}(\mathbf{H}_{\mathbf{Q}_1,\mathbf{Q}_2} - \mathbb{I}). \tag{2.2.29}$$

It is immediate to verify that

$$\mathbf{\Sigma}^{-1} = \frac{2}{\xi_2^2(2\xi_1^2 - 0.5\xi_2^2)} \begin{pmatrix} \xi_1^2\mathbf{I}_{9\times9} & -(\xi_1^2 - 0.5\xi_2^2)\mathbf{I}_{9\times9} \\ -(\xi_1^2 - 0.5\xi_2^2)\mathbf{I}_{9\times9} & \xi_1^2\mathbf{I}_{9\times9} \end{pmatrix}. \tag{2.2.30}$$

Thus,

$$\begin{aligned}
A+B &= \tfrac{1}{\xi_2^2(2\xi_1^2-0.5\xi_2^2)}[\xi_1^2 vec(\mathbf{Q}_1-\mathbf{I}_{3\times3})^{\mathrm{T}} vec(\mathbf{Q}_1-\mathbf{I}_{3\times3})+\\
&\quad (0.5\xi_2^2-\xi_1^2) vec(\mathbf{Q}_1-\mathbf{I}_{3\times3})^{\mathrm{T}} vec(\mathbf{Q}_2-\mathbf{I}_{3\times3})+\\
&\quad \xi_1^2 vec(\mathbf{Q}_2-\mathbf{I}_{3\times3})^{\mathrm{T}} vec(\mathbf{Q}_2-\mathbf{I}_{3\times3})+\\
&\quad (0.5\xi_2^2-\xi_1^2) vec(\mathbf{Q}_2-\mathbf{I}_{3\times3})^{\mathrm{T}} vec(\mathbf{Q}_1-\mathbf{I}_{3\times3})]\\
&= \tfrac{1}{\xi_2^2(2\xi_1^2-0.5\xi_2^2)}[\xi_2^2 vec(\mathbf{Q}_1-\mathbf{I}_{3\times3})^{\mathrm{T}} vec(\mathbf{Q}_2-\mathbf{I}_{3\times3})+\\
&\quad \xi_1^2 vec(\mathbf{Q}_1-\mathbf{Q}_2)^{\mathrm{T}} vec(\mathbf{Q}_1-\mathbf{Q}_2)]\\
&= \tfrac{1}{\xi_1^2\xi_2^2(1+\rho)}[\xi_2^2 trace((\mathbf{Q}_1-\mathbf{I}_{3\times3})(\mathbf{Q}_2-\mathbf{I}_{3\times3})^{\mathrm{T}})+\\
&\quad \xi_1^2 trace((\mathbf{Q}_1-\mathbf{Q}_2)(\mathbf{Q}_1-\mathbf{Q}_2)^{\mathrm{T}})]\\
&= \tfrac{1}{1+\rho}[\tfrac{1}{\xi_1^2} trace((\mathbf{Q}_1-\mathbf{I}_{3\times3})(\mathbf{Q}_2-\mathbf{I}_{3\times3})^{\mathrm{T}})+\\
&\quad \tfrac{1}{\xi_2^2} trace((\mathbf{Q}_1-\mathbf{Q}_2)(\mathbf{Q}_1-\mathbf{Q}_2)^{\mathrm{T}})].
\end{aligned}$$

Hence the natural choice of A and B:

$$\begin{aligned}
A &= \frac{1}{(1+\rho)\xi_1^2} trace((\mathbf{Q}_1-\mathbf{I}_{3\times3})(\mathbf{Q}_2-\mathbf{I}_{3\times3})^{\mathrm{T}}),\\
B &= \frac{1}{(1+\rho)\xi_2^2} trace((\mathbf{Q}_1-\mathbf{Q}_2)(\mathbf{Q}_1-\mathbf{Q}_2)^{\mathrm{T}}).
\end{aligned}$$

Now, the maximum a posteriori estimates of the parameters can be obtained by maximising the log-density (2.2.28) of the posterior distribution as stated below:

$$(\hat{\mathbf{Q}}_1, \hat{\mathbf{Q}}_2, \hat{\sigma_2}^2) = \arg\max_{\mathbf{Q}_1,\mathbf{Q}_2,\sigma_2^2} \left\{ \log P(\mathbf{Q}_1, \mathbf{Q}_2, \sigma_2^2 \mid \mathbf{S}) \right\}. \tag{2.2.31}$$

Then, $\hat{\mathbf{D}}_1 = \hat{\mathbf{Q}}_1\hat{\mathbf{Q}}_1^{\mathrm{T}}$, and $\hat{\mathbf{D}}_2 = \hat{\mathbf{Q}}_2\hat{\mathbf{Q}}_2^{\mathrm{T}}$ are the *maximum a posteriori* estimators. In practice, we maximise the log-posterior distribution (2.2.28) using the function *fminsearch* in MATLAB.

2.2.4 Bayesian model selection

In the multi-tensor model, the choice of the number of compartments m presents a model-selection problem. In the DTI community, it is commonly assumed that for each voxel, there are at most two distinct fibre orientations. Therefore, we have to choose between the single tensor model (M_1) or the double tensor model (M_2), parameterised by $(\mathbf{Q}, \sigma_1^2)$ and $(\mathbf{Q}_1, \mathbf{Q}_2, \sigma_2^2)$ respectively, given as follows:

$$\begin{aligned} M_1: \quad S_i &= S_0 \exp(-b\mathbf{g}_i^{\mathrm{T}}\mathbf{Q}\mathbf{Q}^{\mathrm{T}}\mathbf{g}_i) + \varepsilon_i,\ i = 1, ..., N. \\ M_2: \quad S_i &= S_0 \exp(-b\mathbf{g}_i^{\mathrm{T}}\mathbf{Q}_1\mathbf{Q}_1^{\mathrm{T}}\mathbf{g}_i) + S_0 \exp(-b\mathbf{g}_i^{\mathrm{T}}\mathbf{Q}_2\mathbf{Q}_2^{\mathrm{T}}\mathbf{g}_i) + \varepsilon_i,\ i = 1, ..., N. \end{aligned}$$

Since we have used Bayesian methods to estimate parameters in the multi-tensor model, we can directly use the Bayes factor for model selection on the basis of N measured signal intensities $\mathbf{S} = (S_1, ..., S_N)^{\mathrm{T}}$. The Bayes factor K_B in favour of M_2 over M_1 is given by [Bernardo and Smith, 2009]

$$K_B(\mathbf{S}) = \frac{p(M_2 \mid \mathbf{S})}{p(M_1 \mid \mathbf{S})} \left\{ \frac{p(M_2)}{p(M_1)} \right\}^{-1}, \tag{2.2.32}$$

which is the posterior odds in favour of M_2 divided by the prior odds in favour of M_2. A large value of K_B gives support for M_2 over M_1. Note that Bayesian model selection will not inform about which model is 'true', but rather about the preference for the model given the data and other information. A possible interpretation for Bayes factors [Jeffreys, 1961; Bernardo and Smith, 2009] is listed in Table 2.2 . We have carried out experiments for determining the number of diffusion tensors at voxels in some well-known regions (e.g. the corpus callosum and corana radiata). The thresholds of 3 listed in Table 2.2 are suitable for our study.

Table 2.2: Jeffreys' scale of evidence for Bayes factors

Bayes factor K_B	Strength of evidence
$K_B < 1/10$	Strong evidence for M_1
$1/10 < K_B < 1/3$	Moderate evidence for M_1
$1/3 < K_B < 1$	Weak evidence for M_1
$1 < K_B < 3$	Weak evidence for M_2
$3 < K_B < 10$	Moderate evidence for M_2
$K_B > 10$	Strong evidence for M_2

2.3 MCMC for the Bayesian Single Tensor Framework

As described in the previous section, the posterior distribution quantifies the remaining uncertainty in our current knowledge of the parameters.

In addition to point estimates of the parameters, we would also like to obtain a sample of their typical values and assess our confidence in the obtained estimates. Thus, in this section we use Markov chain Monte Carlo (MCMC) methods to sample parameters from the posterior distribution. Then, sampled fibre orientations are extracted from thus sampled diffusion tensors. Consequently, the sample mean of fibre orientations and corresponding credible cone are drawn which are essential for fibre tractography.

2.3.1 A combination MCMC algorithm

As a class of algorithms for sampling from probability distributions, MCMC methods have had a profound influence on Bayesian inference. MCMC methods provide an approximation of posterior statistics such as the posterior mean and variance. More importantly, we can use MCMC to evaluate posterior probabilities concerning the parameters and to construct corresponding credible intervals. Here, we focus on the single tensor model (the multi-tensor model with $m = 1$).

We use a combination of the Metropolis-Hastings algorithm [Gamerman and Lopes, 2006] and the Gibbs sampler [Gamerman and Lopes, 2006] to sample parameters $\mathbf{Q}$ and σ_1^2 respectively in the single tensor model. For this algorithm, at each iteration t, the next state $\mathbf{Q}^{(t+1)}$ is chosen by first sampling a candidate point $\mathbf{Q}^*$ from a *proposal distribution* $q(\mathbf{Q}^* \mid \mathbf{Q}^{(t)}, (\sigma_1^2)^{(t)})$. We let q be a Gaussian distribution depending on the current state $\mathbf{Q}^{(t)}$, i.e. $N(vec(\mathbf{Q}^{(t)}), \xi^2\mathbf{I}_{9\times 9})$. The candidate point $\mathbf{Q}^*$ is accepted with probability $\alpha(\mathbf{Q}^{(t)}, \mathbf{Q}^*)$, where:

$$\alpha(\mathbf{Q}^{(t)}, \mathbf{Q}^*) = \min\left\{1, \frac{p(\mathbf{Q}^* \mid \mathbf{S}, (\sigma_1^2)^{(t)}) q(\mathbf{Q}^{(t)} \mid \mathbf{Q}^*)}{p(\mathbf{Q}^{(t)} \mid \mathbf{S}, (\sigma_1^2)^{(t)}) q(\mathbf{Q}^* \mid \mathbf{Q}^{(t)}))}\right\}. \tag{2.3.1}$$

With the Gibbs sampler, the next state $(\sigma_1^2)^{(t+1)}$ is generated from $p(\sigma_1^2 \mid \mathbf{Q}, \mathbf{S})$ which is density of the *full conditional distribution* of the posterior distribution $p(\mathbf{Q}, \sigma_1^2 \mid \mathbf{S})$ in (2.2.17). The full conditional distribution of σ_1^2 is given by

$$p(\sigma_1^2 \mid \mathbf{Q}, \mathbf{S}) \propto (\sigma_1^2)^{-\frac{N}{2}-\alpha-1} \exp\left\{-\frac{1}{2\sigma_1^2}(2\beta + RSS_1)\right\}, \tag{2.3.2}$$

where $RSS_1 = \sum_{i=1}^{N} [S_i - S_0 \exp(-b\mathbf{g}_i^{\mathrm{T}}\mathbf{Q}\mathbf{Q}^{\mathrm{T}}\mathbf{g}_i)]^2$. Note that $p(\sigma_1^2 \mid \mathbf{Q}, \mathbf{S})$ is the density of an Inverse-gamma distribution, i.e. Inv-Gamma($\alpha_{\sigma_1^2}, \beta_{\sigma_1^2}$) where

$$\begin{aligned} \alpha_{\sigma_1^2} &= \frac{N}{2} + \alpha \\ \beta_{\sigma_1^2} &= \frac{1}{2}(2\beta + RSS_1). \end{aligned} \tag{2.3.3}$$

Therefore, the algorithm generates samples $(\mathbf{Q}^{(t)}, \left(\sigma_1^2\right)^{(t)}, t = 0, 1, 2, ...)$ through the following steps:

1. Start with an arbitrary initial set of parameter values $\mathbf{Q}^{(0)}, \left(\sigma_1^2\right)^{(0)}$.

2. Update from $\mathbf{Q}^{(t)}$ to $\mathbf{Q}^{(t+1)}$ with Metropolis-Hasting algorithm:

   2.1. Generate $\mathbf{Q}^*$ from $q(\mathbf{Q}^* \mid \mathbf{Q}^{(t)}, \sigma_1^{2(t)})$, where q follows $N_9(vec(\mathbf{Q}^{(t)}), \xi^2\mathbf{I}_{9\times 9})$.

2.2. Evaluate $\alpha(\mathbf{Q}^{(t)}, \mathbf{Q}^*) = \min\left\{1, \frac{p(\mathbf{Q}^*|\mathbf{S},\sigma_1^{2(t)})q(\mathbf{Q}^{(t)}|\mathbf{Q}^*)}{p(\mathbf{Q}^{(t)}|\mathbf{S},\sigma_1^{2(t)})q(\mathbf{Q}^*|\mathbf{Q}^{(t)}))}\right\}$

2.3. Sample a point U from a Uniform(0,1) distribution

2.4. Set

$$\mathbf{Q}^{(t+1)} = \begin{cases} \mathbf{Q}^* & \text{if } U \leq \alpha(\mathbf{Q}^{(i)}, \mathbf{Q}^*) \\ \mathbf{Q}^{(t)} & \text{otherwise.} \end{cases} \tag{2.3.4}$$

3. Update from $\left(\sigma_1^2\right)^{(t)}$ to $\left(\sigma_1^2\right)^{(t+1)}$ with the Gibbs sampler by generating $\left(\sigma_1^2\right)^{(t+1)}$ from $p(\sigma_1^2 \mid \mathbf{Q}^{(t+1)}, \mathbf{S})$.

Since the chain is aperiodic and irreducible, after a large number of iterations we have simulated an observation from the posterior distribution. To ensure that the results from the MCMC algorithm are reliable, i.e. the equilibrium of the chain has been reached, a *burn in* period is taken into account. After the burn in period, at each iteration, we obtain a dependent sample value from the posterior distribution $P(\mathbf{Q}, \sigma_1^2 \mid \mathbf{S})$.

2.3.2 Credible cone of fibre orientations

Once a random sample of $\mathbf{Q}$ is generated with the MCMC scheme, we also obtain a sample of diffusion tensors ($\mathbf{D} = \mathbf{Q}\mathbf{Q}^{\mathrm{T}}$) and consequently a sample of principal eigenvectors extracted from the diffusion tensors. It is commonly assumed that at a given voxel, the principal eigenvector of the diffusion tensor is aligned with the fibre orientation.

Now consider a sample of fibre orientations $\mathbf{q}_1, ..., \mathbf{q}_m$, where each $\mathbf{q}_k \in \mathbb{R}P^2$ is an axial direction ($\mathbf{q}_k \equiv -\mathbf{q}_k$ and $\| \mathbf{q}_k \| = 1$). We consider two methods for obtaining the mean direction.

Method 1

Suppose $\mathbf{q}_1, ..., \mathbf{q}_m$ are all in the same hemisphere after reflection with respect to the origin. The mean direction of the sample is given by [Mardia and Jupp, 2000, p.163]

$$\bar{\mathbf{q}}_d = \| \bar{\mathbf{q}} \|^{-1} \bar{\mathbf{q}} \tag{2.3.5}$$

where

$$\bar{\mathbf{q}} = \frac{1}{m} \sum_{k=1}^{m} \mathbf{q}_k \tag{2.3.6}$$

is the sample mean and $\| \bar{\mathbf{q}} \|$, the Euclidean norm of $\bar{\mathbf{q}}$, is the *mean resultant length*. Note that Method 1 will depend on how to define the hemisphere. Thus, this method is not satisfactory.

Method 2

Let $\bar{\mathbf{D}}$ be the scatter matrix about the origin, and $\bar{\mathbf{D}}$ is defined by

$$\bar{\mathbf{D}} = \frac{1}{m} \sum_{k=1}^{m} \mathbf{q}_k \mathbf{q}_k^{\mathrm{T}}. \tag{2.3.7}$$

Then the extrinsic mean direction is defined as the principal eigenvector $\mathbf{a}_1$ of $\bar{\mathbf{D}}$.

Now let us explain why $\mathbf{a}_1$ can be the mean direction $\bar{\mathbf{q}}_d$ ($\| \bar{\mathbf{q}}_d \| = 1$) of the sample. Define $c_k = \bar{\mathbf{q}}_d^{\mathrm{T}} \mathbf{q}_k$, $k = 1, ..., m$. Note that $c_k \mathbf{q}_k$ is the projection of $\mathbf{q}_k$ onto the line generated by scalar multiples of $\mathbf{q}_d^{\mathrm{T}}$. The aim is to find a mean direction $\bar{\mathbf{q}}_d$ minimising the averaged sum of squared dispersions from each sample point to the line $\bar{\mathbf{q}}_d$. We define the kth dispersion of $\mathbf{q}_k$ as the distance from the point $\mathbf{q}_k$ to $\bar{\mathbf{q}}_d$ (see Figure 2.1).

Then the objective function is given by

$$\min \frac{1}{m} \sum_{k=1}^{m} (1 - c_k^2). \tag{2.3.8}$$

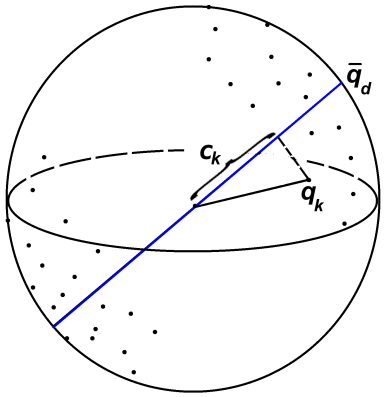

Figure 2.1: Dispersion from the sample point to the mean direction. The sample point $\mathbf{q}_k$. The mean direction $\bar{\mathbf{q}}_d$ (blue line).

It is equivalent to

$$\max \frac{1}{m} \sum_{k=1}^{m} c_k^2. \tag{2.3.9}$$

Now

$$\begin{aligned} \frac{1}{m} \sum_{k=1}^{m} c_k^2 &= \frac{1}{m} \sum_{k=1}^{m} (\bar{\mathbf{q}}_d^{\mathrm{T}} \mathbf{q}_k)^2 \\ &= \bar{\mathbf{q}}_d^{\mathrm{T}} [\frac{1}{m} \sum_{k=1}^{m} (\mathbf{q}_k \mathbf{q}_k^{\mathrm{T}})] \bar{\mathbf{q}}_d \\ &= \bar{\mathbf{q}}_d^{\mathrm{T}} \bar{\mathbf{D}} \bar{\mathbf{q}}_d. \end{aligned} \tag{2.3.10}$$

Since $\bar{\mathbf{D}}$ is symmetric, we can use spectral decomposition as follows

$$\bar{\mathbf{D}} = \mathbf{A} \Xi \mathbf{A}^{\mathrm{T}} = \sum_{k=1}^{3} \rho_k \mathbf{a}_k \mathbf{a}_k^{\mathrm{T}} \tag{2.3.11}$$

where $\mathbf{A}$ is an orthogonal matrix with three eigenvectors $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$ as three columns, and Ξ is a diagonal matrix with eigenvalues $\rho_1 \geq \rho_2 \geq \rho_3$ as three diagonal entries. Then

$$\begin{aligned} \frac{1}{m} \sum_{k=1}^{m} c_k^2 &= \bar{\mathbf{q}}_d^{\mathrm{T}} \bar{\mathbf{D}} \bar{\mathbf{q}}_d \\ &= \bar{\mathbf{q}}_d^{\mathrm{T}} [\sum_{k=1}^{3} \rho_k \mathbf{a}_k \mathbf{a}_k^{\mathrm{T}}] \bar{\mathbf{q}}_d \end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=1}^{3} \rho_k (\bar{\mathbf{q}}_d^{\mathrm{T}} \mathbf{a}_k)^2 \\
&\leq \rho_1 \sum_{k=1}^{3} (\bar{\mathbf{q}}_d^{\mathrm{T}} \mathbf{a}_k)^2 \\
&= \rho_1 \bar{\mathbf{q}}_d^{\mathrm{T}} [\sum_{k=1}^{3} \mathbf{a}_k \mathbf{a}_k^{\mathrm{T}}] \bar{\mathbf{q}}_d \\
&= \rho_1 \bar{\mathbf{q}}_d^{\mathrm{T}} \mathbf{A} \mathbf{I}_{3\times 3} \mathbf{A}^{\mathrm{T}} \bar{\mathbf{q}}_d \\
&= \rho_1. \qquad (2.3.12)
\end{aligned}
$$

Now

$$
\begin{aligned}
\mathbf{a}_1^{\mathrm{T}} \bar{\mathbf{D}} \mathbf{a}_1 &= \mathbf{a}_1^{\mathrm{T}} [\sum_{k=1}^{3} \rho_k \mathbf{a}_k \mathbf{a}_k^{\mathrm{T}}] \mathbf{a}_1 \\
&= \sum_{k=1}^{3} \rho_k (\mathbf{a}_1^{\mathrm{T}} \mathbf{a}_k)^2 \\
&= \rho_1, \qquad (2.3.13)
\end{aligned}
$$

since

$$
\mathbf{a}_1^{\mathrm{T}} \mathbf{a}_k = \begin{cases} 1 & \text{if } k = 1 \\ 0 & \text{if } k \neq 1 \end{cases}. \qquad (2.3.14)
$$

So $\frac{1}{m} \sum_{k=1}^{m} c_k^2$ or equivalently the averaged sum of squared dispersions is maximised when $\bar{\mathbf{q}}_d = \mathbf{a}_1$. Method 2 can directly provide the mean of axial data $\mathbf{q}_1, ..., \mathbf{q}_m$ without reflecting them onto the same hemisphere, since

$$
\bar{\mathbf{D}} = \frac{1}{m} \sum_{k=1}^{m} \mathbf{q}_k \mathbf{q}_k^{\mathrm{T}} = \sum_{k=1}^{m} (-\mathbf{q}_k)(-\mathbf{q}_k^{\mathrm{T}}). \qquad (2.3.15)
$$

Therefore, the mean direction calculated with Method 2 is more preferable for DTI study.

The 95% credible cone for the sample of fibre orientations can be obtained by ordering the distances from the mean direction $\bar{\mathbf{q}}_d$ to the samples and cutting

off at the 95th percentile. Figure 2.2 shows an example of the 95% credible cone for the sample of fibre orientations. Note that with this approach the credible

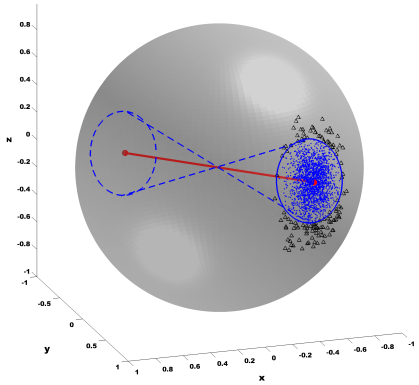

Figure 2.2: An example of the 95% credible cone for the sample of fibre orientations. The red line represents the mean direction. Points in blue are ends of sample directions that fall in the 95% credible cone.

cone is circular. An alternative method is to fit a bivariate Gaussian distribution in the tangent plane and then use the usual 95% confidence ellipse.

2.4 Applications

2.4.1 Simulation study

The purpose of this simulation study is to compare three diffusion gradient direction schemes used with the single and double tensor models, with the Bayesian and LLS estimates. The accuracy of DTI measurements depends on the applied schemes of diffusion gradient directions. The MR scanner has a set of magnetic field gradients applied at directions $\mathbf{g}_1, ..., \mathbf{g}_N \in \mathbb{R}P^2$ ($\mathbf{g}_i \equiv -\mathbf{g}_i$ and $\| \mathbf{g}_i \| = 1$). Details on how to design diffusion gradient direction schemes can be found in Chapter 4. In this study, we consider three diffusion gradient direction schemes, Phillips 15 (Figure 2.3(a)), Phillips 32 (Figure 2.3(b)), and Uniform 32 (Figure 2.3(c)) directions schemes (see Appendix 5.2.6).

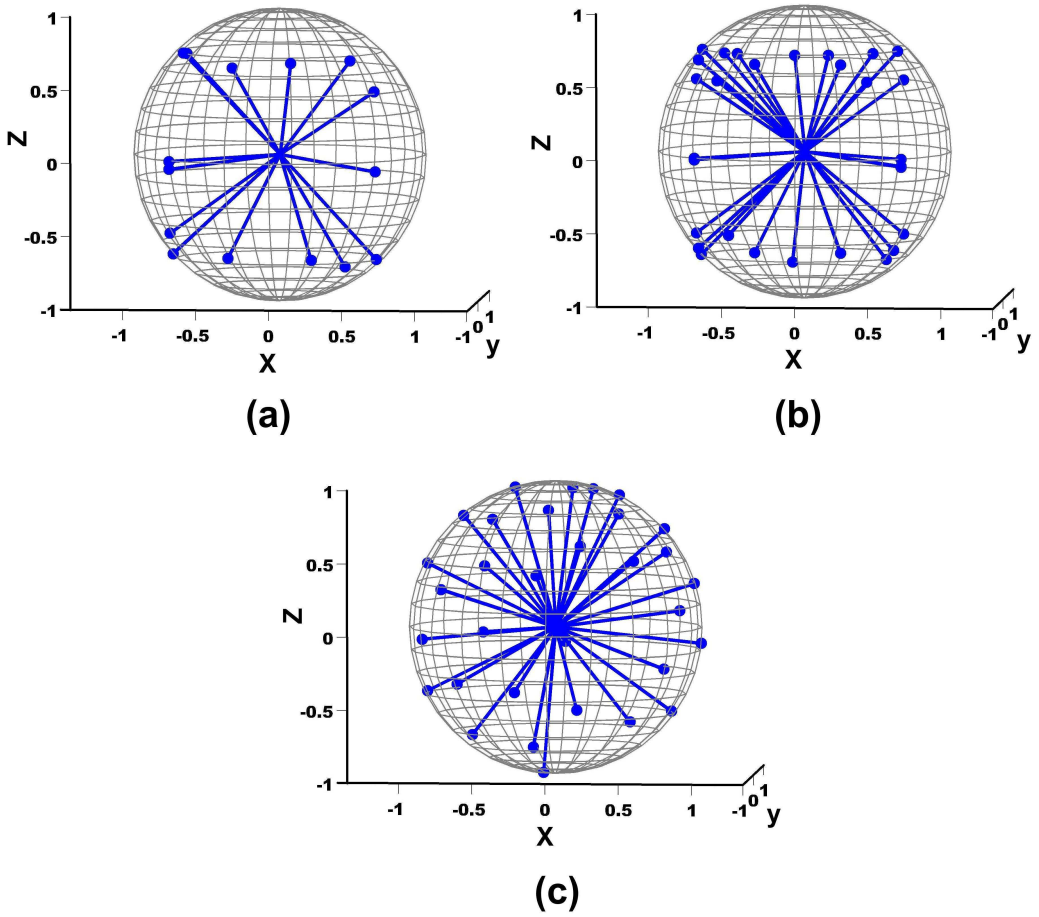

Figure 2.3: Three diffusion gradient direction schemes. (a) Phillips 15 (b) Phillips 32 and (c) Uniform 32.

When analysing the single tensor model, signal vector $\mathbf{S} = (S_1, ..., S_N)$ has been sampled n times from the multivariate Gaussian distribution, $N(\boldsymbol{\mu}, \sigma_1^2 \mathbf{I}_{N\times N})$ with $\boldsymbol{\mu} = (\mu_1, \mu_2, ..., \mu_N)$, for the three diffusion direction schemes. Let n be the sample size, that is, the number of generated values of $\mathbf{S}$ from a given diffusion direction scheme with N directions. Note that in real experiments we usually have $n = 1$ sample at each voxel. Let M be the number of Monte Carlo (MC) simulations. So, from each MC simulation we can estimate one diffusion tensor from $n \times N$ generated signals.

The *root mean squared errors* (RMSE) of estimators $\hat{\mathbf{D}}$ is one such measure of

accuracy, and is given by

$$RMSE(\hat{\mathbf{D}}) = \sqrt{\frac{1}{M}\sum_{k=1}^{M} \parallel \hat{\mathbf{D}}(k) - \mathbf{D} \parallel^2}. \tag{2.4.1}$$

where $\hat{\mathbf{D}}(k)$ is the estimate from the kth MC simulation and $\parallel \mathbf{A} \parallel = \sqrt{\text{trace}\left\{\mathbf{A}^{\mathrm{T}}\mathbf{A}\right\}}$ is the Euclidean norm. We can also define the RMSE of FA values of tensor estimators as

$$RMSE(FA) = \sqrt{\frac{1}{M}\sum_{k=1}^{M}(FA(k) - FA)^2}, \tag{2.4.2}$$

where $FA(k)$ is the FA of $\hat{\mathbf{D}}(k)$ from the kth MC simulation, and FA is the true value.

For the Bayesian single tensor framework, we set the prior parameter $\xi = 1$ for the prior distribution of $\mathbf{Q}$ in (2.2.12), $\alpha = 2.1$ and $\beta = 1000$ (the sum of squared residuals) for the prior distribution of σ_1^2 in (2.2.14).

For the Bayesian double tensor framework, we set the prior parameters $\xi_1 = \xi_2 = 1$, $\alpha = 2.1$ and $\beta = 1000$ for prior distributions of $\mathbf{Q}_1$, $\mathbf{Q}_2$, $\mathbf{Q}_1 - \mathbf{Q}_2$ and σ_2^2 in (2.2.22) to (2.2.27), respectively.

2.4.1.1 Single tensor model

For the single tensor model, a simulation study was performed with the following diffusion tensor:

$$\mathbf{D} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}. \tag{2.4.3}$$

The eigenvalues are 1, 2 and 3, and the corresponding eigenvectors are along x, y and z axis respectively (Figure 2.4). The FA value measuring the diffusion anisotropy is 0.4629.

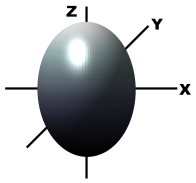

Figure 2.4: The diffusion ellipsoid of **D**.

We carry out two simulation studies for the single tensor as follows:

1. Consider a range of variances $\sigma_1^2 \in \{1, 2, ..., 10\}$. Fix the sample size $n = 1$ and $S_0 = 500$, perform $M = 100$ MC simulations for each σ_1^2. Then, σ_1/S_0 ranges from $\frac{1}{500}$ ($= 0.0020$) to $\frac{\sqrt{10}}{500}$ (≈ 0.0063). That is, SNR $= 500$ to SNR ≈ 158.1.

2. Consider a range of sample sizes $n \in \{1, 2, ..., 30\}$. Fix $\sigma_1^2 = 5$ and $S_0 = 500$, perform $M = 100$ MC simulations for each n. Then, $\sigma_1/S_0 \approx 0.045$ (SNR$= 353.55$).

After performing simulation study 1, Figure 2.5 and Figure 2.6 show $RMSE(\hat{\mathbf{D}})$ and $RMSE(FA)$ respectively with Bayesian and LLS estimations from the Phillips 15, Phillips 32 and Uniform 32 diffusion direction schemes as σ_1^2 increases from 1 to 10. Since the NLS and LLS methods produce similar estimates for diffusion tensor [Zhou, 2006], in this study we only compare the Bayesian estimates with the LLS estimates. In Figure 2.5, as σ_1^2 increases from 1 to 10, $RMSE(\hat{\mathbf{D}})$ becomes larger. It clearly shows that Bayesian estimation performs better than LLS estimation, since given a direction scheme for each σ_1^2, $RMSE(\hat{\mathbf{D}})$ with the Bayesian method is always smaller than that with LLS method. We can also see that the Phillips 15 direction scheme has larger $RMSE(\hat{\mathbf{D}})$ than other schemes. The Uniform 32 direction scheme is more preferable with smaller $RMSE(\hat{\mathbf{D}})$. In Figure 2.6, $RMSE(FA)$ values were classified clearly according to the three direction schemes. The Uniform 32 direction scheme is the best one among the three schemes. Phillips 15 is the worst with similar $RMSE(FA)$ from Bayesian

and LLS estimators.

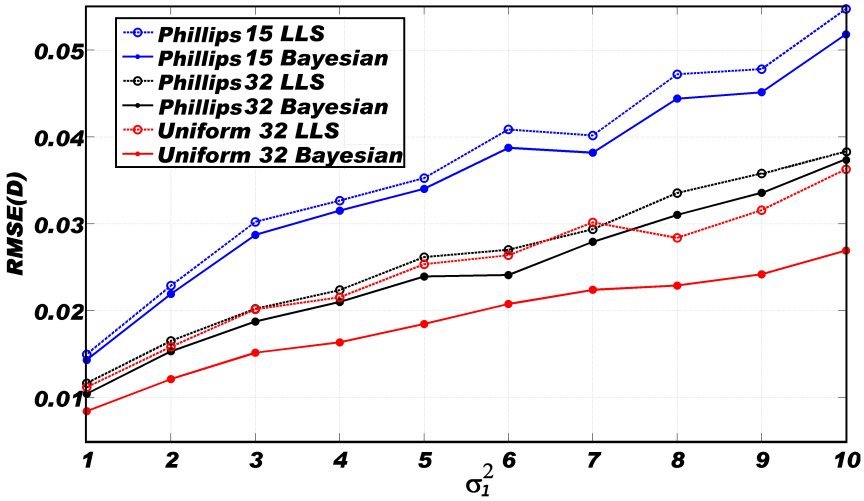

Figure 2.5: Root mean squared errors $RMSE(\hat{\mathbf{D}})$ from three diffusion gradient direction schemes. The Bayesian and LLS estimations are used, and σ_1^2 increases from 1 to 10.

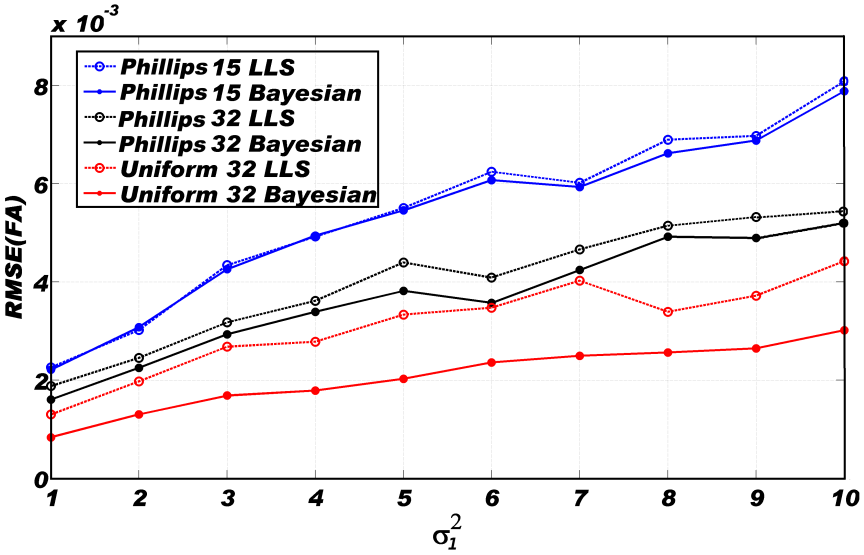

Figure 2.6: Root mean squared errors $RMSE(FA)$ from three direction schemes. The Bayesian and LLS estimations are used, and σ_1^2 increases from 1 to 10.

Figure 2.7 and Figure 2.8 show the results from simulation study 2 as the sample size n increases from 1 to 30. The values of $RMSE(\hat{\mathbf{D}})$ in Figure 2.7 decrease dramatically from $n = 1$ to $n = 15$, and then the improvement is slower. There is not a large difference between Bayesian and LLS methods for the two Phillips schemes. However, the Bayesian estimators from the Uniform 32 scheme is significantly better than LLS estimators. Similar patterns can be

shown in Figure 2.8 where the Bayesian method for Uniform 32 scheme is the best.

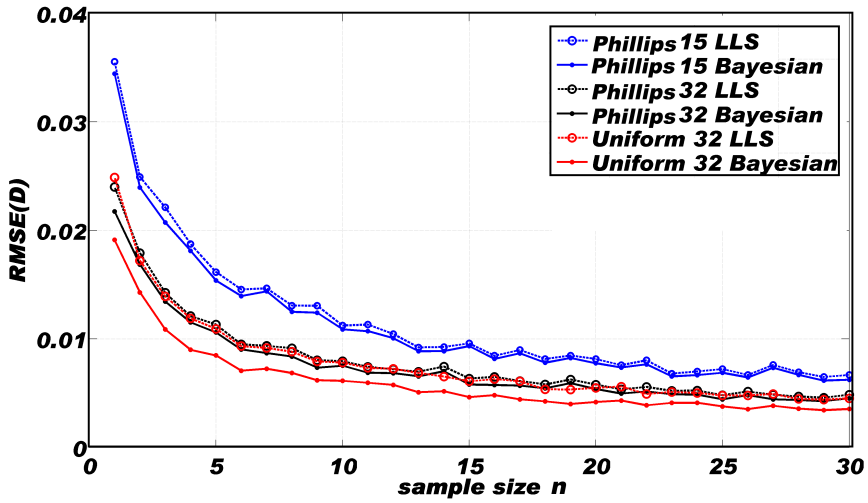

Figure 2.7: Root mean squared errors $RMSE(\hat{\mathbf{D}})$ from three diffusion direction schemes. The Bayesian and LLS estimations are used, and the sample size n increases from 1 to 30.

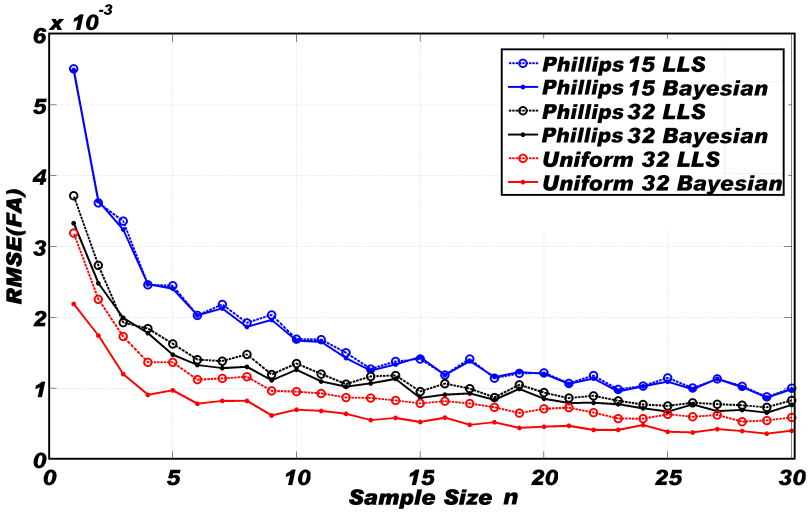

Figure 2.8: Root mean squared errors $RMSE(FA)$ from three diffusion direction schemes. The Bayesian and LLS estimations are used, and the sample size n increases from 1 to 30.

To conclude, the Bayesian method is better than LLS method, and the Uniform 32 direction scheme are clearly preferable to two Phillips schemes in our simulation study. It is not surprising that the 32 direction schemes are better.

2.4.1.2 MCMC sampling

Let us focus on the MCMC sampling from the posterior distribution of the model parameters. Namely, we fit the single tensor model to a real data consisting of $n = 1$ vector S acquired with the Uniform 32 scheme and corresponding to a single voxel. The Bayesian (MAP) estimate of $\mathbf{D}$ is as follows:

$$\mathbf{D}_{MAP} = \begin{pmatrix} 0.00053 & -0.00002 & 0.00003 \\ -0.00002 & 0.00058 & -0.00022 \\ 0.00003 & -0.00022 & 0.00112 \end{pmatrix}.$$

It may be noted that the real scale for the tensor coefficients is different from that in the previous synthetic examples.

Thus, we would also like to sample from the posterior distribution of $\mathbf{D}$ and generate 10^4 samples of $\mathbf{D}$ using our MCMC scheme (see Section 2.3.1). Figure 2.9 shows samples of the three diagonal elements of $\mathbf{D}$. The samples appear to become stationary after around 400 iterations, which we then accept as the burn-in period. Figure 2.10 shows the mean axis and 95% credible cone of fibre orientations from the MCMC samples after the burn-in period in the given voxel.

2.4.1.3 Double tensor model

Now consider the double tensor model which captures two distinct diffusion behaviours at a voxel by fitting two diffusion tensors. Two simulation studies were carried out (1) to measure the goodness of fit of the Bayesian double tensor model with MC simulations for two given diffusion tensors, and (2) to measure the dependence of the estimation on the angle between the two principal diffusion directions.

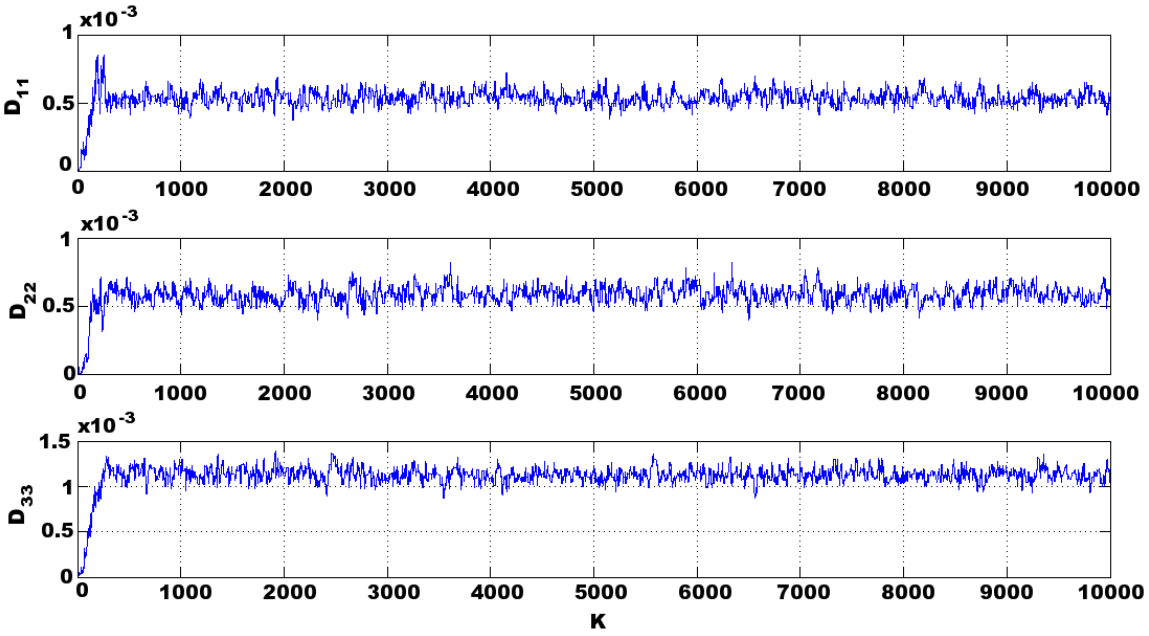

Figure 2.9: Samples of D_{11}, D_{22} and D_{33} with MCMC sampling.

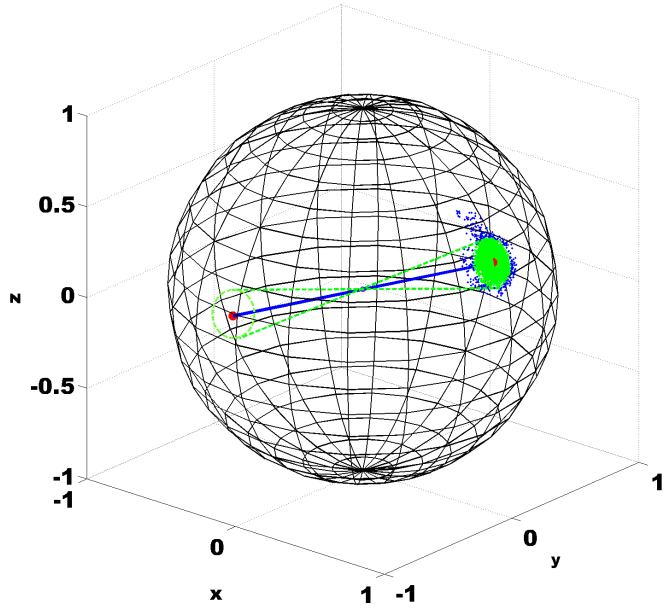

Figure 2.10: Green points are samples in 95% credible cone of fibre orientation. The blue line between red dots is the mean axis of fibre orientation in the voxel.

In the first simulation study, we fix the true $\mathbf{D_1}$ and $\mathbf{D}_2$ as follows:

$$\mathbf{D_1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}, \mathbf{D_2} = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

and corresponding ellipsoids are shown in Figure 2.11.

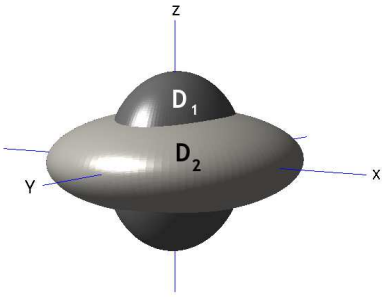

Figure 2.11: Ellipsoids for the two defined diffusion tensors. $\mathbf{D}_1$ is oblate and $\mathbf{D}_2$ is prolate.

Consider a range of sample sizes $n \in \{1, 5, 10, 15, 20, 25, 30\}$. We fix the variance $\sigma_2^2 = 3$, and carry out 100 MC simulations for each n. Figure 2.12 shows the means and standard errors of the estimates of the diffusion coefficients from the MC simulations. For both $\mathbf{D}_1$ and $\mathbf{D}_2$, the Uniform scheme provides excellent estimators. For example, the true value of $\mathbf{D}_1(3,3)$ is 3. The mean of estimates with the Uniform 32 scheme is almost the same as the true value. However, the means of estimates with the Phillips 15 scheme are all higher than the true value. Figure 2.13 shows RMSE plots for the $\mathbf{D}_1$ and $\mathbf{D}_2$ estimators. The Uniform 32 scheme results in much lower RMSE when compared with the Phillips schemes. It is noticeable that Phillips 32 scheme shows growing RMSE at sample size 20.

In simulation study 2, we fix one diffusion tensor as:

$$\mathbf{D_1} = \begin{pmatrix} 1/4 & 0 & 0 \\ 0 & 1/4 & 0 \\ 0 & 0 & 20 \end{pmatrix},$$

which is a strongly linear tensor with eigenvalues 20, 1/4 and 1/4, and with corresponding eigenvectors along z, y and x axis respectively. Let $\mathbf{v}_1$ be the principal eigenvector of $\mathbf{D_1}$, then $\mathbf{v}_1 = [0, 0, 1]^{\mathrm{T}}$. Now consider four settings of

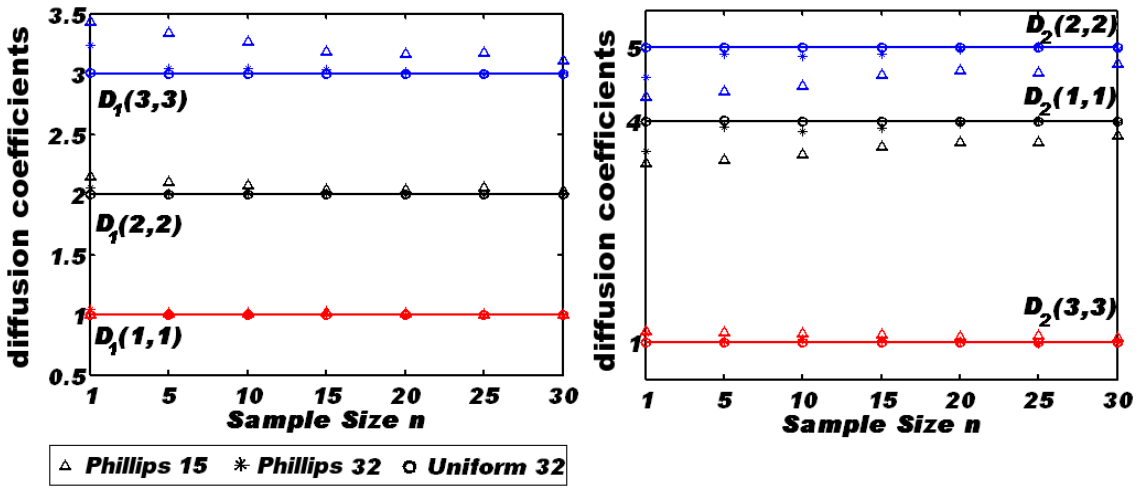

Figure 2.12: The mean of diffusion coefficients from MC simulations. The sample size $n \in \{1, 5, 10, 15, 20, 25, 30\}$.

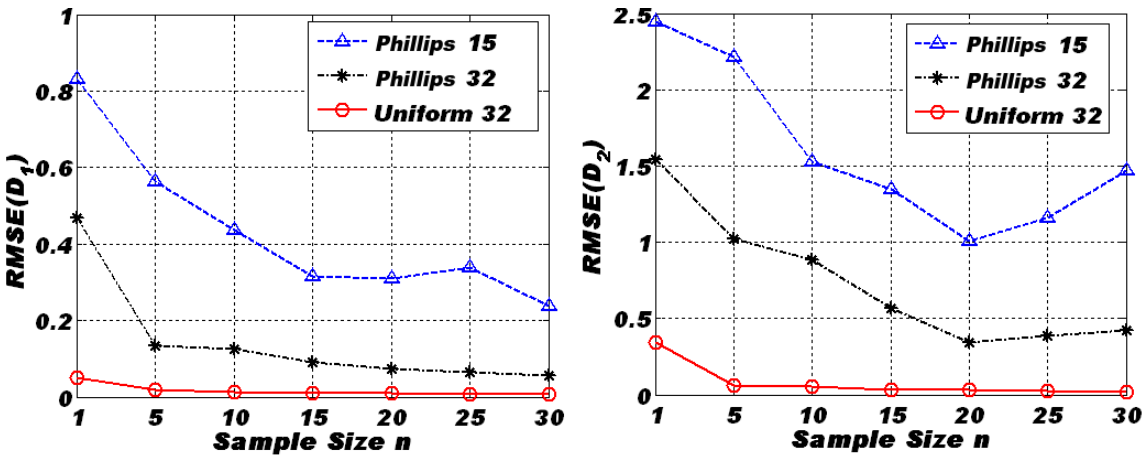

Figure 2.13: RMSE of two tensors. The sample size $n \in \{1, 5, 10, 15, 20, 25, 30\}$.

diffusion tensors $\mathbf{D}_2^a$, $\mathbf{D}_2^b$, $\mathbf{D}_2^c$ and $\mathbf{D}_2^d$. These four diffusion tensors have same eigenvalues 15,1/4 and 1/4, but different eigenvectors. Let $\mathbf{E}_2^i = [\mathbf{v}_1^i, \mathbf{v}_2^i, \mathbf{v}_3^i], i \in \{a, b, c, d\}$, where $\mathbf{v}_1^i$, $\mathbf{v}_2^i$, and $\mathbf{v}_3^i$ are three eigenvectors of $\mathbf{D}_2^i$, with $\mathbf{v}_1^i$ being principal. The angle between $\mathbf{v}_1$ and $\mathbf{v}_1^i, i \in \{a, b, c, d\}$ is denoted as $\phi_i, i \in \{a, b, c, d\}$. The eigenvectors of $\mathbf{D}_2^a$, $\mathbf{D}_2^b$, $\mathbf{D}_2^c$ and $\mathbf{D}_2^d$, and the corresponding angles ϕ_a, ϕ_b, ϕ_c and ϕ_d are listed in Table 2.3. Figure 2.14 shows the diffusion ellipsoids of $\mathbf{D_1}$, $\mathbf{D}_2^a$, $\mathbf{D}_2^b$, $\mathbf{D}_2^c$ and $\mathbf{D}_2^d$.

For each of the three diffusion direction schemes, signal vectors $\mathbf{S} = (S_1, ..., S_N)$ were generated from the multivariate Gaussian distribution according to the

Table 2.3: Settings of eigenvectors of $\mathbf{D}_2^i$, $i = a, b, c, d$, and corresponding angles ϕ_i.

	$[\mathbf{v}_1^i, \mathbf{v}_2^i, \mathbf{v}_3^i]$	ϕ_i
$\mathbf{D}_2^a$	$\begin{bmatrix} 1.0000 & 0.0000 & 0 \\ 0 & 0 & 1.0000 \\ 0 & 1.0000 & 0 \end{bmatrix}$	90°
$\mathbf{D}_2^b$	$\begin{bmatrix} 0.9063 & -0.4226 & 0 \\ 0 & 0 & 1.0000 \\ 0.4226 & 0.9063 & 0 \end{bmatrix}$	65°
$\mathbf{D}_2^c$	$\begin{bmatrix} 0.6428 & -0.7660 & 0 \\ 0 & 0 & 1.0000 \\ 0.7660 & 0.6428 & 0 \end{bmatrix}$	40°
$\mathbf{D}_2^d$	$\begin{bmatrix} 0.2588 & -0.9659 & 0 \\ 0 & 0 & 1.0000 \\ 0.9659 & 0.2588 & 0 \end{bmatrix}$	15°

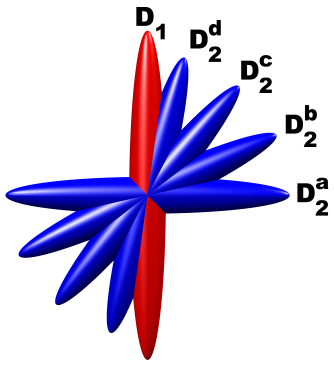

Figure 2.14: Diffusion ellipsoids of $\mathbf{D}_1$ (in red), $\mathbf{D}_2^a$, $\mathbf{D}_2^b$, $\mathbf{D}_2^c$ and $\mathbf{D}_2^d$.

double tensor model with one tensor fixed as $\mathbf{D}_1$ and the second one set to be $\mathbf{D}_2^i, i \in \{a, b, c, d\}$. For each of these 12 ($3 \times 4$) settings, 100 MC simulations were carried out with variance $\sigma^2 = 5$ and sample size $n = 1$. Figure 2.15 shows box plots of the Bayesian estimates of the angles ϕ_a, ϕ_b, ϕ_c and ϕ_d from the Bayesian estimates. It is clear that the Uniform 32 scheme performs the best with the medians of the Bayesian estimates which are almost the same as the true values. However, there are big variations of estimators from the two Phillips schemes. Phillips 15 provides the worst medians and more outliers, which is not surprising as there are much fewer measurements.

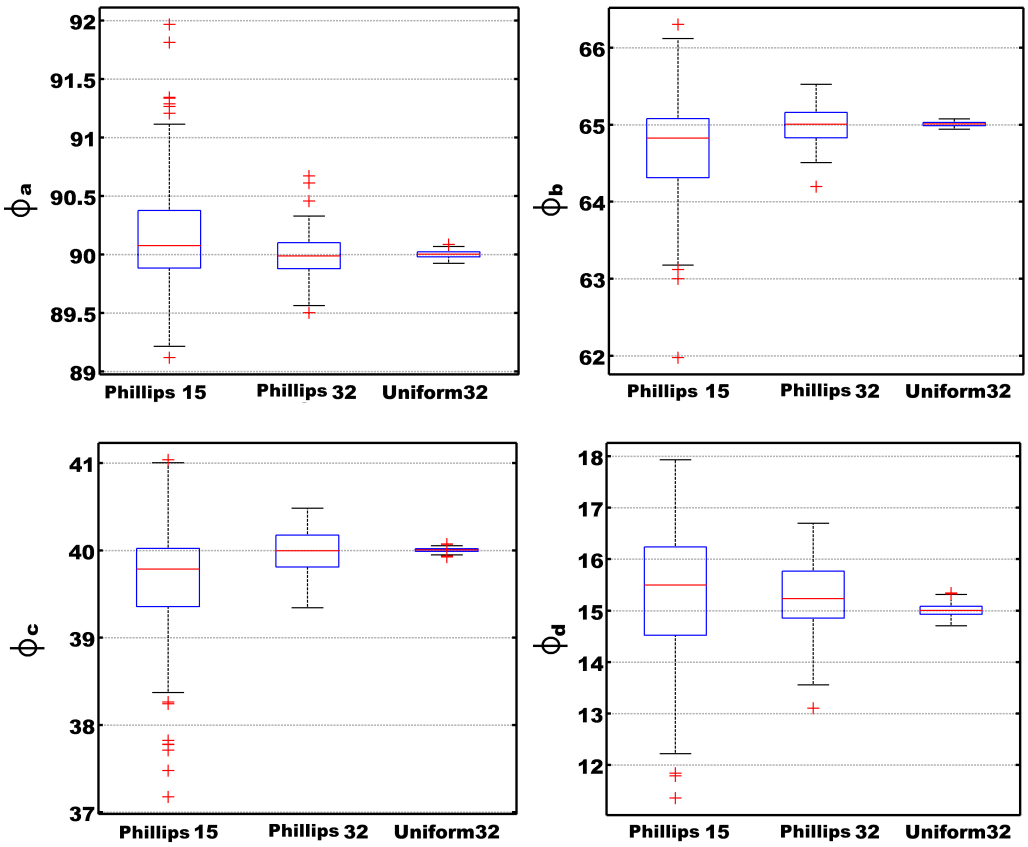

Figure 2.15: Box plots of angles ϕ_a, ϕ_b, ϕ_c and ϕ_d from Bayesian estimates.

Let $\hat{\phi}_i$, $i = a, b, c,$ and d be Bayesian estimators of the angles ϕ_i, $i = a, b, c, d$ respectively. Figure 2.16 shows histograms of the deviations $\hat{\phi}_i - \phi_i$, $i \in \{a, b, c, d\}$ of $\hat{\phi}_i$ from the true angles (90°, 65°, 40° and 15°) using the Uniform 32 direction scheme. It is clear that the mean square error of $\hat{\phi}_d$ is much larger than the variance of the other three estimates. We also estimate the mean of the estimators $\hat{\phi}_i$, $i \in \{a, b, c, d\}$, namely 90.0007, 65.0080, 40.0038 and 15.0028 respectively. Overall, the estimator for $\phi_a = 90°$ performs best, which is expected given that $\phi_a = 90°$ is the maximal possible separation between any two tensors.

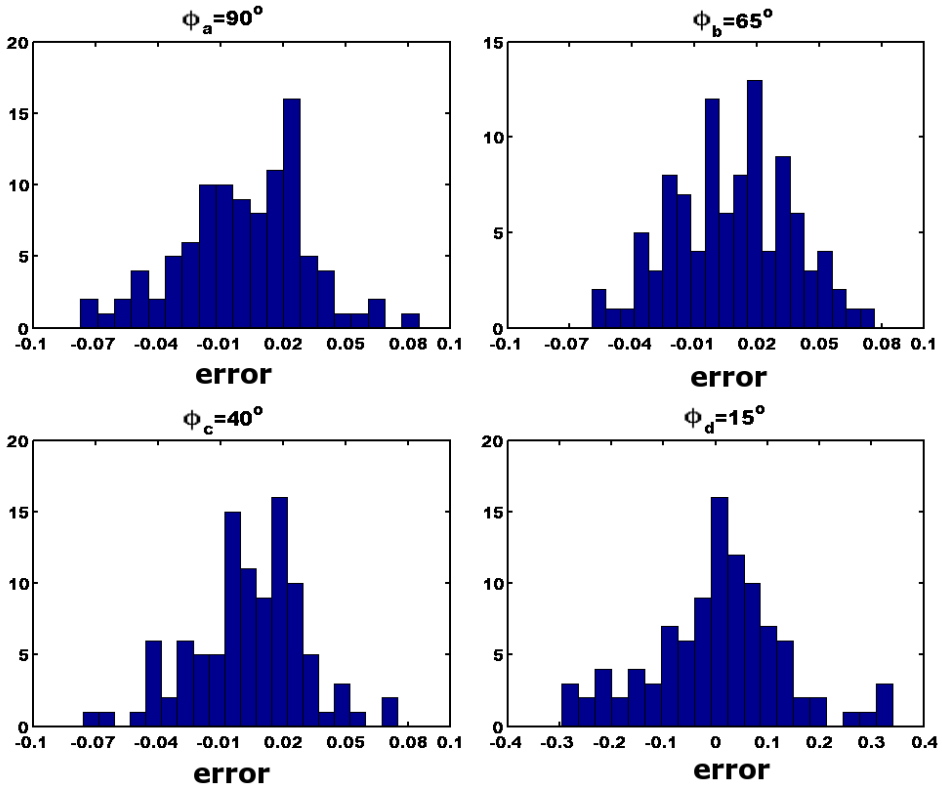

Figure 2.16: Histograms of errors of Bayesian estimates (ϕ_a, ϕ_b, ϕ_c and ϕ_d) applying the Uniform 32 direction scheme.

2.4.2 Real data

2.4.2.1 Materials

A set of diffusion weighted MR images acquired with the Uniform 32 DTI diffusion gradient direction scheme (see Figure 2.3 (c)) from a healthy human brain was provided by the Academic Radiology Department of Queen's Medical Centre, University of Nottingham. The MR images were acquired using a spin echo EPI (echo planar imaging) sequence with diffusion weighting gradients applied with a weighting factor of b=1000 s/mm^2. 52 interleaved contiguous transaxial slices were acquired throughout the subject's head in a matrix of 112x112 (interpolated to 224x224) with an acquisition voxel size of 1x1x2 mm^3. For each slice, the acquisition was repeated for each of the 32 non-collinear directions according to the Uniform 32 direction scheme, and once with no dif-

fusion weighting (b=0).

All methods used in this section have been programmed with MATLAB (R2009a, The MathWorks, Inc., Natick, Massachusetts, USA). We also use MATLAB to obtain figures and graphs for visualisation of our results.

2.4.2.2 Single tensor model

Figure 2.17 shows comparison of FA maps obtained using the LLS (Figure 2.17 a, c and e) and the single tensor Bayesian (Figure 2.17 b, d and f) estimators. FA maps from an axial slice of the brain found with the LLS and the Bayesian methods are displayed in Figure 2.17 a and b, respectively. Voxels containing diffusion tensor estimates which are non-positive-definite (i.e. with at least one negative eigenvalue) are all coloured in red.

It is clear that all the tensor estimators obtained with the Bayesian method are (semi) positive-definite as expected. However, non positive-definite estimates do appear with the LLS method. For example, insets a1 and b1 of Figure 2.17 zoom in on the splenium (sp), and it is clear that a1 contains many non positive-definite estimates. Recall that such estimates violate the basic constraint of diffusion modelling and hence may hardly be useful. We also display FA maps from coronal (Figure 2.17 c and d) and sagittal (Figure 2.17 e and f) slices. Although our Bayesian estimation, unlike LLS, ensures positive definiteness of the tensor, both the methods provide similar brain structures as represented by the anisotropy maps in this example.

In the DTI community, line and colour coded orientation maps are two useful visualisation methods of displaying the principal eigenvector $\mathbf{v}_1$ which represents the main fibre orientation at the voxel. Figure 2.18 (a) shows a ROI in the FA map (coronal view) obtained with the Bayesian method. The ROI contains the following three fibre bundled: the corpus callosum (cc), corana radiata (cr)

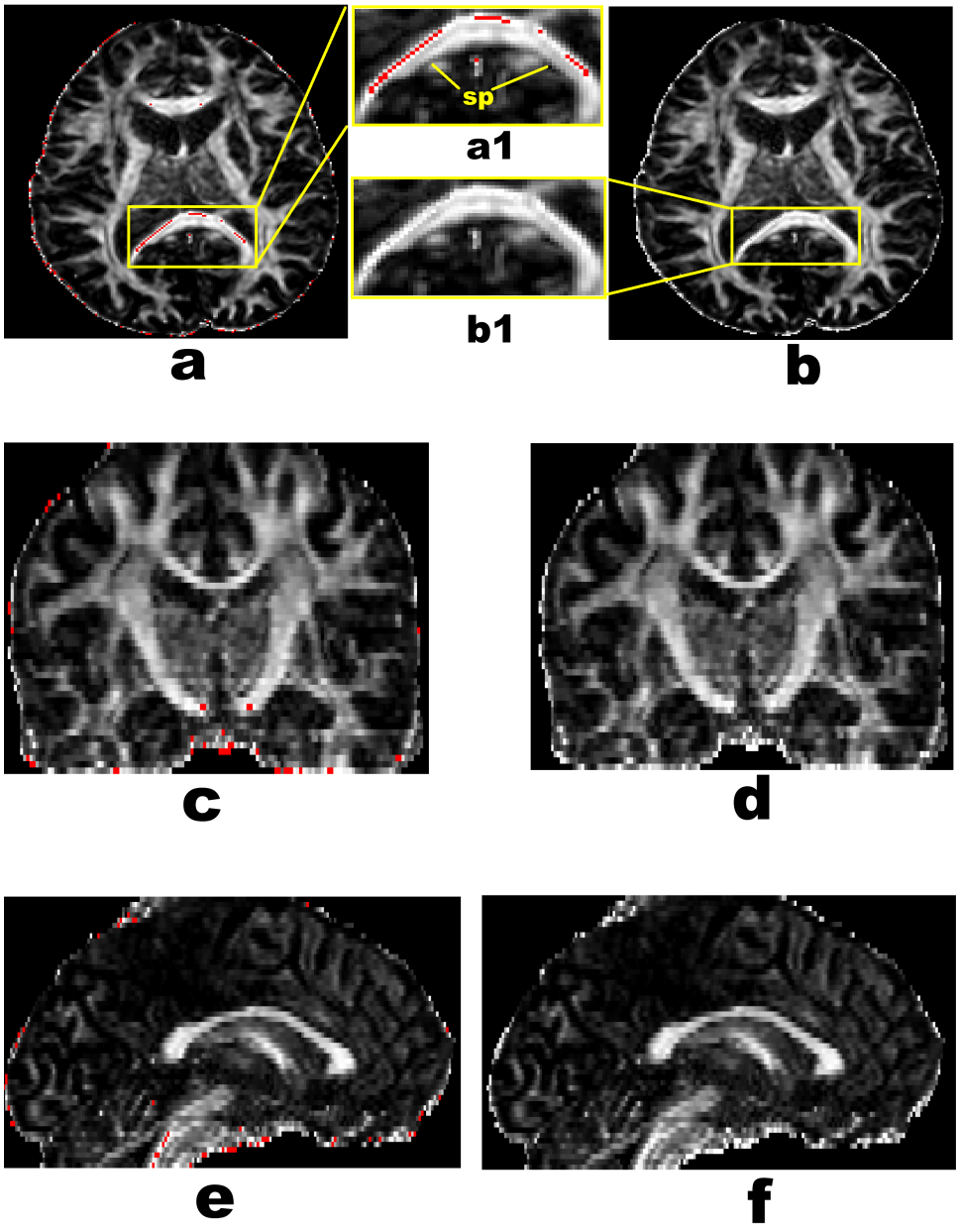

Figure 2.17: Comparison of FA maps from single tensor estimators with LLS and Bayesian methods. Maps a, c and e are from LLS estimates. Maps b, d, and f are from Bayesian estimates. a and b are from an axial slice; a1 and b1 are zoomed insets; c and d are from coronal slice; e and f are from sagittal slice. Red voxels contain non-positive-definite estimators.

and cingulum (cg). Figure 2.18 (b) is the line fibre orientation map of the ROI with FA background. Each line represents the principal diffusion direction at the corresponding voxel. We can see that the cingulum is perpendicular to the plane of the figure, the corona radiata orientates itself vertically, and the corpus

callosum with high anisotropy has a 'U' shape tract. Similar features of three fibre tracts can be seen in Figure 2.18 (c) which is the colour coded orientation map. A colour is assigned to each voxel using $\mathbf{v}_1 = (v_{1x}, v_{1y}, v_{1z})^T$. The absolute value of the v_{1x}, v_{1y}, v_{1z} components are used as red (left-right direction), green (front-back) and blue (feet-head) channels. Here, the colour intensities are additionally scaled by the FA values. Figure 2.18 (d) is the diffusion ellipsoid map, where the ellipsoids are also rescaled by the FA values. The diagonal tract on the lower left is the anterior limb of the internal capsule and on the lower right we see the superior fronto-occipital fasciculus.

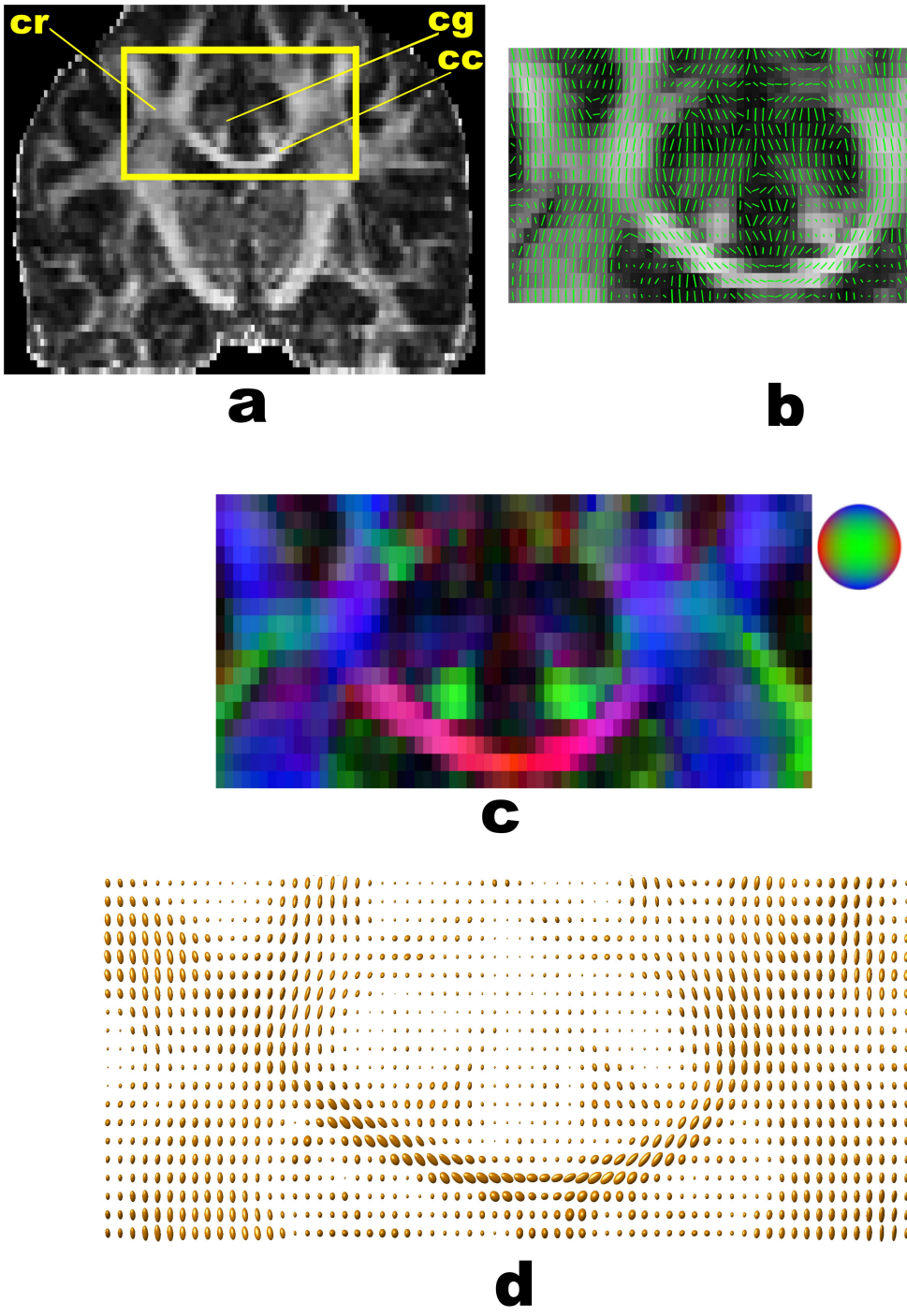

Figure 2.18: Visualisations of Bayesian diffusion anisotropy and fibre orientation. The FA map (coronal view) containing the ROI of three main fibre tracts the corpus callosum (cc), corona radiata (cr) and cingulum (cg) (a), the line (green) fibre orientation map of the ROI with FA background (b), the colour coded orientation map scaled by FA (c) with red (left-right direction), green (front-back) and blue (feet-head) directions, and diffusion ellipsoid map with the volume of the ellipsoids scaled by FA (d) scaled to have volume proportional to FA.

2.4.2.3 Multiple tensors fitting with model selection

The Bayesian single tensor and double tensor (multi-tensor model with $m = 2$) models were fitted to the data in the ROI shown in Figure 2.19 which is a colour coded orientation map (axial slice). The Bayes factor with threshold $K_B = 3$ was used for model selection, i.e., when $K_B < 3$, the single tensor model will be fitted at a given voxel, otherwise, double tensor model will be fitted. In Figure 2.20 (magnified ROI line fibre orientation map with FA background), a crossing of the anterior part of the superior longitudinal fasciculus (with a strong front-back orientation) with a tract of the corona radiata (from inferior regions) is presented. Such a crossing leads to a lower fit at this point from a single tensor model, however the crossing is captured very well in the double tensor model.

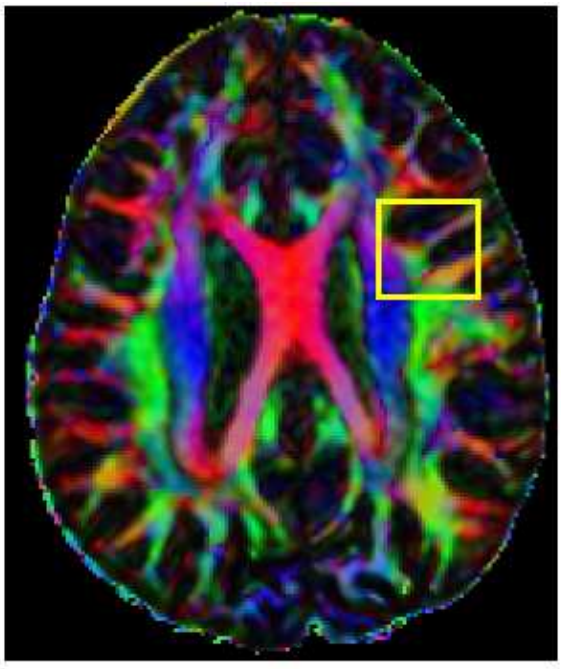

Figure 2.19: An axial colour coded orientation map with a region containing crossing fibre tracts.

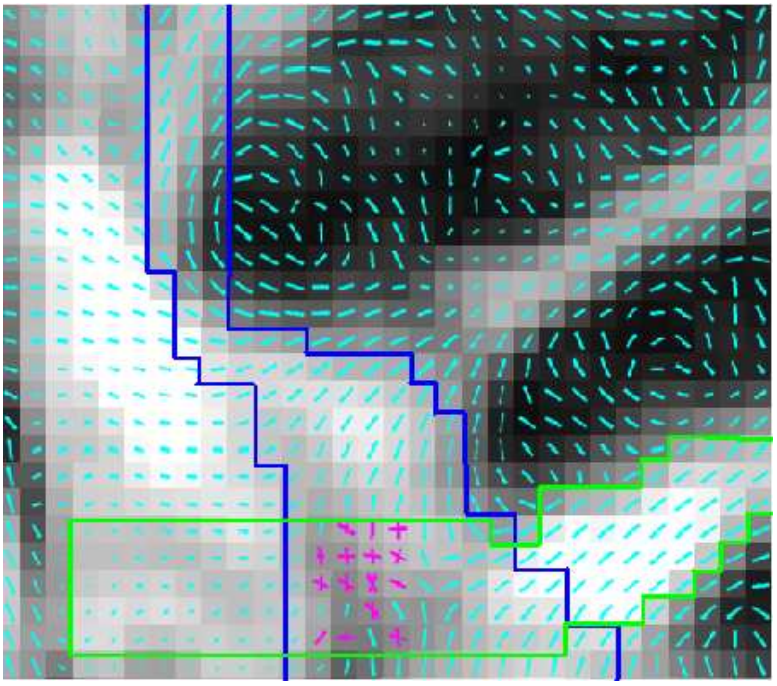

Figure 2.20: The zoomed ROI line fibre orientation map with FA background, containing a crossing of the anterior part of the superior longitudinal fasciculus (with a strong front-back orientation) with a tract of the corona radiata (from inferior regions). Diffusion tensor estimation is carried out by first fitting the the multi-tensor model ($m = 1$ or $m = 2$), and then applying model selection based on the Bayes factor.

2.4.3 Fibre tractographies

The principal eigenvector of a diffusion tensor at a given voxel represents the local fibre orientation. Once diffusion tensors in a ROI have been estimated, the fibre orientation information provided directly by DTI can be used to reconstruct the pathways of major white matter structures through the brain [Hagmann *et al.*, 2003; Parker and Alexander, 2003].

2.4.3.1 Comparison of tractographies

LLS, Bayesian (MAP) and MCMC (the average of MCMC samples from the posterior distribution) methods are used for diffusion tensors estimation with a single tensor model through the whole brain. Then, the freely available pack-

age Camino [Cook *et al*., 2006] and a Matlab program (provided by Mara Cercignani, Institute of Neurology, UCL) are employed for streamline calculations and visualisation. We set the start voxels along the corpus callosum (see Figure 2.21). Three tractographies are shown in Figure 2.22. Several differences of three tractographies are pointed out with circles and arrows. For example, in coronal views (a.1) and (b.1) estimated with the LLS and Bayesian methods respectively, the white circle points out the absence of a minor tract - the *stria terminalis* (st). However, the st tract can be seen in c.1 (MCMC tractography). Differences among the three tractographies can be clearly seen from sagittal views (a.2), (b.2) and (c.2). The white arrow in (a.2) (LLS) points to a short tract from the corpus callosum to the anterior corona radiata. However, two corresponding tracts in b.2 (Bayesian) and c.2 (MCMC) are longer and extend fully to the end of the anterior corona radiata. The pink arrow in (c.2) (MCMC) points to a tract from the corpus callosum to the posterior coronal radiata. The tract disappears in (a.2) and (b.2). In (c.2) (MCMC), a vertical tract (pointed out with a cyan arrow) grows from the splenium of the corpus callosum to the brain stem which is a well-known feature of the human brain. Hence, the MCMC tractography method gives better estimation of the fibre orientation in the corpus callosum.

2.4.3.2 Uncertainty of tractography

To understand the uncertainty of fibre tractography is important for surgical planning and postoperative assessment. To this effect, we have obtained some initial results for the uncertainty study of fibre tractography. For example, Figure 2.23 shows fibre tractographies of the corpus callosum seeded with two vertically neighbouring voxels (red) from four MCMC simulations from the posterior distribution. Two differences are pointed out with white and cyan

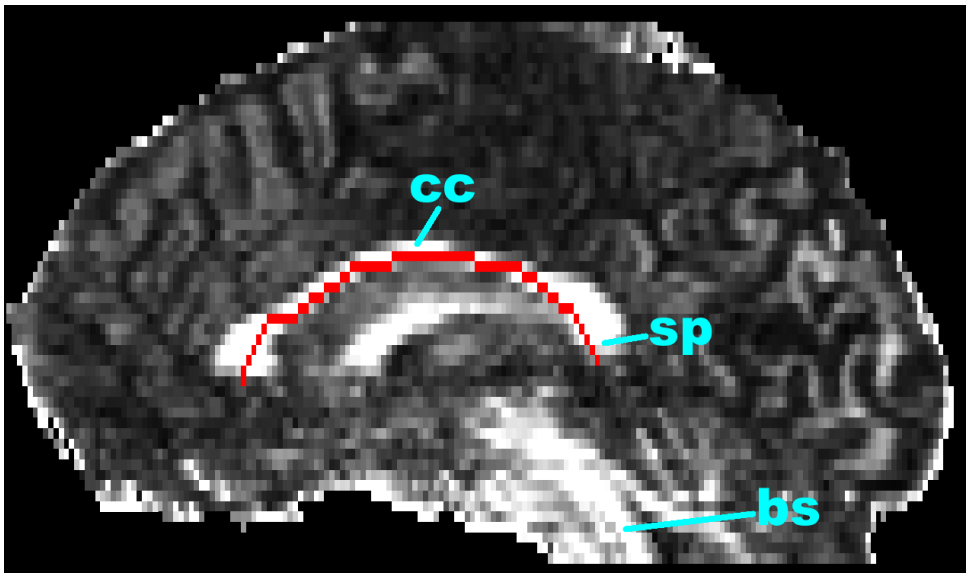

Figure 2.21: Seeding voxels (red) of tractography (sagittal view). cc: corpus callosum, sp: splenium of cc and bs: brain stem.

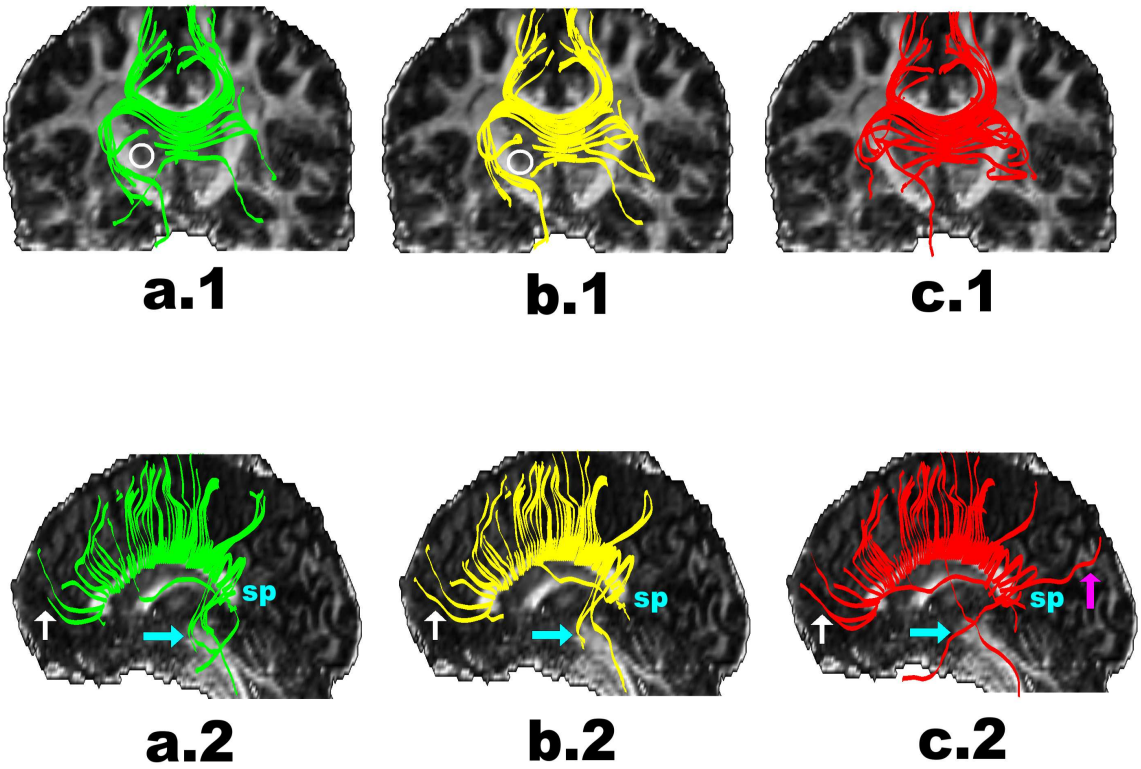

Figure 2.22: Tractographies of the corpus callosum (CC) based on LLS (green), Bayesian (yellow) and MCMC (red) estimators. a.1, b.1 and c.1 are coronal views (from the anterior to the frontal). a.2, b.2 and c.2 are sagittal views (from the left to the right). Differences between methods are pointed out with arrows and circles. Splenium: sp.

arrows. Figure 2.23 (d) gives more information about the brain structure than the other three tractographies. A more complete tool would be to carry out tractography for many more MCMC simulations and then count the proportion of times a streamline passes through a voxel. Alternative methods for averaging the tracts and summarising the variability would be of interest for further study in future work.

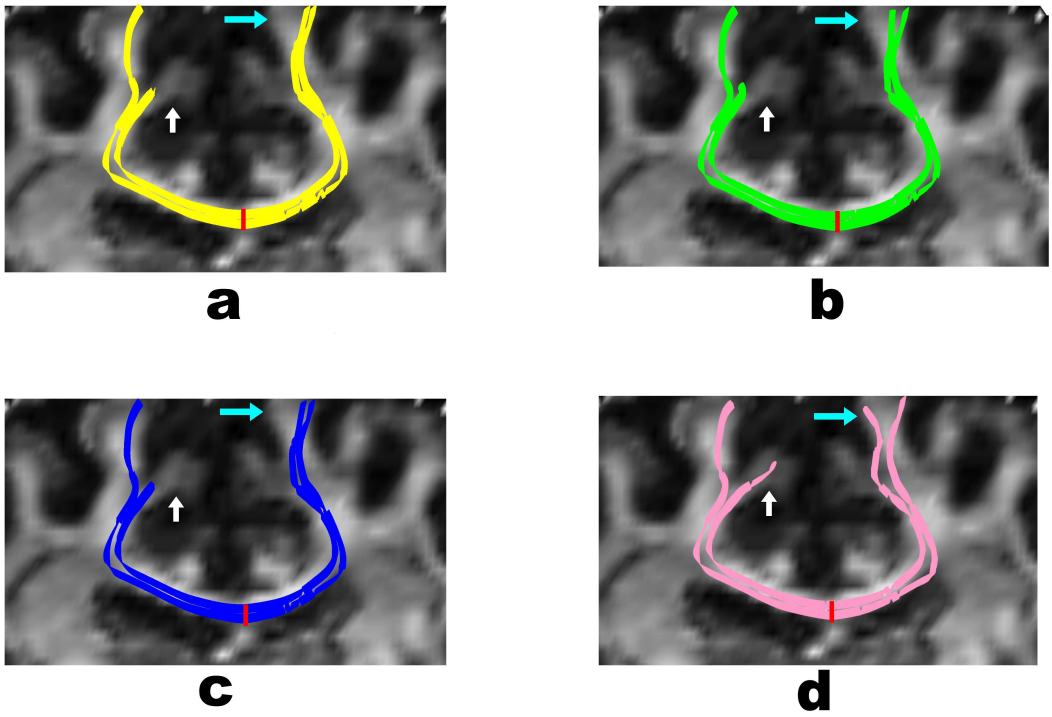

Figure 2.23: Tractographies of the corpus callosum from four MCMC simulations.

2.5 Summary

In this chapter, a multi-tensor model for diffusion MRI measurements of water diffusion at voxels with one or more distinct fibre orientations was developed. We propose a new parametrisation for the symmetry and positive semi-definiteness of the diffusion tensor. A Bayesian framework has been established for estimating diffusion tensors under the single tensor model and the double

tensor model. Inference has been drawn from the posterior distribution. We use the Bayes factor for model selection, i.e., for deciding between the single and double tensor models. An combination MCMC algorithm was used to study the uncertainty of the model parameters based on the posterior distribution of these parameters. We also define the sample mean of diffusion directions and the 95% credible cone around the mean direction. The comparisons of the Bayesian estimation and the LLS method for both the single tensor and double tensor models were made using three simulated datasets with three diffusion direction schemes. The anisotropy maps (FA maps), fibre orientation maps (the line and colour coded orientation maps) and the map of diffusion ellipsoids were obtained for the corpus callosum region. The multi-tensor model with automatic model selection has also been applied to a region containing crossing fibre bundles from a healthy human brain dataset. Fibre tractographies of the corpus callosum were obtained based on the Bayesian, LLS and the mean MCMC sample mean estimators. Uncertainty study of fibre tractography with an MCMC method were also discussed.

Once a sample of diffusion tensors generated using MCMC method, we wish to estimate the mean tensor and then carry out statistical inference. The key point is to consider that the diffusion tensor is positive semi-definite symmetric matrix. In the following chapter (see Chapter 3) we will discuss the non-Euclidean analysis of diffusion tensor data. The accuracy of DTI measurements depends on the diffusion gradient direction scheme applied. Consequently the diffusion gradient direction scheme applied influences the estimation of the diffusion tensor. In Chapter 4, we will design a series of new distributed direction schemes using directional statistics.

Chapter 3

Non-Euclidean Analysis of Diffusion Tensor Data

This chapter is partly based on the paper 'Dryden, I. L., Koloydenko, A., and Zhou, D. (2009). Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Annals of Applied Statistics*, 3(3):1102–1123.' and the paper 'Zhou, D., Dryden, I. L., Koloydenko, A., and Bai, L. (2009). Procrustes analysis of diffusion tensor data. *Proceedings of the International Society for Magnetic Resonance in Medicine*, 17:3584.'

3.1 Introduction

The diffusion tensor corresponding to a covariance matrix in the molecular displacement at a given voxel is required to be 3×3 symmetric and positive semi-definite. A variety of methods have been illustrated in Chapter 2 for estimating the diffusion tensor $\mathbf{D}$ from the DWI data at each voxel. Once a sample of diffusion tensors is available we wish to estimate the population mean and then carry out statistical inference. The difficulty is that the space of diffusion tensors is most naturally not Euclidean. Using the arithmetic mean of the sample to estimate the population mean may not be suitable for diffusion tensor data. In this chapter, the statistical analysis of diffusion tensor data is considered which takes into account the non-Euclidean nature of the space of positive semi-definite symmetric matrices.

3.1.1 Euclidean distance

Now we consider a sample of N diffusion tensors (symmetric and positive semi-definite) $\mathbf{D}_1$,..., $\mathbf{D}_N$. We assume that the $\mathbf{D}_i$ are independent and identically distributed (i.i.d.) from a distribution with mean $\mathbf{T}$ although care must be taken in defining what is a mean in a non-Euclidean space (see Section 3.1.2). The usual approach to estimate the mean covariance matrix in statistics is to assume a scaled Wishart distribution for the sample covariance matrices, and then the maximum likelihood estimator (m.l.e.) of the population covariance matrix is the arithmetic mean of the sample covariance matrices. This estimator can also be obtained if using a least squares approach by minimising the sum of square Euclidean distances. The Euclidean distance between two diffusion tensors is given by

$$d_E(\mathbf{D}_1, \mathbf{D}_2) = \| \mathbf{D}_1 - \mathbf{D}_2 \|, \tag{3.1.1}$$

where $\| \mathbf{A} \| = \sqrt{\text{trace}\left\{\mathbf{A}^{\mathrm{T}}\mathbf{A}\right\}}$ is the Euclidean norm (also known as the Frobenius norm). Then the least squares estimator is given by

$$\hat{\mathbf{T}}_E = \arg\inf_{\mathbf{T}} d_E(\mathbf{D}_i, \mathbf{T})^2 = \frac{1}{N}\sum_{i=1}^{N} \mathbf{D}_i. \tag{3.1.2}$$

However, this Euclidean method is often unsatisfactory for diffusion tensors. One defect with Euclidean calculus is that non-positive semi-definite symmetric matrices can appear during Euclidean computations, e.g. in extrapolation. Other drawbacks will be discussed in Section 3.4.

3.1.2 The Fréchet mean

As the space of diffusion tensors is non-Euclidean, there is growing need to use non-Euclidean metrics to estimate the mean diffusion tensor given the sample

tensors. Actually, it is first of all necessary to define what is meant by a mean diffusion tensor in a non-Euclidean space. Let $f(\mathbf{D})$ be a probability density function of a diffusion tensor $\mathbf{D}$ on a Riemannian metric space. The Fréchet mean is defined as [Fréchet, 1948; Koenker, 2006; Le and Kume, 2000]

$$\mathbf{T} = \arg\inf_{\mathbf{T}} \frac{1}{2} \int d(\mathbf{D}, \mathbf{T})^2 f(\mathbf{D}) d\mathbf{D}, \tag{3.1.3}$$

where $d(\cdot)$ is a non-Euclidean distance. A Fréchet mean is not necessarily unique. However, it is possible to prove the uniqueness with sufficient conditions [Le, 1995]. For example, for non-Euclidean spaces with negative sectional curvature, the Fréchet mean is always unique.

Let us consider a sample of N diffusion tensors $\mathbf{D}_1$,....., $\mathbf{D}_N$. Then, the sample Fréchet mean is given by finding

$$\hat{\mathbf{T}} = \arg\inf_{\mathbf{T}} \sum_{i=1}^{N} d(\mathbf{D}_i, \mathbf{T})^2. \tag{3.1.4}$$

3.2 Non-Euclidean Estimations of Mean Tensor

It is more natural to choose non-Euclidean metrics in order to define and estimate the mean tensor due to the positive semi-definiteness of $\mathbf{D}$.

3.2.1 Logarithm-based estimators

Recently, a log-Euclidean approach has been proposed for mean diffusion tensor estimation using matrix logarithms [Arsigny et al., 2006]. We can write the logarithm of a diffusion tensor $\mathbf{D}$ as follows. The usual spectral decomposition of $\mathbf{D}$ is $\mathbf{D} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^{\mathrm{T}}$ with $\mathbf{E} \in O(3)$ an orthogonal matrix and $\mathbf{\Lambda}$ diagonal with

strictly positive entries. Then, the logarithm of $\mathbf{D}$ is given by

$$\log(\mathbf{D}) = \mathbf{E}\log(\mathbf{\Lambda})\mathbf{E}^{\mathrm{T}} \tag{3.2.1}$$

where $\log(\mathbf{\Lambda})$ is a diagonal matrix with logarithm of the diagonal elements of $\mathbf{\Lambda}$ on the diagonal. Likewise, the exponential of $\mathbf{D}$ is given by

$$\exp(\mathbf{D}) = \mathbf{E}\exp(\mathbf{\Lambda})\mathbf{E}^{T} \tag{3.2.2}$$

where $\exp(\mathbf{\Lambda})$ is a diagonal matrix with exponential of the diagonal elements on the diagonal. The log-Euclidean distance is the Euclidean distance between the logarithm of diffusion tensors which is defined as [Arsigny et al., 2006]

$$d_L(\mathbf{D}_1, \mathbf{D}_2) = \parallel \log(\mathbf{D}_1) - \log(\mathbf{D}_2) \parallel . \tag{3.2.3}$$

The log-Euclidean estimator for the mean diffusion tensor is given by

$$\begin{aligned} \hat{\mathbf{T}}_L &= \exp\left\{ \arg\inf_{\mathbf{T}} \sum_{i=1}^{N} d_L(\mathbf{D}_i, \mathbf{T})^2 \right\} \\ &= \exp\left\{ \frac{1}{N} \sum_{i=1}^{N} \log(\mathbf{D}_i) \right\}. \end{aligned} \tag{3.2.4}$$

Another logarithm-based metric is a Riemannian metric described by many authors [e.g., Fletcher and Joshi, 2007; Pennec *et al.*, 2006; Batchelor *et al.*, 2004] who viewed the space of symmetric positive definite matrices as a Riemannian symmetric space. This Riemannian metric is defined for any pair of positive definite diffusion tensors as

$$d_R(\mathbf{D}_1, \mathbf{D}_2) = \parallel \log(\mathbf{D}_1^{-1/2}\mathbf{D}_2\mathbf{D}_1^{-1/2}) \parallel . \tag{3.2.5}$$

The Riemannian estimator of the mean diffusion tensor is then given by [Pennec *et al.*,

2006]

$$\begin{aligned}\hat{\mathbf{T}}_R &= \arg\inf_{\mathbf{T}} \sum_{i=1}^{N} d_R(\mathbf{D}_i, \mathbf{T})^2 \\ &= \arg\inf_{\mathbf{T}} \sum_{i=1}^{N} \| \log(\mathbf{D}_i^{-1/2}\mathbf{T}\mathbf{D}_i^{-1/2}) \|^2 . \end{aligned} \tag{3.2.6}$$

A gradient descent algorithm has been used to obtain the estimate [Pennec *et al.*, 2006]. Since this Riemannian metric space has negative sectional curvature, the population and sample Fréchet means are unique in this case [Pennec *et al.*, 2006].

3.2.2 Estimators with reparameterisation

An alternative way of analysing positive definite tensors is to use a reparameterisation of the diffusion tensor, such as the Cholesky decomposition [Wang *et al.*, 2004], where $\mathbf{D} = \mathbf{L}\mathbf{L}^{\mathrm{T}}$ and $\mathbf{L} = \mathrm{chol}(\mathbf{D})$ is lower triangular with positive diagonal entries. Then the Cholesky distance is given by

$$d_C(\mathbf{D}_1, \mathbf{D}_2) = \| \mathrm{chol}(\mathbf{D}_1) - \mathrm{chol}(\mathbf{D}_2) \| . \tag{3.2.7}$$

An estimator can be obtained with the least squares method, i.e.,

$$\hat{\mathbf{T}}_C = \hat{\mathbf{\Delta}}_C \hat{\mathbf{\Delta}}_C^{\mathrm{T}}, \tag{3.2.8}$$

where

$$\begin{aligned}\hat{\mathbf{\Delta}}_C &= \arg\inf_{\mathbf{\Delta}} \left\{ \frac{1}{N} \sum_{i=1}^{N} \| \mathrm{chol}(\mathbf{D}_i) - \mathbf{\Delta} \|^2 \right\} \\ &= \frac{1}{N} \sum_{i=1}^{N} \mathrm{chol}(\mathbf{D}_i). \end{aligned} \tag{3.2.9}$$

We propose an alternative decomposition using the matrix square root where $\mathbf{D}^{1/2} = \mathbf{E}\mathbf{\Lambda}^{1/2}\mathbf{E}^{\mathrm{T}}$ which has not been used in this context as far as we are aware. The distance is given by

$$d_H(\mathbf{D}_1, \mathbf{D}_2) = \| \mathbf{D}_1^{1/2} - \mathbf{D}_2^{1/2} \| . \tag{3.2.10}$$

A least squares estimator can be obtained from

$$\hat{\mathbf{T}}_H = \hat{\mathbf{\Delta}}_H \hat{\mathbf{\Delta}}_H^{\mathrm{T}}, \tag{3.2.11}$$

where

$$\begin{aligned} \hat{\mathbf{\Delta}}_H &= \arg\inf_{\mathbf{\Delta}} \left\{ \frac{1}{N} \sum_{i=1}^{N} \| \mathbf{D}_i^{1/2} - \mathbf{\Delta} \|^2 \right\} \\ &= \frac{1}{N} \sum_{i=1}^{N} \mathbf{D}_i^{1/2}. \end{aligned} \tag{3.2.12}$$

The power Euclidean metric is another possible metric, i.e.,

$$d_A(D_1, D_2) = \frac{1}{a} \| \mathbf{D}_1^a - \mathbf{D}_2^a \|, \tag{3.2.13}$$

where $\mathbf{D}^a = \mathbf{E}\mathbf{\Lambda}^a\mathbf{E}^{\mathrm{T}}$. The power a is nonzero and $a \in \mathbb{R}$. Then the estimate of the mean tensor is given by

$$\hat{\mathbf{T}}_A = (\hat{\mathbf{\Delta}}_A)^{1/a}, \tag{3.2.14}$$

where

$$\begin{aligned} \hat{\mathbf{\Delta}}_A &= \arg\inf_{\mathbf{\Delta}} \left\{ \frac{1}{N} \sum_{i=1}^{N} \| \mathbf{D}_i^a - \mathbf{\Delta} \|^2 \right\} \\ &= \frac{1}{N} \sum_{i=1}^{N} \mathbf{D}_i^a. \end{aligned} \tag{3.2.15}$$

Note that the Euclidean ($a = 1$) and root Euclidean ($a = 1/2$) are two special

cases of the power Euclidean metric.

3.2.3 Procrustes-based estimators

Yet another approach is to use Procrustes analysis [Gower, 1975] and the statistical shape theory developed in e.g. Dryden and Mardia [1998].

3.2.3.1 Full ordinary Procrustes analysis

Full ordinary Procrustes analysis (FOPA) is used to match two objects as closely as possible with similarity transformations (translation, rotation and scale). Let us first consider a pair of diffusion tensors $\mathbf{D}_1$ and $\mathbf{D}_2$. To ensure the positive semi-definiteness of $\mathbf{D}_i$, $i = 1, 2$, we use a reparameterisation $\mathbf{D}_i = \mathbf{Q}_i\mathbf{Q}_i^{\mathrm{T}}$, where $\mathbf{Q}_i$ is a 3×3 real matrix. For example, $\mathbf{Q}_i = chol(\mathbf{D}_i)$ is the Cholesky decomposition, or $\mathbf{Q}_i = \mathbf{D}_i^{1/2}$ is the matrix square root. In our computation we shall choose the Cholesky decomposition. Note that $\mathbf{Q}_i$ and any rotation of it $\mathbf{Q}_i\mathbf{R}$ ($\mathbf{R} \in O(3)$) result in the same $\mathbf{D}_i$, i.e. $\mathbf{D}_i = \mathbf{Q}_i\mathbf{Q}_i^{\mathrm{T}} = \mathbf{Q}_i\mathbf{R}(\mathbf{Q}_i\mathbf{R})^{\mathrm{T}}$.

The objective of FOPA is to minimise $S_{FOPA}(\mathbf{D}_1, \mathbf{D}_2)^2$ in (3.2.16) the squared Euclidean distance between $\mathbf{D}_1$ and $\mathbf{D}_2$ under the similarity transformations. The squared Euclidean distance is given by

$$S_{FOPA}(\mathbf{D}_1, \mathbf{D}_2)^2 = \| \mathbf{Q}_1 - \beta\mathbf{Q}_2\mathbf{R} - \mathbf{1}_3\boldsymbol{\gamma}^{\mathrm{T}} \|^2, \tag{3.2.16}$$

where a 3×3 rotation matrix $\mathbf{R} \in O(3)$, a scale parameter $\beta > 0$, and a 3×1 location vector $\boldsymbol{\gamma}$ represent three similarity transformations. Note $\mathbf{1}_3$ is the 3×1 vector of ones.

The solution $(\hat{\boldsymbol{\gamma}}, \hat{\beta}, \hat{\mathbf{R}})$ to the minimisation of Equation (3.2.16) is given by

[Dryden and Mardia, 1998, p.84-85]

$$\hat{\gamma} = \mathbf{0}_3 \tag{3.2.17}$$

where $\mathbf{0}_3$ is the 3×1 vector of zeros,

$$\hat{\mathbf{R}} = \mathbf{U}\mathbf{V}^{\mathrm{T}}, \tag{3.2.18}$$

where $\mathbf{U}$,$\mathbf{V} \in O(3)$ are obtained from a singular value decomposition:

$$\mathbf{Q}_1^{\mathrm{T}}\mathbf{Q}_2 = \mathbf{V}\mathbf{\Delta}\mathbf{U}^{\mathrm{T}}, \tag{3.2.19}$$

with $\mathbf{\Delta}$ a diagonal 3×3 matrix of singular values. Furthermore,

$$\hat{\beta} = \frac{\mathrm{trace}(\mathbf{Q}_1^{\mathrm{T}}\mathbf{Q}_2\hat{\mathbf{R}})}{\mathrm{trace}(\mathbf{Q}_2^{\mathrm{T}}\mathbf{Q}_2)}. \tag{3.2.20}$$

The full Procrustes shape metric between $\mathbf{D}_1$ and $\mathbf{D}_2$ is given by

$$\begin{aligned} d_F(\mathbf{D}_1, \mathbf{D}_2) &= \inf_{\mathbf{R}\in O(3), \beta \in \mathbb{R}} \| \mathbf{Q}_1 - \beta\mathbf{Q}_2\mathbf{R} \| \\ &= \| \mathbf{Q}_1 - \hat{\beta}\mathbf{Q}_2\hat{\mathbf{R}} \| \end{aligned} \tag{3.2.21}$$

3.2.3.2 Procrustes size-and-shape distance

In DTI study, we wish to match $\mathbf{Q}_1$ (from $\mathbf{D}_1$) and $\mathbf{Q}_2$ (from $\mathbf{D}_2$) under location, rotation and reflection while often preserving scale information. Then the joint study of size-and-shape is of interest. Size-and-shape spaces were introduced by Kendall [1989]. The definition of the size-and-shape of a configuration matrix was given by Dryden and Mardia [1998, p.57].

The Procrustes size-and-shape distance between two diffusion tensors is de-

fined as

$$d_S(\mathbf{D}_1, \mathbf{D}_2) = \inf_{\mathbf{R}\in O(3)} \parallel \mathbf{Q}_1 - \mathbf{Q}_2\mathbf{R} \parallel . \tag{3.2.22}$$

The Procrustes solution $\hat{\mathbf{R}}$ for matching $\mathbf{Q}_1$ to $\mathbf{Q}_2$ is

$$\begin{aligned} \hat{\mathbf{R}} &= \arg \inf_{\mathbf{R}\in O(3)} \parallel \mathbf{Q}_1 - \mathbf{Q}_2\mathbf{R} \parallel \\ &= \mathbf{U}\mathbf{V}^{\mathrm{T}} \end{aligned} \tag{3.2.23}$$

where $\mathbf{U}$ and $\mathbf{V}$ are from the singular value decomposition in Equation (3.2.19).

3.2.3.3 Procrustes estimators

Consider the general case where there are $N \geqslant 2$ diffusion tensors $\mathbf{D}_1$,...,$\mathbf{D}_N$, and $\mathbf{D}_i = \mathbf{Q}_i\mathbf{Q}_i^{\mathrm{T}}, i = 1, ...N$. Now the aim is to calculate the Fréchet mean using the full Procrustes shape metric in Equation (3.2.21) and the Procrustes size-and-shape metric in Equation (3.2.22).

The sample Fréchet mean relative to the full Procrustes shape metric $d_F(\cdot)$ is given by

$$\hat{\mathbf{T}}_F = \hat{\mathbf{Q}}_F\hat{\mathbf{Q}}_F^{\mathrm{T}}, \tag{3.2.24}$$

where

$$\hat{\mathbf{Q}}_F = \arg\inf_{\mathbf{Q}} \sum_{i=1}^{N} \inf_{\mathbf{R}_i\in O(k)} \parallel \beta_i\mathbf{Q}_i\mathbf{R}_i - \mathbf{Q} \parallel^2 . \tag{3.2.25}$$

The sample Fréchet mean relative to the Procrustes size-and-shape distance $d_S(\cdot)$ is given by

$$\hat{\mathbf{T}}_S = \arg\inf_{\mathbf{T}} \sum_{i=1}^{N} d(\mathbf{D}_i, \mathbf{T})^2. \tag{3.2.26}$$

where $d(\cdot)$ can be $d_F(\cdot)$ or $d_S(\cdot)$. Specifically,

$$\hat{\mathbf{T}}_S = \hat{\mathbf{Q}}_S\hat{\mathbf{Q}}_S^{\mathrm{T}}, \tag{3.2.27}$$

where

$$\hat{\mathbf{Q}}_S = \arg\inf_{\mathbf{\Delta}} \sum_{i=1}^{N} \inf_{\mathbf{R}_i \in O(k)} \parallel \mathbf{Q}_i\mathbf{R}_i - \mathbf{\Delta} \parallel^2 . \tag{3.2.28}$$

3.2.4 Partial generalised Procrustes analysis with rotations

Consider the general case that there are $N \geqslant 2$ diffusion tensors $\mathbf{D}_1$,...,$\mathbf{D}_N$, and $\mathbf{D}_i = \mathbf{Q}_i\mathbf{Q}_i^{\mathrm{T}}, i = 1, ...N$. Consider the partial generalised Procrustes analysis (PGPA) [Dryden and Mardia, 1998, p.90-91] which minimises the total sum $S_{PGPA}(\mathbf{D}_1, ..., \mathbf{D}_N)$ of the squared Euclidean distances between all pairs $\mathbf{Q}_i\mathbf{R}_i$ and $\mathbf{Q}_j\mathbf{R}_j$ over the orthogonal transformations $\mathbf{R}_1$, $\mathbf{R}_2$,..., $\mathbf{R}_N \in O(3)$. The minimisation of the sum of squares is given by

$$\begin{aligned} S_{PGPA}(\mathbf{D}_1, ..., \mathbf{D}_N) &= \inf_{\mathbf{R}_1,...,\mathbf{R}_N} \frac{1}{N} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \parallel \mathbf{Q}_i\mathbf{R}_i - \mathbf{Q}_j\mathbf{R}_j \parallel^2 \\ &= \inf_{\mathbf{R}_1,...,\mathbf{R}_N} \sum_{i=1}^{N} \parallel \mathbf{Q}_i\mathbf{R}_i - \frac{1}{N} \sum_{j=1}^{N} \mathbf{Q}_j\mathbf{R}_j \parallel^2 . \end{aligned} \tag{3.2.29}$$

Let $\hat{\mathbf{R}}_i, i = 1, ..., N$ be the estimates of rotations which minimising Equation (3.2.29). An algorithm 1 with iterative procedures for estimating $\hat{\mathbf{R}}_i, i = 1, ..., N$ has been adapted from the Generalised Procrustes Algorithm described by Gower [1975, p.35-50] and Dryden and Mardia [1998, p.90-91]. The algorithm converges quickly in the DTI study.

Algorithm 1: Partial Generalised Procrustes Method

Initial setting: $\mathbf{Q}_i^P \leftarrow chol(\mathbf{D}_i)$, $i = 1, ..., N$
S_{PGPA} from previous iteration: $S_p \leftarrow 0$
S_{PGPA} from current iteration: $S_c \leftarrow \sum_{i=1}^{N} \parallel \mathbf{Q}_i^P - \frac{1}{N} \sum_{j=1}^{N} \mathbf{Q}_j^P \parallel^2$
while $|S_p - S_c| >$ tolerance **do**
  for $i = 1$ to N **do**
    $\hat{\mathbf{Q}}_i = \frac{1}{N-1} \sum_{j \neq i} \mathbf{Q}_j^P$
    Calculate the $\hat{\mathbf{R}}_i$ minimising $\parallel \hat{\mathbf{Q}}_i - \mathbf{Q}_i^P \mathbf{R}_i \parallel$ (partial ordinary Procrustes analysis)
    $\mathbf{Q}_i^P \leftarrow \mathbf{Q}_i^P \hat{\mathbf{R}}_i$
  end for
  $S_p \leftarrow S_c$
  $S_c \leftarrow \sum_{i=1}^{N} \parallel \mathbf{Q}_i^P - \frac{1}{N} \sum_{j=1}^{N} \mathbf{Q}_j^P \parallel^2$
end while
$\hat{\mathbf{Q}}_{PGPA} \leftarrow \frac{1}{N} \sum_{i=1}^{N} \mathbf{Q}_i^P$
return $\hat{\mathbf{Q}}_{PGPA}$

Hence, the Fréchet mean of $\mathbf{D}_1$,...,$\mathbf{D}_N$ with generalised Procrustes analysis is given by

$$\hat{\mathbf{T}}_{PGPA} = \hat{\mathbf{Q}}_{PGPA} \hat{\mathbf{Q}}_{PGPA}^{\mathrm{T}}, \tag{3.2.30}$$

where

$$\hat{\mathbf{Q}}_{PGPA} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{Q} \hat{\mathbf{R}}_i. \tag{3.2.31}$$

3.2.5 Comparison of approaches

We have discussed several choices of distances between diffusion tensors which can be applied to DTI studies. For completeness we list the metrics and the mean tensor estimators considered in this study in Table 3.1, and discuss briefly some of their properties.

Mean tensor estimators $\hat{\mathbf{T}}_E$, $\hat{\mathbf{T}}_C$, $\hat{\mathbf{T}}_H$, $\hat{\mathbf{T}}_L$, $\hat{\mathbf{T}}_A$ can be calculated straightfor-

Table 3.1: Notation and definitions of the metrics and mean tensor estimators

Name	Notation	Form	Estimator
Euclidean	$d_E(\mathbf{D}_1, \mathbf{D}_2)$	$\parallel \mathbf{D}_1 - \mathbf{D}_2 \parallel$	$\hat{\mathbf{T}}_E$
Log-Euclidean	$d_L(\mathbf{D}_1, \mathbf{D}_2)$	$\parallel \log(\mathbf{D}_1) - \log(\mathbf{D}_2) \parallel$	$\hat{\mathbf{T}}_L$
Riemannian	$d_R(\mathbf{D}_1, \mathbf{D}_2)$	$\parallel \log(\mathbf{D}_1^{-1/2}\mathbf{D}_2\mathbf{D}_1^{-1/2}) \parallel$	$\hat{\mathbf{T}}_R$
Cholesky	$d_C(\mathbf{D}_1, \mathbf{D}_2)$	$\parallel \text{chol}(\mathbf{D}_1) - \text{chol}(\mathbf{D}_2) \parallel$	$\hat{\mathbf{T}}_C$
Root Euclidean	$d_H(\mathbf{D}_1, \mathbf{D}_2)$	$\parallel \mathbf{D}_1^{1/2} - \mathbf{D}_2^{1/2} \parallel$	$\hat{\mathbf{T}}_H$
Procrustes size-and-shape	$d_S(\mathbf{D}_1, \mathbf{D}_2)$	$\inf_{\mathbf{R}\in O(3)} \parallel \mathbf{Q}_1 - \mathbf{Q}_2\mathbf{R} \parallel$	$\hat{\mathbf{T}}_S$
Full Procrustes shape	$d_F(\mathbf{D}_1, \mathbf{D}_2)$	$\inf_{\mathbf{R}\in O(3),\beta\in\mathbb{R}} \parallel \mathbf{Q}_1 - \beta\mathbf{Q}_2\mathbf{R} \parallel$	$\hat{\mathbf{T}}_F$
Power Euclidean	$d_A(\mathbf{D}_1, \mathbf{D}_2)$	$\frac{1}{a} \parallel \mathbf{D}_1^a - \mathbf{D}_2^a \parallel$	$\hat{\mathbf{T}}_A$

wardly using arithmetic averages. The Procrustes based estimators $\hat{\mathbf{T}}_S$, $\hat{\mathbf{T}}_F$ are to be computed using the Generalised Procrustes Algorithm which works very well in practice. For computing the Riemannian metric estimator $\hat{\mathbf{T}}_R$, a gradient descent algorithm [Pennec *et al.*, 2006] is guaranteed to converge.

All these metrics, except for the Cholesky-based d_C, are unchanged when replacing both $\mathbf{D}_i$ by $\mathbf{W}\mathbf{D}_i\mathbf{W}^{\mathrm{T}}$, $\mathbf{W} \in O(3)$, $i = 1, 2$. This means that these metrics are invariant under the orthogonal change of coordinates of $\mathbb{R}^3$, which is very important for real life applications. Metrics $d_L(\centerdot)$, $d_R(\centerdot)$, $d_F(\centerdot)$ are invariant under simultaneous scaling of $\mathbf{D}_i$, $i = 1, 2$, i.e., replacing both $\mathbf{D}_i$ by $\beta\mathbf{D}_i$. Metric $d_R(\centerdot)$ is also affine invariant, i.e., the metrics are unchanged by replacing both $\mathbf{D}_i$ by $\mathbf{B}\mathbf{D}_i\mathbf{B}^{\mathrm{T}}$, $i = 1, 2$, where $\mathbf{B}$ is a general 3×3 full rank matrix. Metrics $d_L(\centerdot)$, $d_R(\centerdot)$ have the property that

$$d(\mathbf{B}, \mathbf{I}_{3\times 3}) = d(\mathbf{B}^{-1}, \mathbf{I}_{3\times 3}). \tag{3.2.32}$$

Metrics $d_L(\centerdot)$, $d_R(\centerdot)$, $d_F(\centerdot)$ are not valid for comparing rank deficient covariance matrices. Finally, there are problems with extrapolation with metric $d_E(\centerdot)$ extrapolate too far and the result is no longer positive semi-definite.

3.2.6 Anisotropy indices

Anisotropy measures are useful tools to quantitatively capture the diffusion anisotropy in tissues [Basser *et al.*, 2000] (see Chapter 1 Section 1.3.1).

3.2.6.1 Metric-based anisotropy indices

The anisotropy indices AI_E and AI_R [Moakher and Batchelor, 2006, p.291-295] of $\mathbf{D}$ relative to the Euclidean and Riemannian metrics are defined respectively as follows

$$AI_E(\mathbf{D}) = d_E(\mathbf{D}, \frac{\text{trace}(\mathbf{D})}{3}\mathbf{I}_{3\times 3}) \quad (3.2.33)$$

$$AI_R(\mathbf{D}) = d_R(\mathbf{D}, \sqrt[3]{\det(\mathbf{D})}\mathbf{I}_{3\times 3}) \quad (3.2.34)$$

where $\mathbf{I}_{3\times 3}$ is the 3×3 identity matrix, and $\det(\mathbf{D})$ is the determinant of $\mathbf{D}$. The range of both AI_E and AI_R is $[0, \infty)$.

The Fractional Anisotropy (FA) which is a commonly used anisotropy measure in the DTI community, can be derived by normalising AI_E as follow

$$\begin{aligned} FA(\mathbf{D}) &= \sqrt{\frac{3}{2}} AI_E(\mathbf{D}) / \parallel \mathbf{D} \parallel \\ &= \sqrt{\frac{3[(\lambda_1 - \bar{\lambda})^2 + (\lambda_2 - \bar{\lambda})^2 + (\lambda_3 - \bar{\lambda})^2]}{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}} \end{aligned} \quad (3.2.35)$$

where λ_1, λ_2 and λ_3 are three eigenvalues of the diffusion tensor $\mathbf{D}$, and $\bar{\lambda} = \sum_{i=1}^{3} \lambda_i / 3$. FA ranges from 0 (full isotropy) to 1 (complete anisotropy).

The *geodesic anisotropy* (GA) [Batchelor *et al.*, 2004] is defined to be AI_R which in terms of the eigenvalues is given by

$$GA(\mathbf{D}) = AI_R(\mathbf{D})$$

$$= \sqrt{\sum_{i=1}^{3}(\log\lambda_i - \overline{\log\lambda})^2} \tag{3.2.36}$$

where $\overline{\log\lambda} = (\sum_{i=1}^{3}\log\lambda_i)/3$, $0 \leqslant \mathrm{GA} < \infty$. Since it is more convenient to work with the unit range, renormalisation of GA can be defined as given in Equation (3.2.37) below:

$$\tanh(GA) = \frac{\exp\{2GA\} - 1}{\exp\{2GA\} + 1} \tag{3.2.37}$$

and tanh(GA) ranges from 0 to 1.

3.2.6.2 Procrustes anisotropy

We define a new anisotropy measure Procrustes anisotropy (PA) with the full Procrustes shape metric. The definition of PA is given by

$$\begin{aligned} PA(\mathbf{D}) &= \sqrt{\frac{3}{2}}d_F(\frac{\mathbf{I}_{3\times3}}{\sqrt{3}}, \mathbf{D}) \\ &= \sqrt{\frac{3}{2}\sum_{i=1}^{3}(\sqrt{\lambda_i} - \overline{\sqrt{\lambda}})^2 / \sum_{i=1}^{3}\lambda_i} \end{aligned} \tag{3.2.38}$$

where $\overline{\sqrt{\lambda}} = \sum_{i=1}^{3}\sqrt{\lambda_i}/3$. It is clear that PA is a normalisation of the FOPA distance from any given diffusion tensor $\mathbf{D}$ to the identity tensor, representing the case of ideal isotropy. The range of PA is $[0, 1]$ with PA$= 0$ indicating full isotropy and PA ≈ 1 representing the extremely strong anisotropy. PA is invariant to the uniform scaling of a diffusion tensor. Now let us prove Equation (3.2.38).

Proof. The full ordinary Procrustes metric is defined as

$$d_F(\mathbf{D}_1, \mathbf{D}_2) = \inf_{\mathbf{R}\in O(3), \beta\in\mathbb{R}} \parallel \mathbf{Q}_1 - \beta\mathbf{Q}_2\mathbf{R} \parallel .$$

The solution $(\hat{\beta}, \hat{\mathbf{R}})$ has been shown in Section 3.2.3.1.

Now let $\mathbf{D}_1 = \frac{\mathbf{I}_{3\times3}}{\sqrt{3}}$ and $\mathbf{D}_2 = \mathbf{D}$, where $\mathbf{D} = \mathbf{Q}\mathbf{Q}^{\mathrm{T}}$. The full ordinary Procrustes metric between $\frac{\mathbf{I}_{3\times3}}{\sqrt{3}}$ and $\mathbf{D}$ is given by

$$d_F(\frac{\mathbf{I}_{3\times3}}{\sqrt{3}}, \mathbf{D}) = \parallel \frac{\mathbf{I}_{3\times3}}{\sqrt{3}} - \hat{\beta}\mathbf{Q}\hat{\mathbf{R}} \parallel$$

where

$$\hat{\mathbf{R}} = \mathbf{U}\mathbf{V}^{\mathrm{T}},$$

and $\mathbf{U}$,$\mathbf{V} \in SO(3)$ are obtained from a singular value decomposition:

$$\frac{\mathbf{I}_{3\times3}^{\mathrm{T}}}{\sqrt{3}}\mathbf{Q}_2 = \mathbf{V}\mathbf{\Delta}\mathbf{U}^{\mathrm{T}},$$

with $\mathbf{\Delta}$ a diagonal 3×3 matrix of singular values. Furthermore,

$$\begin{aligned} \hat{\beta} &= \frac{\mathrm{trace}(\frac{\mathbf{I}_{3\times3}^{\mathrm{T}}}{\sqrt{3}}\mathbf{Q}\hat{\mathbf{R}})}{\mathrm{trace}(\mathbf{Q}^{\mathrm{T}}\mathbf{Q})} \\ &= \frac{\frac{1}{\sqrt{3}}(\sum_{i=1}^{3}\sqrt{\lambda_i})}{\sum_{i=1}^{3}\lambda_i}. \end{aligned}$$

Then,

$$\begin{aligned} PA(\mathbf{D}) &= \sqrt{\frac{3}{2}}d_F(\frac{\mathbf{I}_{3\times3}}{\sqrt{3}}, \mathbf{D}) \\ &= \sqrt{\frac{3}{2}}\inf_{\mathbf{R}\in O(3),\beta\in\mathbb{R}} \parallel \frac{\mathbf{I}_{3\times3}}{\sqrt{3}} - \beta\mathbf{Q}\mathbf{R} \parallel \\ &= \sqrt{\frac{3}{2}} \parallel \frac{\mathbf{I}_{3\times3}}{\sqrt{3}} - \hat{\beta}\mathbf{Q}\hat{\mathbf{R}} \parallel \\ &= \sqrt{\frac{3}{2}}\left\{\mathrm{trace}\left[(\frac{\mathbf{I}_{3\times3}}{\sqrt{3}} - \hat{\beta}\mathbf{Q}\hat{\mathbf{R}})^{\mathrm{T}}(\frac{\mathbf{I}_{3\times3}}{\sqrt{3}} - \hat{\beta}\mathbf{Q}\hat{\mathbf{R}})\right]\right\}^{1/2} \\ &= \sqrt{\frac{3}{2}}\left[\mathrm{trace}(\frac{\mathbf{I}_{3\times3}}{3}) - \frac{2}{\sqrt{3}}\mathrm{trace}(\hat{\beta}\mathbf{Q}\hat{\mathbf{R}}) + \hat{\beta}^2\mathrm{trace}(\hat{\mathbf{R}}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\mathbf{Q}\hat{\mathbf{R}})\right]^{1/2} \end{aligned}$$

$$
\begin{aligned}
&= \sqrt{\frac{3}{2}}\left[1-\frac{2}{3}\frac{(\sum_{i=1}^{3}\sqrt{\lambda_i})^2}{\sum_{i=1}^{3}\lambda_i}+\frac{1}{3}\frac{(\sum_{i=1}^{3}\sqrt{\lambda_i})^2}{\sum_{i=1}^{3}\lambda_i}\right]^{1/2}\\
&= \sqrt{\frac{3}{2}}\left[\frac{\sum_{i=1}^{3}(\sqrt{\lambda_i})^2-3(\overline{\sqrt{\lambda}})^2}{\lambda_1+\lambda_2+\lambda_3}\right]^{1/2}\\
&= \sqrt{\frac{3}{2}\sum_{i=1}^{3}(\sqrt{\lambda_i}-\overline{\sqrt{\lambda}})^2/\sum_{i=1}^{3}\lambda_i}.
\end{aligned}
$$

□

3.2.6.3 Comparison of anisotropy indices

To be useful, measures of anisotropy should be invariant to uniform scaling of the tensor. Indeed, all the measures considered above (i.e. FA, GA) as well as PA have this invariance.

Figure 3.1 shows a comparison of FA, tanh(GA) and PA values as functions of $t \in [0,1]$ which defines three eigenvalues as $t, (1-t)/2$ and $(1-t)/2$. The diffusion tensor varies from planar to spherical as t increases from 0 to $1/3$. It then changes to a linear tensor as t grows to 1. PA value is always smaller than corresponding FA and tanh(GA) values. Figure 3.2 shows the first derivative of FA, PA and tanh(GA) with respect to t. When approximately $t < 0.075$ or $t > 0.708$, $|\mathrm{PA}'(t)|$ is larger than both $|\mathrm{FA}'(t)|$ and $|\tanh(\mathrm{GA})'(t)|$. This means PA is more sensitive to highly planar or highly linear anisotropy of diffusion.

We have discussed the power Euclidean metric in Equation (3.2.13). The anisotropy index based on the power Euclidean is a generalisation of FA($\mathbf{D}$) given by

$$
FA(\mathbf{D}^a) = \left\{\frac{3}{2}\sum_{i=1}^{3}(\lambda_i^a-\overline{\lambda^a})^2/\sum_{i=1}^{3}\lambda_i^{2a}\right\}^{1/2}, \tag{3.2.39}
$$

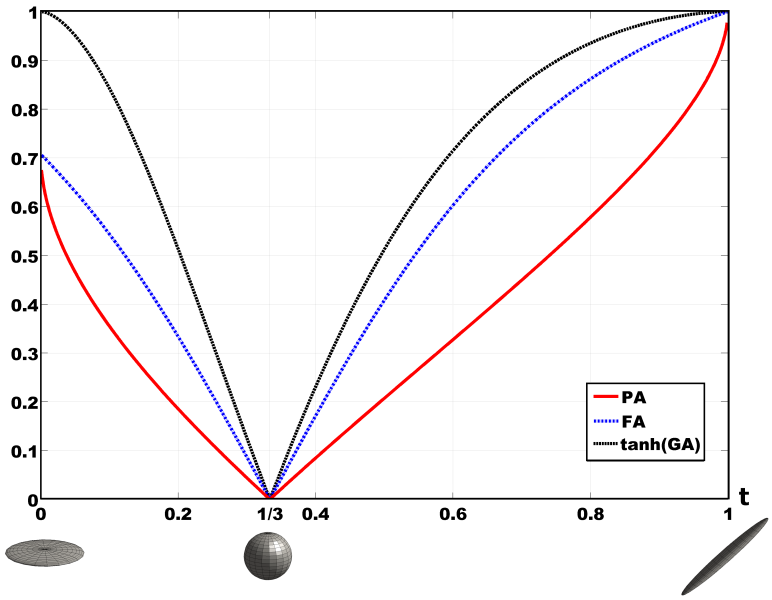

Figure 3.1: Comparison of FA, tanh(GA) and PA values as functions of t. The x-axis is $t \in [0, 1]$ which defines three eigenvalues as t, $(1-t)/2$ and $(1-t)/2$.

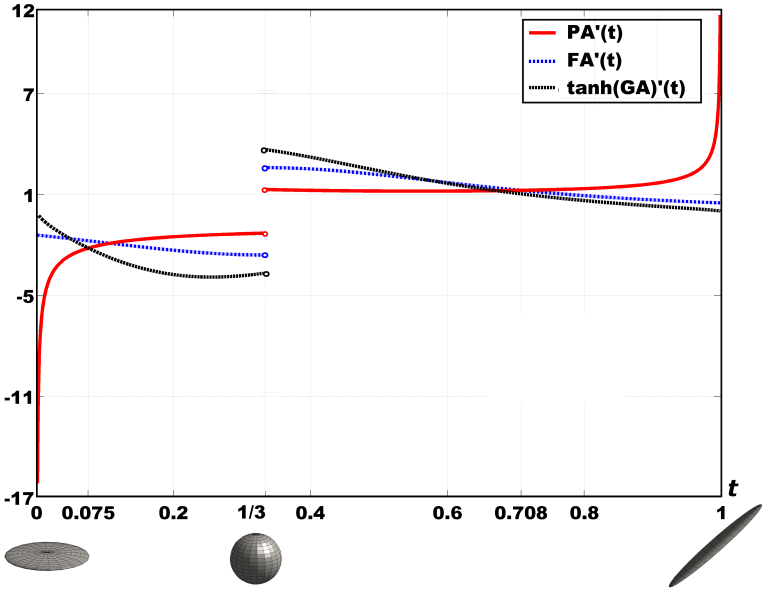

Figure 3.2: Derivatives of FA, PA and tanh(GA) with respect to t. When, approximately, $t < 0.075$ or $t > 0.708$, $|\mathrm{PA}'(t)|$ is larger than both $|\mathrm{FA}'(t)|$ and $|\tanh(\mathrm{GA})'(t)|$.

where $\overline{\lambda^a} = \sum_{i=1}^{3} \lambda_i^a / 3$. It is noted that FA and PA of $\mathbf{D}$ are two members of $FA(\mathbf{D}^a)$ when $a = 1$ and $a = 1/2$ respectively.

Theorem 3.1. For any symmetric semi-positive definite $\mathbf{D}$, $\mathrm{FA}(\mathbf{D}^a)$ is an increasing function of a.

Proof. Let $a, b > 0$ be such that $a < b$. We need to show that

$$\frac{\sum_{i=1}^{3}(\lambda_i^a - \overline{\lambda^a})^2}{\sum_{i=1}^{3}\lambda_i^{2a}} \leq \frac{\sum_{i=1}^{3}(\lambda_i^b - \overline{\lambda^b})^2}{\sum_{i=1}^{3}\lambda_i^{2b}}$$
$$\iff [\overline{\lambda^b}]^2\overline{\lambda^{2a}} \leq [\overline{\lambda^a}]^2\overline{\lambda^{2b}} \tag{3.2.40}$$

Let $q = \frac{2b-a}{b}$, $p = \frac{2b-a}{b-a}$, so that $\frac{1}{q} + \frac{1}{p} = 1$, $p \geq 1$, $q \geq 1$, then write $\lambda_i^b = \lambda_i^{s+t}$, $i \in \{1, 2, 3\}$ where $s = \frac{a}{q} = \frac{ab}{2b-a}$, $t = \frac{2b}{p} = \frac{2b(b-a)}{2b-a}$.

Apply Hölder's inequality [Mirsky, 1955] with $g_i = \lambda_i^s$, $f_i = \lambda_i^t$,$i \in 1, 2, 3$,

$$\sum_{i=1}^{3}|g_i f_i| \leq [\sum_{i=1}^{3} g_i^q]^{1/q}[\sum_{i=1}^{3} f_i^p]^{1/p} \tag{3.2.41}$$

to obtain

$$\overline{\lambda^b} \leq [\overline{\lambda^a}]^{\frac{b}{2b-a}}[\overline{\lambda^{2b}}]^{\frac{b-a}{2b-a}}$$
$$\implies [\overline{\lambda^b}]^2 \leq [\overline{\lambda^a}]^{\frac{2b}{2b-a}}[\overline{\lambda^{2b}}]^{\frac{2b-2a}{2b-a}}. \tag{3.2.42}$$

Next, take $q' = \frac{2b-a}{2b-2a}$, $p' = \frac{2b-2a}{a}$, note that $p', q' \geq 1$, $\frac{1}{q'} + \frac{1}{p'} = 1$. then write $\lambda_i^{2a} = \lambda_i^{s'+t'}$, and take $s' = \frac{a}{q'} = \frac{2a(b-a)}{2b-a}$, $t' = \frac{2b}{p'} = \frac{2ab}{2b-a}$. Again, let $g_i' = \lambda_i^{s'}$, $f_i' = \lambda_i^{t'}$, $i \in \{1, 2, 3\}$ and apply Hölder's inequality to get

$$\overline{\lambda^{2a}} = \overline{\lambda^{s'+t'}} \leq [\overline{\lambda^a}]^{\frac{a(b-a)}{2b-a}}[\overline{\lambda^{2b}}]^{\frac{a}{2b-a}} \tag{3.2.43}$$

Finally, multiply both sides of inequality (3.2.42) by respective sides of inequality (3.2.43) to obtain inequality (3.2.40). □

Corollary 3.1. $\forall \mathbf{D}$ ($\mathbf{D}$ is symmetric semi-positive definite), $PA(\mathbf{D}) \leq FA(\mathbf{D})$.

Corollary 3.1 implies that PA maps will be darker that FA maps in the same ROI, but in some regions of high diffusion anisotropy PA provides better con-

tract than FA (see Section 3.4.2 for a comparison of PA and FA maps in a real data study).

3.2.7 Geodesic interpolation

3.2.7.1 Interpolation of two tensors

The minimal geodesic can be used for interpolating between two diffusion tensors. Geodesic paths with different metrics starting at $\mathbf{D}_1$ and ending at $\mathbf{D}_2$ are listed in Table 3.2.

Table 3.2: Geodesic paths with different metrics between two diffusion tensors.

Metric	Geodesic path
Euclidean	$w_1\mathbf{D}_1 + w_2\mathbf{D}_2$
Log-Euclidean	$\exp\{w_1 \log(\mathbf{D}_1) + w_2 \log(\mathbf{D}_2)\}$
Riemannian	$\arg\inf_{\mathbf{T}} w_1 \parallel \log(\mathbf{D}_1^{-\frac{1}{2}}\mathbf{T}\mathbf{D}_1^{-\frac{1}{2}}) \parallel + w_2 \parallel \log(\mathbf{D}_2^{-\frac{1}{2}}\mathbf{T}\mathbf{D}_2^{-\frac{1}{2}}) \parallel$
Cholesky	$[w_1\text{chol}(\mathbf{D}_1) + w_2\text{chol}(\mathbf{D}_2)][(w_1\text{chol}(\mathbf{D}_1) + [w_2\text{chol}(\mathbf{D}_2)]^{\mathrm{T}}$
Root Euclidean	$(w_1\mathbf{D}_1^{1/2} + w_2\mathbf{D}_2^{1/2})(w_1\mathbf{D}_1^{1/2} + w_2\mathbf{D}_2^{1/2})^{\mathrm{T}}$
Procrustes size-and-shape	$(w_1\mathbf{Q}_1 + w_2\mathbf{Q}_2\hat{\mathbf{R}})(w_1\mathbf{Q}_1 + w_2\mathbf{Q}_2\hat{\mathbf{R}})^{\mathrm{T}}$

In Table 3.2, $w_1 + w_2 = 1$, $w_i \geq 0$, $i = 1, 2$, and $\hat{\mathbf{R}}$ for Procrustes path is the Procrustes solution given in Equation (3.2.23).

3.2.7.2 Multiple tensor interpolation

In some regions of the human brain, a voxel may contain more than one fibre bundle orientated in distinct directions. Let us consider two voxels (voxel a and voxel b) shown in Figure 3.3. Each voxel contains multiple fibre bundles (two fibre bundles in this example). We can model the diffusion behaviour in each voxel with our Bayesian multi-tensor model in Section 2.2.3, then estimators of two diffusion tensors can be obtained at each voxel ($\mathbf{D}_{a1}$ and $\mathbf{D}_{a2}$ at voxel a, and

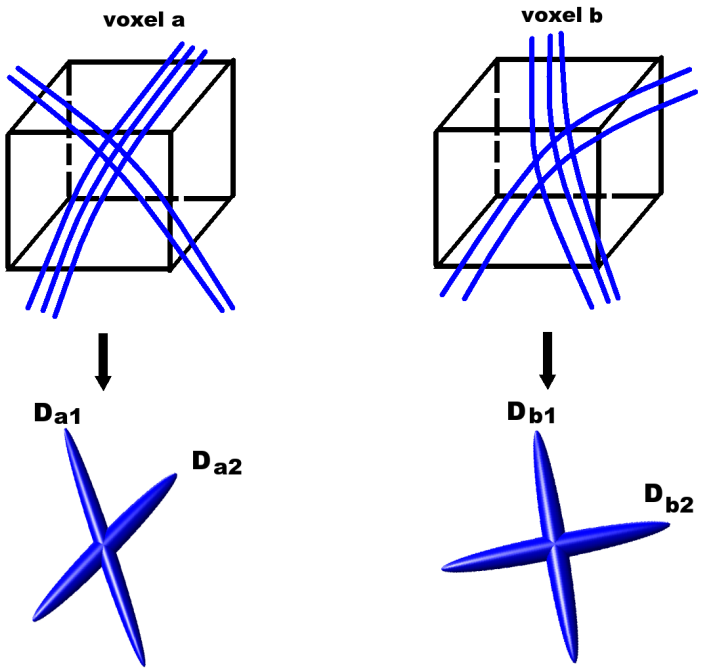

Figure 3.3: Two voxels containing multiple fibre bundles. Diffusion behaviours are modelled with diffusion tensors $\mathbf{D}_{a1}$ and $\mathbf{D}_{a2}$ at voxel a, and $\mathbf{D}_{b1}$ and $\mathbf{D}_{b2}$ at voxel b.

$\mathbf{D}_{b1}$ and $\mathbf{D}_{b2}$ at voxel b). Now the aim is to interpolate two tensors at voxel a and two tensors at voxel b with different metrics.

Here, we propose an algorithm for multiple tensor interpolation:

1. Find out the optimal match.

    1.1. Calculate distances of possible matches, i.e.,

$$d(\mathbf{D}_{a1}, \mathbf{D}_{b1}) \text{ and } d(\mathbf{D}_{a2}, \mathbf{D}_{b2}) \quad \text{for match 1}$$
$$d(\mathbf{D}_{a1}, \mathbf{D}_{b2}) \text{ and } d(\mathbf{D}_{a2}, \mathbf{D}_{b1}) \quad \text{for match 2}$$

    where $d(\cdot)$ is a metric.

    1.2. Find out the minimal distance, and the corresponding match is the optimal. Let $\{\mathbf{D}_{ai}, \mathbf{D}_{bj}\}$ and $\{\mathbf{D}_{ak}, \mathbf{D}_{bl}\}$ be the optimal match, where $i, j \in \{1,2\}$ and $k \in \{1,2\} - i$, $j \in \{1,2\} - j$.

2. Interpolation.

    2.1. Plot the geodesic path between $\mathbf{D}_{ai}$ and $\mathbf{D}_{bj}$ with metric $d(\cdot)$.

    2.2. Plot the geodesic path between $\mathbf{D}_{ak}$ and $\mathbf{D}_{bl}$ with metric $d(\cdot)$.

In Section 3.4.3.3, some experiments on multiple tensor interpolation will be discussed.

3.3 Weighted Generalised Procrustes Interpolation and Smoothing

Since diffusion MRI is a relatively low resolution modality, advanced tensor processing methods such as non-Euclidean interpolation have been considered. Yet, reliable and accurate estimation of the highly complex white matter architecture of the brain remains a challenge despite the many advances in modelling, processing, and analysis of diffusion MRI data (Lenglet *et al.*, 2009). Moreover, further inference, e.g., analysis of variance across groups, depends critically on tensor processing methods such as interpolation [Chao *et al.*, 2009]. At the same time, the recently introduced DT processing methods based on Procrustes analysis [Dryden *et al.*, 2009a] have shown promising performance and deserve further investigation. Thus, this study explores *weighted generalised Procrustes analysis (WGPA)* [Zhou *et al.*, 2009b] in which an arbitrary number of tensors can be interpolated or smoothed efficiently with the additional flexibility of controlling their individual contributions.

3.3.1 Weighted generalised Procrustes method

Given a suitable distance function d, the weighted Fréchet sample mean of $\mathbf{D}_1, \ldots, \mathbf{D}_N$ is defined by:

$$\mathbf{T} \;=\; \arg\inf_{\mathbf{D}} \sum_{i=1}^{N} w_i d(\mathbf{D}_i, \mathbf{D})^2, \tag{3.3.1}$$

where the weights w_i satisfy $w_i \geq 0$ and $\sum_{i=1}^{N} w_i = 1$, and in applications can be, for example, a function of the Euclidean distance from the location of interest to the sampling locations (e.g., voxels).

Weighted generalised Procrustes analysis (WGPA) is proposed to estimate $\hat{\mathbf{T}}$ when d $= d_S$ is the size-and-shape distance [Dryden *et al.*, 2009a]. It can then be shown that the *WGPA mean tensor* is given by

$$\hat{\mathbf{T}}_{WGPA} = \hat{\mathbf{Q}}_{WGPA} \hat{\mathbf{Q}}_{WGPA}^{T}, \tag{3.3.2}$$

where $\hat{\mathbf{Q}}_{WGPA} = \sum_{i=1}^{N} w_i \mathbf{Q}_i \hat{\mathbf{R}}_i$ and the orthogonal matrices $\hat{\mathbf{R}}_i, i = 1, \ldots, N$ minimise S_{WGPA}, the sum of weighted squared Euclidean norms, which is given by

$$\begin{aligned} S_{WGPA}(\mathbf{D}_1, ..., \mathbf{D}_N) &= \inf_{\mathbf{R}_1,\ldots,\mathbf{R}_N} \sum_{i=1}^{N} w_i \parallel \mathbf{Q}_i \mathbf{R}_i - \sum_{j=1}^{n} w_j \mathbf{Q}_j \mathbf{R}_j \parallel^2 \\ &= \inf_{\mathbf{R}_1,\ldots,\mathbf{R}_N} \sum_{i=1}^{N} w_i \parallel (1 - w_i) \mathbf{Q}_i \mathbf{R}_i - \sum_{j \neq i} w_j \mathbf{Q}_j \mathbf{R}_j \parallel^2 \\ &= \inf_{\mathbf{R}_1,\ldots,\mathbf{R}_N} \sum_{i=1}^{n} \frac{w_i}{(1 - w_i)^2} \parallel \mathbf{Q}_i \mathbf{R}_i - \frac{1}{(1 - w_i)} \sum_{j \neq i} w_j \mathbf{Q}_j \mathbf{R}_j \parallel^2 . \end{aligned} \tag{3.3.3}$$

Below we give Algorithm 2 for computing $\hat{\mathbf{Q}}_{WGPA}$:

Algorithm 2: Weighted Generalised Procrustes Method

Initial setting: $\mathbf{Q}_i^P \leftarrow chol(\mathbf{D}_i)$, $i = 1, ..., N$
S_{WGPA} from previous iteration: $S_p \leftarrow 0$
S_{WGPA} from current iteration: $S_c \leftarrow \sum_{i=1}^{N} w_i \parallel \mathbf{Q}_i^P - \sum_{j=1}^{N} w_j \mathbf{Q}_j^P \parallel^2$
while $|S_p - S_c| >$ tolerance **do**
  for $i = 1$ to N **do**
    $\hat{\mathbf{Q}}_i = \frac{1}{1-w_i} \sum_{j \neq i} w_j \mathbf{Q}_j^P$
    Calculate the $\hat{\mathbf{R}}_i$ minimising $\parallel \hat{\mathbf{Q}}_i - \mathbf{Q}_i^P \mathbf{R}_i \parallel$ (partial ordinary Procrustes analysis)
    $\mathbf{Q}_i^P \leftarrow \mathbf{Q}_i^P \hat{\mathbf{R}}_i$
  end for
  $S_p \leftarrow S_c$
  $S_c \leftarrow \sum_{i=1}^{N} w_i \parallel \mathbf{Q}_i^P - \sum_{j=1}^{N} w_j \mathbf{Q}_j^P \parallel^2$
end while
$\hat{\mathbf{Q}}_{WGPA} \leftarrow \sum_{i=1}^{N} w_i \mathbf{Q}_i^P$
return $\hat{\mathbf{Q}}_{WGPA}$

3.3.2 Weights

In WGPA we assume that the weights $w_i, i = 1, ..., N$ are a function of the Euclidean distance from the voxel of interest to the sampling voxel. The simplest setting for the weights is with the inverse distance function given by

$$w_i = \frac{d_i^{-1}}{\sum_{j=1}^{N} d_j^{-1}}, i = 1, ..., N \tag{3.3.4}$$

where d_i is the Euclidean distance from the voxel containing the weighted mean to the ith voxel with $\mathbf{D}_i$.

For more flexibility of weight setting, an exponential weight function is pro-

posed as follows:

$$w_i = \frac{\exp(-Ad_i^2) + B}{\sum_{j=1}^{N} [\exp(-Ad_j^2) + B]}, i = 1, ..., N \tag{3.3.5}$$

where $A, B \geq 0$. A comparison of the inverse distance and exponential weight functions is shown in Figure (3.4). As the distance d_i increases, the exponential

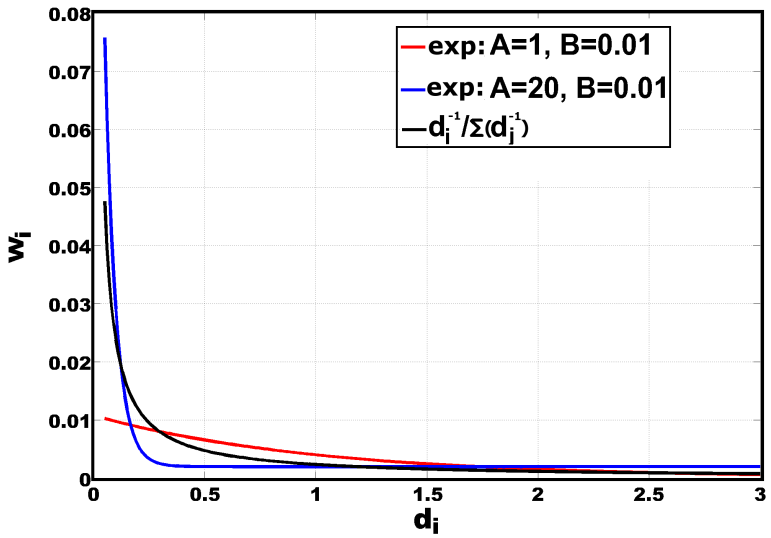

Figure 3.4: Comparison of the inverse distance and exponential weight functions. Black line: linear weights. Blue line: exponential weights with $A = 20$ and $B = 0.01$. Red line: exponential weights with $A = 1$ and $B = 0.01$.

weight with $A = 20$ and $B = 0.01$ (blue line) decreases more significantly than the inverse distance weight (black line). However, with $A = 1$ and $B = 0.01$ the exponential weight changes more steadily than the inverse distance weight.

In summary, the two parameter exponential weight family is evidently very flexible. In our experiments (see Section 3.4.5), we will use the exponential weight function. In the region of homogeneous diffusion behaviour, a set of equally likely weights is preferable. But for the region of complex diffusion behaviour, the choice of the weight function is application dependent. Future work related to automatic selection of weights for tensor smoothing will be discussed in Section 5.2.5.

3.3.3 Smoothing

Weighted generalised Procrustes framework can be adapted to smooth the diffusion tensor data. Let V_s be the voxel location in (x,y,z) coordinates. Let $\mathbf{D}_s$ be the original diffusion tensor in voxel V_s. Neighbour voxels of V_s can be defined by

$$\{V_1, V_2, ...V_m\} = \underset{V}{\arg} \quad \parallel V - V_s \parallel \leq d^* \tag{3.3.6}$$

where $d^* \geq 0$ is a constant.

Given $\mathbf{D}_1$,...,$\mathbf{D}_m$ at voxels V_1,...,V_m, the weighted mean $\bar{\mathbf{D}}_s$ is the weighted generalised Procrustes mean of $\mathbf{D}_1$,...,$\mathbf{D}_m$ and $\mathbf{D}_{m+1}$, where $\mathbf{D}_{m+1} = \mathbf{D}_s$. It is natural to let $\mathbf{D}_s$ contribute to the weighted mean, and let V_s be a neighbour of itself, i.e.$V_s = V_{m+1}$. Weights of each diffusion tensor can be set with a weight function. For example, the exponential weights are given by

$$w_i = \frac{\exp\left\{-A \parallel V_i - V_s \parallel^2\right\} + B}{\sum_{j=1}^{m+1} [\exp\left\{-A \parallel V_j - V_s \parallel^2\right\} + B]}, i = 1, ..., m+1. \tag{3.3.7}$$

In particular, since $\parallel V_{m+1} - V_s \parallel = 0$, $\mathbf{D}_s$ with the largest weight w_{m+1} contributes most.

In a diffusion tensor dataset given each diffusion tensor $\mathbf{D}_s$ at voxel V_s and $\mathbf{D}_1, ..., \mathbf{D}_N$ at neighbour voxels $V_1, ..., V_N$, we can calculate the weighted mean tensor $\bar{\mathbf{D}}_s$ which will replace each $\mathbf{D}_s$. The weights $w_1, ..., w_N$ and w_s are set as proposed in Equation (3.3.5).

3.3.4 Interpolation

By interpolation of the tensor data we mean construction of new diffusion tensors based on the original data. More specifically, we mesh the three-dimensional

volume containing diffusion tensor data with regular rigid. For each new born subvoxel V^*, we will sample a weighted generalised Procrustes mean of diffusion tensors at V^*'s neighbours, and allocate this mean to V^*.

3.4 Applications

3.4.1 Materials

A set of diffusion weighted-MR images from a healthy human brain is used (see Chapter 2, Section 2.4.2.1).

A Bayesian estimation method [Zhou *et al.*, 2008] discussed in Section 2.2.3 has been carried out to compute the tensor field and all methods considered in this study are programed with MATLAB (R2009a, The MathWorks, Inc., Natick, Massachusetts, USA).

3.4.2 Anisotropy study

Now let us compare FA, PA and tanh(GA) maps from real data. Figure 3.5 is FA, PA and tanh(GA) maps (axial slices) computed from the healthy human brain data described in Section 2.4.2.1. Since PA of diffusion tensor is always smaller than FA and tanh(GA) values, the PA map gives a darker colour overall. The splenium in corpus callosum is one of the regions where the overall anisotropy is strongly high [Lee *et al.*, 2009]. We take FA, PA and tanh(GA) values along the green line in the splenium and show them in Figure 3.6. PA has significantly higher variation than FA and tanh(GA). In general, PA offers better contrast in highly anisotropic regions.

We have discussed that $PA(\mathbf{D}) = FA(\mathbf{D}^{1/2})$ is a member of the family

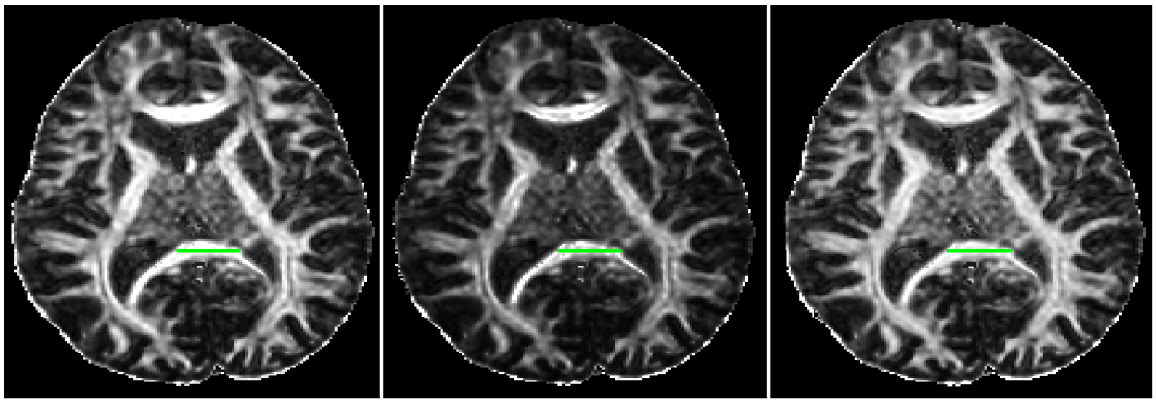

Figure 3.5: Anisotropy maps from axial view. Left: FA map. Middle: PA map. Right: tanh GA map.

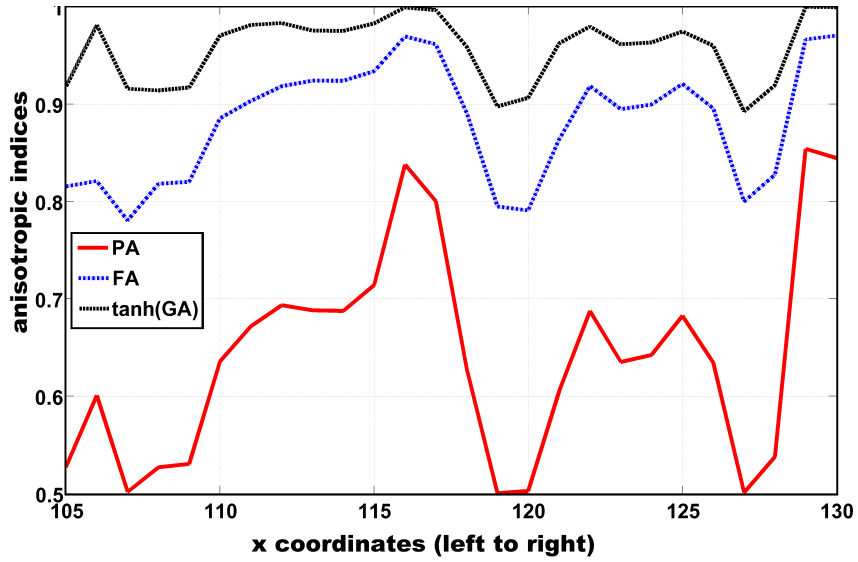

Figure 3.6: Comparison of FA, PA and tanh(GA) values. FA, PA and tanh(GA) values are from tensors at voxels along the green line in Figure 3.5. PA value is more sensitive than FA and tanh(GA) with significantly higher variation.

FA($\mathbf{D}^a$) where $a \in [0,1]$. Now let $a \in \{1,3/4,1/2,1/4\}$, and corresponding FA($\mathbf{D}^a$) maps are shown in Figure 3.7. We prove that FA($\mathbf{D}^a$) is a increasing function of a (see Theorem 3.1 in Section 3.2.6.3). Therefore, as we decreases a from 1 to 1/4, the image of FA($\mathbf{D}^a$) becomes darker. The highlighted cutouts in the four maps are the cerebral peduncle (cp) which has strongly anisotropic diffusion. It is obvious that FA($\mathbf{D}^{3/4}$) and FA($\mathbf{D}^{1/2}$) provide better contrast in this region (cp) than FA($\mathbf{D}$). To investigate how to fix the a value for regional anisotropy study is a potential topic in our future work. Certainly, the choice of

a can also be made dynamically by the end user. To develop a practical visualisation tool where a can vary will be helpful for choosing a suitable anisotropy index.

3.4.3 Geodesic interpolation

3.4.3.1 Interpolation of two diffusion tensors

Now we carry out five experiments to investigate the geometric nature of geodesic paths obtained with different metrics. Descriptions of the five experiments are listed in Table 3.3.

Table 3.3: Description of five experiments for investigating the geometric nature of geodesic path obtained with different metrics.

Experiment	Description
1	$\mathbf{D}_1$ is fully isotropic with small size. $\mathbf{D}_2$ is anisotropic with larger size
2	$\mathbf{D}_1$ is fully isotropic. $\mathbf{D}_2$ is anisotropic. Both tensors have same size.
3	$\mathbf{D}_1$ and $\mathbf{D}_2$ have same orientation (eigenvectors). Both tensors are anisotropic but with different size.
4	$\mathbf{D}_1$ and $\mathbf{D}_2$ have same shape (eigenvalues) and are anisotropic. Orientations of $\mathbf{D}_1$ and $\mathbf{D}_2$ are orthogonal.
5	A general case: $\mathbf{D}_1$ and $\mathbf{D}_2$ are not orthogonal with different shape and size.

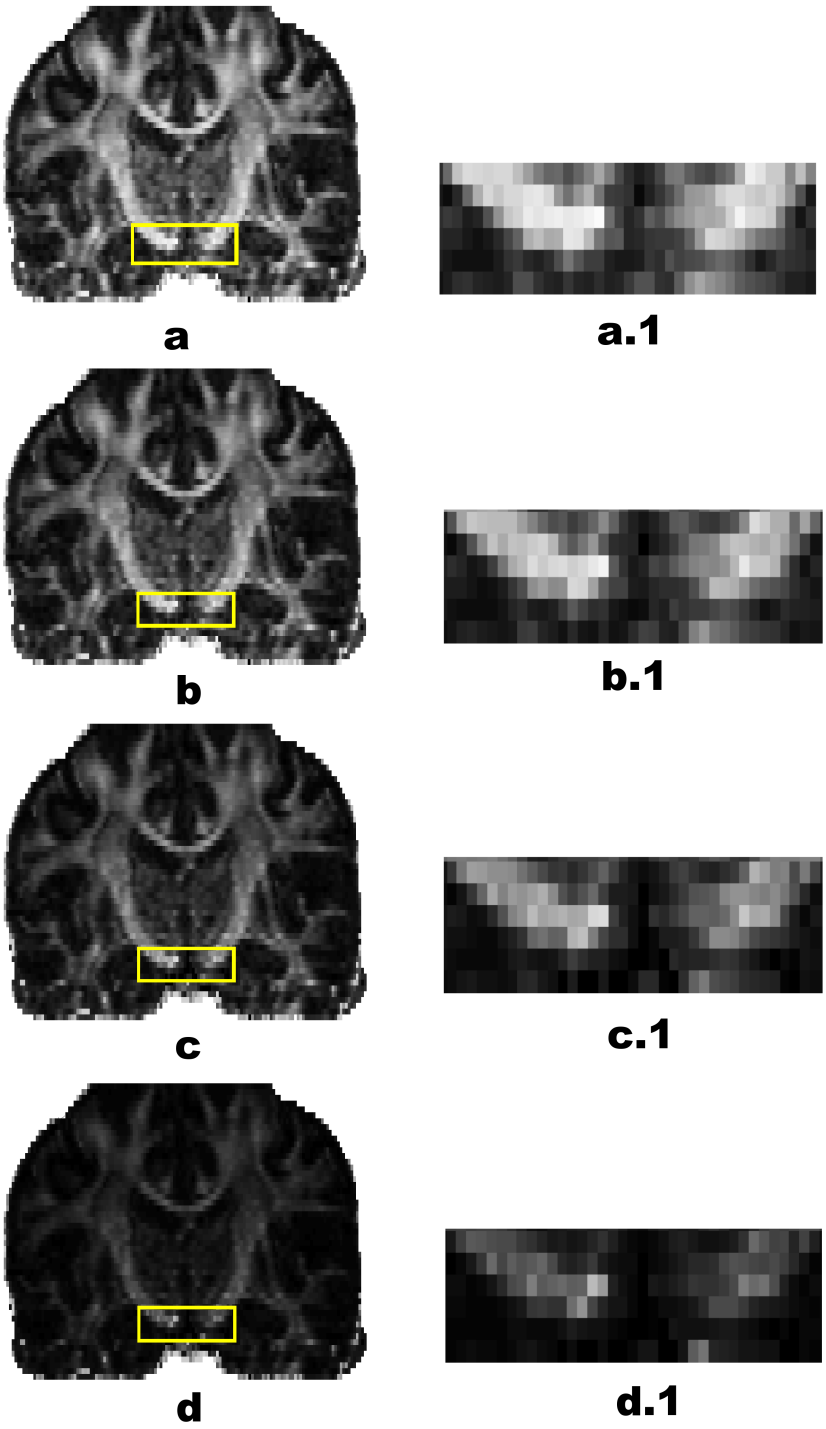

Figure 3.7: $FA(\mathbf{D}^q)$ maps (coronal slices) from human brain. a: FA($\mathbf{D}$) map, b: FA($\mathbf{D}^{3/4}$) map, c: FA($\mathbf{D}^{1/2}$)=PA($\mathbf{D}$) map and d: FA($\mathbf{D}^{1/4}$) map. The highlight cutout in the yellow rectangle is the cerebral peduncle. a.1, b.1, c.1 and d.1 are zoomed cutouts from corresponding maps.

The settings of $\mathbf{D}_1$ and $\mathbf{D}_2$ for each experiment are listed as follows.

$$\text{Experiment 1: } \mathbf{D}_1 = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \mathbf{D}_2 = \begin{pmatrix} 25.75 & 42.8683 & 0 \\ 42.8683 & 75.25 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

$$\text{Experiment 2: } \mathbf{D}_1 = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix}, \mathbf{D}_2 = \begin{pmatrix} 8.5 & 7.5 & 0 \\ 7.5 & 8.5 & 0 \\ 0 & 0 & 4 \end{pmatrix}.$$

$$\text{Experiment 3: } \mathbf{D}_1 = \begin{pmatrix} 21.9472 & 12.9878 & 0 \\ 12.9878 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \mathbf{D}_2 = \begin{pmatrix} 8.2500 & -12.5574 & 0 \\ -12.5574 & 22.75 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

$$\text{Experiment 4: } \mathbf{D}_1 = \begin{pmatrix} 46.506 & 28.2149 & 0 \\ 28.2149 & 18.494 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \mathbf{D}_2 = \begin{pmatrix} 16.75 & -27.2798 & 0 \\ -27.2798 & 48.25 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

$$\text{Experiment 5: } \mathbf{D}_1 = \begin{pmatrix} 5.5 & 4.5 & 0 \\ 4.5 & 5.5 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \mathbf{D}_2 = \begin{pmatrix} 4.7242 & -11.4618 & 0 \\ -11.4618 & 36.2758 & 0 \\ 0 & 0 & 4 \end{pmatrix}.$$

To compare interpolations with different metrics in size, orientation and anisotropy of tensor, we use four measures: the determinant $|\mathbf{D}|$, ϕ, FA and PA, where the angle ϕ measures the difference of orientations from the synthetic $\mathbf{D}_1$ to a interpolated tensor in the geodesic path. The angle ϕ is the smaller angle between the principal eigenvectors of $\mathbf{D}_1$ and the interpolated tensor. The angle ϕ is defined as

$$\phi = \arcsin(\parallel \mathbf{pv}_1 \times \mathbf{pv}_i \parallel), i = 1, ..., 9 \tag{3.4.1}$$

where $\mathbf{pv}_1$ is the principal eigenvector of $\mathbf{D}_1$, and $\mathbf{pv}_i$ is the principal eigenvector of the ith interpolated tensor (including two synthetic diffusion tensors), and $i = 1, ..., 9$, with $i = 1$ and $i = 9$ corresponding to the synthetic tensors $\mathbf{D}_1$ and $\mathbf{D}_2$, respectively. .

In experiment 1, $\mathbf{D}_1$ is fully isotropic and $\mathbf{D}_2$ is strongly anisotropic. Also, $\mathbf{D}_1$ has smaller size. Geodesic interpolation between $\mathbf{D}_1$ and $\mathbf{D}_2$ is carried out with different metrics and results are shown in Figure 3.8. Geodesic paths with the log-Euclidean and the Riemannian metrics are very similar. The tensors interpolated with the Euclidean metric are larger. There is no significant difference in the graphs of ϕ, FA and PA. Note the most abrupt transitions in FA and PA occur with the Euclidean distance.

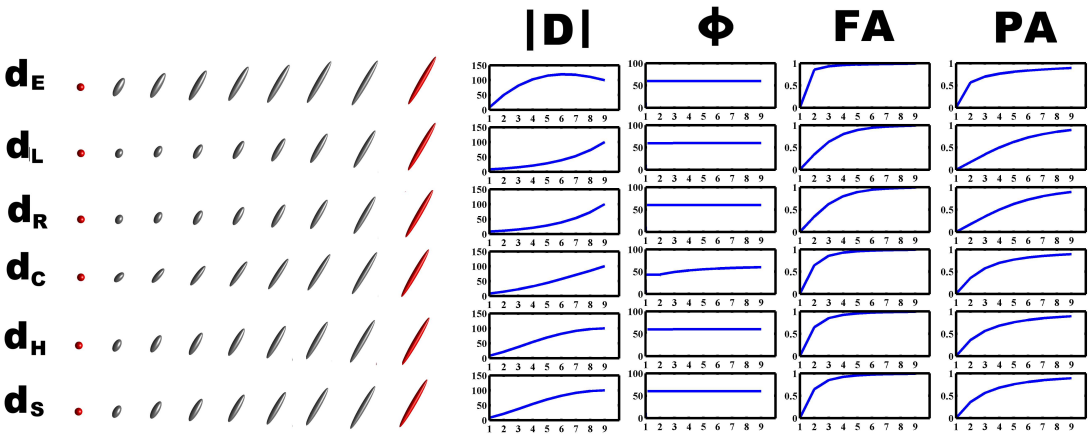

Figure 3.8: Geodesic paths in experiment 1 between an isotropic tensor and an strongly anisotropic tensor. The size of the isotropic tensor $\mathbf{D}_1$ (left in red) is much smaller than the size of the strongly anisotropic tensor $\mathbf{D}_2$ (right in red). The geodesic paths are obtained with $d_E(.)$, $d_L(.)$, $d_R(.)$, $d_C(.)$, $d_H(.)$ and $d_S(.)$.

Figure 3.9 shows six different geodesic paths in experiment 2. In this experiment, $\mathbf{D}_1$ is fully isotropic, whereas $\mathbf{D}_2$ is strongly anisotropic. The two tensors have the same volume. In Figure 3.9, it is clear that the Euclidean metric is prone to swelling. With the Euclidean, log-Euclidean, Riemannian, root Euclidean and Procrustes metrics, all interpolated tensors in the second column orientate into the angle which is same as the orientation of the defined tensor

in the right (red). With the Cholesky metric, however, the angle ϕ increases gradually along the geodesic path, it is worth recalling at this point that the Cholesky interpolation is also problematic in practice due to its lack of rotation invariance.

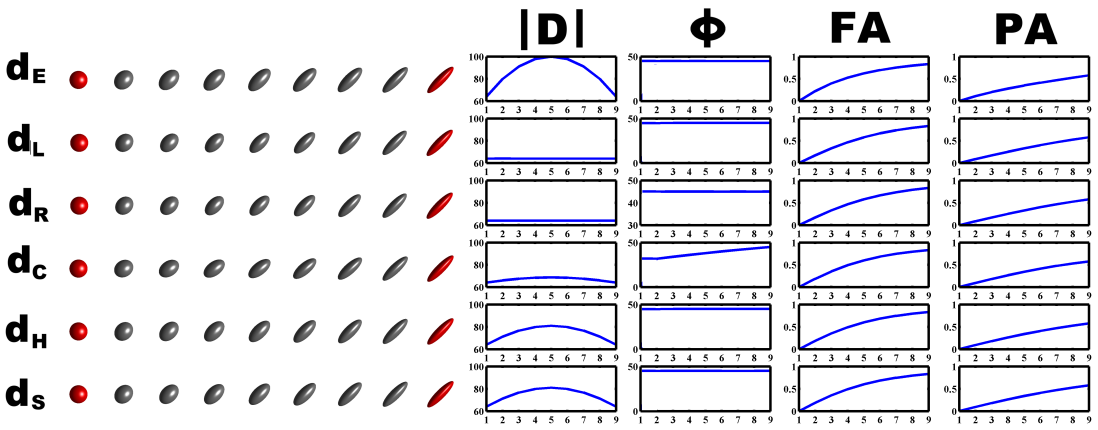

Figure 3.9: Geodesic paths in experiment 2 between two tensors with the same size. The tensor $\mathbf{D}_1$ (left in red) is fully isotropic and the tensor $\mathbf{D}_2$ is strongly anisotropic (right in red). The geodesic paths are obtained with $d_E(.)$, $d_L(.)$, $d_R(.)$, $d_C(.)$, $d_H(.)$ and $d_S(.)$.

Diffusion tensors $\mathbf{D}_1$ and $\mathbf{D}_2$ in experiment 3 have same orientation (eigenvectors). Both the tensors are anisotropic but with different size. In Figure 3.10, the geodesic paths with the six metrics are similar, and there is little difference shown in graphs of $|\mathbf{D}|$, ϕ, FA and PA.

In experiment 4 (Figure 3.11), $\mathbf{D}_1$ and $\mathbf{D}_2$ have the same shape (eigenvalues) and are anisotropic. Orientations of $\mathbf{D}_1$ and $\mathbf{D}_2$ are mutually orthogonal. In Figure 3.11, there are big differences between geodesic paths with different metrics. We can see that the Euclidean, log-Euclidean, Riemannian, root Euclidean and Procrustes metrics provide nearly isotropic tensors in the middle column (5th column from left). This means that the average of two orthogonal tensors of the same shape and size is isotropic. However, the Cholesky path looks rather different from the Procrustes path, due to the lack of rotation invariance. The swelling effect of the Euclidean metric is more obvious in this example. The

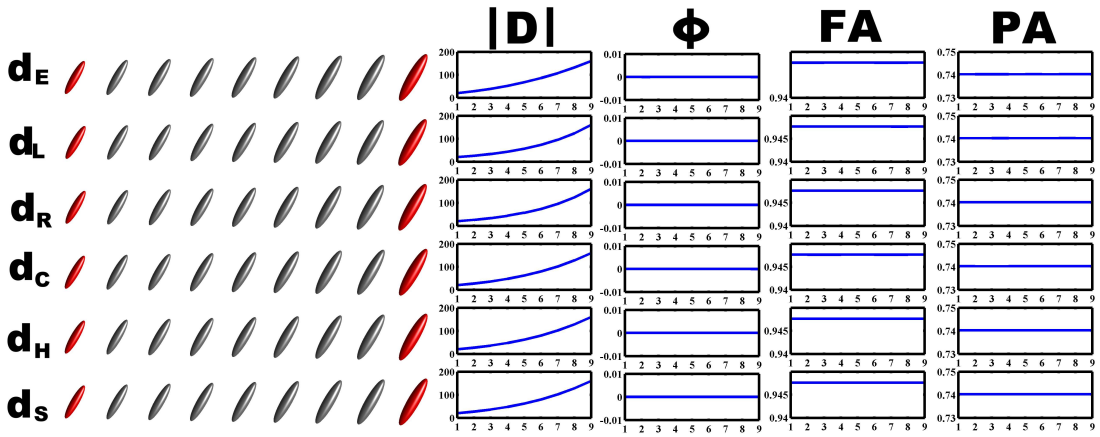

Figure 3.10: Geodesic paths in experiment 3 between two tensors with same orientation. Both tensors (in red) are anisotropic but with different sizes. The geodesic paths are obtained with $d_E(.)$, $d_L(.)$, $d_R(.)$, $d_C(.)$, $d_H(.)$ and $d_S(.)$.

log-Euclidean and Riemannian metrics give strong weights to small volumes. All but the Cholesky path undergo abrupt angular transitions.

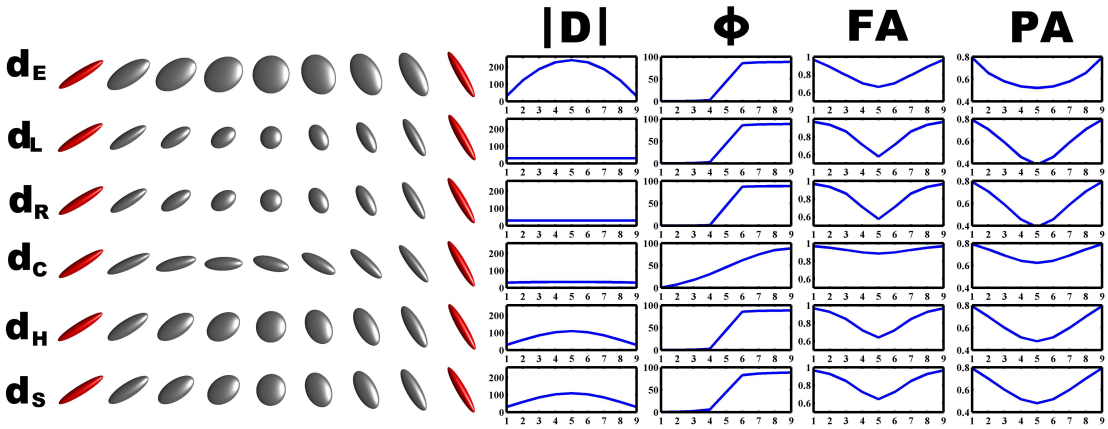

Figure 3.11: Geodesic paths in experiment 4 between two orthogonal tensors. Both tensors (in red) have same shape (eigenvalues) and are anisotropic. The geodesic paths are obtained with $d_E(.)$, $d_L(.)$, $d_R(.)$, $d_C(.)$, $d_H(.)$ and $d_S(.)$.

In experiment 5, the setting of $\mathbf{D}_1$ and $\mathbf{D}_2$ is more general. The tensors $\mathbf{D}_1$ and $\mathbf{D}_2$ are not orthogonal and are of different shape and size. Figure 3.12 shows six different geodesic paths between $\mathbf{D}_1$ and $\mathbf{D}_2$. From a variety of examples it does seem clear that the Euclidean metric is very problematic, especially due to the parabolic interpolation of the determinant. The Procrustes metric offers somewhat better interpolation in the tensor's orientation and anisotropy

(see graphs of $|\mathbf{D}|$ and ϕ). The Cholesky path has a significant cusp in the tensor's volume, orientation and anisotropy in this example. In general, the log-Euclidean and Procrustes size-and-shape methods seem preferable.

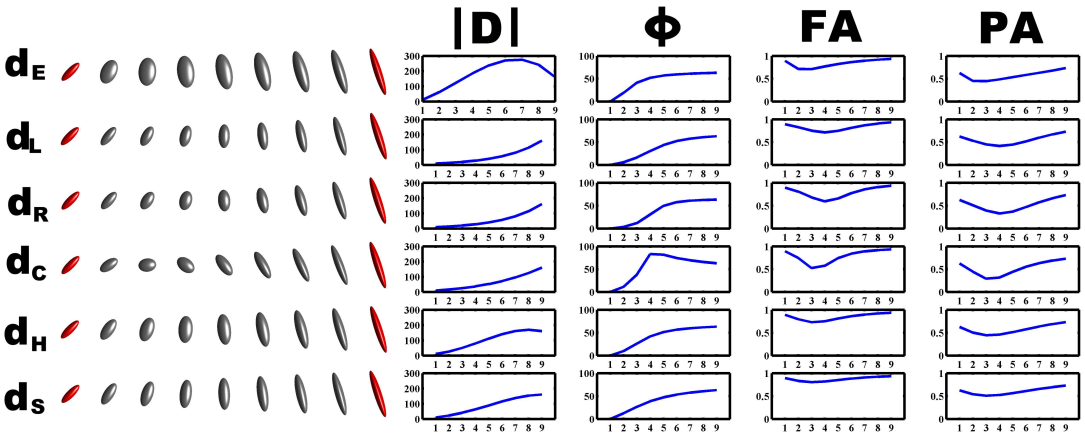

Figure 3.12: Geodesic paths in experiment 5 between two general tensors. Two tensors $\mathbf{D}_1$ (left in red) and $\mathbf{D}_2$ (right in red) with general (i.e. non-collinear non-orthogonal) orientation, different shape and size. The geodesic paths are obtained with $d_E(.)$, $d_L(.)$, $d_R(.)$, $d_C(.)$, $d_H(.)$ and $d_S(.)$.

3.4.3.2 Interpolation under simultaneous rotation

In this study, we also carry out experiments to investigate the geometry of geodesic path under rotation. The experiment is described as follows:

1. Obtain six pairs of diffusion tensors by simultaneous rotation.

   1.1 Start with two orthogonal tensors:

$$\mathbf{D_1} = \begin{pmatrix} 37.3875 & 9.75 & 0 \\ 9.75 & 3.6125 & 0. \\ 0 & 0 & 2 \end{pmatrix}, \mathbf{D_2} = \begin{pmatrix} 3.6125 & -9.75 & 0 \\ -9.75 & 37.3875 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

   with the same setting of eigenvalues 40, 2, 1. Since the largest eigenvalue is much greater than other two eigenvalues, $\mathbf{D_1}$ and $\mathbf{D_2}$ are two linear tensors. Their principal eigenvectors are orthogonal.

1.2 Rotate eigenstructures of $\mathbf{D_1}$ and $\mathbf{D_2}$ simultaneously along z-axis in clockwise direction with step size of 15°.

2. Plot geodesic paths of each pair of tensors with different metrics.

Figure 3.13 and Figure 3.14 show geodesic paths and graphs of measures of each pair of tensors (red) with different metrics respectively. It is clear that the geometric of geodesic path obtained with the Cholesky approach is not invariant under simultaneous rotation of $\mathbf{D_1}$ and $\mathbf{D_2}$, although all the other are rotationally invariant.

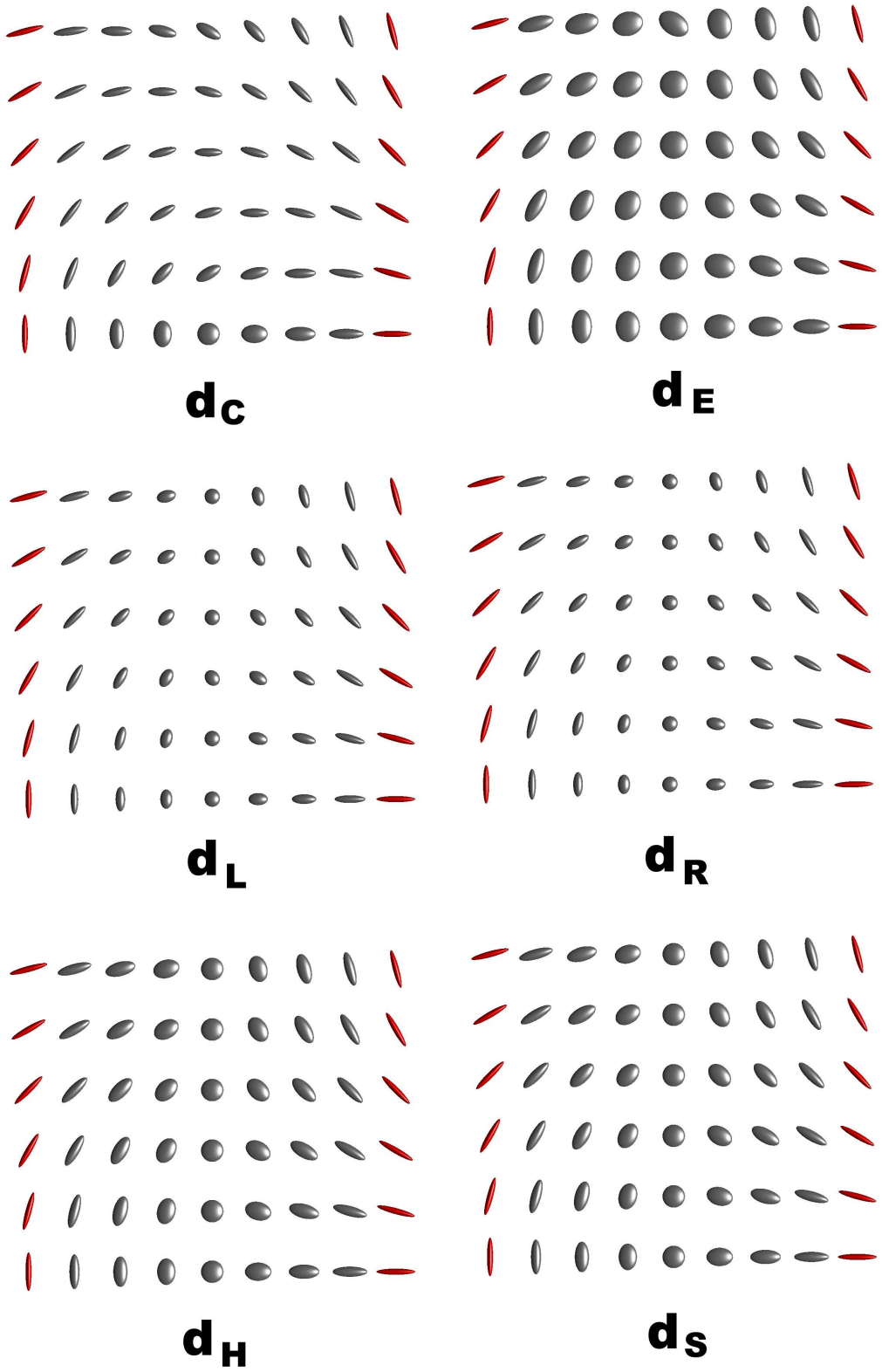

Figure 3.13: Geodesic paths under simultaneous rotation of two tensors (red) with different metrics.

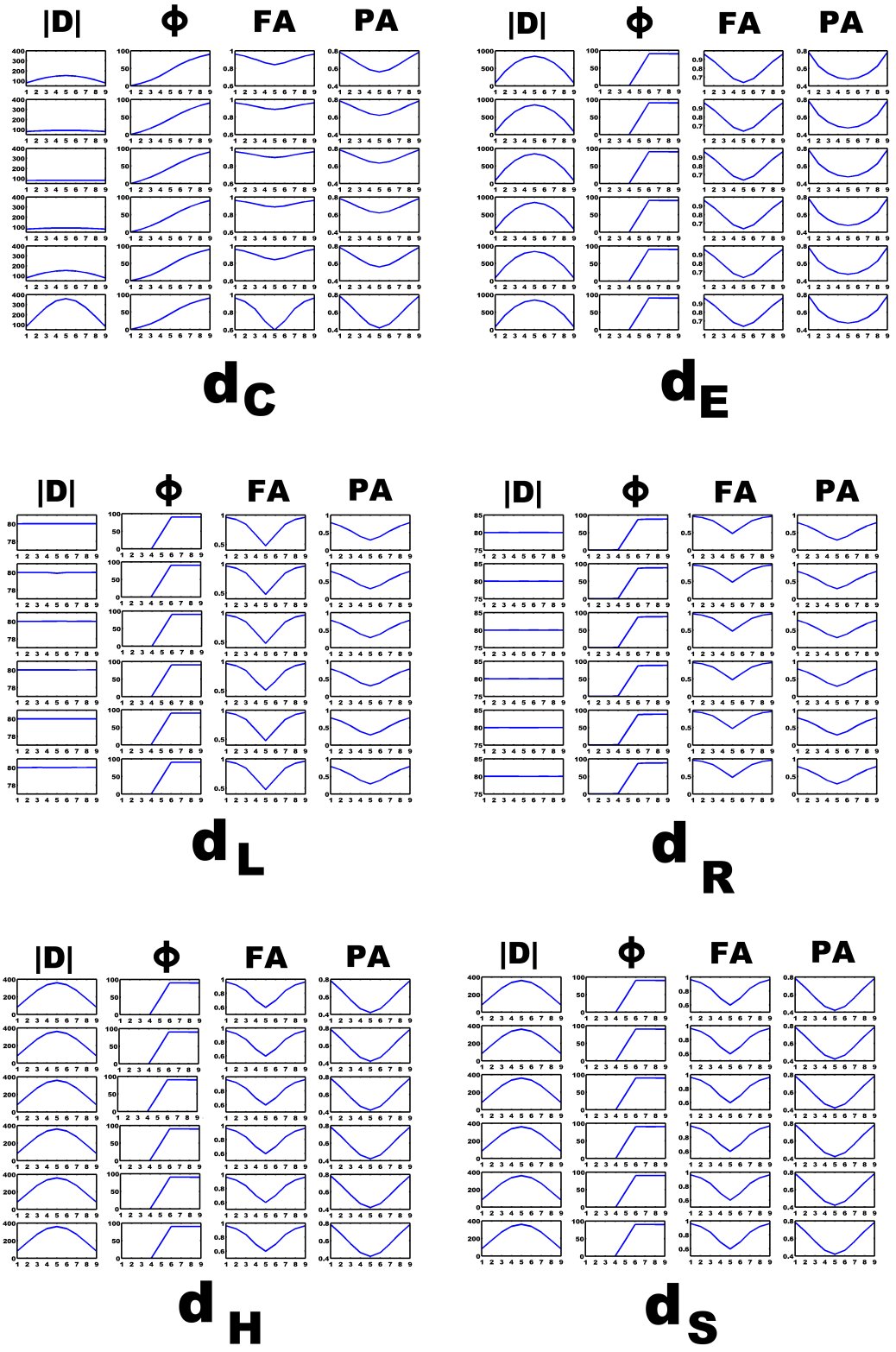

Figure 3.14: Graphs of size, orientation and anisotropy measures of interpolated tensors with different metrics. Two tensors (red) are rotated simultaneously.

3.4.3.3 Multiple tensor interpolation

Now consider multiple tensor interpolation with two tensors. We define two pairs of tensors as follows

$$\text{Pair a: } \mathbf{D}_{a1} = \begin{pmatrix} 40 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \mathbf{D}_{a2} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 40 & 0 \\ 0 & 0 & 2 \end{pmatrix},$$

where $\mathbf{D}_{a1}$ and $\mathbf{D}_{a2}$ are two linear tensor (by setting the largest eigenvalue much greater than other two eigenvalues), and

$$\text{Pair b: } \mathbf{D}_{b1} = \begin{pmatrix} 30.25 & 16.8875 & 0 \\ 16.8875 & 10.75 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \mathbf{D}_{b2} = \begin{pmatrix} 10.75 & -16.8875 & 0 \\ -16.8875 & 30.25 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

Within each pair the two tensors have orthogonal orientations and the same eigenvalues. Let us denote the smaller angle between the principal eigenvectors of $\mathbf{D}_{ai}$ and $\mathbf{D}_{bj}$ by ϕ_{ij}, where $i, j \in \{1, 2\}$. In this example,

$$\phi_{11} = 30^\circ, \phi_{22} = 30^\circ, \tag{3.4.2}$$

and

$$\phi_{12} = 60^\circ, \phi_{21} = 60^\circ. \tag{3.4.3}$$

Figure 3.15 shows six geodesic paths between the two pairs of tensors obtained with the different metrics. There is a clear swelling effect in Euclidean case again in this example. The log-Euclidean and Riemannian paths are more different in this example. Namely, the Riemannian method clearly deflates the middle ellipsoids and also makes them more isotropic, and thus the log-

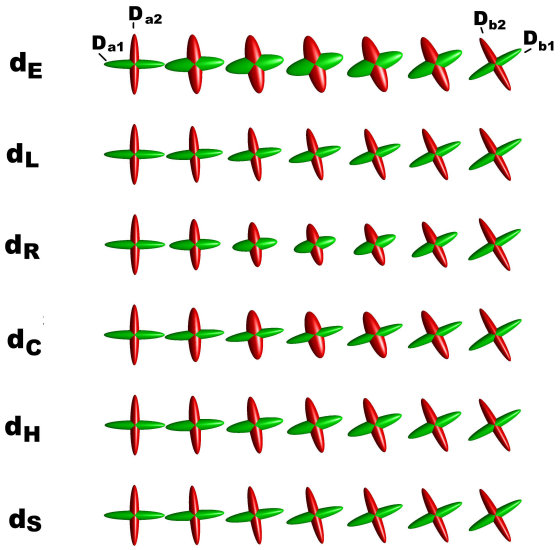

Figure 3.15: Six geodesic paths between two pairs of orthogonal tensors. Pair a: $\mathbf{D}_{a1}$ and $\mathbf{D}_{a2}$. Pair b: $\mathbf{D}_{b1}$ and $\mathbf{D}_{b2}$. There are two geodesic paths (green path or red path) obtained with a specific metric in each row.

Euclidean method provides better interpolation in tensor anisotropy and size. The red path with the Cholesky method tends to be more isotropic in the three middle columns (3rd, 4th and 5th column) than the green path. This clearly reveals the lack of invariance of the Cholesky method to the orthogonal change of basis. The Procrustes and log-Euclidean methods seems most preferable here in this example.

Let us next consider two pairs of tensors given as follows:

$$\text{Pair a: } \mathbf{D}_{a1} = \begin{pmatrix} 1.6029 & 2.25 & 0 \\ 2.25 & 9.3971 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \mathbf{D}_{a2} = \begin{pmatrix} 1.6029 & 2.25 & 0 \\ 2.25 & 9.3971 & 0 \\ 0 & 0 & 2 \end{pmatrix},$$

and

$$\text{Pair b: } \mathbf{D}_{b1} = \begin{pmatrix} 30.25 & 16.8875 & 0 \\ 16.8875 & 10.75 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \mathbf{D}_{b2} = \begin{pmatrix} 10.75 & -16.8875 & 0 \\ -16.8875 & 30.25 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

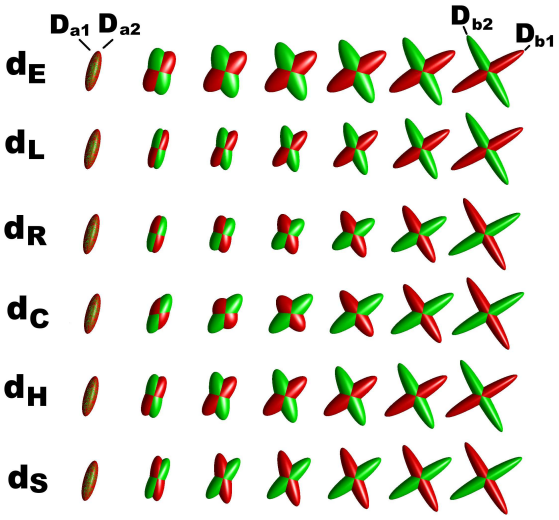

Figure 3.16: Six geodesic paths between two pairs of different tensors. Pair a: $\mathbf{D}_{a1}$ and $\mathbf{D}_{a2}$. Pair b: $\mathbf{D}_{b1}$ and $\mathbf{D}_{b2}$. There are two geodesic paths (green path or red path) obtained with a specific metric in each row.

This is an important special case that $\mathbf{D}_{a1} = \mathbf{D}_{a2}$ and can arise around locations where fibre bundles branch, or, form 'fanning'. In pair b, $\mathbf{D}_{b1}$ and $\mathbf{D}_{b2}$ are orthogonal and have the same eigenvalues. In this example, $\phi_{ij} = 45°, i, j \in \{1, 2\}$.

Figure 3.16 shows six geodesic paths between these two pairs of tensors computed with the six metrics. The Euclidean geodesic path is problematic again. The Cholesky geodesic path is very different from the other paths, i.e., for each step of the path (especially from column 2 to column 3), the green and red tensors have different shape and size. This is, of course, expected since the method is not invariant to the orthogonal change of basis. Recall (Section 3.2.5) that the other methods do have this mode of invariance, hence in the other geodesic paths, the green and red tensors have the same shape and size, although they have different orientations. The Procrustes approach again appears preferable, giving the most uniform interpolation of size and orientation and the least distortion of shape.

3.4.4 Cross validation

For assessing the different metrics, a cross-validation procedure is carried out over a region of interest (Figure 3.17 (a)). Voxels V_1,...,V_K ($K = 36$ for this example) with yellow edges are shown in Figure 3.17 (b) containing the validating tensors $\mathbf{D}_1$,..., $\mathbf{D}_K$ estimated from the Bayesian framework described in Section 2.2.3 [Zhou *et al.*, 2008], and the training tensors are at voxels without edges. We estimate the tensor at voxel V_i, $i = 1$,...,K by computing the weighted mean of six neighbouring tensors (see Figure 3.18) with the Euclidean, log-Euclidean and Procrustes methods. Let $\mathbf{D}_i^*$, ..., $\mathbf{D}_K^*$ be tensors estimated with a certain method. To validate results, we can compute the root mean squared metric (RMSM) between the estimated tensor and the true tensor which is given by

$$RMSM(d) = \sqrt{\frac{1}{K}\sum_{i=1}^{K} d(\mathbf{D}_i, \mathbf{D}_i^*)^2} \tag{3.4.4}$$

where d can be any metric, such as d_E, d_L and d_S.

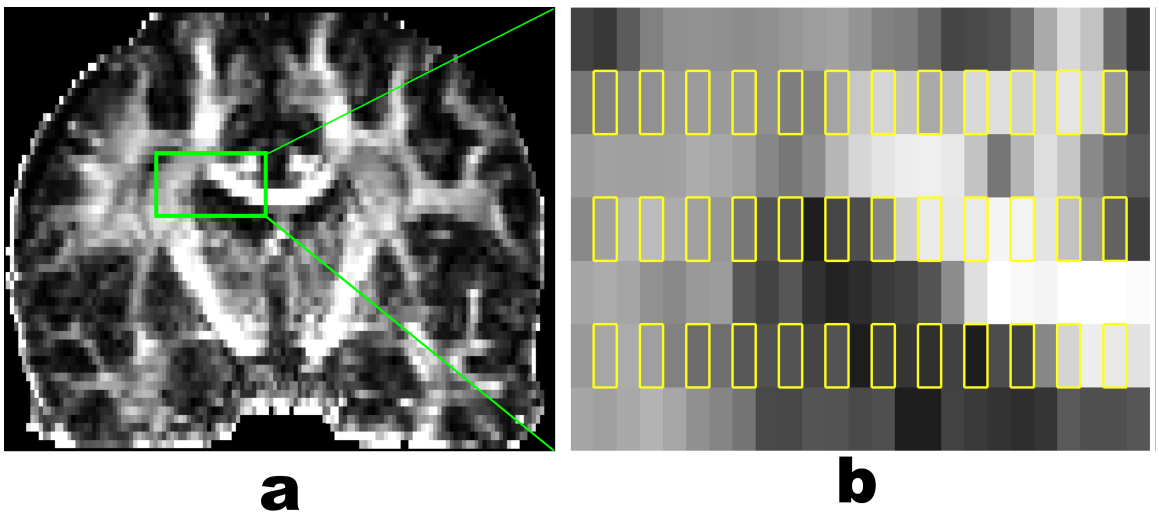

Figure 3.17: Region for cross validation. a: a coronal slice of FA map. b: zoomed in region for cross-validation. Testing tensor data are from voxels with yellow edges.

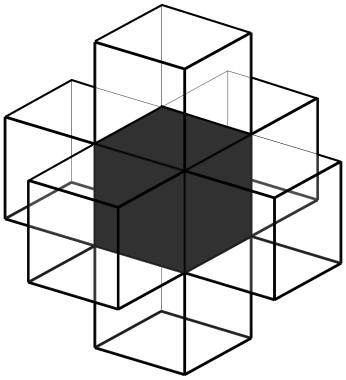

Figure 3.18: The voxel containing validating tensor (black) and its six neighbouring voxels (white).

We can also calculate the root mean squared error (RMSE) of any tensor-derived function, such as $FA(\mathbf{D})$. Then, $RMSE(FA)$ is given by

$$RMSE(FA) = \sqrt{\frac{1}{K}\sum_{i=1}^{K}[FA(\mathbf{D}_i) - FA(\mathbf{D}_i^*)]^2} \tag{3.4.5}$$

Table 3.4 shows $RMSM(d_E)$, $RMSM(d_L)$, $RMSM(d_E)$ and $RMSE(FA)$ obtained with the Euclidean, log-Euclidean and Procrustes methods respectively. By comparing $RMSM(d_L)$, $RMSM(d_S)$ and $RMSE(FA)$, we note that the Euclidean method performs the worst in general. However, it is clear that the Procrustes method is the best for it provides smaller $RMSM(d_L)$, $RMSM(d_S)$ and $RMSE(FA)$ than the other two methods, although there is not such a large difference between the Procrustes and log-Euclidean methods.

Table 3.4: Measures of the validating results with different methods.

	$RMSM(d_E)$	$RMSM(d_L)$	$RMSM(d_S)$	$RMSE(FA)$
Euclidean	$\mathbf{1.95 \times 10^{-4}}$	2.26×10^{-1}	2.12×10^{-3}	6.225×10^{-3}
Log-Euclidean	5.82×10^{-4}	1.99×10^{-1}	1.30×10^{-3}	5.38×10^{-3}
Procrustes	2.81×10^{-4}	$\mathbf{1.62 \times 10^{-1}}$	$\mathbf{8.86 \times 10^{-4}}$	$\mathbf{5.31 \times 10^{-3}}$

If the diffusion tensor at a certain voxel is significantly different from its

neighbours, then the estimate calculated by smoothing its neighbours need not be accurate regardless of the method being used. It is worth observing the histogram of the K individual errors. Four measures could be considered, namely ε_{FA} (the error between FA values of the estimate and the true, $\varepsilon_{FA} = |FA(\mathbf{D}_i) - FA(\mathbf{D}_i^*)|$) and three distances $d_M(\mathbf{D}_i, \mathbf{D}_i^*)$ ($M = E$ for the Euclidean distance, $M = L$ for the log-Euclidean distance, and $M = S$ for the Procrustes distances) between the true tensors $\mathbf{D}_i$ and the estimated tensors $\mathbf{D}_i^*$ (with the Euclidean, log-Euclidean and Procrustes methods). Histograms of these four measures are shown in Figure 3.19. Now, as far as the histograms are concerned, there is little difference among the three methods. The log-Euclidean and Procrustes seem to be more similar to each other than either to the Euclidean. One observation that might be of some interest is that the most significant outliers under the d_E occur with the log-Euclidean method, whereas under d_S and d_L, those come from the Euclidean method. To choose a suitable measure for assessing different diffusion processing methods is important for real data study. The cross validation study again illustrates that analysis with the log-Euclidean and Procrustes methods are rather different from the Euclidean method.

3.4.5 Interpolation and smoothing of real data

We smooth and interpolate (with 2 interpolations between each pair of original voxels) the diffusion tensor data from the human brain, and calculate the FA and PA maps shown in Figure 3.20. Obviously, FA and PA maps from the processed tensor data are much smoother than the ones without processing. The feature that the cingulum (cg) is distinct from the corpus callosum (cc) is clearer in the anisotropy maps from the processed data than those without processing in Figure 3.20.

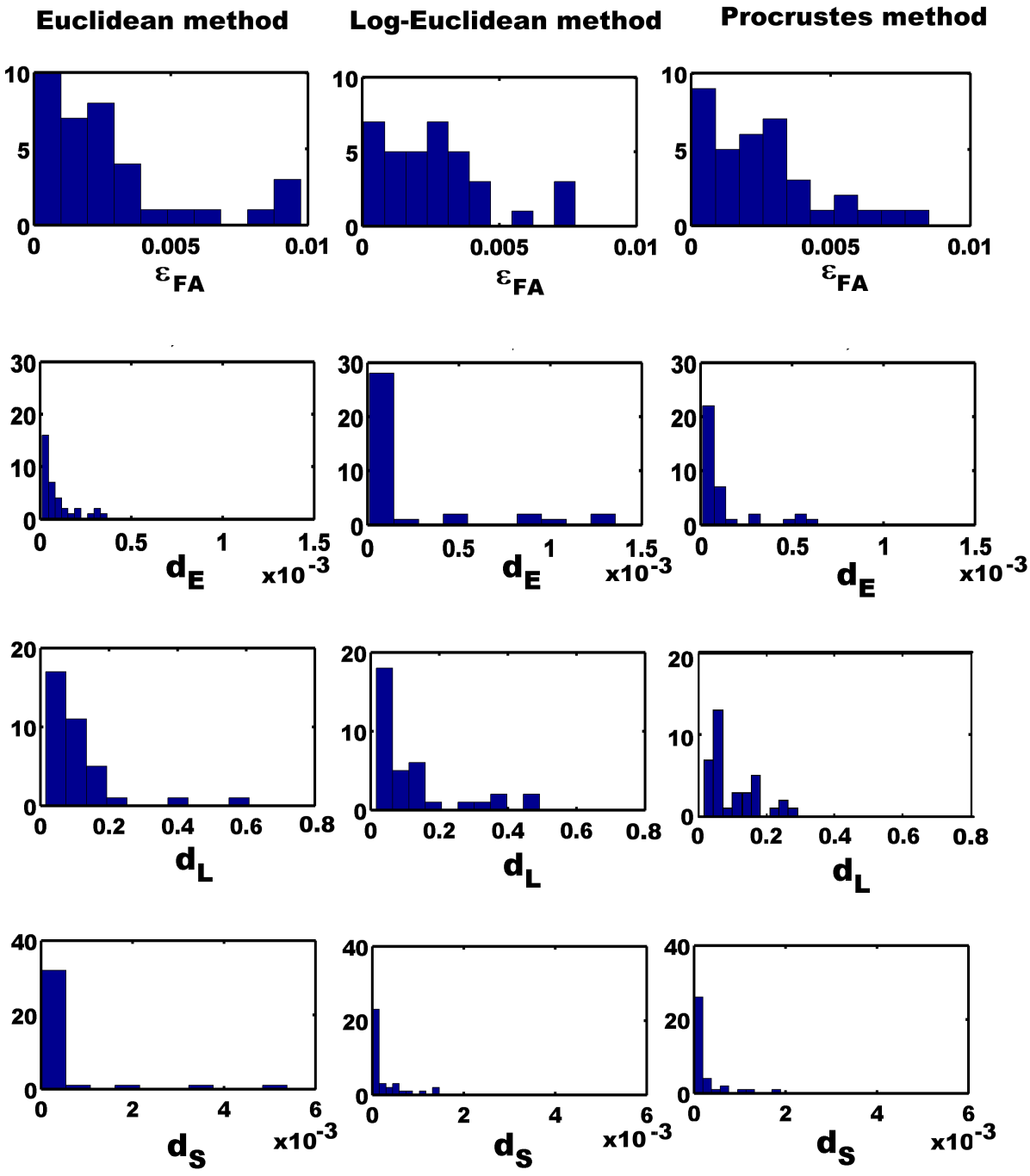

Figure 3.19: Histograms of measures for cross validation. The Euclidean (first column), log-Euclidean (second column) and Procrustes (third column) methods are used for estimating the mean tensors. The FA errors ε_{FA} (first row). The Euclidean distance d_E (second row) between the true and estimated tensors. The log-Euclidean d_L (third row). The Procrustes d_S (fourth row). The unit of each vertical axis is the number of validating tensors.

Results of fibre tractography for the brain stem of a healthy human have been shown in Figure 3.21. Evidently, the tractography based on the WGPA processed tensor field is different from the tractography based on the other methods. For example, in the Euclidean two tracts in Figure 3.21 (b) (Euclidean) and (c) (WGPA) grow from the corpus callosum to the posterior corona radiata (pointed out with a green arrow). This feature can not be found in Figure 3.21 (a) (Bayesian estimates). The corticospinal tract grows to the bottom of the brain stem (pointed out with a pink arrow) with WPGA method. The WGPA method seems preferable for fibre tractography in this example.

3.5 Summary

In Chapter 3, we defined the Fréchet sample mean of diffusion tensor data considering the symmetric positive semi-definiteness of the diffusion tensor. Six non-Euclidean metrics (the Log-Euclidean, affine invariant Riemannian, Cholesky, root Euclidean, power Euclidean and Procrustes metrics) were proposed and tested for estimation of population mean tensor. We also developed the weighted generalised Procrustes analysis (WGPA) in which an arbitrary number of tensors can be interpolated or smoothed efficiently with the additional flexibility of controlling their individual contributions. A new anisotropy measure, Procrustes Anisotropy (PA), was defined and compared with Fractional Anisotropy (FA) and Geodesic Anisotropy (GA). We carried out simulation studies for comparing different estimations of mean tensor with different metrics. The Euclidean approach is problematic due to its swelling effect in the course of tensor interpolation. The Procrustes and Log-Euclidean methods are shown to be preferable. A simulation study showed that PA provided slightly better contrast in highly anisotropic regions that FA. We applied the WGPA for smooth-

ing and interpolation of tensor fields from a real human brain. Initial results of fibre tractography using WGPA were also discussed.

The accuracy of DTI measurements depends on the diffusion gradient direction scheme applied. In Chapter 4 we will discuss optimal designs for diffusion gradient directions.

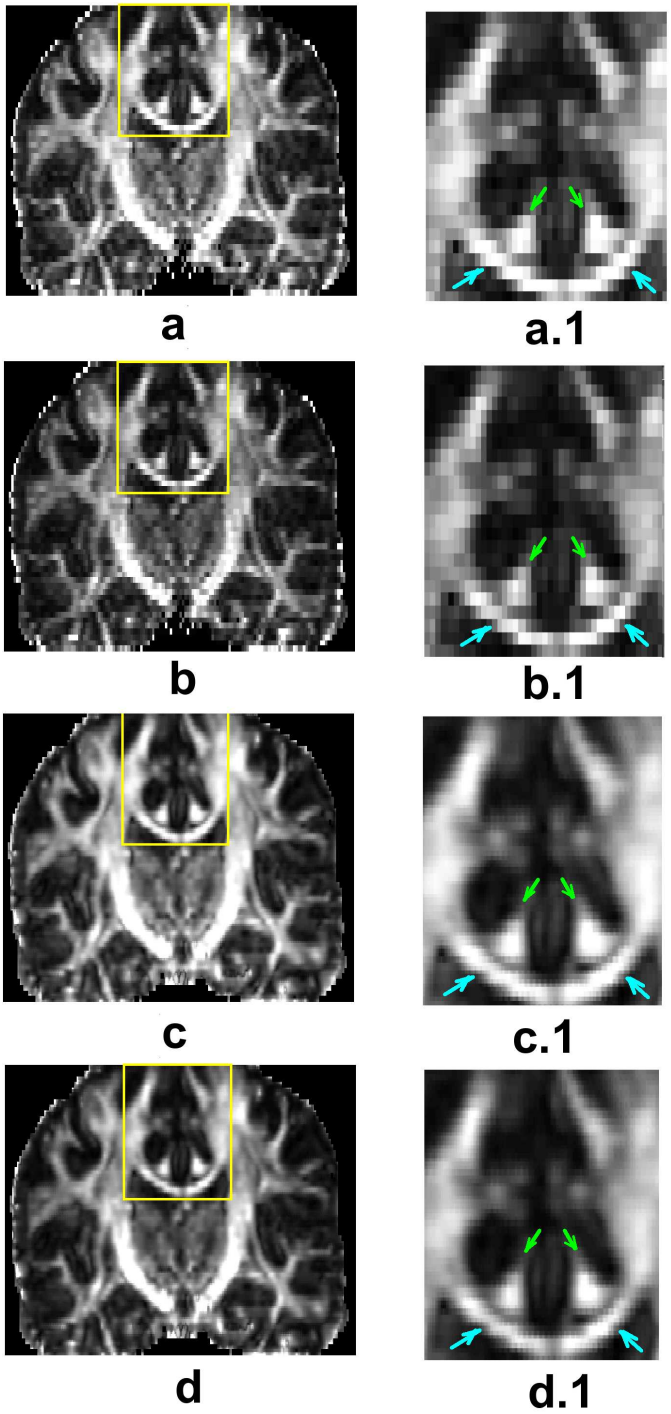

Figure 3.20: Smoothing and interpolation of the human brain diffusion tensor data. FA (a) and PA (b) maps based on Bayesian estimates without post-processing. FA (c) and PA (d) maps from smoothed and interpolated tensor data obtained with the weighted generalised Procrustes method. (a.1), (b.1), (c.1) and (d.1) are zoomed inset regions in yellow box. It clearly shows that the cingulum (pointed by green arrows) is clearly shown to be distinct from the corpus callosum (the 'U' shape tract, cyan arrows) in the anisotropy maps with smoothed and interpolated tensor estimation (c.1, d.1).

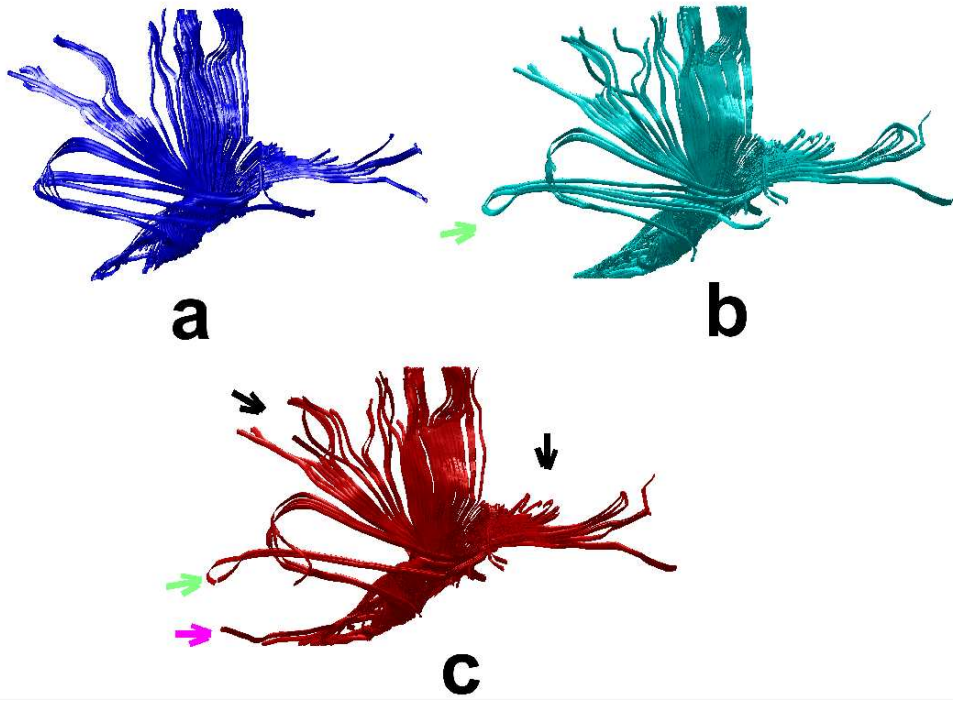

Figure 3.21: Fibre tractographies based on different estimated tensors. The Bayesian estimates (a), Euclidean smoothing (b) and WGPA smoothing (c) are used. The arrows point out some obvious differences when the WGPA tracts are compared with the other methods.

Chapter 4

Optimal Designs for Diffusion Gradient Directions

4.1 Introduction

The measured signal intensity of a DTI image is obtained by applying diffusion gradients along diffusion gradient directions. The accuracy of DTI measurements depends on the diffusion gradient direction scheme applied. In this chapter, we will focus on designing diffusion gradient directions which are spatially uniform using directional statistics as mentioned in Section 1.5.

Recall that by a gradient direction scheme we mean a set of three dimensional vectors $\mathbf{g}_i$, $i = 1, 2, ..., N$ (see Section 1.1.3). The zero vector is special and corresponds to no diffusion weighting, and most schemes contain at least one such setting. Furthermore, we discuss only spherical schemes i.e. in which the magnitude of all non-zero vectors is constant and is hence conveniently scaled to unity. The "discarded" magnitude known as the "b - value" is set separately and characterises the strength of the diffusion gradient. Below, we focus on non-zero vectors $\mathbf{g}_i$, in effect assuming that the baseline MRI signal corresponding to the zero gradient, is known precisely. It is further assumed that directions $\mathbf{g}$ and $-\mathbf{g}$ produce the same signal (in the absence of measurement noise). Thus, let $\mathbf{g}_i$ be the ith diffusion gradient direction, where $\mathbf{g}_i \in \mathbb{R}P^2$, $i = 1, ..., N$, where $\mathbb{R}P^k$ is the real-projective space of axial directions ($\mathbf{g}_i \equiv -\mathbf{g}_i$,

and $\| \mathbf{g}_i \| = 1$). A minimum of six diffusion gradient directions are necessary to apply in a DTI experiment in order to obtain unambiguous estimates of the diffusion tensor (mentioned in Chapter 2).

4.2 Uniform Schemes

It is commonly assumed that the optimum scheme will have uniformly distributed gradient directions $\mathbf{g}_1$, $\mathbf{g}_2$,..., $\mathbf{g}_N$ over a sphere. These diffusion gradient direction schemes involve approximately equal noise levels in any direction. However, there appear to be several distinct notions of uniformity in this context. Most currently used diffusion gradient direction schemes can be divided into three categories: heuristic, numerical, and polyhedral.

Heuristic schemes employ a base set of directions corresponding to the faces of a cube [Basser and Pierpaoli, 2005]. The cube with an edge length = 2, originating at $(0, 0, 0)$ defines 13 non-collinear directions at the face centres, edge bisectors, and body diagonal directions. These base sets are listed in Table 4.1 and shown in Figure 4.1).

Table 4.1: Base sets of heuristic schemes.

Name	Directions
B_0	$[1,0,0]; [0,1,0]; [0,0,1]$
B_1	$[0,1/\sqrt{2},1/\sqrt{2}]; [1/\sqrt{2},0,1/\sqrt{2}]; [1/\sqrt{2},1/\sqrt{2},0]$
B_2	$[0,-1/\sqrt{2},1/\sqrt{2}]; [1/\sqrt{2},0,-1/\sqrt{2}]; [1/\sqrt{2},-1/\sqrt{2},0]$
B_3	$[1/\sqrt{3},1/\sqrt{3},1/\sqrt{3}]; [1/\sqrt{3},1/\sqrt{3},-1/\sqrt{3}];$ $[-1/\sqrt{3},1/\sqrt{3},1/\sqrt{3}]; [1/\sqrt{3},-1/\sqrt{3},1/\sqrt{3}]$

Orthogonal (ORTH,$N = 6$), oblique double (ODG, $N = 6$), orthogonal/tetrahedral hybrid (S7, $N = 7$), decahedral (S10, $N = 10$), and complete heuristic (S13, $N = 13$) schemes, are five commonly used heuristic diffusion gradient direction schemes [Hasan *et al.*, 2001]. Each scheme is a combination of several heuristic bases. Combinations of bases for these five schemes are listed in Table 4.2.

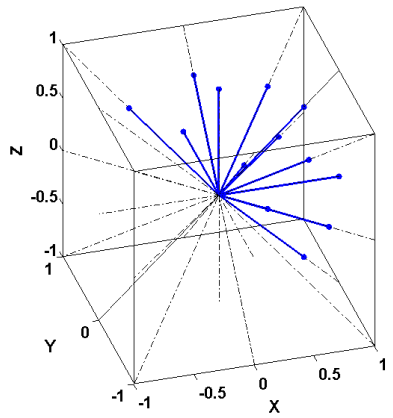

Figure 4.1: Heuristic bases in a corresponding cube.

Table 4.2: Heuristic schemes: ORTH, ODG, S7, S10 and S13.

	B_0	B_1	B_2	B_3
ORTH	✓	✓		
ODG		✓	✓	
S7	✓		✓	
S10		✓	✓	✓
S13	✓	✓	✓	✓

4.2.1 Numerically optimised schemes

Another widely used approach is to optimise the set of gradient directions for any $N \geq 6$ based on some specified criterion. These criteria are determined by DTI acquisition and processing. DTI acquisition and processing schemes can be summarised into two main steps [Basser *et al.*, 1994; Papadakis *et al.*, 1999a]: (a) First, at least six DW images are acquired by applying diffusion gradient along $N \geqslant 6$ directions; these DW images are used for estimating a diffusion tensor **D** at each voxel. (b) Second, tensor-derived quantities are calculated from **D**, such as three main diffusivities represeted by three eigenvalues of the diffusion tensor, mean value of the trace of **D**, and the anisotropy indices. Maps of these tensor-derived quantities are important visualisation results in DTI study. Papadakis *et al.* [1999a] proposed that an efficient diffusion gradient direction scheme should minimise noise propagation at each of the two steps.

Minimum total variance

To minimise the sum of all variances in the estimated tensor elements D_{ij} is one of the optimisation criteria [Papadakis *et al.*, 1999b]. The total variance is given by

$$TV = \sum_{i=1}^{3} \sum_{j=1}^{3} \mathrm{Var}(\hat{D}_{ij}). \tag{4.2.1}$$

In the same study, they hypothesised that the variance decreased as N was increased.

Minimum force

A widely used numerical approach proposed by Jones *et al.* [1999] is to use the analogy of electrostatic repulsion in chemical hybrid orbitals for selecting gradient directions. Pairs of gradient directions distributed on a unit sphere are regarded as pairs of electrons repelling each other in the valence shell of a central atom. Jones' algorithm is to arrange the directions uniformly in 3-dimensional space until the sum of the repulsive forces between every possible pair of electrons is minimised. Let $r_{ij} \geq 0$ be the distance between $\mathbf{g}_i$ and $\mathbf{g}_j$ (see Figure 4.2). Then $r'_{ij} = \sqrt{1 - r_{ij}^2}$ is the distance between $\mathbf{g}_i$ and $-\mathbf{g}_j$. According to Coulomb's law, the repulsive force between $\mathbf{g}_i$ and $\mathbf{g}_j$ is inversely proportional to r_{ij}^2. Hence, the objective to minimise the total forces is equivalent to minimising the function given by

$$J_N = \sum_{1 \leq i < j \leq N} \frac{1}{\min\{r_{ij}^2, r'^2_{ij}\}}. \tag{4.2.2}$$

Jones' J_N diffusion gradient direction schemes [Jones *et al.*, 1999; Skare *et al.*, 2000] with $N = 6, 10, 20, 30$ are listed in Appendix 5.2.6: Table A-3, A-4, A-5, A-6 respectively.

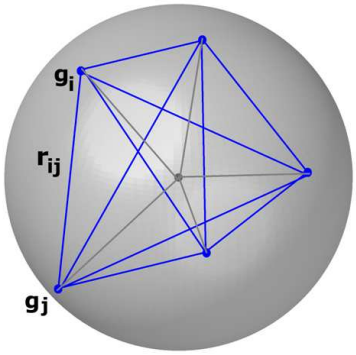

Figure 4.2: Diffusion gradient directions on a sphere. Grey lines: axial directions $\mathbf{g}_i$. Blue lines: distance r_{ij} between $\mathbf{g}_i$ and $\mathbf{g}_j$.

4.2.2 Polyhedra schemes

An alternative method for designing the gradient direction scheme is to use the highest symmetric polyhedra. A series of icosahedral (ICOSA) direction schemes [Muthupallai *et al.*, 1999] have been generated using icosahedral polyhedra. Other polyhedra generated by the regular icosahedra such as the dodecahedron with pentagonal faces [Hasan *et al.*, 2001] can also be used to obtain diffusion direction schemes.

4.3 Tests of Uniformity

In this section, two test statistics of uniformity are proposed for designing diffusion gradient direction schemes.

4.3.1 Bingham test

The Bingham test is the simplest test of uniformity of axial data [Mardia and Jupp, 2000; Bingham, 1974]. In the Bingham test, we reject uniformity of axial direc-

tions ($\mathbf{g}_i \in \mathbb{R}P^{p-1}$, $i = 1, 2, ..., N$) for a large value of

$$B = \frac{p(p+2)N}{2}\left\{\text{trace}(\bar{\mathbf{T}}^2) - \frac{1}{p}\right\} \tag{4.3.1}$$

where p is the dimension and $\bar{\mathbf{T}}$ is the scatter matrix, i.e.

$$\bar{\mathbf{T}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}. \tag{4.3.2}$$

We apply the Bingham test to design the diffusion gradient direction scheme, i.e. $p = 3$, $\mathbf{x}_i = \mathbf{g}_i$, $i = 1, 2, ..., N$. Then the objective is to minimise the Bingham test statistic given by

$$B(\mathbf{g}) = \frac{15N}{2}\left\{\text{trace}(\bar{\mathbf{T}}^2) - \frac{1}{3}\right\}. \tag{4.3.3}$$

Mardia and Jupp [2000, p.233] pointed out a disadvantage of the Bingham test that it is not consistent against all alternatives to $E[\mathbf{x}\mathbf{x}^{\mathrm{T}}] = p^{-1}\mathbf{I}_{p\times p}$ where $\mathbf{I}_{p\times p}$ is a $p \times p$ identity matrix. That means the power of the Bingham test for all alternatives to $E[\mathbf{x}\mathbf{x}^{\mathrm{T}}] = p^{-1}\mathbf{I}_{p\times p}$ does not approach 1 as the sample size becomes infinitely large.

4.3.2 Giné's G_N test

A consistent test of uniformity for axial data is the Giné's G_N test [Mardia and Jupp, 2000; Giné, 1975]. Giné's G_N test rejects uniformity for large values of

$$G_N = C_1 - C_2 \sum_{1\le i<j\le N} \sin\phi_{ij}, \tag{4.3.4}$$

where

$$C_1 = \frac{N}{2}, \tag{4.3.5}$$

$$C_2 = \frac{p-1}{2N}\left\{\frac{\Gamma((p-1)/2)}{\Gamma(p/2)}\right\}^2 \tag{4.3.6}$$

and $\phi_{ij} \in [0, 90^\circ]$ is the smaller angle between $\mathbf{g}_i$ and $\mathbf{g}_j$. In our study, $p = 3$, then

$$C_2 = \frac{1}{2N}\left\{\frac{1}{\Gamma(3/2)}\right\}^2. \tag{4.3.7}$$

We apply G_N to arrange the diffusion gradient directions, and name it the G_N scheme with N directions.

4.3.3 Objective functions of Jones' and Giné's methods

In this study, Jones' method and our Giné's method are of interest for arranging diffusion gradient directions with fixed number of directions N. In this subsection, we will summarise three expressions of the objective functions of both methods.

For simplicity, we assume that $r_{ij} = min\left\{r_{ij}, r'_{ij}\right\}$. Then the objective of Jones' method is to minimise

$$J_N = \sum_{1\le i<j\le N} \frac{1}{r_{ij}^2}, \text{ where } r_{ij} \ge 0. \tag{4.3.8}$$

Let $\phi_{ij} \in [0, 90^\circ]$ be the smaller angle between two directions $\mathbf{g}_i$ and $\mathbf{g}_j$, $1 \le i < j \le N$ (see Figure 4.3). Then function J_N can also be represented by ϕ_{ij}, $1 \le i < j \le N$, and J_N is given by

$$\begin{aligned} J_N &= \sum_{1\le i<j\le N} \frac{1}{4\sin^2(\phi_{ij}/2)} \\ &= \sum_{1\le i<j\le N} \frac{1}{2(1-\cos\phi_{ij})} \end{aligned} \tag{4.3.9}$$

It is necessary to express the function J_N with diffusion gradient directions

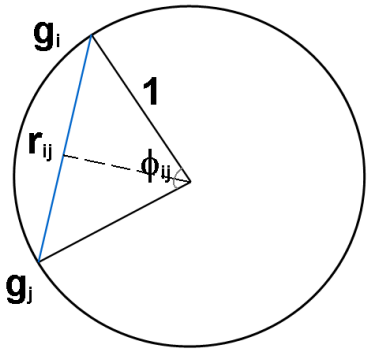

Figure 4.3: The angle between diffusion gradient directions. ϕ_{ij} is the smaller angle between two directions $\mathbf{g}_i$ and $\mathbf{g}_j$.

$\mathbf{g}_i$, $i = 1, 2, ..., N$ for computational purposes. Since

$$\cos\phi_{ij} = |\mathbf{g}_i \cdot \mathbf{g}_j| \tag{4.3.10}$$

where $a \cdot b$ is the dot product of vectors a and b with same size, we have

$$\begin{aligned} J_N &= \sum_{1\le i<j\le N} \frac{1}{2(1-\cos\phi_{ij})} \\ &= \sum_{1\le i<j\le N} \frac{1}{2(1-|\mathbf{g}_i \cdot \mathbf{g}_j|)}. \end{aligned} \tag{4.3.11}$$

For Giné's method, we can also present the objective function G_N with r_{ij}. For a fixed N, the objective of Giné's method is to minimise

$$G_N = C_1 - C_2 \sum_{1\le i<j\le N} \sin\phi_{ij}. \tag{4.3.12}$$

Since

$$\sin\phi_{ij} = 2\sin\frac{\phi_{ij}}{2}\cos\frac{\phi_{ij}}{2}, \tag{4.3.13}$$

and

$$\sin\frac{\phi_{ij}}{2} = \frac{r_{ij}}{2}, \text{ and } \cos\frac{\phi_{ij}}{2} = \sqrt{1-\frac{r_{ij}^2}{4}}, \tag{4.3.14}$$

we obtain

$$G_N = C_1 - C_2 \sum_{1 \le i < j \le N} \frac{r_{ij}}{2} \sqrt{4 - r_{ij}^2}. \tag{4.3.15}$$

Finally, since

$$\sin \phi_{ij} = |\mathbf{g}_i \times \mathbf{g}_j| \tag{4.3.16}$$

where $|a \times b|$ is the magnitude of the cross product of vectors a and b with same size, the function G_N can be represented by the directions $\mathbf{g}_i$ $i = 1, 2, ..., N$ as follows

$$G_N = C_1 - C_2 \sum_{1 \le i < j \le N} |\mathbf{g}_i \times \mathbf{g}_j|. \tag{4.3.17}$$

We summarise all of the above expressions of the objective functions J_N and G_N in Table 4.3. Note that for both Jones' J_N and Giné's G_N arrangements, the

Table 4.3: Objective functions of Jones' and Giné's methods.

	Jones' J_N	Giné's G_N
Function 1	$\min \sum_{i \neq j} \frac{1}{4 \sin^2(\phi_{ij}/2)}$	$\min C_1 - C_2 \sum_{i \neq j} \sin \phi_{ij}$
Function 2	$\min \sum_{i \neq j} \frac{1}{r_{ij}^2}$	$\min C_1 - C_2 \sum_{i \neq j} \frac{r_{ij}}{2} \sqrt{4 - r_{ij}^2}$
Function 3	$\min \sum_{i \neq j} \frac{1}{2(1 - \lvert \mathbf{g}_i \cdot \mathbf{g}_j \rvert)}$	$\min C_1 - C_2 \sum_{i \neq j} \lvert \mathbf{g}_i \times \mathbf{g}_j \rvert$

solution to the optimisation problem in practice is not unique. For example, any simultaneous rotations of $\mathbf{g}_1$, ..., $\mathbf{g}_N$, i.e. $\mathbf{g}_1\mathbf{R}$, ..., $\mathbf{g}_N\mathbf{R}$ where $\mathbf{R} \in O(3)$ can minimise J_N and G_N.

4.4 Results

4.4.1 Giné's G_N and Jones' J_N schemes

We obtain Giné's G_N schemes ($N = 6, 10, 20$ and 30) by minimising Giné's G_N statistic. The optimisation function $fminsearch()$ which adapts the Nelder-

Mead simplex method [Lagarias *et al.*, 1998] in MATLAB (R2009a, The MathWorks, Inc., Natick, Massachusetts, USA) is used to minimise Giné's objective functions. Jones' diffusion gradient directions are used as the starting values for minimising Giné's G_N functions. We list optimised Giné's G_N schemes in Appendix 5.2.6: Tables A-3, A-4, A-5, A-6 respectively. Giné's G_N and Jones' J_N schemes are plotted in Figure 4.4. It is clear that when $N = 6$ and 10, the G_N- and J_N-based schemes are similar. For $N = 20$ and 30, G_N and J_N schemes become more different.

We substitute Giné's and Jones' schemes into both the objective functions $G_N()$ and $J_N()$. The comparison of these objective function values is made in Table 4.4. It is noticeable that Giné's G_N schemes provide very slightly smaller values for both the objective functions $G_N()$ and $J_N()$ than Jones' J_N schemes, which is surprising. This may indicate that our numerical optimisation was carried out to higher precision than Jones' schemes.

Table 4.4: Giné's and Jones' objective function values.

	$G_N()$ value	$J_N()$ value
G_6 scheme	**0.1529**	0.4523
J_6 scheme	0.1530	0.4523
G_{10} scheme	**0.1182**	**0.5452**
J_{10} scheme	0.1183	0.5459
G_{20} scheme	**0.0830**	**0.7049**
J_{20} scheme	0.0842	0.7076
G_{30} scheme	**0.0669**	**0.7961**
J_{30} scheme	0.0690	0.8012

4.4.2 Simulation study

Monte Carlo simulations are carried out using Matlab to evaluate the effect of G_N schemes on tensor estimation and tensor-derived measures. First, synthetic diffusion tensors $\mathbf{D}_1{}^*$, $\mathbf{D}_2{}^*$, ..., $\mathbf{D}_K{}^*$ are generated by rotating the principal

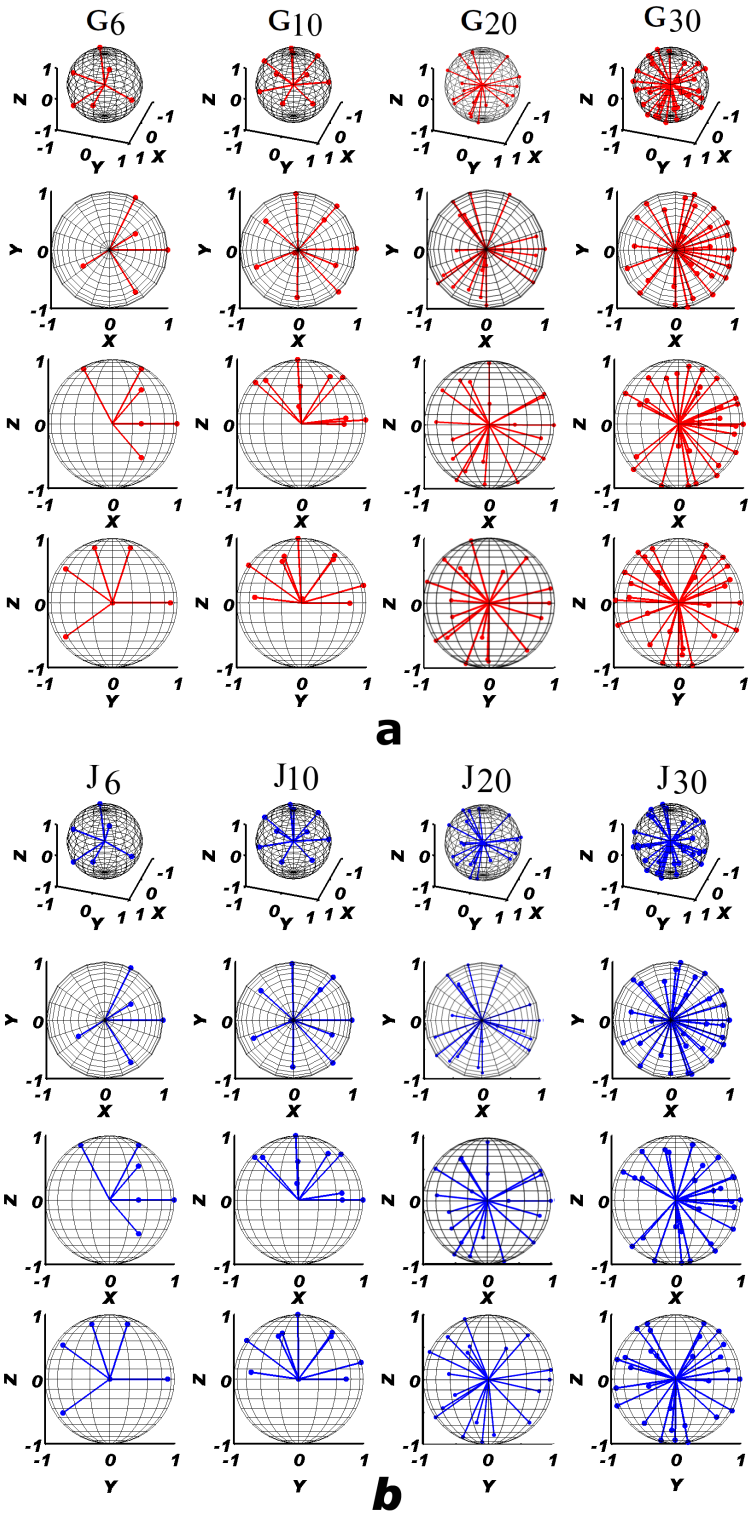

Figure 4.4: Giné's G_N and Jones' J_N diffusion gradient direction schemes. Giné's G_N and Jones' optimised directions are plotted in a and b. Their 3D, x-y, x-z and y-z projection views are in the first, second, third and fourth row, respectively in each figure. The same viewing angles are used for both schemes.

eigenvector $\mathbf{v}_1$ on a hemisphere (with radius=1). Let $(\theta, \psi, 1)$ be the spherical coordinates of $\mathbf{v}_1$, where θ is the azimuthal angle in the x-y plane, ψ is the polar angle from the positive z semi-axis, and 1 is the radius. We rotate $\mathbf{v}_1$ by changing values of θ and ψ. Let θ increase from -180° to 180° with step size 18°. For a fixed θ, ψ increases from 0 to 90° with step size 18°. The settings of $\mathbf{v}_1$ are shown in Figure 4.5. Since the three eigenvectors of $\mathbf{D}^*$ are orthogonal, we take $\mathbf{v}_2 = (1, \theta + 90^\circ, \psi)$, and $\mathbf{v}_3 = \mathbf{v}_1 \times \mathbf{v}_2$. Let $\mathbf{D}_1{}^*$, $\mathbf{D}_2{}^*$, ..., $\mathbf{D}_K{}^*$ ($K = 126$) have the same eigenvalues $(\lambda_1, \lambda_2, \lambda_3) = (12, 2, 1)$, then they have same mean diffusivity $MD = 5$ and Fractional Anisotropy $FA = 0.8631$.

We repeat $M = 500$ Monte Carlo simulations for each synthetic diffusion tensor $\mathbf{D}_j{}^*, j = 1, 2, ..., K$ with each Giné's G_N direction scheme $\mathbf{g}_i = (g_{ix}, g_{iy}, g_{iz})$, N =6, 10, 20, 30. For a fixed N and a specified synthetic diffusion tensor $\mathbf{D}_j{}^*$, the Monte Carlo simulations are carried out as follows:

1. Calculate the noise-free signals $\boldsymbol{\mu} = (\mu_1, \mu_2, ..., \mu_N)$ from the single tensor model

$$\mu_i = S_0 \exp(-b\mathbf{g_i}^{\mathrm{T}}\mathbf{D}_j{}^*\mathbf{g_i}), \text{where } i = 1, 2, ..., N, j = 1, 2, ..., K. \qquad (4.4.1)$$

2. Simulate the measured signals $\mathbf{S} = (S_1, S_2, ..., S_N)$ by adding independent Gaussian random noise with zero mean and a fixed variance $\delta^2 = 5$ to each of the N noise-free signals.

3. Estimate the diffusion tensor. In this study, we use our Bayesian single tensor estimation method discussed in Section 2.2.3.

4. Calculate various summary statistics.

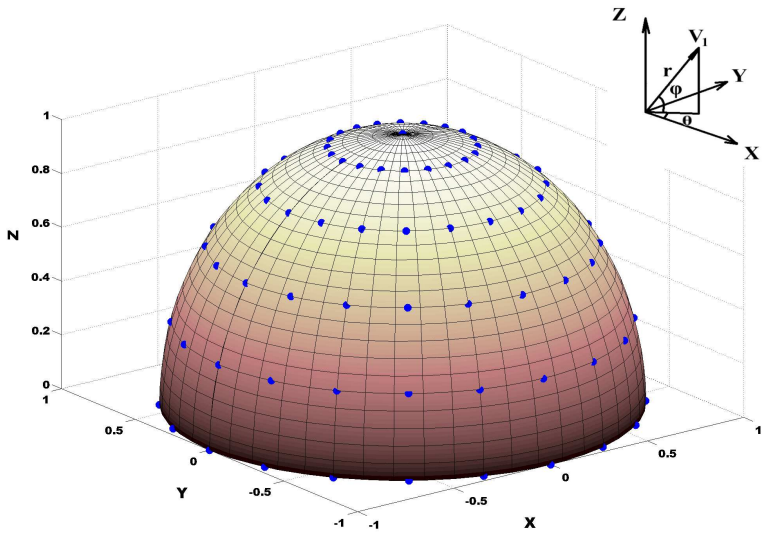

Figure 4.5: Each of the blue point on the hemisphere represents the principal eigenvector of $\mathbf{D}_i{}^*$, for some $i = 1, 2, ..., K$.

4.4.2.1 Effects of varying scheme on tensor estimations

The aim of this simulation study is to compare effects of Giné's G_N and Jones' J_N schemes on tensor estimation. We consider the following four measures of estimation error: The root mean squared error (RMSE($\hat{\mathbf{D}}$)), the 95th percentile of the angular dispersion (95% AD), standard deviation std(FA) and standard deviation std(MD). The RMSE($\hat{\mathbf{D}}$) measures the Euclidean distance between the estimated and true diffusion tensors and is given by

$$RMSE(\hat{\mathbf{D}}) = \sqrt{\frac{1}{M}\sum_{k=1}^{M} \parallel \hat{\mathbf{D}}(k) - \mathbf{D} \parallel^2}. \tag{4.4.2}$$

where $\hat{\mathbf{D}}(k)$ is the kth Monte Carlo estimate of $\mathbf{D}$. Note we are using the Euclidean distance between tensors here. The 95% AD of the estimated principal eigenvector $\hat{\mathbf{v}}_1$ from $\mathbf{v}_1$ is also used in Jones [2003] to determine the respective confidence cone for $\mathbf{v}_1$. The angular dispersion is given by

$$\gamma = \text{acos}(\hat{\mathbf{v}}_1 \cdot \mathbf{v}_1). \tag{4.4.3}$$

Figure 4.6 shows the surfaces of RMSE($\hat{\mathbf{D}}$) from J_N and G_N schemes with N =6, 1, 20 and 30. It is clear that overall $RMSE(\hat{\mathbf{D}})$ values decrease as N increases. Surfaces of $RMSE(\hat{\mathbf{D}})$ from the G_6 and J_6 schemes are very similar. For higher N (N = 10, 20 and 30), G_N and J_N provide slightly different RMSE($\hat{\mathbf{D}}$) surfaces. The same pattern can be shown in the top-down $(\theta - \phi)$ views of RMSE($\hat{\mathbf{D}}$) surfaces (see Figure 4.7).

In Figure 4.8 and Figure 4.9, both G_6 and J_6 schemes yield large variations of the 95% AD from approximately 5° to 20°. As N increases, the variations in orientational uncertainty decrease. There is a marked reduction (approximately 10° reduced) in the mean CI when N increases from 6 to 10. G_{20} and G_{30} schemes provide further improvement in reducing the angular dispersions of principal eigenvectors estimations. This improvement can also be found in reducing the variance of FA and MD in Figure 4.10 and 4.12. The top-down $(\theta - \phi)$ views of $std(FA)$ and $std(MD)$ surfaces are shown in Figure 4.11 and Figure 4.13. The surfaces of FA and MD are shown in Appendix 5.2.6.

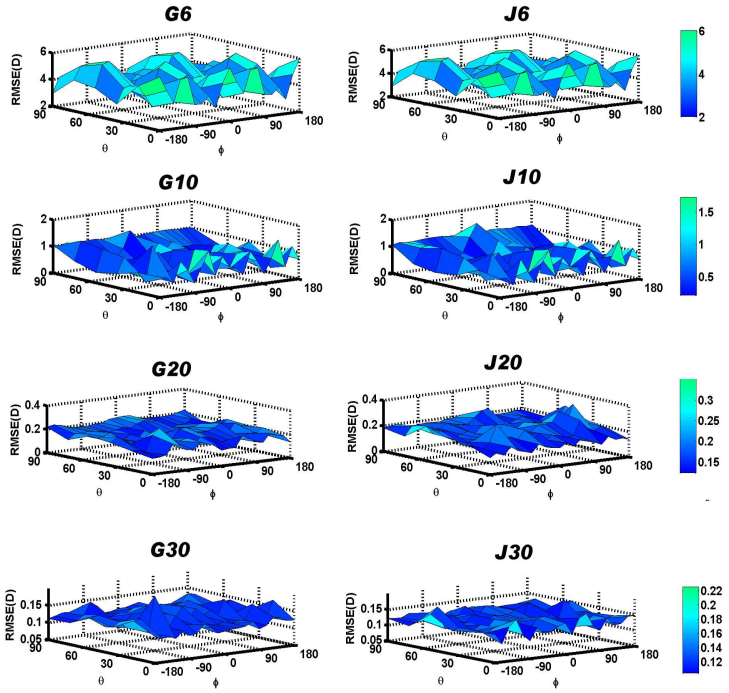

Figure 4.6: The surfaces of RMSE($\hat{\mathbf{D}}$) for G_N and J_N schemes. $N = 6, 10, 20, 30$.

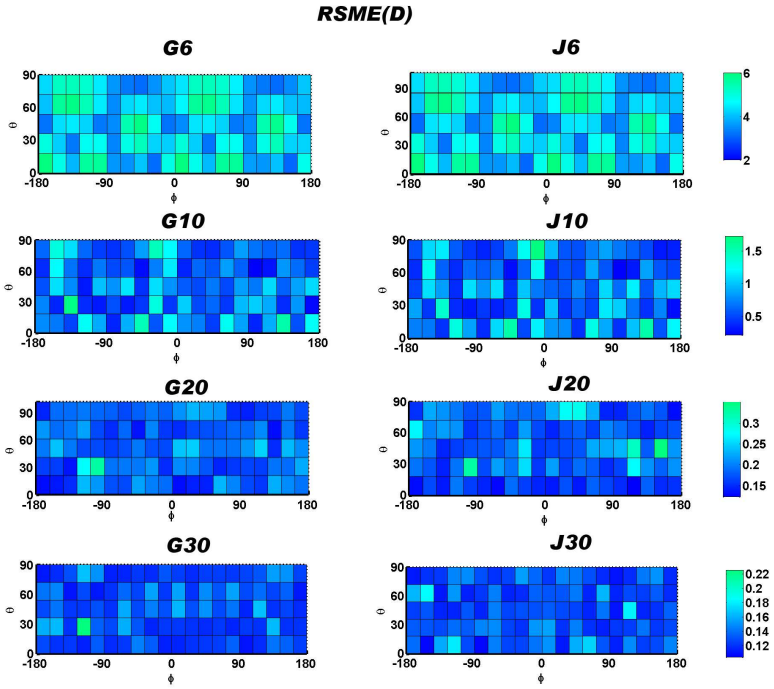

Figure 4.7: The top-down $(\theta - \phi)$ views of RMSE($\hat{\mathbf{D}}$) surfaces for G_N and J_N schemes. $N = 6, 10, 20, 30$.

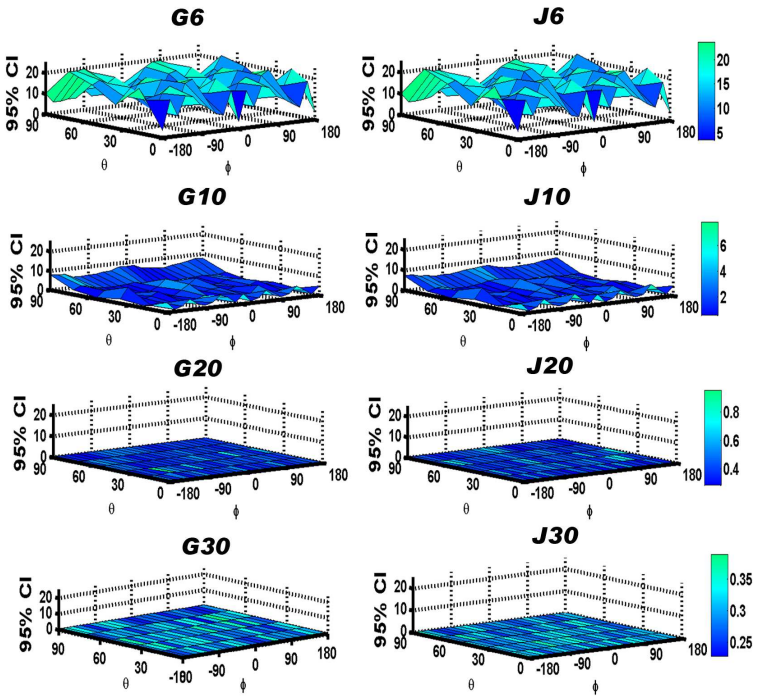

Figure 4.8: Surfaces of 95% AD for J_N and G_N schemes. $N = 6, 10, 20, 30$.

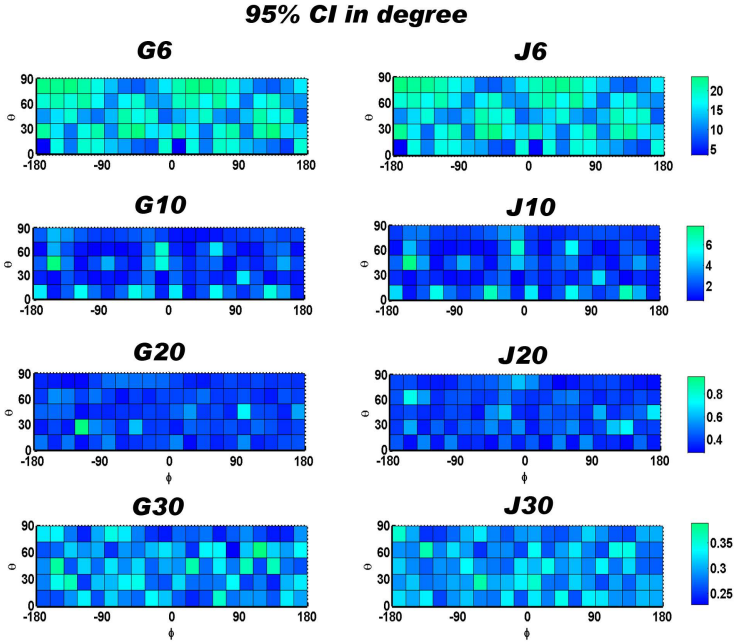

Figure 4.9: The top-down $(\theta - \phi)$ views of 95% AD surfaces for G_N and J_N schemes. $N = 6, 10, 20, 30$.

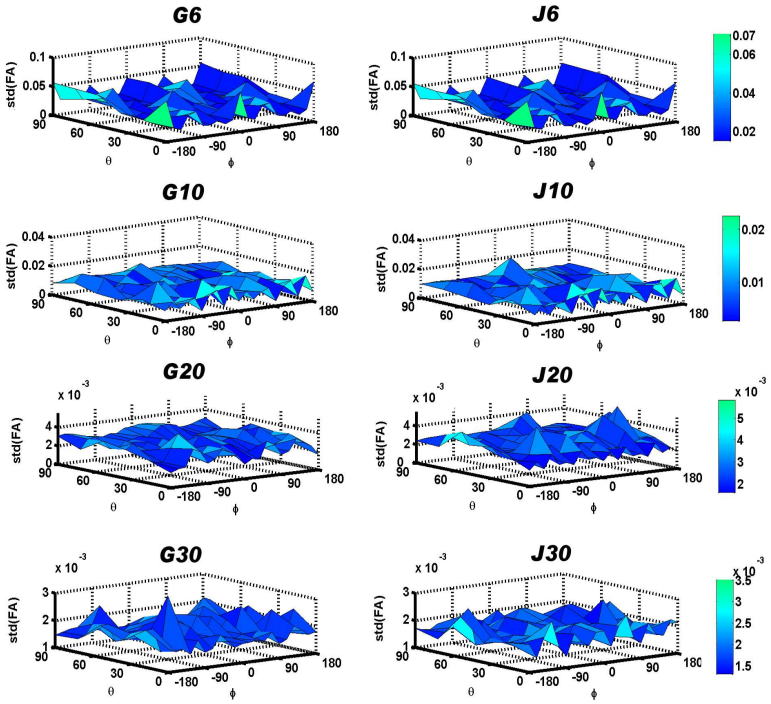

Figure 4.10: The surfaces of std(FA) for G_N and J_N schemes. $N = 6, 10, 20, 30$.

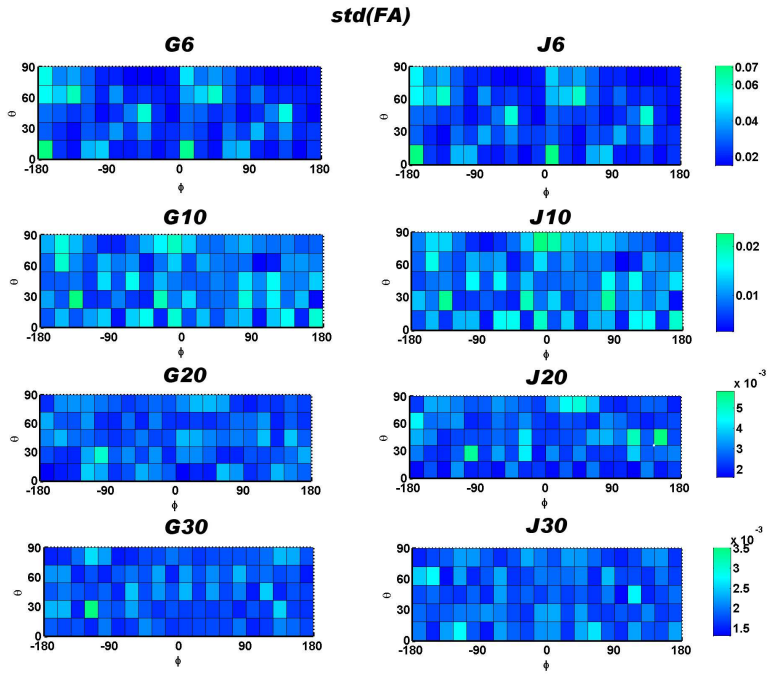

Figure 4.11: The top-down $(\theta - \phi)$ views of $std(FA)$ surface for G_N and J_N schemes,.$N = 6, 10, 20, 30$.

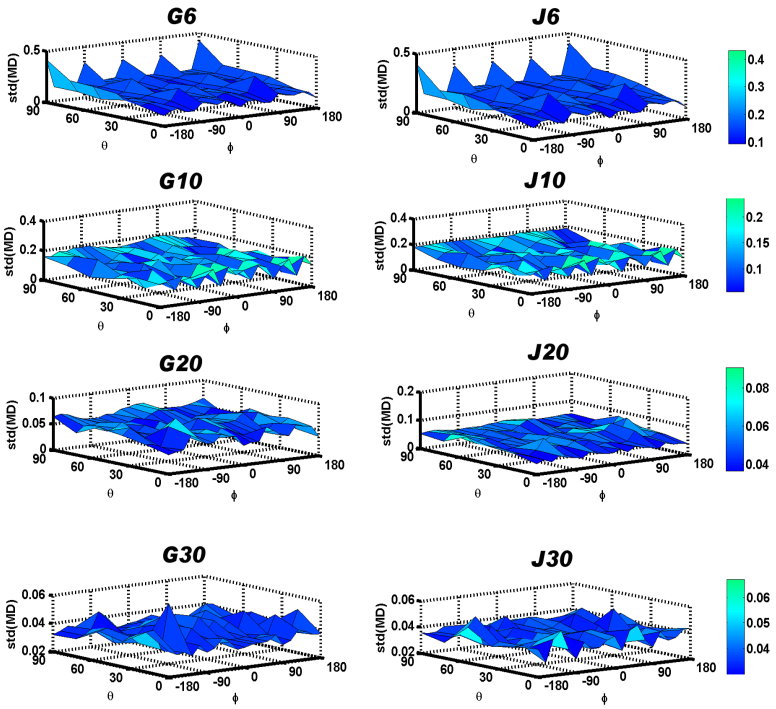

Figure 4.12: The surfaces of std(MD) for G_N and J_N schemes. $N = 6, 10, 20, 30$.

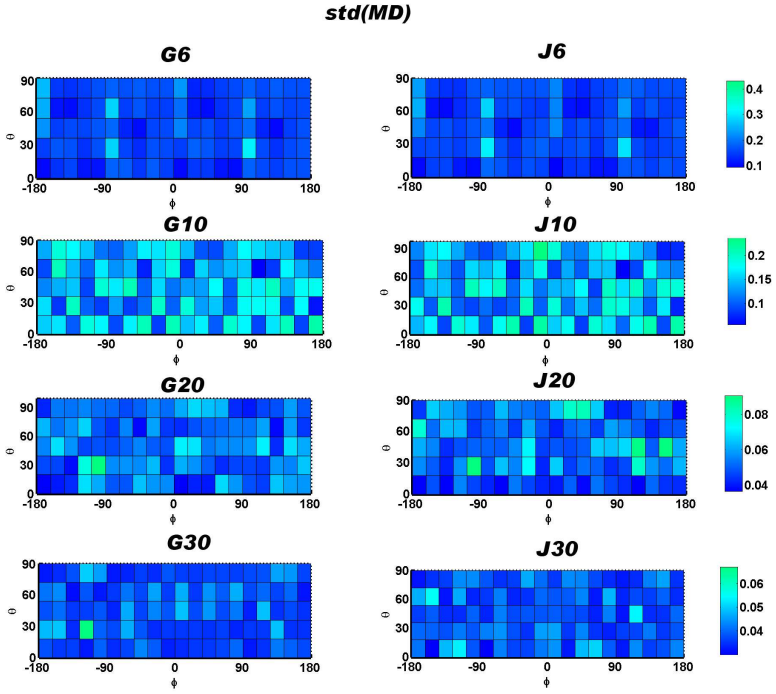

Figure 4.13: The top-down $(\theta - \phi)$ views of $std(MD)$ surfaces for G_N and J_N schemes. $N = 6, 10, 20, 30$.

4.4.2.2 Overall measures of schemes

To easily understand the performance of each direction scheme, we derive 5 overall measures based on the std(FA), std(MD) and RMSE$(\hat{\mathbf{D}})$. They are given by

$$\begin{aligned}
\text{Maximum of } s_k \quad \Psi_1 &= \max_{\theta,\phi} s_k(\theta,\phi), \\
\text{Sum of } s_k \quad \Psi_2 &= \int\int s_k(\theta,\phi) d\theta d\phi, \\
\text{Marginal sum of } s_k \text{ given } \phi = \pi/2 \quad \Psi_3 &= \int s_k(\theta, \phi = \pi/2) d\theta, \\
\text{Sum of} s_k \cos\phi \quad \Psi_4 &= \int\int s_k(\theta,\phi)\cos\phi d\theta d\phi, \\
\text{Standard deviation of } s_k \quad \Psi_5 &= \text{std}(s_k(\theta,\phi)),
\end{aligned}$$

where $k = 1, 2, 3$ and

$$\begin{aligned}
s_1(\theta,\phi) &= \text{std}(FA), \\
s_2(\theta,\phi) &= \text{std}(MD), \\
s_3(\theta,\phi) &= \text{RMSE}(\hat{\mathbf{D}}).
\end{aligned}$$

We summarise the overall measures in Table 4.5, 4.6, 4.7. For both G_6 and J_6 schemes, the maximum stdFA and maximum RMSE$(\hat{\mathbf{D}})$ i.e. $\Psi_1(\text{std}FA)$ were achieved for a fibre with $(\theta = 0, \phi = -180^\circ)$ in Table 4.5 and Table 4.7. The maximum stdMD obtained with G_6 and J_6 schemes was achieved for a fibre with $(\theta = 90^\circ, \phi = 180^\circ)$ and $(\theta = 90^\circ, \phi = -180^\circ)$ respectively in Table 4.6. For $N = 6$, Giné's G_N and Jones' J_N schemes have similar effects on tensor estimation.

We find that Giné's G_{10} scheme is slightly better than Jones' J_{10} scheme through all measures based on std(FA), std(MD) and RMSE($\hat{\mathbf{D}}$). However, there is not a large difference between Giné's and Jones' schemes with higher numbers of directions ($N = 10$). It is worth noting two points: (1) the J_{20} scheme performs better than G_{20} for Ψ_2, Ψ_3 and Ψ_4 through all measures, and (2) the J_{30} scheme performs better than G_{30} for Ψ_1 and Ψ_5 through all measures. We run the computation several times, and obtain similar results.

Table 4.5: Overall measures on $s_1(\theta, \phi) = \text{std}(FA)$.

schemes	$\Psi_1, (\theta^*, \phi^*)$	Ψ_2	Ψ_3	Ψ_4	Ψ_5
G_6	0.0705,(0, -180°)	**3.6285**	0.6583	**2.2297**	0.0131
J_6	**0.0703,(0, -180°)**	3.6495	**0.6568**	2.2393	**0.0130**
G_{10}	**0.0225, (18°,-144°)**	**1.2349**	**0.2200**	**0.7707**	**0.0038**
J_{10}	0.0250,(72°, -18°)	1.2589	0.2234	0.7798	0.0041
G_{20}	**0.0054,(18°, -108°)**	0.3508	0.05436	0.2075	**0.0006**
J_{20}	0.0061,(36°,144°)	**0.3376**	**0.0462**	**0.2071**	0.0008
G_{30}	0.0035,(18°,-126°)	**0.2276**	**0.0372**	**0.1436**	0.0003
J_{30}	**0.0028 ,(0,-126°)**	0.2326	0.0417	0.1457	0.0003

Table 4.6: Overall measures on $s_2(\theta, \phi) = \text{std}(MD)$.

schemes	$\Psi_1, (\theta^*, \phi^*)$	Ψ_2	Ψ_3	Ψ_4	Ψ_5
G_6	0.4251, (90°,180°)	**21.2961**	**2.9868**	**12.2927**	**0.0545**
J_6	**0.4181,(90°,-180°)**	21.4583	2.9895	12.3886	0.0551
G_{10}	**0.2131, (0,162°)**	**18.4709**	**3.2312**	**11.0011**	**0.0384**
J_{10}	0.2341,(72°, -18°)	18.8294	3.2741	11.1060	0.0417
G_{20}	**0.0969,(18°,-108°)**	7.0116	1.0881	4.1277	**0.01085**
J_{20}	0.1061,(36°,144°)	**6.7887**	**0.9349**	**4.0938**	0.0115
G_{30}	0.0669,(18°,-126°)	**4.7677**	**0.7746**	**2.9818**	0.0054
J_{30}	**0.0549,(54°,-162°)**	4.8556	0.8536	3.0212	**0.0049**

For future work it would be interesting to see how the Jones statistic performs in a test for uniformity compared to Giné's test.

Table 4.7: Overall measures on $s_3(\theta, \phi) = RMSE(\hat{\mathbf{D}})$.

schemes	$\Psi_1, (\theta^*, \phi^*)$	Ψ_2	Ψ_3	Ψ_4	Ψ_5
G_6	5.9950, (0,-180°)	**541.7997**	**96.7310**	340.8865	**0.8261**
J_6	**5.9890, (0,-180°)**	541.9041	96.7483	**340.8255**	0.8322
G_{10}	**1.6785,(18°,-144°)**	**96.3082**	**17.9253**	**56.0071**	**0.3140**
J_{10}	1.7197,(72°, -18°)	98.8145	18.2655	56.5059	0.3343
G_{20}	**0.3382,(18°,-108°)**	24.2810	3.7647	14.2870	**0.0376**
J_{20}	0.3738,(36°,144°)	**23.5110**	**3.2362**	**14.2200**	0.0407
G_{30}	0.2252,(18°,-126°)	**16.1889**	**2.6222**	**10.1012**	0.0173
J_{30}	**0.1836,(54°, -162°)**	16.4597	2.8890	10.2283	**0.0157**

4.5 Summary

In this chapter, we designed a series of new uniformly distributed direction schemes with the number of directions $N = 6, 10, 20$ and 30 using Giné's statistic. A Monte Carlo simulation study has been carried out to compare effects of the Giné's and widely used Jones' schemes on tensor estimation. It is clear that all surfaces of measures reduce as N increases. Giné's and Jones' schemes turned out to be quite similar in terms of performance. Five overall measures based on the std(FA), std(MD) and RMSE($\hat{\mathbf{D}}$) have been proposed for the evaluation of the diffusion gradient direction scheme.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

In this thesis, we have developed some statistical methodologies for diffusion tensor imaging and covariance matrix analysis. In Chapter 2, we have developed the multi-tensor model for diffusion MRI measurements of water diffusion at voxels with one or more distinct fibre orientations. Our model substantially alleviates the non-identifiability issue present in the standard multi-tensor model. A new parametrisation was proposed to ensure the symmetry and positive semi-definiteness of the diffusion tensor in the multi-tensor model. A Bayesian framework has been established in Section 2.2.3 to estimate diffusion tensors under the single tensor model ($m = 1$ in the multi-tensor model) and the double tensor model ($m = 2$) assumptions. Inference has been drawn from the posterior distribution. Also, the Bayes factor was adapted for model selection, i.e., for deciding between the single and double tensor models. We developed an MCMC algorithm, which combines the Metropolis-Hasting algorithm and Gibbs sampler in Section 2.3, to study the uncertainty of the model parameters (i.e. the diffusion tensor and the variance of the model noise) based on the posterior distribution of these parameters. We defined the mean fibre orientation, or direction, to be the extrinsic mean identifying directions with axes. We then sampled fibre orientations from the posterior distribution and estimated the true diffusion direction by the sample mean. We also provided the 95%

credible cone around the mean direction by truncating at the 95th percentile of the Euclidean distances from the sampled directions to the their mean. The comparisons of the Bayesian estimation and the LLS method for both the single tensor and double tensor models were made using three simulated datasets with Phillips 15, Phillips 32, and the Uniform 32 diffusion direction schemes. It was established that the Bayesian approach performed better than the LLS method, and the Uniform 32 direction scheme was the best among the three schemes. For real data, we applied Bayesian estimation using the single tensor model. The anisotropy maps (FA maps), fibre orientation maps (the line and colour coded orientation maps) and the map of diffusion ellipsoids were obtained for the corpus callosum region. The multi-tensor model with automatic model selection has also been applied to a region containing crossing fibre bundles from a healthy human brain dataset. Fibre tractographies of the corpus callosum were obtained based on the Bayesian, LLS and the mean MCMC sample mean estimators. Uncertainty in fibre tractography was studied using an MCMC method and results were also discussed.

In Chapter 3, we defined the sample mean of diffusion tensor data to be the Fréchet mean relative to the symmetric positive semi-definite matrix space and empirical distribution of the sample. Non-Euclidean metrics, namely, the Log-Euclidean, affine invariant Riemannian, Cholesky, root Euclidean, power Euclidean and Procrustes metrics were proposed and tested for estimation of population mean tensor. The notion of geodesic has been used to interpolate two diffusion tensors and two pairs of tensors. We also developed the weighted generalised Procrustes analysis (WGPA) in which an arbitrary number of tensors can be interpolated or smoothed efficiently with the additional flexibility of controlling their individual contributions. An algorithm for computing the weighted mean of tensors was also proposed. A new anisotropy

measure, Procrustes Anisotropy (PA), was defined and compared with Fractional Anisotropy (FA) and Geodesic Anisotropy (GA). Simulation studies were carried out to compare different estimations of mean tensor with different metrics. The Euclidean approach is problematic due to its swelling effect in the course of tensor interpolation. From a variety of experiments, the Procrustes and Log-Euclidean methods are shown to be preferable. A simulation study was performed for comparing PA, FA and GA. Evidently, PA provides slightly better contrast in highly anisotropic regions, such as the corpus callosum in our examples. We applied the WGPA for smoothing and interpolation of tensor fields from a real human brain. From the results, FA and PA maps from the processed tensor data were much smoother than the one without processing. The feature that the cingulum (cg) is distinct from the corpus callosum (cc) was clearer in the anisotropy maps from the processed data than those without processing. Initial results of fibre tractography using WGPA were also discussed.

The accuracy of DTI measurements depends on the diffusion gradient direction scheme applied. In Chapter 4, we reviewed some widely used schemes such as heuristic and numerical approaches. Some new uniformly distributed direction schemes with the number of directions $N = 6, 10, 20$ and 30 have been designed using Giné's statistic. A Monte Carlo simulation study has been carried out to compare effects of the Giné's and Jones' schemes on tensor estimation. Surfaces of RMSE($\hat{\mathbf{D}}$), 95% CI, std(FA) and std(MD) were obtained for the comparison. It is clear that all surfaces of measures reduce as N increases. Giné's and Jones' schemes turned out to be quite similar in terms of performance. Five overall measures based on the std(FA), std(MD) and RMSE($\hat{\mathbf{D}}$) have been proposed for the evaluation of the diffusion gradient direction scheme.

The main contributions of this work have appeared in several publications,

notably, Dryden *et al*. [2009a], Zhou *et al*. [2008] and Zhou *et al*. [2009a]. Moreover, there are various issues that require further study in this fascinating topic of study.

5.2 Future Work

5.2.1 Multiple tensor model and model selection

In DTI study, how to model the diffusion behaviour at the region containing more than one distinct fibre orientation is still an open question. In this thesis, we have developed the Bayesian multi-tensor model (in Chapter 2) to capture the multiple diffusion properties at the given voxels. We applied this Bayesian multi-tensor model into a real data from the human brain, and the initial result of multi-tensor estimation in a region contain crossing fibre bundles has been obtained (see Section 2.4.2.3). More recent work has involved estimating the orientation distribution function nonparametrically, although the disadvantage is that many more diffusion gradient directions are needed [see Tuch, 2004]. The Bayes factor B has been adapted for determining the number of fibre orientations at a voxel. Other measures could also be used for model selection, e.g. FA and σ^2. It is noted that to manually set the thresholds for B, FA and σ^2 was not simple. To develop an approach for auto model selection could be one of our future work.

There is an increasing trend to study multi-modal structures in white matter using high-angular-resolution diffusion imaging (HARDI) [e.g. Tuch *et al*., 1999]. The q-ball method [Tuch, 2004], persistent angular structure method (PAS) [Jansons and Alexander, 2003] and multiple compartment model [Alexander, 2005] are three main implementations of the HARDI approach to capture the

multi-modal nature of diffusion at the voxel level. Olhede and Whitcher [2008] developed a statistical framework using a q-space density estimator based on a wavelet lifting scheme for HARDI data to characterise the diffusion non-parametrically in terms of its symmetry and directionality. For future work it would be interesting to use our Bayesian multi-tensor model for HARDI data analysis, and to compare our results with others.

5.2.2 Power Euclidean approach

In Chapter 3, we have discussed the power Euclidean metric given in Equation (3.2.13). Recall that the power Euclidean metric is given by

$$d_A(D_1, D_2) = \frac{1}{a} \parallel \mathbf{D}_1^a - \mathbf{D}_2^a \parallel, \tag{5.2.1}$$

where $\mathbf{D}^a = \mathbf{E}\mathbf{\Lambda}^a\mathbf{E}^{\mathrm{T}}$. The power a is nonzero, and $a \in \mathbb{R}$. We have considered $a \in \{1/2, 1\}$ earlier. Additional flexibility can be derived by varying a. As $a \to 0$, the power Euclidean metric approaches the Log-Euclidean metric. The estimate of the mean tensor is given by

$$\hat{\mathbf{T}}_A = (\hat{\mathbf{\Delta}}_A)^{1/a}, \text{ where } \hat{\mathbf{\Delta}}_A = \arg\inf_{\mathbf{\Delta}} \left\{ \frac{1}{N} \sum_{i=1}^{N} \parallel \mathbf{D}_i^a - \mathbf{\Delta} \parallel^2 \right\} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{D}_i^a. \tag{5.2.2}$$

For positive a the estimators become more resistant to outliers for smaller a, and for larger a the estimators become less resistant to outliers. For negative a one is working with powers of the inverse covariance matrix. The anisotropy index based on the power Euclidean metric (in Equation (3.2.39)) is a generalisation of FA given by

$$FA(a) = \left\{ \frac{3}{2} \sum_{i=1}^{3} (\lambda_i^a - \overline{\lambda^a})^2 / \sum_{i=1}^{3} \lambda_i^{2a} \right\}^{1/2}, \tag{5.2.3}$$

where $\overline{\lambda^a} = \sum_{i=1}^{3} \lambda_i^a / 3$. A practical visualisation tool is to vary a in order for a neurologist to help interpret the white fibre tracts in the images. A question of interest for further work is estimation of the metric from the data itself, e.g., estimation of a.

5.2.3 Fibre tractography assessment

In this thesis, we have obtained the fibre tractographies from Bayesian estimators (in Chapter 2), from MCMC simulations and from processed tensor field with the Procrustes method (in Chapter 3). To develop a statistical assessment tool for fibre tractography and to create measures of fibre orientation uncertainty would be of great benefit to neuroscientists. In particular it is of great interest to consider the connectivity of the brain, and how connectivity analysis is affected by using different metrics and methods.

5.2.4 Tensor regularisation

We have used the weighted generalised Procrustes method to smooth and interpolate the tensor field. We wish to carry out regularisation. Consider a grid of tensors $\mathbf{D}_1$,...,$\mathbf{D}_N$ at voxels V_1, ..., V_N and we wish to predict the tensor at a new site V. We could use the weighted penalised predictor [Dryden *et al.*, 2009b] obtained by minimising, with respect to $\mathbf{D}$,

$$\hat{\mathbf{T}}_{\tau,\omega}(\varrho) = \sum_{i=1}^{N} w_i d(\mathbf{D}_i, \mathbf{D})^{\tau} + \varrho d(\mathbf{D}, \upsilon)^{\omega} \tag{5.2.4}$$

where the weights $w_i \geq 0$, $\sum_{i=1}^{N} w_1 = 1$ are functions of the distance from V_i to the site V, $\varrho > 0$, $\tau > 0$ and $\omega > 0$ are a regularisation parameters, and υ is a reference matrix, such as the identity matrix, zero matrix or an overall average.

For example we could use $w_i \propto \exp\left\{-\gamma \parallel V - V_i \parallel^2\right\}, i = 1, ..., n$.

Consider now smoothing across an image at the voxels $V_1, ..., V_n$, and so we need to minimise, with respect to $\mathbf{D}_j, j = 1, ..., n$,

$$\left\{ \sum_{j=1}^{n} \sum_{i=1}^{n} w_{ij} d(\mathbf{D}_i, \mathbf{D}_j)^{\tau} \right\} + \varrho \sum_{j=1}^{n} d(\mathbf{D}_j, v)^{\omega}, \tag{5.2.5}$$

and w_{ij} is the weight as a function of the distance between site i and site j. By varying (τ, ω), different approaches are obtained (see Table 5.1). Also, fur-

Table 5.1: Tensor regularisation methods

(τ, ω)	Method
(2,0)	Weighted Fréchet mean
(τ,0)	M-estimator [Dryden and Mardia, 1998, p.298]
(1,0)	Geometric median [Fletcher *et al.*, 2009]
(2,1)	Non-Euclidean type of ridge-regression
(2,2)	A non-Euclidean type of LASSO [Tibshirani, 1996]

ther tensor processing techniques could be carried out by using various non-Euclidean approaches. More work is required to understand the implication of choosing different parameters.

5.2.5 Automatic selection of weights for tensor smoothing

How to process images with some level of optimality is an important and generally difficult task in many computer vision application. For diffusion tensor smoothing, the main difficulty lies in the ability to predict how much smoothing needs to applied at various locations. So, it could be interesting to investigate algorithms which would modify the weights which determine contribution of individual voxels. We could firstly start with some large region of interest. We would then try to alternate iteratively between the two phases. First, to segment the region into more or less homogeneous subregions, which will gen-

erally be irregular. Some standard classification tool can be used, eventually trying also to assure some spatial continuity. These classification/segmentation can be "soft", i.e. we would not actually be drawing any boundaries, but rather estimating class-conditional probabilities. Then we would do smoothing separately for each of the subregions, using these probabilities as our weights. So, the voxels that have higher probability of being part of the currently chosen subregion will be contributing more. In the beginning we start by allowing our segmentation algorithm to have many distinct segments/subregions and very peaked distributions of weights within each. So, we would have very little smoothing in the beginning. Once the smoothed values are computed, we rerun the segmentation algorithm, and so on, gradually influencing the segmentation to be increasingly more conservative in terms of the number of distinct subregions. We might even allow for disconnected segments/subregions, which could be important to prevent oversmoothing of structure. So, ideally we would like to be able to smooth distinct fibre bundles independently of one another. Or, at least, to be able to smooth a segment of a major fibre tract preserving its boundaries.

5.2.6 Diffusion direction scheme for selected fibre orientation

Most current diffusion gradient direction schemes aim to provide equal noise levels for fibre in any direction. Therefore, diffusion gradient directions are uniformly distributed in 3D space. But, it is also necessary to design diffusion direction schemes for specific fibre bundles with selected orientations such as the corticospinal tract (CST) or parts of fibre bundles [Peng and Arfanakis, 2007]. Directional statistics could be applied to arrange diffusion gradient directions for fibres with any orientation or with selected orientations.

With all the aspects of the thesis it must be borne in mind that advances

in technology are always occurring. Methods based on diffusion tensor have been popular because of the long acquisition time of gradient directions in diffusion weighted imaging. It is increasingly possible to sample more directions, more quickly, and so nonparametric models for diffusion directions will become more commonplace. Nevertheless, methods developed in this thesis that apply to covariance matrices in general are envisaged to be useful indefinitely.

Appendices

Appendix 1

Three diffusion gradient direction schemes are tabulated as follows.

Phillips 15		
gx	gy	gz
0.5185	-0.5194	0.6792
0.4830	-0.4775	-0.7339
-0.7056	-0.7086	-0.0033
0.1292	-0.7271	0.6743
-0.2715	-0.7194	0.6393
-0.6090	-0.3529	0.7104
-0.6391	-0.2958	0.7100
0.6860	-0.5551	0.4705
-0.7156	-0.4775	-0.5098
0.6763	-0.2312	-0.6994
0.7069	-0.7041	-0.0664
0.2670	-0.6868	-0.6760
-0.3036	-0.6828	-0.6645
-0.7056	-0.7063	-0.0569
-0.7179	-0.1865	-0.6707

Table A-1: Phillips 15 diffusion gradient direction scheme

Phillips 32		
gx	gy	gz
0.6533	-0.2706	0.7071
0.2087	-0.6756	0.7071
-0.0197	-0.7068	0.7071
-0.4212	-0.5679	0.7071
-0.6899	-0.1549	0.7071
0.6535	-0.2707	0.7069
0.5000	-0.5000	0.7071
0.2929	-0.7071	0.6436
-0.2945	-0.7064	0.6436
-0.5150	-0.4861	0.7061
-0.7071	-0.2929	0.6436
0.7071	-0.4725	0.5261
0.4725	-0.7071	0.5261
-0.5555	-0.6440	0.5261
-0.7071	-0.4725	0.5261
0.7071	-0.7071	0.0002
-0.7071	-0.7071	0
0.7071	-0.4725	-0.5261
-0.7071	-0.4725	-0.5261
-0.4725	-0.7071	-0.5261
-0.7071	-0.4725	-0.5261
0.6364	-0.4252	-0.6436
0.7060	-0.7060	-0.0547
0.2929	-0.7071	-0.6436
-0.2929	-0.7071	-0.6436
-0.7071	-0.7071	-0.0078
-0.7071	-0.2929	-0.6436
0.5847	-0.3977	-0.7071
0.7063	-0.7063	-0.0489
-0.0347	-0.7063	-0.7071
-0.7071	-0.7071	-0.0115
-0.7071	0	-0.7071

Uniform 32		
gx	gy	gz
0.9888	0.0961	-0.1142
0.0897	0.7766	-0.6235
-0.8228	0.5263	0.2143
0.4690	0.5569	-0.6855
0.4758	0.7849	0.3970
-0.7137	0.5512	-0.4322
-0.0038	0.9861	-0.1661
0.0922	0.3580	0.9292
-0.6416	0.1708	0.7478
-0.4705	0.5419	0.6964
0.8055	0.5880	0.0737
-0.2881	0.1460	0.9464
0.7747	0.2338	-0.5875
-0.8898	0.1794	0.4197
-0.5393	0.7598	0.3631
-0.3393	0.7935	-0.5053
0.9309	0.2272	0.2860
-0.9366	0.3323	-0.1112
0.1060	0.8641	0.4920
-0.8847	0.1268	-0.4485
0.7209	0.4982	0.4818
0.4250	0.1289	0.8959
-0.5892	0.2769	-0.7590
0.7319	0.1509	0.6645
0.6949	0.6389	-0.3301
-0.0952	0.6510	0.7531
-0.0848	0.0241	-0.9961
-0.5547	0.8268	-0.0934
-0.2020	0.9373	0.2838
0.2357	0.2968	0.9254
0.3918	0.5531	0.7352
-0.1869	0.4818	-0.8561

Table A-2: Phillips 32 and Uniform 32 diffusion gradient direction schemes

Appendix 2

Jones' J_6			Giné's G_6		
gx	gy	gz	gx	gy	gz
1.0000	0.0000	0.0000	1.0000	0.0000	0.0001
0.4460	0.8950	0.0000	0.4472	0.8944	0.0001
0.4470	0.2750	0.8510	0.4472	0.2764	0.8507
0.4480	-0.7230	-0.5250	0.4473	-0.7236	-0.5257
0.4470	-0.7240	0.5260	0.4472	-0.7236	0.5257
-0.4490	-0.2770	0.8500	-0.4472	-0.2765	0.8506

Table A-3: Jones' J_6 and Giné's G_6 diffusion gradient direction schemes

Jones' J_{10}			Giné's G_{10}		
gx	gy	gz	gx	gy	gz
1.0000	0.0000	0.0000	0.9980	0.0218	0.0590
0.6780	0.7350	0.0000	0.6624	0.7492	-0.0061
-0.5560	0.5040	0.6610	-0.5516	0.4931	0.6728
0.6720	-0.7330	0.1060	0.6891	-0.7196	0.0860
-0.0120	-0.8010	0.5980	-0.0166	-0.8129	0.5822
-0.6800	-0.3100	0.6640	-0.7098	-0.2941	0.6400
-0.0450	-0.0110	0.9990	-0.0578	-0.0518	0.9970
-0.0240	0.9660	0.2570	-0.0371	0.9619	0.2710
0.4580	0.5210	0.7210	0.4465	0.5150	0.7318
0.6580	-0.2500	0.7100	0.6416	-0.2646	0.7200

Table A-4: Jones' J_{10} and Giné's G_{10} diffusion gradient direction schemes

Jones' J_{20}			Giné's G_{20}		
gx	gy	gz	gx	gy	gz
1.0000	0.0000	0.0000	0.9999	0.0007	-0.0119
0.3360	0.9420	0.0000	0.4139	0.9103	0.0080
-0.4050	0.6060	0.6850	-0.4728	0.5422	0.6946
0.8250	-0.5130	-0.2360	0.8258	-0.5349	-0.1789
0.0060	-0.3630	0.9320	-0.0004	-0.2129	0.9771
-0.8110	-0.2870	0.5100	-0.7068	-0.4933	0.5070
0.8520	-0.3200	0.4140	0.8162	-0.3346	0.4711
-0.2400	0.9590	0.1490	-0.3614	0.9041	0.2280
0.8350	0.2720	0.4780	0.8460	0.2311	0.4805
0.0090	-0.9040	0.4270	0.0059	-0.9596	0.2813
-0.0630	-0.8120	-0.5800	-0.0615	-0.8636	-0.5005
-0.2690	-0.3900	-0.8810	-0.0247	-0.4027	-0.9150
-0.4220	-0.6240	0.6580	-0.2652	-0.6872	0.6763
-0.6010	0.7790	-0.1770	-0.5572	0.7814	-0.2810
-0.5160	0.0860	-0.8520	-0.4882	-0.0301	-0.8722
-0.7900	-0.6070	0.0870	-0.8180	-0.5752	-0.0033
0.7290	-0.1810	-0.6610	0.8459	-0.0747	-0.5281
0.2650	-0.0960	-0.9600	0.4677	0.0370	-0.8831
-0.5610	-0.7010	-0.4400	-0.5046	-0.6244	-0.5962
-0.4050	0.6310	-0.6620	-0.3228	0.5847	-0.7443

Table A-5: Jones' J_{20} and Giné's G_{20} diffusion gradient direction schemes

Jones' J_{30}					
gx	gy	gz	gx	gy	gz
1.0000	0.0000	0.0000	0.1660	0.9860	0.0000
-0.1100	0.6640	0.7400	0.9010	-0.4190	-0.1100
-0.1690	-0.6010	0.7810	-0.8150	-0.3860	0.4330
0.6560	0.3660	0.6600	0.5820	0.8000	0.1430
0.9000	0.2590	0.3500	0.6930	-0.6980	0.1780
0.3570	-0.9240	-0.1400	0.5430	-0.4880	-0.6830
-0.5250	-0.3960	0.7530	-0.6390	0.6890	0.3410
-0.3300	-0.0130	-0.9440	-0.5240	-0.7830	0.3350
0.6090	-0.0650	-0.7910	0.2200	-0.2330	-0.9470
-0.0040	-0.9100	-0.4150	-0.5110	0.6270	-0.5890
0.4140	0.7370	0.5350	-0.6790	0.1390	-0.7210
0.8840	-0.2960	0.3620	0.2620	0.4320	0.8630
0.0880	0.1850	-0.9790	0.2940	-0.9070	0.3020
0.8870	-0.0890	-0.4530	0.2570	-0.4430	0.8590
0.0860	0.8670	-0.4910	0.8630	0.5040	-0.0250

Giné's G_{30}					
gx	gy	gz	gx	gy	gz
0.9999	0.0087	-0.0066	0.8851	0.4640	-0.0359
0.3231	0.9464	0.0000	-0.1890	0.6839	0.7047
0.8626	-0.4794	-0.1615	-0.0258	-0.6247	0.7804
-0.8234	-0.3174	0.470	0.5916	0.3902	0.7055
0.6533	0.7110	0.2603	0.8840	0.2500	0.3950
0.6372	-0.7661	0.0840	0.2068	-0.9775	0.0407
0.6187	-0.4440	-0.6481	-0.4674	-0.5247	0.7115
-0.4944	0.7915	0.3592	-0.2678	-0.0052	-0.9635
-0.6063	-0.7371	0.2985	0.5912	0.0547	-0.8047
0.2681	-0.2471	-0.9311	-0.0033	-0.9381	-0.3463
-0.6562	0.5537	-0.5127	0.3317	0.7526	0.5689
-0.7008	0.0654	-0.7104	0.9094	-0.2815	0.3062
0.1589	0.4256	0.8908	0.1508	0.2106	-0.9659
0.2969	-0.8309	0.4705	0.8847	-0.1352	-0.4462
0.3664	-0.4055	0.8374	0.1658	0.8916	-0.4215

Table A-6: Jones' J_{30} and Giné's G_{30} diffusion gradient direction schemes

Appendix 3

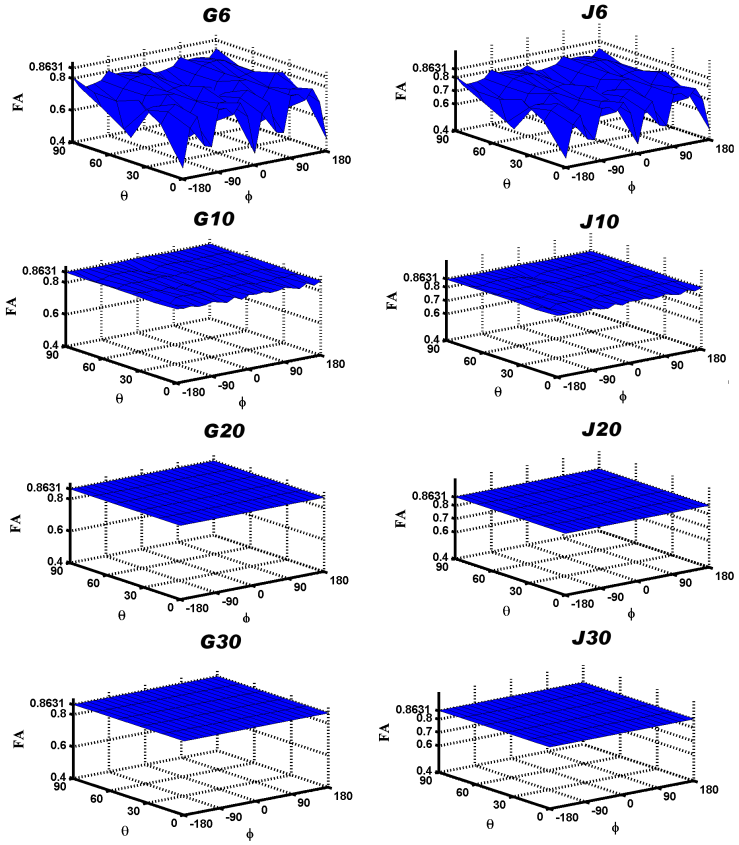

Figure B-1: The FA surfaces for G_N and J_N schemes. $N = 6, 10, 20, 30$.

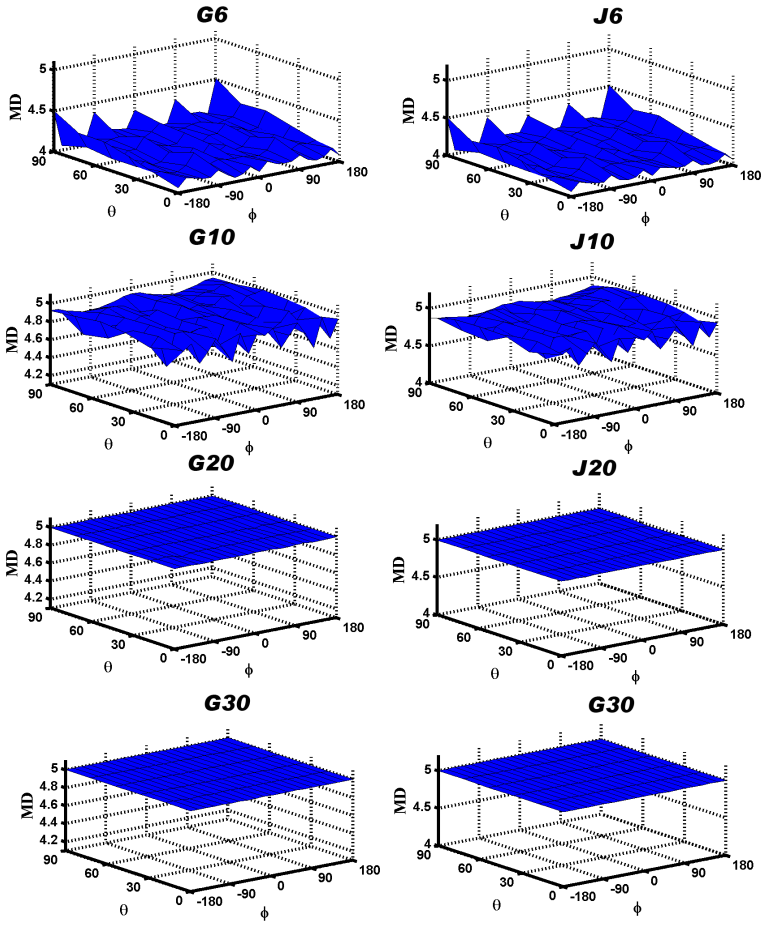

Figure B-2: The MD surfaces for G_N and J_N schemes. $N = 6, 10, 20, 30$.

References

Alexander, D. (2006). *Visualization and image processing of tensor fields*, chapter 5: An introduction to computational diffusion MRI: the diffusion tensor and beyond, pages 83–103. Springer, New York.

Alexander, D. C. (2005). Multiple-fibre reconstruction algorithms for diffusion MRI. *Annals of the New York Academy of Sciences*, 1046:113–133.

Anderson, A. W. (2001). Theoretical analysis of the effects of noise on diffusion tensor imaging. *Magnetic Resonance in Medicine*, 46:1174–1188.

Arsigny, V., Fillard, P., Pennec, X., and Ayache, N. (2006). Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine*, 56:411–421.

Basser, P. J. (2002). Relationships between diffusion tensor and q-space MRI. *Magnetic Resonance in Medicine*, 47:392–397.

Basser, P. J., Mattiello, J., and Le Bihan, D. (1994). Estimation of the effective self-diffusion tensor from the NMR spin echo. *Journal of Magnetic Resonance. Series B*, 103:247–254.

Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J., and Aldroubi, A. (2000). In vivo fiber tractography using DT-MRI data. *Magnetic Resonance in Medicine*, 44:625–632.

Basser, P. J. and Pierpaoli, C. (1996). Microstructural and physiological features of tissues elucidated by quantitative diffusion tensor MRI. *Journal of Magnetic Resonance, Series B*, 111:209–219.

Basser, P. J. and Pierpaoli, C. (2005). A simplified method to measure the diffusion tensor from seven MR images. *Magnetic Resonance in Medicine*, 39(6):928 – 934.

Basu, S., Fletcher, T., and Whitaker, R. (2006). Noise removal in diffusion tensor MRI. In Larsen, R., Nielsen, M., and Sporring, J., editors, *MICCAI 2006, LNCS*, volume 4190, pages 117–125.

Batchelor, P. G., Moakher, M., Atkinson, D., Calamante, F., and Connelly, A. (2004). A rigorous framework for diffusion tensor calculus. *Magnetic Resonance in Medicine*, 53(1):221–225.

Behrens, T., Woolrich, M., Jenkinson, M., Johansen-Berg, H., Nunes, R., Clare, S., Matthews, P., Brady, J., and Smith, S. (2003). Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magnetic Resonance in Medicine*, 50:1077–1088.

Behrens, T. E. J., Johansen Berg, H., Jbabdi, S., Rushworth, M. F. S., and Woolrich, M. W. (2007). Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *NeuroImage*, 34:144–155.

Bernardo, J. and Smith, A. F. (2009). *Bayesian theory*. John Wiley, Chichester.

Bingham, C. (1974). An antipodally symmetric distribution on the sphere. *Annals of Statistics*, 2:1201–1225.

Buchsbaum, M. S., Friedman, J., Buchsbaum, B. R., Chu, K. W., Hazlett, E. A., Newmark, R., Schneiderman, J. S., Torosjan, Y., Tang, C., Hof, P. R., Stew-

art, D., Davis, K. L., and Gorman, J. (2006). Diffusion tensor imaging in schizophrenia. *Biological Psychiatry*, 60:1181–1187.

Callaghan, P. T. (1995). *Principles of Nuclear Magnetic Resonance Microscopy*. Oxford Press, Oxford.

Casten, R. (1990). *Nuclear Structure From a Simple Perspective*. Oxford University Press, Oxford.

Chao, T., Chou, M., Yang, P., Chung, H., and Wu, M. (2009). Effects of interpolation methods in spatial normalization of diffusion tensor imaging data on group comparison of fractional anisotropy. *Magnetic Resonance in Medicine*, 27(5):681–690.

Ciccarelli, O., Behrens, T. E., Altmann, D. R., Orrell, R. W., Howard, R. S., Johansen-Berg, H., Miller, D. H., Matthews, P. M., and Thompson, A. J. (2006). Probabilistic diffusion tractography: a potential tool to assess the rate of disease progression in amyotrophic lateral sclerosis. *Brain*, 129(7):1859–1871.

Cobelli, C. and DiStefano, J. J. (1980). Parameter and structural identifiability concepts and ambiguities: a critical review and analysis. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, 239(1):R7–24.

Conturo, T. E., Lori, N. F., Cull, T. S., Akbudak, E., Snyder, A. Z., Shimony, J. S., McKinstry, R. C., Burton, H., and Raichle, M. E. (1999). Tracking neuronal fiber pathways in the living human brain. *Proceedings of the National Academy of Sciences*, 96:10422–10427.

Cook, P. A., Bai, Y., Nedjati-Gilani, S., Seunarine, K. K., Hall, M. G., Parker, G. J., and Alexander, D. C. (2006). Camino: Open-source diffusion-MRI reconstruction and processing. *the 14th scientific meeting of the international society for magnetic resonance in medicine*, Seattle, USA.

Coyne, K. E. (1999). MRI: A guided tour [online]. *Accessed 22/07/2009*, URL: http://www.magnet.fsu.edu/education/tutorials/magnetacademy/mri/.

Crank, J. (1975). *The Mathematics of Diffusion*. Oxford University Press, Oxford.

Daniel, J. W. (1967). Convergence of the conjugate gradient method with computationally convenient modifications. *Numerische Mathematik*, 10:125–131.

Dong, Q., Welsh, R. C., Chenevert, T. L., Carlos, R. C., Maly-Sundgren, P., Gomez-Hassan, D. M., and Mukherji, S. K. (2004). Clinical applications of diffusion tensor imaging. *Journal of Magnetic Resonance Imaging*, 19:6–18.

Dryden, I. L., Koloydenko, A., and Zhou, D. (2009a). Non-euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Annals of Applied Statistics*, 3(3):1102–1123.

Dryden, I. L., Koloydenko, A., Zhou, D., and Bai, L. (2009b). Non-Euclidean statistical analysis of convarience matrices and diffusion tensors, The 57th Session of the International Statistical Institute (ISI), Durban, South Africa.

Dryden, I. L. and Mardia, K. V. (1998). *Statistical Shape Analysis*. Wiley & Sons, Chichester.

Fletcher, P. T. and Joshi, S. (2007). Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Process*, 87(2):250–262.

Fletcher, P. T., Venkatasubramanian, S., and Joshi, S. (2009). The geometric median on Riemannian manifolds with application to robust atlas estimation. *Neuroimage*, 45:S143–S152.

Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'institut Henri Poincaré*, 10 (4):215–310.

Gamerman, D. and Lopes, H. F. (2006). *Markov chain Monte Carlo : stochastic simulation for Bayesian inference*. Taylor & Francis, London.

Giné, E. M. (1975). Invariant tests for uniformity on compact Riemannian manifolds based on Sobolev norms. *The Annals of Statistics*, 3:1243–1266.

Gower, J. C. (1975). Generalized Procrustes analysis. *Psychometrika*, 40:33–50.

Gudbjartsson, H. and Patz, S. (1995). The Rician distribution of noisy MRI data. *Magnetic Resonance in Medicine*, 34:910–914.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C., Wedeen, V., and Sporns, O. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biology*, 6.

Hagmann, P., Jonasson, L., Maede, P., Thiran, J.-P., van Wedeen, J., and Meuli, R. (2006). Understanding diffusion MR imaging techniques: From scalar diffusion-weighted imaging to diffusion tensor imaging and beyond. *RadioGraphics*, 26:S205–S223.

Hagmann, P., Thiran, J. P., Jonasson, L., P., V., Clarke, S., Maeder, P., and Meuli, R. (2003). DTI mapping of human brain connectivity: statistical fibre tracking and virtual dissection. *NeuroImage*, 19:545–554.

Hänggi, P. and Marchesoni, F. (2005). Introduction: 100 years of Brownian motion. *Chaos*, 15(2):026101–026101–5.

Hasan, K. M., Parker, D. L., and Alexander, A. L. (2001). Comparison of gradient encoding schemes for diffusion-tensor MRI. *Journal of Magnetic Resonance Imaging*, 13:769–780.

Hesseltine, S. M., Law, M., Babba, J., Rad, M., Lopez, S., Ge, Y., Johnson, G., and Grossman, R. I. (2006). Diffusion tensor imaging in multiple sclerosis: As-

sessment of regional differences in the axial plane within normal-appearing cervical spinal cord. *American Journal of Neuroradiology*, 27:1189–1193.

Hiroshi (2009). Modeling human head in 3D max. *Accessed 22/07/2009*, URL: http://www.techmynd.com/category/3d-studio-max/.

Hornak, J. P. (2008). The basics of MRI [online]. *Accessed 26/09/2009*, URL: http://www.cis.rit.edu/htbooks/mri/.

Hsiao, C. (1983). *Handbook of Econometrics*, chapter 4 Identification, pages 223–283. Elsevier, Amsterdam.

Hüppi, P. S., Maier, S. E., Peled, S., Zientara, G. P., Barnes, P. D., Jolesz, F. A., and Volpe, J. J. (1998). Microstructural development of human newborn cerebral white matter assessed in vivo by diffusion tensor magnetic resonance imaging. *Pediatric Radiology*, 44(4):584–590.

Jansons, K. M. and Alexander, D. C. (2003). Persistent angular structure: new insights from diffusion mri data. *Inverse Problems*, 19:1031–1046.

Jeffreys, H. (1961). *The Theory of Probability*. Oxford University Press, Oxford.

Jones, D., Horsfield, M., and Simmons, A. (1999). Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magnetic Resonance in Medicine*, 42(3):515–525.

Jones, D. K. (2003). Determining and visualizing uncertainty in estimates fiber orientation from diffusion tensor MRI. *Magnetic Resonance in Medicine*, 49:7–12.

Kang, N., Zhang, J., Carlson, E. S., and Gembris, D. (2005). White matter fiber tractography via anisotropic diffusion simulation in the human brain. *IEEE Transactions on Medical Imaging*, 24 (9):1127–1137.

Kendall, D. G. (1989). A survey of the statistical theory of shape. *Statistical Science*, 4:87–120.

Kingsley, P. B. (2006a). Introduction to diffusion tensor imaging mathematics: Part i. tensors, rotations, and eigenvectors. *Concepts in Magnetic Resonance Part A*, 28A:101–122.

Kingsley, P. B. (2006b). Introduction to diffusion tensor imaging mathematics: Part ii. anisotropy, diffusion-weighting factors, and gradient encoding schemes. *Concepts in Magnetic Resonance Part A*, 28A:123–154.

Kingsley, P. B. (2006c). Introduction to diffusion tensor imaging mathematics: Part iii. tensor calculation, noise, simulations, and optimization. *Concepts in Magnetic Resonance Part A*, 28A:155–179.

Koay, C. G., Carew, J. D., Alexander, A. L., Basser, P. J., and Meyerand, M. E. (2006). Investigation of anomalous estimates of tensor-derived quantities in diffusion tensor imaging. *Magnetic Resonance in Medicine*, 55:930–936.

Koenker, R. (2006). The median is the message: toward the Fréchet median. *Journal de la Société Francaise de Statistique*, 147(2):61–64 (2007).

Lagarias, J., Reeds, J. A., Wright, M. H., and Wright, P. E. (1998). Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM Journal of Optimization*, 9(1):112–147.

Lazar, M., Alexander, A., Thottakara, P., Badie, B., and Field, A. (2006). White matter reorganization after surgical resection of brain tumors and vascular malformations. *American Journal of Neuroradiology*, 27:1258–1271.

Le, H. (1995). Mean size-and-shapes and mean shapes: A geometric point of view. *Advances in Applied Probability*, 27 (1):44–55.

Le, H. and Kume, A. (2000). The Fréchet mean shape and the shape of the means. *Advances in Applied Probability*, 32(1):101–113.

Le Bihan, D., Mangin, J. F., Poupon, C., Clark, C. A., Pappata, S., Molko, N., and Chabriat, H. (2001). Diffusion tensor imaging: concepts and applications. *Journal of Magnetic Resonance Imaging*, 13:534–546.

Lee, C. E. C., Danielian, L. E., Thomasson, D., and Baker, E. H. (2009). Normal regional fractional anisotropy and apparent diffusion coefficient of the brain measured on a 3T MR scanner. *Neuroradiology*, 51:3–9.

Lee, P. M. (1991). *Bayesian Statistics: An Introduction*. Oxford University Press, Oxford.

Lenglet, C., Campbell, J., Descoteaux, M., Haro, G., Savadjiev, P., Wassermann, D., Anwander, A., Deriche, R., Pike, G., Sapiro, G., Siddiqi, K., and Thompson, P. (2009). Mathematical methods for diffusion MRI processing. *NeuroImage*, 45:S111 – S122.

Lindgren, B. W. (2003). *Statistical Theory*. Chapman & Hall/CRC, Florida.

Lori, N. F., Akbudak, E., Shimony, J. S., Cull, T. S., Snyder, A. Z., Guillory, R. K., and Conturo, T. E. (2002). Diffusion tensor fiber tracking of human brain connectivity: aquisition methods, reliability analysis and biological results. *NMR in Biomedicine*, 15:494–515.

Mäkelä, J. P., Kirveskari, E., Seppä, M., Hämäläinen, M., Forss, N., Avikainen, S., Salonen, O., Salenius, S., Kovala, T., Randel, T., Jääskeläinen, J., and Hari, R. (2001). Three-dimensional integration of brain anatomy and function to facilitate intraoperative navigation around the sensorimotor strip. *Human Brain Mapping*, 12:180–192.

Mansfield, P. and Grannell, P. K. (1973). NMR 'diffraction' in solids? *Journal of Physics C: Solid State Physics*, 6:L422.

Mardia, K. and Jupp, P. (2000). *Directional Statistics*. J. Wiley & Sons, Chichester.

McGraw, T., Vemuri, B. C., Chen, Y., Rao, M., and Mareci, T. (2004). DT-MRI denoising and neuronal fiber tracking. *Medical Image Analysis*, 8:95–111.

Medina, D. A. and Gaviria, M. (2008). Diffusion tensor imaging investigations in Alzheimer's disease: the resurgence of white matter compromise in the cortical dysfunction of the aging brain. *Neuropsychiatric Disease and Treatment*, 4:737–742.

Melnick, E. L. and Tenenbein, A. (1982). Misspecifications of the normal distribution. *The American Statistician*, 36(4):372–373.

Mirsky, L. (1955). *An introduction to linear algebra*. Clarendon Press, Oxford.

Moakher, M. and Batchelor, P. G. (2006). *Visualization and Processing of Tensor Fields*. Springer, New York.

Mori, S. (2007). *Introduction to Diffusion Tensor Imaging*. Elsevier, Amsterdam.

Mori, S., Frederiksen, K., van Zijl, P. C., Stieltjes, B., Kraut, M. A., Solaiyappan, M., and Pomper, M. G. (2002). Brain white matter anatomy of tumor patients evaluated with diffusion tensor imaging. *Annals of Neurology*, 51:377–380.

Muthupallai, R., Holder, C. A., Song, A. W., and Dixon, W. T. (1999). Navigator-aided, multishot EPI diffusion images of brain with complete orientation and anisotropic information. *The 8th Scientific Meeting of the International Society for Magnetic Resonance in Medicine.*

Narr, K. L., Hageman, N., Woods, R. P., Hamilton, L. S., Clark, K., Phillips, O., Shattuck, D. W., Asarnow, R. F., Toga, A. W., and Nuechterlein, K. H. (2009).

Mean diffusivity: a biomarker for CSF-related disease and genetic liability effects in schizophrenia. *Psychiatry Research*, 171:20–32.

Nucifora, P. G. P., Verma, R., Lee, S.-K., and Melhem, E. R. (2007). Diffusion-tensor MR imaging and tractography: Exploring brain microstructure and connectivity. *Radiology*, 245:367–384.

Olesen, P. J., Nagy, Z., Westerberg, H., and Klingberg, T. (2003). Combined analysis of DTI and fMRI data reveals a joint maturation of white and grey matter in a fronto-parietal network. *Cognitive Brain Research*, 18(1):48–57.

Olhede, S. and Whitcher, B. (2008). A statistical framework to characterise microstructure in high angular resolution diffusion imaging. *Biomedical Imaging: From Nano to Macro. ISBI 2008, Paris, France*.

Papadakis, N. G., Xing, D., Houston, G. C., Smith, J. M., Smith, M. I., James, M. F., Parsons, A. A., Huang, C. L. H., Hall, L. D., and Carpenter, T. A. (1999a). A study of rotationally invariant and symmetric indices of diffusion anisotropy. *Magnetic Resonance Imaging*, 17:881–892.

Papadakis, N. G., Xing, D., Huang, C. L. H., Halla, L. D., and Carpenter, T. A. (1999b). A comparative study of acquisition schemes for diffusion tensor imaging using MRI. *Journal of Magnetic Resonance*, 137:67–82.

Parker, G. J. and Alexander, D. C. (2003). *Probabilistic Monte Carlo Based Mapping of Cerebral Connections Utilising Whole-Brain Crossing Fibre Information*, volume 2732/2003 of *Lecture Notes in Computer Science: Information Processing in Medical Imaging*. Springer, Berlin.

Paulino, C. D. M. and de Bragança Pereira, C. (1994). On identifiability of parametric statistical models. *Statistical Methods and Applications*, 3 (1):125–151.

Paus, T., Zijdenbos, A., Worsley, K., Collins, D. L., Blumenthal, J., Giedd, J. N., Rapoport, J. L., and Evans, A. C. (1999). Structural maturation of neural pathways in children and adolescents: In vivo study. *Science*, 283 (5409):1908–1911.

Peng, H. and Arfanakis, K. (2007). Diffusion tensor encoding schemes optimized for white matter fibers with selected orientations. *Magnetic Resonance Imaging*, 25 (2):147–153.

Pennec, X., Fillard, P., and Ayache, N. (2006). A Riemannian framework for tensor computing. *International Journal of Computer Vision*, 66:41–66.

Pierpaoli, C., Jezzard, P., Basser, P., A, B., and Chiro, G. D. (1996). Diffusion tensor MR imaging of the human brain. *Radiology*, 201:637–648.

Pomara, N., Crandalla, D. T., Choia, S. J., Johnsonb, G., and Lim, K. O. (2001). White matter abnormalities in HIV-1 infection: A diffusion tensor imaging study. *Psychiatry Research: Neuroimaging*, 106(1):15–24.

Schempp, W. J. (1998). *Magnetic Resonance Imaging (Mathematical Foundations and Applications)*. WileyLiss, New York.

Simmonds, J. G. (1994). *A Brief on Tensor Analysis*. Springer-Verlag New York, Inc.

Skare, S., Hedehus, Maj andMoseley, M. E., and Li, T.-Q. (2000). Condition number as a measure of noise performance of diffusion tensor data acquisition schemes with MRI. *Journal of Magnetic Resonance*, 147(2):340 – 352.

Smith, S., Jenkinson, M., Woolrich, M., Beckmann, C., Behrens, T., Johansen-Berg, H., Bannister, P., De Luca, M., Drobnjak, I., Flitney, D., Niazy, R., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J., and Matthews, P.

(2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23(S1):208–219.

Stejskal, E. O. and Tanner, J. E. (1965). Spin diffusion measurements: Spin echoes in the presence of a time - dependent field gradient. *Journal of Chemical Physics*, 42:288–292.

Sundgren, P. C., Dong, Q., Gómez-Hassan, D., Mukherji, S. K., Maly, P., and Welsh, R. (2004). Diffusion tensor imaging of the brain: review of clinical applications. *Neuroradiology*, 46:339–350.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58 (1):267–288.

Tievsky, A. L., Ptak, T., and Farkas, J. (1999). Investigation of apparent diffusion coefficient and diffusion tensor anisotropy in acute and chronic multiple sclerosis lesions. *American Journal of Neuroradiology*, 20:1491–1499.

Tuch, D. S. (2004). Q-ball imaging. *Magnetic Resonance in Medicine*, 52:1358–1372.

Tuch, D. S., Reese, T. G., Wiegell, M. R., and van Wedeen, J. (2003). Diffusion MRI of complex neural architecture. *Neuron*, 48:885–895.

Tuch, D. S., Weisskoff, R. M., Belliveau, J. W., and Wedeen, V. J. (1999). High angular resolution diffusion imaging of the human brain, The 7th Annual Meeting of the ISMRM, Philadelphia, PA.

van den Hof, J. M. (1998). Structural identifiability of linear compartmental systems. *IEEE Transactions on Automatic Control*, 43:800–818.

Vlaardingerbroek, M. T. and den Boer, J. A. (1999). *Magnetic Resonance Imaging: Theory and Practice*. Springer, New York.

Wang, Z., Vemuri, B., Chen, Y., and Mareci, T. (2004). A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from complex DWI. *IEEE Transactions on Medical Imaging*, 23 (8):930–939.

Waxman, S. (2005). *From Neuroscience to Neurology: Neuroscience, Molecular Medicine, and the Therapeutic Transformation of Neurology*. Elsevier Academic Press, London.

Westin, C., Maier, S., Mamata, H., Nabavi, A., Jolesz, F. A., and Kikinis, R. (2002). Processing and visualization for diffusion tensor MRI. *Medical Image Analysis*, 6:93–108.

Westin, C., Peled, S., Gudbjartsson, H., Kikinis, R., and Jolesz, F. (1997). Geometrical diffusion measures for MRI from tensor basis analysis, The 5th International Society for Magnetic Resonance in Medicine, Vancouver.

Wheeler-Kingshott, C. A., Barker, G. J., Steens, S. C., and van Buchem, M. A. (2003). *Quantitative MRI of the Brain: measuring changes caused by disease*, chapter 7: The Diffusion of Water, pages 203–256. Edited by Tofts, Paul S. John Wiley & Sons Ltd, Chichester.

Woodward, P. (2001). *MRI for technologists*. McGraw-Hill, New York.

Yudofsky, S. C. and Hales, R. E. (2004). *Essentials of neuropsychiatry and clinical neurosciences*. American Psychiatric Publishig, Inc., Arlington.

Zhou, D. (2006). Exploring diffusion tensor imaging. Msc. in statistics, School of Mathematical Sciences, University of Nottingham.

Zhou, D., Dryden, I. L., Koloydenko, A., and Bai, L. (2008). A Bayesian method with reparameterisationfor diffusion tensor imaging. In Reinhardt, J. M. and Pluim, J. P. W., editors, *Proceedings of SPIE Medical Imaging 2008: Image Processing*, page 69142J.

Zhou, D., Dryden, I. L., Koloydenko, A., and Bai, L. (2009a). Procrustes analysis of diffusion tensor data, The 17th Scientific Meeting of International Society for Magnetic Resonance in Medicine, Hawaii, USA.

Zhou, D., Dryden, I. L., Koloydenko, A., and Bai, L. (2009b). Weighted generalised Procrustes analysis of diffusion tensors. In Gusnanto, A., Mardia, K., and Fallaize, C. J., editors, *Proceedings of LASR 2009–Statistical Tools for Challenges in Bioinformatics*, pages 120–123. Leeds University Press, Leeds.